**MÁRCIO TADEU GODINHO**

# COEXISTÊNCIA E EVOLUÇÃO MOLECULAR DE POPULAÇÕES DE BEGOMOVÍRUS NA PLANTA NÃO-CULTIVADA *Sida acuta*

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

VIÇOSA
MINAS GERAIS – BRASIL
2014

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

T
G585c
2014

Godinho, Márcio Tadeu, 1981-
Coexistência e evolução molecular de populações de begomovírus na planta não-cultivada *Sida acuta* / Márcio Tadeu Godinho. – Viçosa, MG, 2014.
x, 71f. : il. (algumas color.) ; 29 cm.

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

1. Begomovirus. 2. Evolução. 3. Geminiviridae. 4. Nova espécie. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Fitopatologia. II. Título.

CDD 22. ed. 571.9928

**MÁRCIO TADEU GODINHO**

# COEXISTÊNCIA E EVOLUÇÃO MOLECULAR DE POPULAÇÕES DE BEGOMOVÍRUS NA PLANTA NÃO-CULTIVADA *Sida acuta*

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

APROVADA: 25 de fevereiro de 2014

| | |
|---|---|
| Prof$^{a}$. Claudine Márcia Carvalho | Prof$^{a}$. Elizabeth Pacheco Batista Fontes |
| Prof. Luis Cláudio Vieira da Cunha | Prof$^{a}$. Renate Krause Sakate |

Prof. Francisco Murilo Zerbini Júnior
(Orientador)

# AGRADECIMENTOS

A Deus, por me guiar em meu caminho, iluminando e permitindo a conquista de mais essa vitória;

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

Ao Professor Francisco Murilo Zerbini, pelo suporte profissional, orientação, dedicação e amizade durante a realização deste trabalho;

Aos funcionários Joaquim e Renan pelo apoio para a realização deste trabalho e à amizade durante este tempo;

Aos amigos do Laboratório de Virologia Vegetal Molecular, Alison, Adriana, Larissa, Diogo, Silvia, César, André, Diego, Talita, Osvaldo, Roberto, Fábio, Glória, Tathiana, Renan, Fernanda, Karel, Hermano, João, Josiane, Vitor, Marcos, Pedro, Chaianne, Daniel e especialmente a Camila pela amizade e pelo agradável ambiente de trabalho;

Ao pessoal das repúblicas Toca do Tatu, Vaca Véia, Os Cavalo deitado, Alambique, Mói-Cana, Grozop, Os Morangueiros e Quarentena, por me mostrarem o que é a vida em uma república de verdade.

Aos amigos, Éder (Tidé), Igor (Carrapis), Leonardo (Leo), Geraldo (Gera), Débora, Hélvio, Tiago (Jaja), Teófilo (Teo), Ismael, Bruno (Brunerva), Helen e muitos outros, aos quais agradeço em especial pela longa amizade, momentos de descontração, incentivo, pelas muitas histórias geradas, e por me aguentarem durante todos esses anos de Viçosa;

Aos meus tios e tias, Aurinha, Ana Maria, Elza, Paula, Marco Antônio, Paulo Vítor, Ângelo, Maria das Graças, Rubens, Cida, Marli (in memorian) e meus respectivos primos e primas por serem essa minha família maravilhosa em todos os sentidos;

Às minhas avós, Áurea e Doralice (in memorian), e meu avô Paulo Amâncio (in memorian) pelo carinho, amor e por me acolher sempre que precisei;

A minha namorada Marcela, por acreditar em mim, pelo amor e ser minha companheira em todos os momentos, tanto bons quanto ruins, estando sempre ao meu lado;

Aos meus irmãos, Karina e Guilherme, por acreditarem em mim e estarem serem presentes ao longo de minha vida;

Aos meus pais, Augusto e Cármina, por acreditarem em mim, pelo suporte em todos os sentidos, pelo incentivo constante, pelo amor, compreensão, atenção e por me ensinarem seus valores que me guiam durante toda minha vida;

A todos que ajudaram e apoiaram durante esses anos e contribuíram direta ou indiretamente para a realização deste trabalho.

# BIOGRAFIA

MÁRCIO TADEU GODINHO, filho de Augusto Tadeu Godinho e Cármina Maria Amâncio Godinho, nasceu em Três Pontas, Minas Gerais, no dia 03 de novembro de 1981.

Residiu e estudou todos os anos do ensino fundamental e médio na cidade de Três Pontas, Minas Gerais, sendo concluído em 1999. Graduou-se em Agronomia pela Universidade Federal de Viçosa (UFV), na cidade de Viçosa, Minas Gerais, em maio de 2006. Ingressou, em março de 2008, no curso de pós-graduação em Fitopatologia em nível de Mestrado na Universidade Federal de Viçosa, obtendo o título de *Magister Scientiae* em fevereiro de 2010. Ingressou no mês seguinte no curso de pós-graduação em Fitopatologia em nível de Doutorado na Universidade Federal de Viçosa e, em fevereiro de 2014, submeteu-se à defesa de tese.

# SUMÁRIO

# RESUMO

GODINHO, Marcio Tadeu, D.Sc., Universidade Federal de Viçosa, fevereiro de 2014. **Coexistência e evolução molecular de populações de begomovírus na planta não-cultivada *Sida acuta*.** Orientador: Francisco Murilo Zerbini Júnior Co-orientador: Eduardo Seiti Gomide Mizubuti.

A família *Geminiviridae* inclui vírus cujos genomas são compostos por uma ou duas moléculas de DNA fita simples circular, encapsidadas por uma única proteína estrutural em partículas icosaédricas geminadas. A família é composta por sete gêneros definidos com base no tipo de inseto vetor, gama de hospedeiros, relacionamento filogenético e organização genômica. Os vírus pertencentes ao gênero *Begomovirus* possuem um ou dois componentes genômicos, são transmitidos pela mosca-branca *Bemisia tabaci* e infectam plantas dicotiledôneas. A incidência de begomovírus em plantas cultivadas e não-cultivadas vem sendo relatada no Brasil desde a década de 1950, inicialmente em malváceas e leguminosas, e a partir da década de 1990, após a introdução do biótipo B de *B. tabaci*, também em solanáceas. Diversos estudos já foram realizados para caracterizar os begomovírus que ocorrem em tomateiro e feijoeiro no Brasil, e para avaliar a variabilidade genética das populações virais em plantas cultivadas. Estudos determinando a estrutura genética e dinâmica de populações de begomovírus em hospedeiros não-cultivados são menos frequentes, embora seja aceito que estes hospedeiros atuem como fonte de inóculo, podendo portanto contribuir para epidemias em hospedeiros cultivados. Este trabalho teve como objetivos caracterizar e estudar a variabilidade genética de populações de begomovírus infectando a planta não-cultivada *Sida acuta* em uma área de 10.000 m$^2$. Os resultados indicam uma alta variabilidade entre os isolados, com três novas espécies de begomovírus cujos nomes propostos são *Sida acuta* mosaic virus (SAMV), *Sida* golden yellow mosaic virus (SiGYMV) e *Sida* yellow spot virus (SiYSV). A população de SAMV está subdividida em três estirpes (S1/S2/S3), com mutações pontuais nos genes CP e Rep separando as três estirpes, inclusive com uma possível diferença fenotípica entre as estirpes S1 e S2. Mesmo na pequena área amostrada foram encontradas plantas com infecção mista e pseudo-recombinação entre estirpes. Os clones pertencentes às espécies SiGYMV e SiYSV apresentam uma organização genômica relacionada a begomovírus presentes no "Velho Mundo" (Europa, Ásia e África) com a presença de uma ORF tipo AV2 existente apenas em begomovirus daquela região, e de motivos conservados nas sequências dos genes CP e Rep também existentes apenas nos begomovírus do Velho Mundo. Conclui-se que a grande diversidade de espécies e variabilidade genética dos begomovírus pode ser amostrada mesmo em pequenas áreas.

# ABSTRACT

GODINHO, Márcio Tadeu, D.Sc., Universidade Federal de Viçosa, February 2014. **Coexistence and molecular evolution of begomovirus populations in the non-cultivated plant *Sida acuta*.** Advisor: Francisco Murilo Zerbini Júnior. Co-advisor: Eduardo Seiti Gomide Mizubuti.

The *Geminiviridae* family includes viruses with genomes consisting of one or two molecules of single-stranded circular DNA, encapsidated by a single structural protein in twinned icosahedral particles. The family includes seven genera defined based on the type of insect vector, host range, genome organization and phylogenetic relationships. Viruses belonging to the genus *Begomovirus* have one or two genomic components, are transmitted by the whitefly *Bemisia tabaci* and infect dicotyledonous plants. The occurrence of begomoviruses in cultivated and non-cultivated plants has been reported in Brazil since the 1950s, initially in malvaceous and leguminous hosts, and since the mid-1990s, following the introduction of the B biotype of *B. tabaci*, also in solanaceous hosts. Several studies have been conducted to characterize the begomoviruses occurring in tomato and bean crops in Brazil, and to evaluate the genetic variability of viral populations in cultivated plants. Studies determining the genetic structure and population dynamics of begomoviruses in non-cultivated hosts are less frequent, although it is accepted that these hosts act as a source of inoculum, and thus may contribute to epidemics in cultivated hosts. This study aimed to characterize and study the genetic variability of begomoviruses infecting the non-cultivated plant *Sida acuta*, sampled in an area of 10,000 $m^2$. The results showed a high variability among isolates, with the detection of three new begomovirus speceis whose proposed names are *Sida acuta* mosaic virus (SAMV), *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV). The SAMV population is subdivided into three different strains (S1/S2/S3), with point mutations in the CP and Rep genes separating the three strains, and with a possible phenotypic difference between strains S1 and S2. Even in the small sampled area, several plants had mixed infection and pseudorecombination between the different strains. The clones belonging to the species SiGYMV and SiYSV have a genomic organization related to begomoviruses present in the "Old World" (Europe, Asia and Africa), with the presence of an AV2-like ORF present only in begomovirus from that region, and with conserved motifs in the CP and Rep proteins also existing only in the Old World. Thus, the high species diversity and genetic variability of begomoviruses can be sampled even in small areas.

# INTRODUÇÃO

A família *Geminiviridae* inclui vírus cujos genomas são compostos de uma ou duas moléculas de DNA fita simples circular, encapsidadas por uma única proteína estrutural em partículas icosaédricas geminadas. A família é composta pelos gêneros *Begomovirus, Becurtovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus* e *Turncurtovirus,* definidos com base no tipo de inseto vetor, gama de hospedeiros, relacionamento filogenético e organização genômica (Varsani *et al.*, 2014). Os begomovírus possuem um ou dois componentes genômicos, são transmitidos pela mosca-branca *Bemisia tabaci* e infectam plantas dicotiledôneas.

A grande maioria das espécies de begomovírus que ocorrem no "Velho Mundo" (Europa, Ásia e África) possuem um componente genômico (monossegmentados), e frequentemente estão associadas a DNAs satélites denominados alfa- ou betassatélites (Mansoor *et al.*, 2003). Begomovírus encontrados no "Novo Mundo" (Américas) e alguns encontrados no Velho Mundo possuem dois componentes, denominados DNA-A e DNA-B (bissegmentados). O DNA-A contém genes envolvidos na replicação e encapsidação da progênie viral. O DNA-B contém os genes requeridos para o movimento intra- e intercelular na planta (Rojas *et al.*, 2005). Ambos os componentes genômicos são requeridos para a infecção sistêmica do hospedeiro. Uma diferença entre os begomovírus bissegmentados do Velho e do Novo Mundo é a presença de um gene no DNA-A (AV2) apenas nas espécies do Velho Mundo.

Diversos relatos nos últimos três anos vêm levando a uma reavaliação de conceitos anteriormente estabelecidos em relação aos begomovírus, como a associação de DNAs satélites apenas com begomovírus monossegmentados no Velho Mundo e a presença apenas de begomovírus bissegmentados no Novo Mundo. Assim, a associação de begomovírus do Novo Mundo com alfassatélites foi demonstrada no Brasil e na Venezuela

(Paprotka *et al.*, 2010; Romay *et al.*, 2010), e a existência de uma terceira classe de DNAs satélites associados a begomovírus infectando plantas daninhas foi relatada em Cuba (Fiallo-Olive *et al.*, 2012). Além disso, recentemente foi demonstrada na América do Sul a emergência de um begomovírus do Novo Mundo (sem o gene AV2) possuindo apenas um componente genômico (Melgarejo *et al.*, 2013). Esses relatos sugerem que a diversidade genotípica e a variabilidade genética dos begomovírus é muito maior do que se acreditava. O fato de muitos desses relatos terem sido feitos a partir da análise de plantas não-cultivadas (Paprotka *et al.*, 2010; Fiallo-Olive *et al.*, 2012) ressalta a necessidade de se investigar essas plantas como um reservatório da diversidade viral e como uma fonte de novos vírus que podem causar doenças em plantas cultivadas.

Populações de geminivírus, incluindo os begomovírus, possuem uma elevada variabilidade genética, equivalente a de vírus com genoma de RNA (Ge *et al.*, 2007; Prasanna *et al.*, 2010; Rocha *et al.*, 2013). As principais fontes de variabilidade genética de vírus em plantas são mutação, recombinação e pseudo-recombinação (García-Arenal *et al.*, 2003; Seal *et al.*, 2006). Eventos frequentes de recombinação (Padidam *et al.*, 1999), a ocorrência de pseudo-recombinação entre vírus com genoma bissegmentado (Andrade *et al.*, 2006), e a alta taxa de evolução molecular (Duffy e Holmes, 2008) são fatores que promovem a elevada variabilidade observada.

As frequências de mutação e as taxas de substituição de nucleotídeos observadas para os begomovírus apresentam valores similares aos verificados para vírus de RNA, apesar da expectativa de que fossem menores devido à utilização, pelos geminivírus, da maquinaria de replicação do hospedeiro, o que em teoria permitiria corrigir as incorporações incorretas de nucleotídeos, aumentando a fidelidade da replicação (Duffy e Holmes, 2008). É possível que os geminivírus não utilizem os mecanismos de correção de erro da síntese de DNA do hospedeiro, de forma que mutações não deletérias sejam mantidas (Sanz *et al.*, 1999) e, ou que o genoma de DNA de fita simples seja particularmente suscetível a estresses oxidativos que aumentam a taxa basal de mutação (Monjane *et al.*, 2012).

A recombinação intermolecular é considerada fundamental para a variabilidade genética dos geminivírus. O conhecimento da existência e frequência de recombinação em uma população de vírus pode ajudar a entender quais genes são intercambiados e o surgimento de novas espécies de vírus. Esta informação é essencial, por exemplo, para determinar a durabilidade da resistência genética, pois novas variantes recombinantes poderiam ser formadas com incremento da adaptabilidade a genótipos resistentes (Monci *et al.*, 2002; Awadalla, 2003; Sattar *et al.*, 2013).

A estrutura genética de uma população reflete a história evolutiva e o potencial da população para evoluir (Pinel *et al.*, 2003; Moreno *et al.*, 2004; Font *et al.*, 2007). O grau de variabilidade genética representa o potencial de um dado organismo em se adaptar ao ambiente. O entendimento da dinâmica da variabilidade de populações de vírus de plantas, tanto em hospedeiros cultivados como não-cultivados, é necessário para entender como as populações evoluem, bem como as implicações para a durabilidade de medidas de manejo (Seal *et al.*, 2006).

Alguns estudos já foram realizados para investigar a estrutura genética de populações de geminivírus em plantas cultivadas (Zhou *et al.*, 1997; Fondong *et al.*, 2000; Legg e Thresh, 2000; Pita *et al.*, 2001; Owor *et al.*, 2007). Com o recente advento de técnicas de amplificação do genoma viral completo ("rolling-circle amplification", RCA), novas possibilidades foram criadas para a análise de populações virais a nível genômico (Haible *et al.*, 2006). Tais estudos de "genômica de populações", envolvendo grandes conjuntos de genomas completos, elucidaram as origens e a emergência do mastrevírus

*Maize streak virus* (MSV) e do begomovírus *Tomato yellow leaf curl virus* (TYLCV) como patógenos (Varsani *et al.*, 2008; Harkins *et al.*, 2009; Lefeuvre *et al.*, 2010).

O Brasil é um centro de diversidade genética de begomovírus, com relatos de sua detecção desde a década de 1950 em malváceas e leguminosas (Costa e Bennett, 1950; Costa, 1955). Mais recentemente, um grande número de novas espécies de begomovírus infectando o tomateiro tem sido caracterizadas (Ribeiro *et al.*, 2003; Fernandes *et al.*, 2006; Calegario *et al.*, 2007; Ribeiro *et al.*, 2007; Castillo-Urquiza *et al.*, 2008; Fernandes *et al.*, 2008). Estes vírus emergiram após a introdução do biótipo B de *B. tabaci* em meados da década de 1990 (Ribeiro *et al.*, 1998). Trabalhos mais recentes utilizando a técnica de RCA descreveram diversas novas espécies infectando plantas cultivadas e não cultivadas (Fernandes *et al.*, 2009; Silva *et al.*, 2011; Albuquerque *et al.*, 2012; Silva *et al.*, 2012; Tavares *et al.*, 2012).

Populações de begomovírus infectando o tomateiro e plantas daninhas associadas no Brasil foram objeto de um estudo aprofundado que avaliou sua estrutura e variabilidade genética (Rocha *et al.*, 2013). Os resultados indicaram que as populações brasileiras são altamente recombinantes, possuem um elevado grau de variabilidade genética e uma rápida taxa de evolução molecular. Além disso, verificou-se que uma população associada a uma planta não-cultivada (*Blainvillea rhomboidea*) possui maior variabilidade genética em relação a populações associada a plantas cultivadas. Por fim, demonstrou-se que as populações são estruturadas com base em localização geográfica, o que explica a predominância de determinadas espécies virais nos diferentes estados brasileiros (Ribeiro *et al.*, 2003; Fernandes *et al.*, 2008). Os resultados desse estudo reforçam a hipótese de que os begomovírus encontrados em tomateiro são originados de vírus nativos presentes em plantas não-cultivadas, e que após a transferência para o tomateiro as populações virais evoluíram rapidamente, originando novas espécies mais adaptadas ao novo hospedeiro.

A análise comparativa de populações dos begomovírus *Macroptilium yellow spot virus* (MaYSV) e *Tomato severe rugose virus* (ToSRV), provenienentes de plantas não-cultivadas e cultivadas, respectivamente, também indicou maior variabilidade genética para a população provenienente de plantas não-cultivadas, e demonstrou que a recombinação, e não a seleção adaptativa, explica essa maior variabilidade (Lima *et al.*, 2013a).

Esses estudos foram os primeiros a analisar a estrutura genética e a dinâmica de populações de begomovírus em hospedeiros não-cultivados, embora seja aceito que estes hospedeiros atuem como fonte de inóculo, podendo contribuir para epidemias em hospedeiros cultivados (Idris *et al.*, 2003; Jovel *et al.*, 2004; Castillo-Urquiza *et al.*, 2008; Barreto *et al.*, 2013). Graças à disponibilidade de ferramentas para a rápida clonagem de um grande número de genomas virais a partir de amostras foliares dessecadas, é possível atualmente analisar de forma detalhada e comparativa populações dos mesmos begomovírus em diferentes hospedeiros (cultivados e não-cultivados). Dentre outras informações relevantes, essa análise pode identificar hospedeiros que atuem como reservatórios (incluindo hospedeiros assintomáticos) ou como "mixing vessels", ou seja, plantas que sejam hospedeiras de várias espécies virais favorecendo a ocorrência de infecções mistas e consequentes eventos de recombinação (de forma análoga aos suínos para o vírus da influenza A). Além disso, um entendimento da dinâmica das populações virais em plantas não-cultivadas pode auxiliar na previsão e consequente prevenção de novas viroses em plantas cultivadas.

Os estudos realizados no Brasil e em outros países indicam uma elevada diversidade de espécies de begomovírus em plantas não-cultivadas, principalmente em espécies de *Macroptilium* e *Sida* (Frischmuth *et al.*, 1997; Roye *et al.*, 1997; Idris *et al.*, 2003; Jovel *et al.*, 2004; Amarakoon *et al.*, 2008; Silva *et al.*, 2012; Tavares *et al.*, 2012).

Algumas dessas espécies virais também infectam plantas cultivadas como o feijoeiro e o tomateiro (Barreto *et al.*, 2013; Rocha *et al.*, 2013), reforçando o papel de plantas não-cultivadas como reservatórios de vírus. Uma característica dos trabalhos conduzidos no Brasil tem sido a amostragem em grandes áreas (centenas ou milhares de km$^2$). Seria interessante conduzir estudos em áreas menores, a fim de verificar se a diversidade de espécies também ocorre nessas condições. Isso também facilitaria estudos de coexistência de diferentes espécies e da evolução das populações virais ao longo do tempo. Este trabalho teve como objetivos caracterizar e estudar a variabilidade genética de populações de begomovírus infectando a planta não-cultivada *Sida acuta* em uma área de 10.000 m$^2$.

## LITERATURA CITADA

ALBUQUERQUE, L.C.; VARSANI, A.; FERNANDES, F.R.; PINHEIRO, B.; MARTIN, D.P.; OLIVEIRA FERREIRA, P.D.T.; LEMOS, T.O.; INOUE-NAGATA, A.K. Further characterization of tomato-infecting begomoviruses in Brazil. Archives of Virology, v. 157, p. 747-752, 2012.

AMARAKOON, I.I.; ROYE, M.E.; BRIDDON, R.W.; BEDFORD, I.D.; STANLEY, J. Molecular and biological characterization of *Macroptilium yellow mosaic virus* from Jamaica. Plant Pathology, v. 57, p. 417-426, 2008.

ANDRADE, E.C.; MANHANI, G.G.; ALFENAS, P.F.; CALEGARIO, R.F.; FONTES, E.P.B.; ZERBINI, F.M. *Tomato yellow spot virus*, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida* sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*. Journal of General Virology, v. 87, p. 3687-3696, 2006.

AWADALLA, P. The evolutionary genomics of pathogen recombination. Nature Reviews Genetics, v. 4, p. 50-60, 2003.

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

CALEGARIO, R.F.; FERREIRA, S.S.; ANDRADE, E.C.; ZERBINI, F.M. Characterization of *Tomato yellow spot virus*, (ToYSV), a novel tomato-infecting begomovirus from Brazil. Pesquisa Agropecuaria Brasileira, v. 42, p. 1335-1343, 2007.

CASTILLO-URQUIZA, G.P.; BESERRA JR., J.E.A.; BRUCKNER, F.P.; LIMA, A.T.M.; VARSANI, A.; ALFENAS-ZERBINI, P.; ZERBINI, F.M. Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Archives of Virology, v. 153, p. 1985-1989, 2008.

COSTA, A.S. Studies on *Abutilon* mosaic in Brazil. Phytopathologische Zeitschrift, v. 24, p. 97-112, 1955.

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

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

FERNANDES, F.R.; ALBUQUERQUE, L.C.; GIORDANO, L.B.; BOITEUX, L.S.; ÁVILA, A.C.; INOUE-NAGATA, A.K. Diversity and prevalence of Brazilian bipartite begomovirus species associated to tomatoes. Virus Genes, v. 36, p. 251-258, 2008.

FERNANDES, F.R.; CRUZ, A.R.R.; FARIA, J.C.; ZERBINI, F.M.; ARAGÃO, F.J.L. Three distinct begomoviruses associated with soybean in central Brazil. Archives of Virology, v. 154, p. 1567-1570, 2009.

FERNANDES, J.J.; CARVALHO, M.G.; ANDRADE, E.C.; BROMMONSCHENKEL, S.H.; FONTES, E.P.B.; ZERBINI, F.M. Biological and molecular properties of *Tomato rugose mosaic virus* (ToRMV), a new tomato-infecting begomovirus from Brazil. Plant Pathology, v. 55, p. 513-522, 2006.

FIALLO-OLIVE, E.; MARTINEZ-ZUBIAUR, Y.; MORIONES, E.; NAVAS-CASTILLO, J. A novel class of DNA satellites associated with New World begomoviruses. Virology, v. 426, p. 1-6, 2012.

FONDONG, V.N.; PITA, J.S.; REY, M.E.C.; KOCHKO, A.; BEACHY, R.N.; FAUQUET, C.M. Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. Journal of General Virology, v. 81, p. 287-297, 2000.

FONT, M.I.; RUBIO, L.; MARTINEZ-CULEBRAS, P.V.; JORDA, C. Genetic structure and evolution of natural populations of viruses causing the tomato yellow leaf curl disease in Spain. Virus Research, v. 128, p. 43-51, 2007.

FRISCHMUTH, T.; ENGEL, M.; LAUSTER, S.; JESKE, H. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, *Sida*-infecting bipartite geminiviruses in Central America. Journal of General Virology, v. 78, p. 2675-2682, 1997.

GARCÍA-ARENAL, F.; FRAILE, A.; MALPICA, J.M. Variation and evolution of plant virus populations. International Microbiology, v. 6, p. 225-232, 2003.

GE, L.M.; ZHANG, J.T.; ZHOU, X.P.; LI, H.Y. Genetic structure and population variability of tomato yellow leaf curl China virus. Journal of Virology, v. 81, p. 5902-5907, 2007.

HAIBLE, D.; KOBER, S.; JESKE, H. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. Journal of Virological Methods, v. 135, p. 9-16, 2006.

HARKINS, G.W.; DELPORT, W.; DUFFY, S.; WOOD, N.; MONJANE, A.L.; OWOR, B.E.; DONALDSON, L.; SAUMTALLY, S.; TRITON, G.; BRIDDON, R.W.; SHEPHERD, D.N.; RYBICKI, E.P.; MARTIN, D.P.; VARSANI, A. Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. Virology Journal, v. 6, p. 104, 2009.

IDRIS, A.M.; HIEBERT, E.; BIRD, J.; BROWN, J.K. Two newly described begomoviruses of *Macroptilium lathyroides* and common bean. Phytopathology, v. 93, p. 774-783, 2003.

JOVEL, J.; RESKI, G.; ROTHENSTEIN, D.; RINGEL, M.; FRISCHMUTH, T.; JESKE, H. *Sida micrantha* mosaic is associated with a complex infection of begomoviruses different from *Abutilon mosaic virus*. Archives of Virology, v. 149, p. 829-841, 2004.

LEFEUVRE, P.; MARTIN, D.P.; HARKINS, G.; LEMEY, P.; GRAY, A.J.A.; MEREDITH, S.; LAKAY, F.; MONJANE, A.; LETT, J.M.; VARSANI, A.; HEYDARNEJAD, J. The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathogens, v. 6, p. e1001164, 2010.

LEGG, J.P.; THRESH, J.M. Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. Virus Research, v. 71, p. 135-149, 2000.

LIMA, A.T.M.; SOBRINHO, R.R.; GONZALEZ-AGUILERA, J.; ROCHA, C.S.; SILVA, S.J.C.; XAVIER, C.A.D.; SILVA, F.N.; DUFFY, S.; ZERBINI, F.M. Synonymous site variation due to recombination explains the higher genetic variability in begomovirus populations infecting non-cultivated hosts. Journal of General Virology, v. 94, p. 418-431, 2013.

MANSOOR, S.; BRIDDON, R.W.; ZAFAR, Y.; STANLEY, J. Geminivirus disease complexes: an emerging threat. Trends in Plant Science, v. 8, p. 128-134, 2003.

MELGAREJO, T.A.; KON, T.; ROJAS, M.R.; PAZ-CARRASCO, L.; ZERBINI, F.M.; GILBERTSON, R.L. Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. Journal of Virology, v. 87, p. 5397-5413, 2013.

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

MONJANE, A.L.; PANDE, D.; LAKAY, F.; SHEPHERD, D.N.; VAN DER WALT, E.; LEFEUVRE, P.; LETT, J.M.; VARSANI, A.; RYBICKI, E.P.; MARTIN, D.P. Adaptive evolution by recombination is not associated with increased mutation rates in *Maize streak virus*. BMC Evolutionary Biology, v. 12, p. 252, 2012.

MORENO, I.M.; MALPICA, J.M.; DIAZ-PENDON, J.A.; MORIONES, E.; FRAILE, A.; GARCIA-ARENAL, F. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. Virology, v. 318, p. 451-460, 2004.

OWOR, B.E.; MARTIN, D.P.; SHEPHERD, D.N.; EDEMA, R.; MONJANE, A.L.; RYBICKI, E.P.; THOMSON, J.A.; VARSANI, A. Genetic analysis of *Maize streak virus* isolates from Uganda reveals widespread distribution of a recombinant variant. Journal of General Virology, v. 88, p. 3154-3165, 2007.

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

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

PINEL, A.; ABUBAKAR, Z.; TRAORE, O.; KONATE, G.; FARGETTE, D. Molecular epidemiology of the RNA satellite of *Rice yellow mottle virus* in Africa. Archives of Virology, v. 148, p. 1721-1733, 2003.

PITA, J.S.; FONDONG, V.N.; SANGARE, A.; OTIM-NAPE, G.W.; OGWAL, S.; FAUQUET, C.M. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. Journal of General Virology, v. 82, p. 655-665, 2001.

PRASANNA, H.C.; SINHA, D.P.; VERMA, A.; SINGH, M.; SINGH, B.; RAI, M.; MARTIN, D.P. The population genomics of begomoviruses: global scale population structure and gene flow. Virology Journal, v. 7, p. 220, 2010.

RIBEIRO, S.G.; AMBROZEVICIUS, L.P.; ÁVILA, A.C.; BEZERRA, I.C.; CALEGARIO, R.F.; FERNANDES, J.J.; LIMA, M.F.; MELLO, R.N.; ROCHA, H.; ZERBINI, F.M. Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. Archives of Virology, v. 148, p. 281-295, 2003.

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

RIBEIRO, S.G.; MARTIN, D.P.; LACORTE, C.; SIMÕES, I.C.; ORLANDINI, D.R.S.; INOUE-NAGATA, A.K. Molecular and biological characterization of *Tomato chlorotic mottle virus* suggests that recombination underlies the evolution and diversity of Brazilian tomato begomoviruses. Phytopathology, v. 97, p. 702-711, 2007.

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

ROJAS, M.R.; HAGEN, C.; LUCAS, W.J.; GILBERTSON, R.L. Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. Annual Review of Phytopathology, v. 43, p. 361-394, 2005.

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

ROYE, M.E.; MCLAUGHLIN, W.A.; NAKHLA, M.K.; MAXWELL, D.P. Genetic diversity among geminiviruses associated with the weed species *Sida* spp., *Macroptilium lathyroides,* and *Wissadula amplissima* from Jamaica. Plant Disease, v. 81, p. 1251-1258, 1997.

SANZ, A.I.; FRAILE, A.; GALLEGO, J.M.; MALPICA, J.M.; GARCÍA-ARENAL, F. Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. Journal of Molecular Evolution, v. 49, p. 672-681, 1999.

SATTAR, M.N.; KVARNHEDEN, A.; SAEED, M.; BRIDDON, R.W. Cotton leaf curl disease - an emerging threat to cotton production worldwide. Journal of General Virology, v. 94, p. 695-710, 2013.

SEAL, S.E.; JEGER, M.J.; VAN DEN BOSCH, F. Begomovirus evolution and disease management. Advances in Virus Research, v. 67, p. 297-316, 2006a.

SEAL, S.E.; VAN DEN BOSCH, F.; JEGER, M.J. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Critical Reviews in Plant Sciences, v. 25, p. 23-46, 2006b.

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

SILVA, S.J.C.; CASTILLO-URQUIZA, G.P.; HORA-JÚNIOR, B.T.; ASSUNÇÃO, I.P.; LIMA, G.S.A.; PIO-RIBEIRO, G.; MIZUBUTI, E.S.G.; ZERBINI, F.M. High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in northeastern Brazil. Archives of Virology, v. 156, p. 2205-2213, 2011.

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

VARSANI, A.; NAVAS-CASTILLO, J.; MORIONES, E.; HERNÁNDEZ-ZEPEDA, C.; IDRIS, A.; BROWN, J.K.; ZERBINI, F.M.; MARTIN, D.P. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. Archives of Virology, DOI 10.1007/s00705-014-2050-2, 2014.

VARSANI, A.; SHEPHERD, D.N.; MONJANE, A.L.; OWOR, B.E.; ERDMANN, J.B.; RYBICKI, E.P.; PETERSCHMITT, M.; BRIDDON, R.W.; MARKHAM, P.G.; OLUWAFEMI, S.; WINDRAM, O.P.; LEFEUVRE, P.; LETT, J.M.; MARTIN, D.P. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. Journal of General Virology, v. 89, p. 2063-2074, 2008.

ZHOU, X.; LIU, Y.; CALVERT, L.; MUNOZ, C.; OTIM-NAPE, G.W.; ROBINSON, D.J.; HARRISON, B.D. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. Journal of General Virology, v. 78, p. 2101-2111, 1997.

# CAPITULO 1

# TWO NOVEL BEGOMOVIRUS SPECIES FROM THE NEW WORLD WITH FEATURES RECALLING OLD WORLD BEGOMOVIRUSES

Godinho, M.T., Lima, A.T.M., Xavier, C.A.D., Zerbini, F.M. Two novel begomovirus species from the New World with features recalling Old World begomoviruses. Plant Pathology, *in preparation*.

# TWO NOVEL BEGOMOVIRUS SPECIES FROM THE NEW WORLD WITH FEATURES RECALLING OLD WORLD BEGOMOVIRUSES

M.T. Godinho, A.T.M. Lima, C.A.D. Xavier, F.M. Zerbini[#]

Dep. de Fitopatologia/BIOAGRO and National Research Institute for Plant-Pest Interactions, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

[#]Corresponding author: Francisco Murilo Zerbini

Phone: (+55-31) 3899-2935; Fax: (+55-31) 3899-2240; E-mail: zerbini@ufv.br

## Abstract

Begomoviruses (family *Geminiviridae*) have a circular, ssDNA genome encapsidated in twinned icosahedral particles. In Brazil, a number of begomoviruses have been described infecting weeds. Here, we describe two novel begomovirus species infecting *Sida acuta* plants collected from a small area (about 10,000 $m^2$) at Viçosa, state of Minas Gerais in December 2011. Total DNA was extracted from *S. acuta* samples and the viral genome was amplified by RCA, cloned and sequenced. A total of 12 full-length DNA-A component were obtained from four samples, and the ICTV-established 89% DNA-A identity threshold was used for taxonomic placement. This analysis indicated that the cloned components correspond to two novel species, for which the names *Sida* golden yellow mosaic virus and *Sida* yellow spot virus (SiGYMV and SiYSV, respectively) are proposed. The DNA-A components exhibited a highly divergent 5' half , including part of the intergenic region, the putative CP gene and an AV2-like ORF (present only in Old Word begomoviruses). The deduced amino acid sequence of the CP had very low identity with other begomoviruses, but the presence of conserved motifs in the CP and Rep coding regions, characteristic of OW begomoviruses, was detected. Although New World-like begomoviruses have been found in the OW, this is the first time that OW-like begomoviruses are found naturally in the NW.

## Introduction

The *Geminiviridae* family includes viruses whose genomes are composed of one or two molecules of circular, single-stranded DNA, encapsulated within one structural protein forming geminated icosahedral particles. The family includes the genera *Becurtovirus, Begomovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus and Turncurtovirus,* defined based on insect vector, host range, phylogenetic relations and genomic organization (Varsani *et al.*, 2014). Begomoviruses are transmitted by the whitefly *Bemisia tabaci* and infect dicotyledoneous plants.

Based on phylogenetic studies and genome features, begomoviruses have been broadly divided into two groups: Old World (Eastern hemisphere: Europe, Africa, Asia and Oceania) and New World (Western hemisphere: the Americas) (Rybicki, 1994; Padidam *et al.*, 1999; Paximadis *et al.*, 1999). Old World (OW) begomoviruses can be either mono- or bipartite, with a DNA-A (or a single component) containing six genes: AV1/V1 (coat protein, CP) and AV2/V2 in the virion-sense strand, and AC1/C1 (replication initiation protein, Rep), AC2/C2 (transcriptional activator, Trap), AC3/C3 (replication enhancer, Ren) and AC4/C4 (AC4/C4 protein) in the complementary-sense strand. The majority of species are monopartite and associated with satellite DNAs (Mansoor *et al.*, 2003).

New World (NW) begomoviruses have the same DNA-A organization except they lack the AV2/V2 gene, and all viruses except *Tomato leaf deformation virus* (Melgarejo *et al.*, 2013) are bipartite. The second component, denominated DNA-B, contains two genes encoding movement proteins: BV1 (nuclear shuttle protein, NSP) in the virion-sense strand and BC1 (movement protein, MP) in the complementary-sense strand (Rojas *et al.*, 2005).

The divergence of bipartite viruses is thought to have occurred before continental separation, due to the presence of bipartite viruses in both the OW and NW (Rojas *et al.*, 2005). All NW viruses lack the AV2 gene, and it has been proposed that they evolved from a common ancestor that had lost the AV2 gene after continental separation (Rybicki, 1994). However, the occurrence of *Corchorus yellow vein virus* (CoYVV) and *Corchorus golden mosaic virus* (CoGMV), bearing features similar to NW viruses in Vietnam, suggests that NW-like viruses were present in the OW prior to continental separation (Ha *et al.*, 2006; 2008).

In Brazil, a large number of indigenous begomoviruses associated with leguminous and solanaceous plants have been characterized. (Gilbertson *et al.*, 1991; Ribeiro *et al.*, 2003; Fernandes *et al.*, 2006; Calegario *et al.*, 2007; Ribeiro *et al.*, 2007; Castillo-Urquiza *et al.*, 2008; Fernandes *et al.*, 2008; Paprotka *et al.*, 2010; Silva *et al.*, 2012; Tavares *et al.*, 2012; Rocha *et al.*, 2013). The recent advent of techniques for the unbiased amplification of circular DNA genomes (specially rolling circle amplification, RCA; Inoue-Nagata *et al.*, 2004), created new possibilities for the discovery of novel begomoviruses, and also of divergent ssDNA viruses (Krenz *et al.*, 2012; Loconsole *et al.*, 2012). Many of these novel begomoviruses have been found in non-cultivated plants (Castillo-Urquiza *et al.*, 2008; Paprotka *et al.*, 2010; Fiallo-Olive *et al.*, 2012; Silva *et al.*, 2012; Tavares *et al.*, 2012), highlighting the need to investigate these plants as a reservoir of viral diversity and as a source of new viruses which may cause diseases in crops.

Here we describe two new begomovirus species isolated from the non-cultivated host *Sida acuta*. These two viruses have a highly divergent 5' half of the DNA-A, including the CP, with low identity with any other geminivirus. Furthermore, they show a genomic organization resembling OW begomoviruses, with the presence of an AV2-like ORF and conserved motifs present in OW begomoviruses.

## Materials and Methods

### Sample collection and cloning of begomovirus genomes

Foliar samples from *Sida acuta* plants (n=50) showing symptoms of yellow mosaic were collected in the municipality of Viçosa, state of Minas Gerais, Brazil, on December 2011. The sampling area had approximately 10,000 $m^2$ (1 ha) in size and no cultivated plants were around. Total DNA was extracted from fresh tissue or preserved (pressed and dried) samples as described by Doyle & Doyle (1987).

Full-length viral genomes were amplified by rolling-circle amplification, as described by Inoue-Nagata *et al.* (2004). Single genome-length fragments were excised with *Apa*I, *Bam*HI, *Cla*I, *Eco*RV, *Kpn*I, *Pst*I, *Sac*I or *Spe*I, and ligated into the pBLUESCRIPT-KS+ plasmid vector (Stratagene), previously cleaved with the same enzyme. Viral inserts were sequenced commercially (Macrogen Inc.) by primer walking. All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5'-TAATATT//AC-3').

PCR was carried out using the universal DNA-B primers CRC1 and 2040 (Rojas *et al.*, 1993), since the *Msp*I restriction pattern of the RCA product showed bands that were not predicted by simulation on the APE program (González-Aguilera *et al.*, 2012) (data not shown). The PCR reaction was carried out with 3 µl of DNA, 5 µl of 5× buffer, 2.5 µl of 25mM $MgCl_2$, 2 µl of dNTPs, 0.5 µl of each primer and 11.5 µl of $H_2O$, totaling 25 µl. The reaction consisted of 30 cycles at 94°C for 1 min, 52°C for 1 min and 72°C for 1 min with a final extension at 72°C for 10 min.

### Multiple sequence alignments and phylogenetic analysis

Sequences were initially analyzed with the BLAST*n* algorithm (Altschul *et al.*, 1990) to determine the viral species with which they shared greatest similarity. Multiple sequence alignments were prepared for the full-length DNA-A and for the CP and Rep coding sequences using MUSCLE (Edgar, 2004). Phylogenetic trees were constructed using Bayesian inference performed with MrBayes v. 3.0b4 (Ronquist & Huelsenbeck, 2003), with the model selected by MrModeltest v. 2.2 (Nylander, 2004) in the Akaike Information Criterion (AIC). The analyses were carried out running 20,000,000 generations and excluding the first 2,000,000 generations as burn-in. Trees were visualized using Fig Tree (http://tree.bio.ed.ac.uk/software/figtree/).

## Virus taxonomy

Pairwise comparisons from all DNA-A sequences generated, plus the others with greatest similarity, were performed with the program SDT (Muhire *et al.*, 2013) using the Muscle alignment option (Edgar, 2004). Demarcation of new viral species was based on a threshold of 89% nucleotide sequence identity for the full length DNA-A (Brown *et al.*, 2012). New virus species names were assigned using the guidelines proposed by the ICTV *Geminiviridae* Study Group (Fauquet *et al.*, 2008).

## Protein prediction analysis

The program Interpro (Quevillon *et al.*, 2005), available at the EMBOSS site (http://www.ebi.ac.uk/Tools/emboss/), was used to determine conserved structural domains from the observed AV1 (CP) and AV2-like ORFs.

## Recombination analysis

Recombination analysis was performed using the rdp, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods as implemented in Recombination Detection Program (RDP) version 3.44 (Martin *et al.*, 2010). Alignments were scanned with default settings for the different methods. Statistical significance was inferred by *P*-values lower than a Bonferroni-corrected cut-off of 0.05. Only recombination events detected by at least three of the analysis methods available in the program were considered reliable.

### Infectivity assay

To associate the virus with the disease an infectivity assay was conducted, despite the fact that no DNA-B was cloned. To construct the infectious clone, the viral insert (30 µg) in clone VIC39D-1P was excised from the vector with *Pst*I and then religated. The infectivity assay was conducted in *Sida acuta* plants (grown from seeds) using biolistics (Aragão *et al.*, 1996), with the pressure adjusted to 50 psi. Ten plants were inoculated with infectious clones of *Tomato yellow spot virus* (ToYSV; Andrade *et al.*, 2006) as a positive control. Six plants were bombarded with tungsten particles alone as a negative control. The plants were evaluated at 14 and 30 days post inoculation (dpi). Infection was assessed by visual observation of symptoms, and confirmed by RCA-RFLP.

## Results

### Viral detection and sequence comparisons

A total of 50 samples were collected, and 47 were preliminarily positive for the presence of a begomovirus based on the detection of a 2,600-bp band after digestion of the RCA products with restriction enzymes (data not shown). Twelve complete genomic components, cloned from four samples, all corresponding to a DNA-A component (Table 1) were analyzed in this work.

Pairwise comparisons were perfomed with the most closely related viruses (based on BLAST*n* analysis), plus two OW begomoviruses (*Ludwigia yellow vein virus* and

*Sauropus leaf curl virus*) that have a small portion of the genome similar to these new species and one OW begomovirus that has characteristics of a NW begomovirus (*Corchorus golden mosaic virus*, CoGMV) (Figure 1). The results indicated that the 12 cloned components can be divided into two clades, with within-clade identities of 93-99% and 96-99%, and between-clade identities of 85-87%. Clones in both clades have the highest sequence identity (76-77%) with *Sida yellow mosaic virus* (SiYMV). Therefore, these two clades correspond to two new begomovirus species, for which the names *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) are proposed (Table 1).

The genomic organization of both species is the same. All 12 clones have a region of approximately 1,100 nt with very low similarity to other begomoviruses. This region includes the CP gene (the most conserved gene in begomovirus genomes) and an AV2-like ORF, as well as part of the common region. The genomic organization of one representative SiYSV clone and the pairwise identities of different regions of the genome are shown in Figure 2. The deduced amino acid sequences of the CP and AV2-like proteins were further analyzed with the program Interpro. The analysis indicated the presence of a domain related to geminivirus coat proteins in the CP gene. No functional domains were predicted in the AV2-like protein.

### Detection of motifs related to OW begomoviruses

Strikingly, and despite the low similarity of the 5' half of the viral genome of both SiGYMV and SiYSV with other begomoviruses, conserved motifs that are characteristic of OW begomoviruses could be found in the deduced amino acid sequences of their proteins (Tables 2 and 3). These include the KVRRR motif in the N-terminal region of the CP (Ha *et al.*, 2008), present in both viruses (Table 2), and the iteron-related domain (IRD)

PKRFQI in the Rep protein with the corresponding iteron GGTAC in the common region (Arguello-Astorga & Ruiz-Medrano, 2001) detected in the SiGYMV isolates (Table 3). Furthermore, both viruses have an AV2-like ORF, which is present only in OW begomoviruses.

### Detection of interspecies recombination events

Analysis with the RDP 3 program detected a strongly supported interspecies recombination event in the SiGYMV isolate VIC43D_1S (lowest *p*-value $3.717\times10^{-36}$ with SiScan). The major (SiYSV-[VIC43D_5S]) and minor (SiGYMV-[VIC25D_1P]) parents have 99.6% and 98.2% identity, respectively, with VIC43D_1S. The portion that corresponds to VIC25D_1P is between nucleotide positions 122 to 1142 in the alignment and includes the entire CP and AV2-like genes as well as 16 nucleotides of the common region.

### SiGYMV and SiYSV are more related to NW begomoviruses in phylogenetic analysis

The Bayesian phylogenetic tree based on the full-length DNA-A divides the isolates into two distinct clades corresponding to the two species (data not shown). Phylogenetic analysis (DNA-A) including SiGYMV and SiYSV isolates plus eight NW begomoviruses and seven OW begomoviruses (Figure 3) separates the viruses into two major clades, with the SiGYMV and SiYSV isolates clustering with NW begomoviruses. The CP-based tree with the same data set (Figure 4) separated the SiGYMV and SiYSV isolates from both OW and NW begomoviruses, but in the Rep-based tree (Figure 4) both species clustered with NW begomoviruses, with SiYSV showing a slightly closer relationship with other NW begomoviruses compared with SiGYMV.

The intraspecific CP- and Rep-based trees (Figure 5) were not congruent, but the topological differences between them were due to the strongly supported recombination event near the CP gene of SiGYMV-[VIC43D_1S], mentioned above. Without this clone, the clustering of isolates in the CP and Rep trees is identical (data not shown), mimicking the clustering found in the full-length genome tree. All Bayesian trees showed strong support for all branches (posterior probabilities higher than 0.85).

### Detection of DNA-B

We found evidence that a DNA-B is present in all samples that originated the clones obtained in this work (Figure 6). However we were unable to clone any DNA-B, despite numerous attempts. Additional indirect evidence of the bipartite nature of both SiGYMV and SiYSV was obtained  in the infectivity assay, in which no plants bombarded with SiYSV-[VIC39D_1P] showed symptoms at either 14 or 30 dpi, with RCA-RFLP results also negative (data now shown).

## Discussion

Old World (OW) and New World (NW) begomovirus genomes have a number of distinguishing characteristics. Almost all indigenous NW begomoviruses are bipartite [a recent study by Melgarejo *et al.* (2013), reports the first monopartite NW begomovirus], whereas both bipartite and monopartite begomoviruses are present in the OW. Old World begomoviruses have and AV2 gene, which is absent in NW viruses. In turn, NW begomoviruses have an N-terminal (P/S)WRxMxGT motif in the CP that is absent in OW viruses (Harrison, 2002). Interestingly, two "NW-like" begomoviruses have been described in Vietnam, *Corchorus yellow vein virus* (CoYVV) and *Corchorus golden mosaic virus* (CoGMV). These viruses lack both the AV2 gene and the N-terminal CP motif, and are phylogenetically more closely related to NW than to OW begomoviruses (Ha *et al.*, 2006; 2008). However, to this date, no "OW-like" begomoviruses had ever been found in natural infections in the NW.

Here we describe two new begomovirus species from the NW with several features recalling OW begomoviruses: (*i*) the presence of an AV2-like ORF; (*ii*) the presence of the basic KVRRR motif in the CP; (*iii*) the absence of the (P/S)WRxMxGT motif in the CP; (*iv*) the presence of an iteron (GGTAC) and its respective IRD (PKRFQI) found in several OW begomoviruses but in only one NW begomovirus.

It is generally accepted that NW begomoviruses arose more recently than the OW viruses, diverging after continental separation (Rybicki, 1994). However, not only have NW-like viruses been found in the OW (CoYVV and CoGMV), but now also OW-like viruses have been found in the NW. Thus, it seems that the divergence of these two major clades predates continental separation. Further work is necessary to assess the distribution of SiGYMV and SiYSV, as well as the presence of additional OW-like viruses in other non-cultivated hosts.

Diversification of begomoviruses is greatly accelerated by frequent recombination (Padidam *et al.*, 1999; Lefeuvre *et al.*, 2009; Rocha *et al.*, 2013), particularly in the context of adaptation to new host species and vector biotypes (Padidam *et al.*, 1999; Sanz *et al.*, 2000; Berrie *et al.*, 2001; Monci *et al.*, 2002). Confirming previous reports of frequent interspecies recombination among Brazilian begomoviruses (Inoue-Nagata *et al.*, 2006; Ribeiro *et al.*, 2007; Rocha *et al.*, 2013), we found evidence of interspecies recombination involving SiGYMV and SiYSV.

The protein encoded by the AV2 gene is involved in symptom development, efficient viral movement and viral DNA accumulation (Padidam *et al.*, 1996; Rojas *et al.*, 2001). The AV2 gene overlaps the CP N-terminal motif present in OW viruses, suggesting that functions normally attributed to this CP N-terminal motif and to AV2 are encoded by the DNA-B component, which is present in all but one NW viruses. Curiously, we could not find a predicted structure for the AV2-like proteins from SiGYMV and SiYSV with the algorithms used. One explanation could be that this ORF lost its function over time since it is not needed in the presence of the DNA-B, and therefore mutations in this region would not impact virus fitness. In support of this hypothesis, we detected the presence of a DNA-B using degenerated primers in a PCR assay. Additionally, the pattern from RCA-RFLP indicates either (*i*) the presence of a second DNA-A in a mixed infection, (*ii*) the presence of a DNA satellite, or (*iii*) the presence of a DNA-B. Further attempts to clone a DNA-B are ongoing.

Besides their OW-like features, both SiGYMV and SiYSV have a highly divergent CP, with very low identity with other begomoviruses. This is unexpected considering that the CP is the most conserved gene among begomoviruses, being essential not only for particle formation but also for vector transmission (but not for replication or movement). Although their deduced amino acid sequences include a gemini-like coat protein domain, it is not unreasonable to suppose that these viruses could be defective in particle formation (and therefore also in vector transmission). If one of these hypothesis is shown to be true, it would bring up the even more interesting possibility that these viruses may be seed borne in *Sida acuta* or transmitted by a vector other than *Bemisia tabaci*.

## References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-10.

Andrade EC, Manhani GG, Alfenas PF, Calegario RF, Fontes EPB, Zerbini FM, 2006. *Tomato yellow spot virus*, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida* sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*. *Journal of General Virology* **87**, 3687-96.

Aragão FJL, Barros LMG, Brasileiro ACM, *et al.*, 1996. Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. *Theoretical and Applied Genetics* **93**, 142-50.

Arguello-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**, 1465-85.

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**, 53-8.

Brown JK, Fauquet CM, Briddon RW, Zerbini FM, Moriones E, Navas-Castillo J, 2012. Family *Geminiviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, eds. *Virus Taxonomy. 9th Report of the International Committee on Taxonomy of Viruses.* London, UK: Elsevier Academic Press, 351-73.

Calegario RF, Ferreira SS, Andrade EC, Zerbini FM, 2007. Characterization of *Tomato yellow spot virus*, (ToYSV), a novel tomato-infecting begomovirus from Brazil. *Pesquisa Agropecuaria Brasileira* **42**, 1335-43.

Castillo-Urquiza GP, Beserra Jr. JEA, Bruckner FP, *et al.*, 2008. Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. *Archives of Virology* **153**, 1985-9.

Doyle JJ, Doyle JL, 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-5.

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

Fernandes FR, Albuquerque LC, Giordano LB, Boiteux LS, Ávila AC, Inoue-Nagata AK, 2008. Diversity and prevalence of Brazilian bipartite begomovirus species associated to tomatoes. *Virus Genes* **36**, 251-8.

Fernandes JJ, Carvalho MG, Andrade EC, Brommonschenkel SH, Fontes EPB, Zerbini FM, 2006. Biological and molecular properties of *Tomato rugose mosaic virus* (ToRMV), a new tomato-infecting begomovirus from Brazil. *Plant Pathology* **55**, 513-22.

Fiallo-Olive E, Navas-Castillo J, Moriones E, Martinez-Zubiaur Y, 2012. Begomoviruses infecting weeds in Cuba: increased host range and a novel virus infecting *Sida rhombifolia*. *Archives of Virology* **157**, 141-6.

Gilbertson RL, Hidayat SH, Martinez RT, *et al.*, 1991. Diferentiation of bean-infecting geminiviruses by nucleic acid hybridization probes and aspects of bean golden mosaic in Brazil. *Plant Disease* **75**, 336-42.

González-Aguilera J, Tavares SS, Sobrinho RR, *et al.*, 2012. Genetic structure of a Brazilian population of the begomovirus *Tomato severe rugose virus* (ToSRV). *Tropical Plant Pathology* **37**, 346-53.

Ha C, Coombs S, Revill P, Harding R, Vu M, Dale J, 2006. Corchorus yellow vein virus, a New World geminivirus from the Old World. *Journal of General Virology* **87**, 997-1003.

Ha C, Coombs S, Revill P, Harding R, Vu M, Dale J, 2008. Molecular characterization of begomoviruses and DNA satellites from Vietnam: additional evidence that the New World geminiviruses were present in the Old World prior to continental separation. *Journal of General Virology* **89**, 312-26.

Harrison BD, 2002. Virus variation in relation to resistance breaking in plants. *Euphytica* **124**, 181-92.

Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T, 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *Journal of Virological Methods* **116**, 209-11.

Inoue-Nagata AK, Martin DP, Boiteux LS, Giordano LD, Bezerra IC, De Avila AC, 2006. New species emergence via recombination among isolates of the Brazilian tomato-infecting begomovirus complex. *Pesquisa Agropecuária Brasileira* **41**, 1329-32.

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.

Lefeuvre P, Lett JM, Varsani A, Martin DP, 2009. Widely conserved recombination patterns among single-stranded DNA viruses. *Journal of Virology* **83**, 2697-707.

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**, 162-72.

Mansoor S, Briddon RW, Zafar Y, Stanley J, 2003. Geminivirus disease complexes: an emerging threat. *Trends in Plant Science* **8**, 128-34.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P, 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-3.

Melgarejo TA, Kon T, Rojas MR, Paz-Carrasco L, Zerbini FM, Gilbertson RL, 2013. Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *Journal of Virology* **87**, 5397-413.

Monci F, Sanchez-Campos S, Navas-Castillo J, Moriones E, 2002. A natural recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* **303**, 317-26.

Muhire B, Martin DP, Brown JK, *et al.*, 2013. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). *Archives of Virology* **158**, 1411-24.

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

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

Padidam M, Sawyer S, Fauquet CM, 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-24.

Paprotka T, Boiteux LS, Fonseca MEN, *et al.*, 2010a. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Research* **149**, 224-33.

Paprotka T, Metzler V, Jeske H, 2010b. The first DNA 1-like alpha satellites in association with New World begomoviruses in natural infections. *Virology* **404**, 148-57.

Paximadis M, Idris AM, Torres-Jerez I, Villarreal A, Rey MEC, Brown JK, 1999. Characterization of tobacco geminiviruses in the Old and New world. *Archives of Virology* **144**, 703-17.

Quevillon E, Silventoinen V, Pillai S, *et al.*, 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**, W116-W20.

Ribeiro SG, Ambrozevicius LP, Ávila AC, *et al.*, 2003. Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Archives of Virology* **148**, 281-95.

Ribeiro SG, Martin DP, Lacorte C, Simões IC, Orlandini DRS, Inoue-Nagata AK, 2007. Molecular and biological characterization of *Tomato chlorotic mottle virus* suggests that recombination underlies the evolution and diversity of Brazilian tomato begomoviruses. *Phytopathology* **97**, 702-11.

Rocha CS, Castillo-Urquiza GP, Lima ATM, *et al.*, 2013. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *Journal of Virology* **87**, 5784-99.

Rojas A, Kvarnheden A, Marcenaro D, Valkonen JPT, 2005. Sequence characterization of *Tomato leaf curl Sinaloa virus* and *Tomato severe leaf curl virus*: phylogeny of New World begomoviruses and detection of recombination. *Archives of Virology* **150**, 1281-99.

Rojas MR, Gilbertson RL, Russell DR, Maxwell DP, 1993. Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* **77**, 340-7.

Rojas MR, Jiang H, Salati R, *et al.*, 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, tomato yellow leaf curl virus. *Virology* **291**, 110-25.

Ronquist F, Huelsenbeck JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.

Rybicki EP, 1994. A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Archives of Virology* **139**, 49-77.

Sanz AI, Fraile A, García-Arenal F, *et al.*, 2000. Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *Journal of General Virology* **81**, 1839-49.

Silva SJC, Castillo-Urquiza GP, Hora-Junior BT, *et al.*, 2012. Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathology* **61**, 457-67.

Tavares SS, Ramos-Sobrinho R, Gonzalez-Aguilera J, Lima GSA, Assunção IP, Zerbini FM, 2012. Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on *Sida* spp. *Planta Daninha* **30**, 305-15.

Varsani A, Navas-Castillo J, Moriones E, *et al.*, 2014. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Archives of Virology* **159**, in press, DOI 10.1007/s00705-014-2050-2.

**Table 1.** Clones generated in this work.

| **Sample** | **Clone** | **Size (nt)** | **Enzyme** | **Species** |
|---|---|---|---|---|
| 25D | VIC25D_1P | 2810 | *Pst*I | SiGYMV |
| 25D | VIC25D_2A | 2813 | *Apa*I | SiGYMV |
| 25D | VIC25D_2P | 2810 | *Pst*I | SiGYMV |
| 43D | VIC43D_1S | 2823 | *Spe*I | SiGYMV |
| 25D | VIC25D_1S | 2826 | *Spe*I | SiYSV |
| 26D | VIC26D_3S | 2829 | *Spe*I | SiYSV |
| 26D | VIC26D_4S | 2828 | *Spe*I | SiYSV |
| 39D | VIC39D_1C | 2828 | *Cla*I | SiYSV |
| 39D | VIC39D_1P | 2828 | *Pst*I | SiYSV |
| 39D | VIC39D_1S | 2828 | *Spe*I | SiYSV |
| 39D | VIC39D_5S | 2828 | *Spe*I | SiYSV |
| 43D | VIC43D_5S | 2828 | *Spe*I | SiYSV |

**Table 2.** Comparison of deduced amino acid sequences from the N-terminal coat protein of *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) isolates determined in this study, with a number of begomoviruses from the Old and New Worlds.

| Virus | N-terminal CP amino acid sequence | Type | Origin |
|---|---|---|---|
| SiGYMV-[VIC25D_1P] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiGYMV-[VIC25D_2A] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiGYMV-[VIC25D_2P] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiGYMV-[VIC43D_1S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC25D_1S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC26D_3S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC26D_4S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC39D_1C] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC39D_1P] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC39D_1S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC39D_5S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| SiYSV-[VIC43D_5S] | MDYRKRKRSYLTPLTPDQIAKRFRQIQSGIGVRRRLSYPYVPQGIPRLPVKSR | OW | NW (this study) |
| *Corchorus yellow vein mosaic virus* (KC223600.1) | MPKRPLDMSISTPGWKVRRRLNFDSPYSARAAVPIVPVISKSR | OW | OW |
| *Ludwigia yellow vein virus* (NC_007210.2) | MSKRPADIVISTPVSKVRRRLNFDSPGMNRAAARTVLG | OW | OW |
| *Tomato yellow leaf curl virus* (AJ489258.1) | MSKRPGDIIISTPVSKVRRRLNFDSPYSSRAAVPIVQGTNKRRSW | OW | OW |
| *African cassava mosaic virus* (HE814065.1) | MSKRPGDIIISTPVSKVRRRLNFDSPYSSRAAVPIVQGTNKRRSW | OW | OW |
| *Indian cassava mosaic virus* (NC_001932.1) | MSKRPGDIIISTPGSKVRRRLNFDSPYRNRATAPTVHVTNRKRAW | OW | OW |
| *Loofa yellow mosaic virus* (NC_004824.1) | MSKRPADIIISTPGSKVRRRLNFDSPYSSRAAVPTVRVTKRQSW | OW | OW |
| *Squash leaf curl China virus* (AY184487.3) | MSKRPADIIISTPASKVRRRLNFDSPYVSRAVVPIARVTKGKAW | OW | OW |
| *Corchorus yellow vein virus* (NC_006358.1) | MPKRDAPWRLMAGTSKVSRSSNYSPRGGVSDSGSYLPRRFSRASL | NW | OW (Vietnam) |
| *Sida mottle virus* (AY090555.1) | MPKRDPSWRQMAGTSKVSRSSNFSPRGGIGPKFNKASEW | NW | NW |
| *Cabbage leaf curl virus* (NC_003866.1) | MPKRDAPWRSMAGTSKVSRNANYSPRAGMIHKFDKAAAW | NW | NW |
| *Squash leaf curl virus* (JX131281.1) | MVKRDAPWRLMAGTSKVSRSANFSPREGMGPKFNKAAAW | NW | NW |
| *Tomato golden mottle virus* (NC_008058.1) | MPKRDAPWRLMGGTSKVSRSFNQVSRTGTGPKFDKAHAW | NW | NW |
| *Macroptilium mosaic Puerto Rico virus* (NC_004097.1) | MPKRDAPWRSSAGTSKVSRNLNYSPGGGPKSNRANAW | NW | NW |
| *Rhynchosia golden mosaic virus* (NC_010294.1) | MPKRDAPWRLSAGTSKVSRSANYSPGGGMGPKSNRANAW | NW | NW |

**Table 3.** Comparison of deduced amino acid sequences of the N-terminal Rep protein of *Sida* golden yellow mosaic virus (SiGYMV) isolates determined in this study, with begomoviruses from the Old and New Worlds.

| **Virus** | **N-terminal Rep amino acid sequence** | **Iteron** | **Origin** |
|---|---|---|---|
| SiGYMV-[VIC25D_1P] | MPPPKRFQINAKNYFLTY | GGTAC | NW (this study) |
| SiGYMV-[VIC25D_2A] | MPPPKRFQINAKNYFLTY | GGTAC | NW (this study) |
| SiGYMV-[VIC25D_2P] | MPPPKRFQINAKNYFLTY | GGTAC | NW (this study) |
| SiGYMV-[VIC43D_1S] | MPPPKRFQINAKNYFLTY | GGTAC | NW (this study) |
| AYVV | MAPPKRFQINAKNYFLTY | GGTAC | OW |
| ICMV | MSPPKRFQINAKNYFLTY | GGTAC | OW |
| BGMV | MPPXPKRFKINAKNYFLT | GGAG | NW |
| ABMV | MQEWADDYFGSGSAARPD | GGAG | NW |

## Figure legends

**Figure 1.** Pairwise sequence identities of the full-length DNA-A of *Sida* golden mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) isolates with the most closely related begomoviruses, plus *Corchorus golden mosaic virus*, a "NW-like" OW begomovirus, and two Old World begomoviruses (*Ludwigia yellow vein virus* and *Sauropus leaf curl virus*) that have a small region of similarity (E-value 1.2) in the AV2-like ORF.

**Figure 2.** Genome organization of *Sida* yellow spot virus (SiYSV) and pairwise nucleotide sequence identities of the virion- and complementary-sense genes with the most closely related begomoviruses.

**Figure 3.** Bayesian phylogenetic tree for the full length DNA-A of *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) isolates (indicated in green), plus a number of NW (blue) and OW (red) begomoviruses.

**Figure 4.** Bayesian phylogenetic trees for the CP and Rep deduced amino acid sequences of *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) isolates (indicated in green), plus a number of NW (blue) and OW (red) begomoviruses.

**Figure 5.** Bayesian phylogenetic trees for the CP and Rep deduced amino acid sequences of *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) isolates. The recombinant isolate VIC43D_1S is indicated in green.

**Figure 6.** PCR-based detection of a DNA-B component in the *Sida acuta* samples from which the DNA-A components of *Sida* golden yellow mosaic virus (SiGYMV) and *Sida* yellow spot virus (SiYSV) were cloned. **M,** Size marker (1 kb plus DNA ladder, in base pairs); **25D, 26D, 39D, 43D,** *Sida acuta* samples; **+**, positive control (DNA-B from *Tomato yellow spot virus*); **-**, negative control (no DNA). The 500 bp band corresponding to a fragment of the DNA-B is indicated.

**Figure 1**

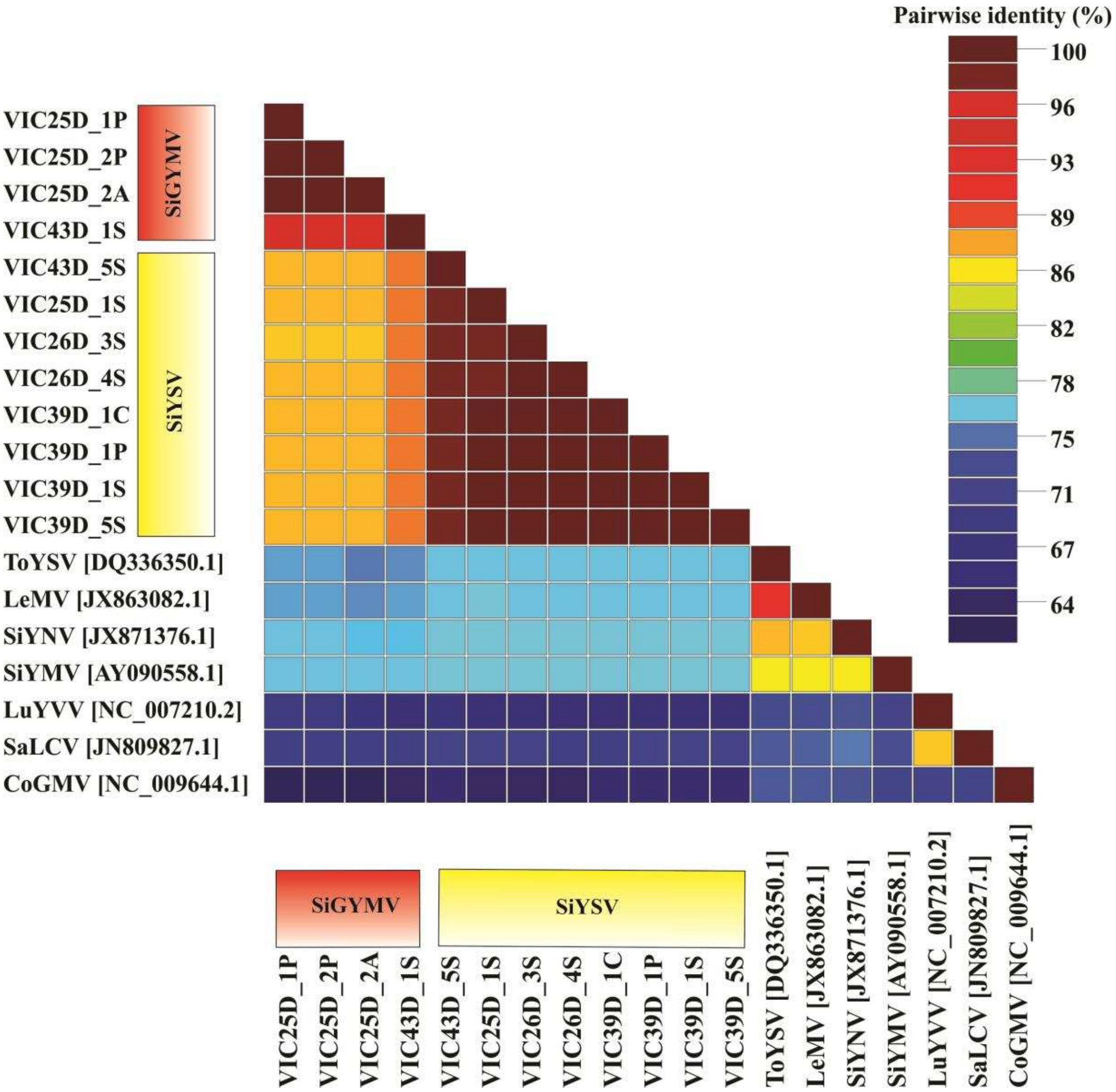

Pairwise identity (%)
100
96
93
89
86
82
78
75
71
67
64
SiGYMV
SiYSV
VIC25D_1P
VIC25D_2P
VIC25D_2A
VIC43D_1S
VIC43D_5S
VIC25D_1S
VIC26D_3S
VIC26D_4S
VIC39D_1C
VIC39D_1P
VIC39D_1S
VIC39D_5S
ToYSV [DQ336350.1]
LeMV [JX863082.1]
SiYNV [JX871376.1]
SiYMV [AY090558.1]
LuYVV [NC_007210.2]
SaLCV [JN809827.1]
CoGMV [NC_009644.1]

**Figure 2**

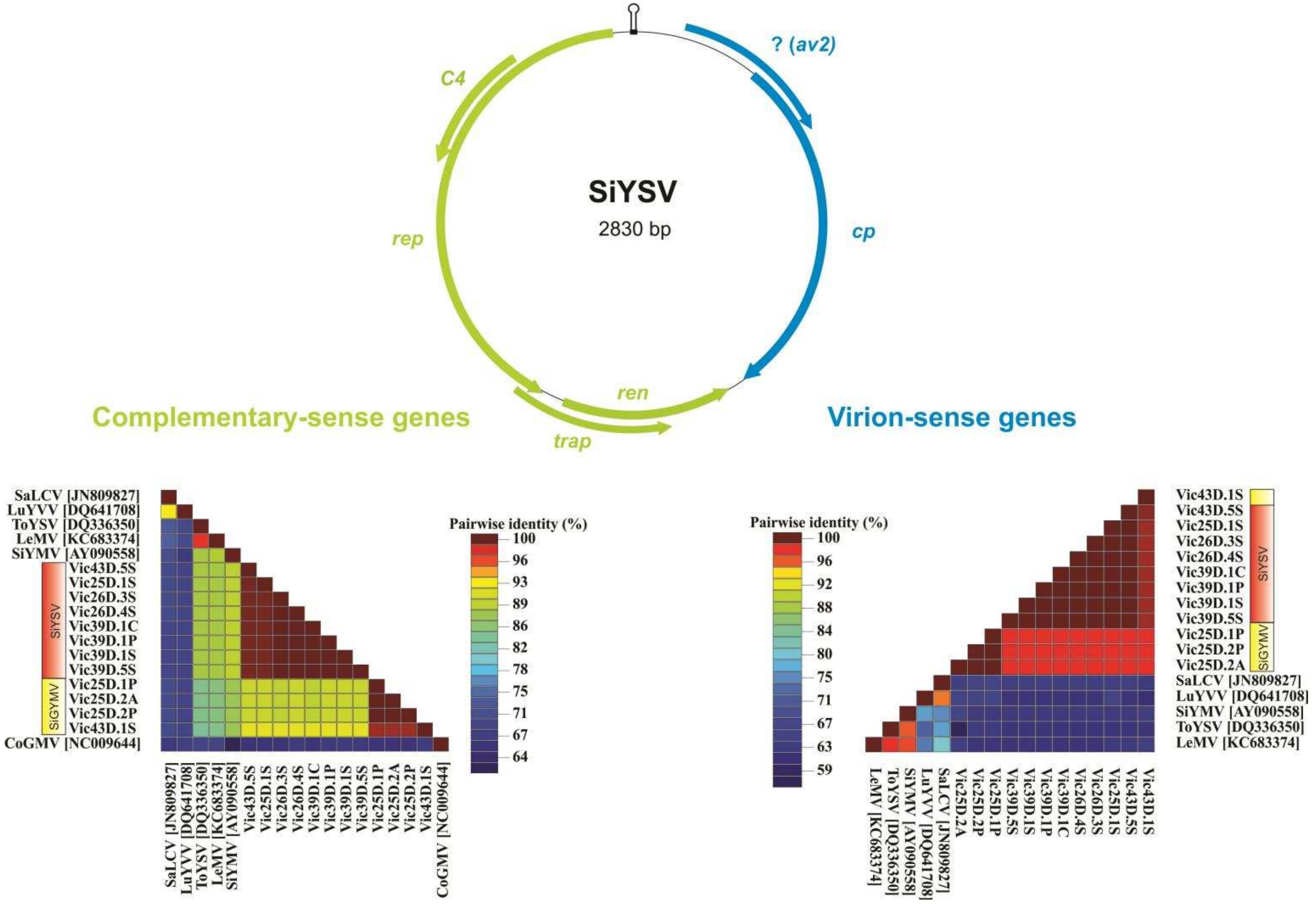

SiYSV
2830 bp
? (av2)
C4
rep
cp
ren
trap
Complementary-sense genes
Virion-sense genes
Pairwise identity (%)
100
96
93
89
86
82
78
75
71
67
64
SaLCV [JN809827]
LuYVV [DQ641708]
ToYSV [DQ336350]
LeMV [KC683374]
SiYMV [AY090558]
Vic43D.5S
Vic25D.1S
Vic26D.3S
Vic26D.4S
Vic39D.1C
Vic39D.1P
Vic39D.1S
Vic39D.5S
Vic25D.1P
Vic25D.2A
Vic25D.2P
Vic43D.1S
CoGMV [NC009644]
SiYSV
SiGYMV
Pairwise identity (%)
100
96
92
88
84
80
75
71
67
63
59
Vic43D.1S
Vic43D.5S
Vic25D.1S
Vic26D.3S
Vic26D.4S
Vic39D.1C
Vic39D.1P
Vic39D.1S
Vic39D.5S
Vic25D.1P
Vic25D.2P
Vic25D.2A
SaLCV [JN809827]
LuYVV [DQ641708]
SiYMV [AY090558]
ToYSV [DQ336350]
LeMV [KC683374]
SiYSV
SiGYMV

Figure 3

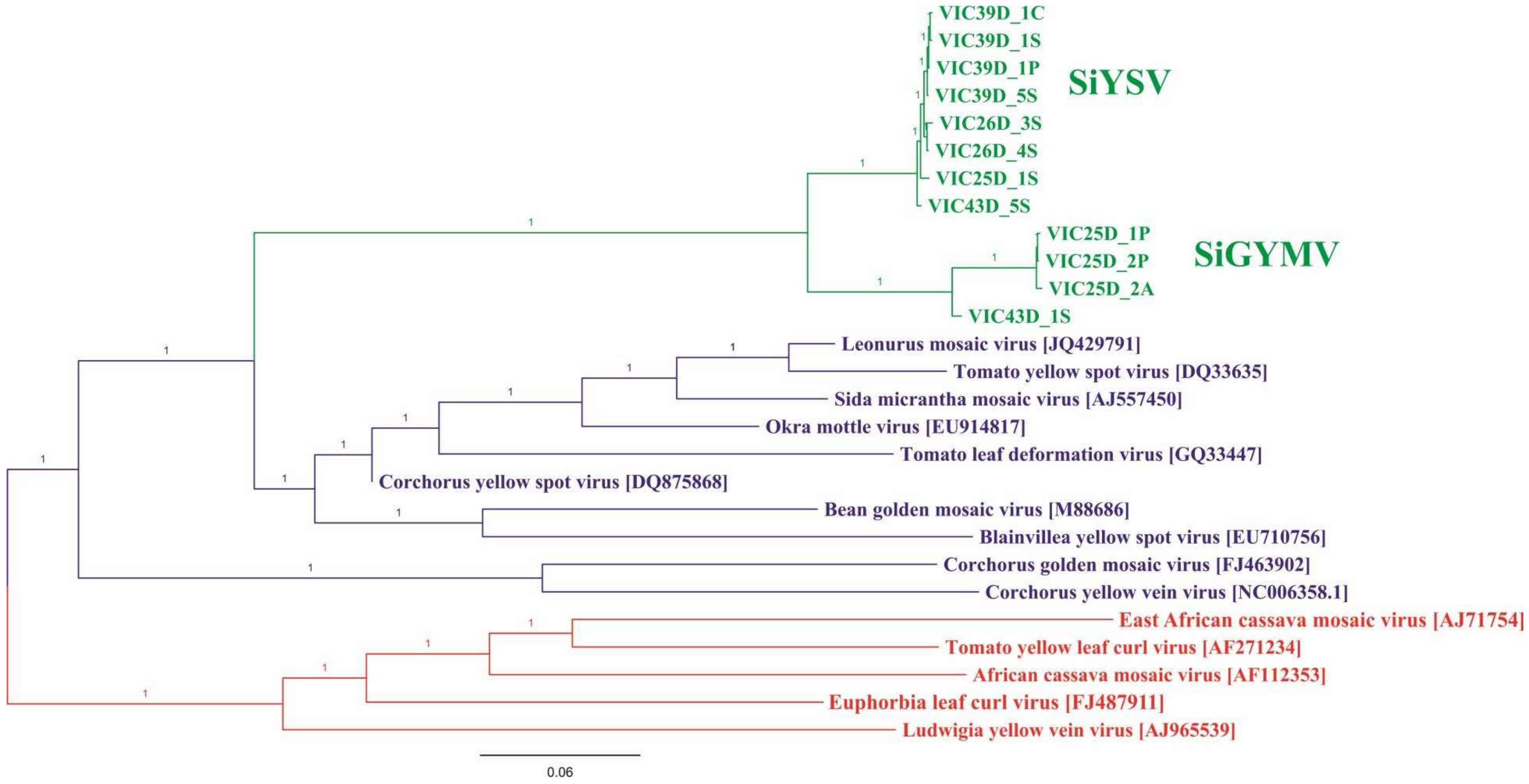

VIC39D_1C
VIC39D_1S
VIC39D_1P
VIC39D_5S
VIC26D_3S
VIC26D_4S
VIC25D_1S
VIC43D_5S
SiYSV
VIC25D_1P
VIC25D_2P
VIC25D_2A
VIC43D_1S
SiGYMV
Leonurus mosaic virus [JQ429791]
Tomato yellow spot virus [DQ33635]
Sida micrantha mosaic virus [AJ557450]
Okra mottle virus [EU914817]
Tomato leaf deformation virus [GQ33447]
Corchorus yellow spot virus [DQ875868]
Bean golden mosaic virus [M88686]
Blainvillea yellow spot virus [EU710756]
Corchorus golden mosaic virus [FJ463902]
Corchorus yellow vein virus [NC006358.1]
East African cassava mosaic virus [AJ71754]
Tomato yellow leaf curl virus [AF271234]
African cassava mosaic virus [AF112353]
Euphorbia leaf curl virus [FJ487911]
Ludwigia yellow vein virus [AJ965539]
0.06

**Figure 4**

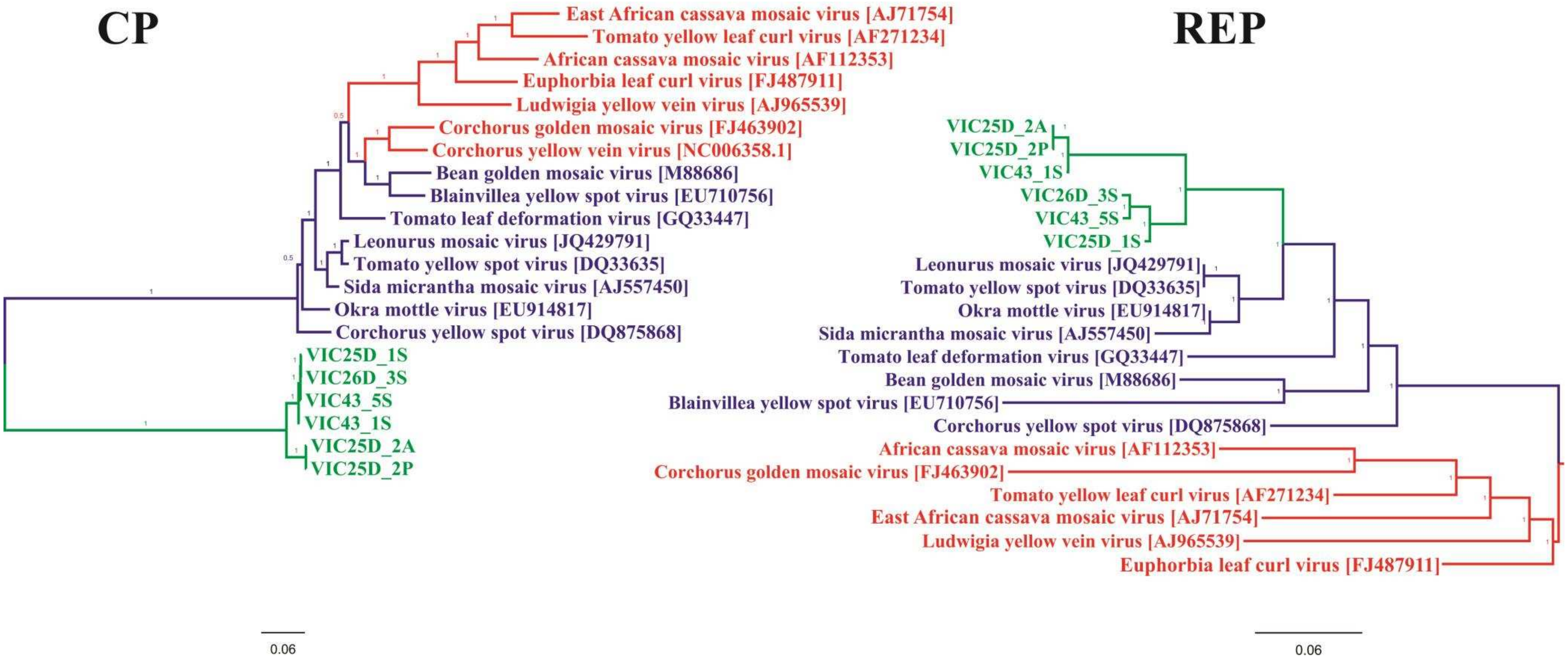

CP
East African cassava mosaic virus [AJ71754]
Tomato yellow leaf curl virus [AF271234]
African cassava mosaic virus [AF112353]
Euphorbia leaf curl virus [FJ487911]
Ludwigia yellow vein virus [AJ965539]
Corchorus golden mosaic virus [FJ463902]
Corchorus yellow vein virus [NC006358.1]
Bean golden mosaic virus [M88686]
Blainvillea yellow spot virus [EU710756]
Tomato leaf deformation virus [GQ33447]
Leonurus mosaic virus [JQ429791]
Tomato yellow spot virus [DQ33635]
Sida micrantha mosaic virus [AJ557450]
Okra mottle virus [EU914817]
Corchorus yellow spot virus [DQ875868]
VIC25D_1S
VIC26D_3S
VIC43_5S
VIC43_1S
VIC25D_2A
VIC25D_2P
0.5
0.5
0.06
REP
VIC25D_2A
VIC25D_2P
VIC43_1S
VIC26D_3S
VIC43_5S
VIC25D_1S
Leonurus mosaic virus [JQ429791]
Tomato yellow spot virus [DQ33635]
Okra mottle virus [EU914817]
Sida micrantha mosaic virus [AJ557450]
Tomato leaf deformation virus [GQ33447]
Bean golden mosaic virus [M88686]
Blainvillea yellow spot virus [EU710756]
Corchorus yellow spot virus [DQ875868]
African cassava mosaic virus [AF112353]
Corchorus golden mosaic virus [FJ463902]
Tomato yellow leaf curl virus [AF271234]
East African cassava mosaic virus [AJ71754]
Ludwigia yellow vein virus [AJ965539]
Euphorbia leaf curl virus [FJ487911]
0.06

Figure 5

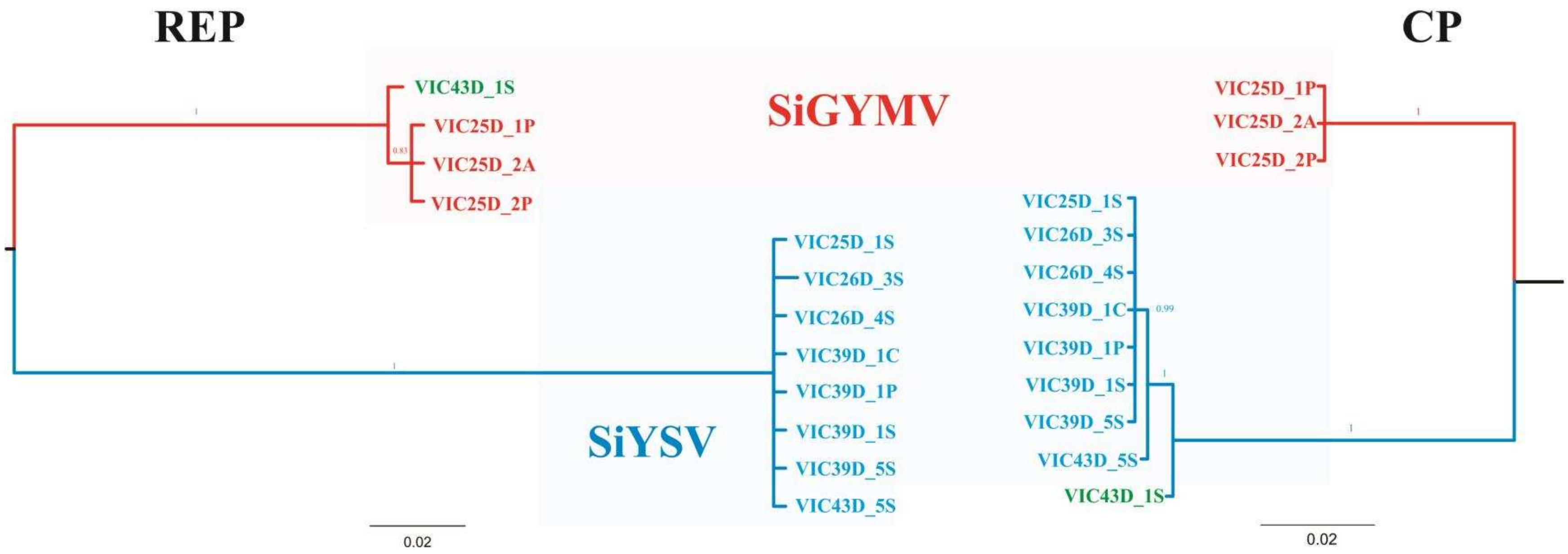
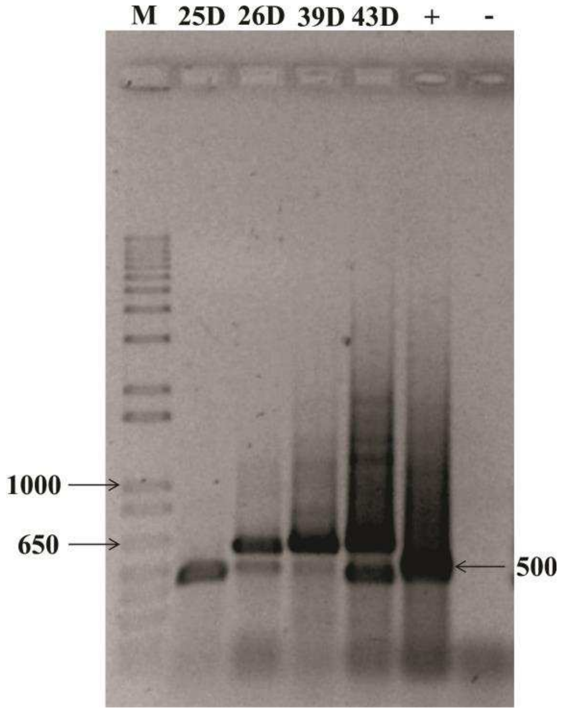

REP
CP
SiGYMV
SiYSV
VIC43D_1S
VIC25D_1P
VIC25D_2A
VIC25D_2P
1
0.83
VIC25D_1S
VIC26D_3S
VIC26D_4S
VIC39D_1C
VIC39D_1P
VIC39D_1S
VIC39D_5S
VIC43D_5S
1
0.02
VIC25D_1P
VIC25D_2A
VIC25D_2P
1
VIC25D_1S
VIC26D_3S
VIC26D_4S
VIC39D_1C
VIC39D_1P
VIC39D_1S
VIC39D_5S
VIC43D_5S
VIC43D_1S
0.99
1
1
0.02

**Figure 6**



# CAPÍTULO 2

# A BEGOMOVIRUS EXISTING AS A COMPLEX OF WELL DEFINED SUBPOPULATIONS IN A NON-CULTIVATED HOST

Godinho, M.T., Lima, A.T.M., Xavier, C.A.D., Zerbini, F.M. A begomovirus existing as a complex of well defined subpopulations in a non-cultivated host. Virology, *in preparation.*

# A BEGOMOVIRUS EXISTING AS A COMPLEX OF WELL DEFINED SUBPOPULATIONS IN A NON-CULTIVATED HOST

M.T. Godinho, A.T.M. Lima, C.A.D. Xavier, F.M. Zerbini[#]

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

[#]Corresponding author: Francisco Murilo Zerbini

Phone: (+55-31) 3899-2935; Fax: (+55-31) 3899-2240; E-mail: zerbini@ufv.br

## Abstract

DNA plant viruses, particularly the begomoviruses, cause serious epidemics in economically important crops worldwide. In Brazil several indigenous begomoviruses have been described infecting tomatoes following the introduction of a novel biotype of the whitefly vector in the mid-1990s. Non-cultivated plants harbor many begomoviruses, and it is believed that these hosts may act as reservoirs and as mixing vessels where recombination may occur. Begomoviruses also display nucleotide substitution rates equivalent to those of RNA viruses. In this work we sampled *Sida acuta* plants showing typical viral symptoms in a small area (aprox. 1 ha) in the municipality of Viçosa, MG, Brazil. Total DNA was extracted from fifty samples and the viral genome was amplified by RCA, cloned and sequenced. A total of 65 full-length genomes (33 DNA-A and 32 DNA-B components, from 26 samples) were obtained and the 89% DNA-A identity threshold established by the ICTV was used for taxonomic placements. Sequence analysis indicated that the clones correspond to a novel species for which the name *Sida acuta* mosaic virus (SAMV) is proposed. Additionally, the analyses indicated the coexistence of three well-defined SAMV strains, with mixed infections and pseudorecombination among them. We reconstructed the phylogenetic relationships for full-length genomic components, CP and Rep genes using Bayesian inference (BI). Well-supported clades (posterior probabilities higher than 0.99) were observed in all phylogenetic trees representing each distinct SAMV strain. Our results indicate a complex evolutionary interplay amongst begomovirus isolates even in small populations.

## Introduction

Geminivirus populations, including the whitefly-transmitted begomoviruses, possess a high genetic variability (Ge *et al.*, 2007; Prasanna *et al.*, 2010; Rocha *et al.*, 2013). The occurrence of frequent recombination (Padidam *et al.*, 1999), of pseudorecombination events between viruses with bipartite genomes (Andrade *et al.*, 2006) and high molecular evolution rates (Duffy and Holmes, 2008) all contribute to the high variability observed. Mutation frequencies and rates of nucleotide substitution observed for geminiviruses are similar to those observed for RNA viruses, despite the expectation that they would be lower due to the use of the host's replication machinery, which in theory would allow for the correction of misincorporated nucleotides, increasing the fidelity of replication (Duffy and Holmes, 2008).

Brazil is a center of genetic diversity of begomoviruses, with reports of their detection dating back to the 1950's (Costa, 1955; Costa and Bennett, 1950). More recently, a large number tomato-infecting begomoviruses have been characterized (Calegario *et al.*, 2007; Castillo-Urquiza *et al.*, 2008; Fernandes *et al.*, 2008; Fernandes *et al.*, 2006; Ribeiro *et al.*, 2003; Ribeiro *et al.*, 2007). These viruses, which are indigenous to Brazil, emerged after the introduction of the B biotype of *Bemisia tabaci* in the mid-1990s (Ribeiro *et al.*, 1998). More recent work using the non-sequence-biased rolling-circle amplification technique described several new species infecting cultivated and non-cultivated hosts (Albuquerque *et al.*, 2012b; Fernandes *et al.*, 2009; Silva *et al.*, 2012; Silva *et al.*, 2011; Tavares *et al.*, 2012). These reports suggest that the genotypic diversity and genetic variability of begomoviruses is even higher than previously believed. The fact that many of these reports have been made from the analysis of non-cultivated plants (Paprotka *et al.*, 2010; Silva *et al.*, 2012; Tavares *et al.*, 2012) highlights the need to investigate these plants as a reservoir of viral diversity and as a source of new viruses which may cause diseases in crops.

The key evolutionary aspects of DNA and RNA viruses are still only partly understood (Pagan *et al.*, 2010). This is of both academic and practical importance, as virus evolution may compromise disease control strategies, including the rapid generation of genotypes that are able to evade host immune responses, that are resistant to antivirals, or that can overcome crop genetic resistance (García-Arenal *et al.*, 2003; Holmes, 2009; McDonald and Linde, 2002; Pagan *et al.*, 2010).

Most of our knowledge of the rapidity of virus evolution comes from the study of animal viruses, for which estimates of nucleotide substitution rates normally fall within an order of magnitude of $1\times10^{-3}$ nucleotide substitutions per site per year (subs/site/year) and largely reflect the background mutation rate (Drake, 1991; Duffy *et al.*, 2008; Pagan *et al.*, 2010; Ruiz-Ferrer *et al.*, 2004). According to Pagan *et al.* (2010), there is a growing body of data on intraspecific evolutionary processes in plant RNA viruses, including nucleotide substitution rates, however there has been a general neglect of long term evolutionary patterns, including the determinants of viral speciation.

Most data on virus evolution are related to RNA viruses, mainly due to their elevated rate of evolution and their importance as human and crop pathogens. In the context of RNA viruses, allopatric speciation can be thought of as the genetic diversification that occurs when viruses jump to new host species and thereafter evolve independently, as is commonly associated with the process of "viral emergence". In contrast, sympatric speciation occurs when viruses diversify within a single host species, perhaps by exploiting different cell types (Holmes, 2009).

The vast majority of works assessing the genetic variability and the molecular evolution of begomoviruses are based on large sampling areas (Castillo-Urquiza *et al.*, 2008; Lima *et al.*, 2013; Rocha *et al.*, 2013), assuming that the larger the sampling area, the greater the chance of accessing the real variability. In these studies, a small number of samples (sometimes a single sample) are collected at a given location. It would be relevant to conduct similar studies in which a larger number of samples are collected in a small area, to verify whether the high genetic varialibility found over large areas is reflected in a smaller environment.

Here, we investigated the characteristics of a population of *Sida acuta* mosaic virus (SAMV, a new species in the genus *Begomovirus*), infecting the non-cultivated host *Sida acuta* and sampled in an area of aprox. 1 ha. Even in such small area the population is comprised of three strains, with cases of mixed infections and pseudorecombination. Thus, begomoviruses in non-cultivated hosts display a high degree of genetic variability even in small areas.

## Materials and methods

### Sample collection, cloning and sequencing of viral genomes

The dataset comprised isolates obtained from *Sida acuta* plants showing symptoms of begomovirus infection, collected at a single site in the municipality of Viçosa, state of Minas Gerais (MG), in December 2011. The sampling area has approximately 10,000 $m^2$ (1 ha).

Total DNA was extracted from fresh tissue or herbarium-like (pressed and dried) samples as described by Doyle and Doyle (1987). Viral genomes were amplified by rolling-circle amplification using the bacteriophage phi29 DNA polymerase (New England Biolabs) according to the method described by Inoue-Nagata *et al.* (2004). Aliquots of amplification reactions were subjected to cleavage with eight restriction enzymes, and the products were analyzed on 0.7% agarose gels stained with ethidium bromide. Fragments of

2,600 nucleotides (nt), corresponding approximately to the size of the genomic components of bipartite begomoviruses, were cloned in the pBluescript KS+ vector (Stratagene) and sequenced commercially (Macrogen Inc.) by primer walking. All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5'-TAATATT//AC-3').

### Multiple sequence alignments and phylogenetic analysis

Sequences were initially analyzed with the BLAST*n* algorithm (Altschul *et al.*, 1990) to determine the viral species with which they shared the greatest similarity. Multiple sequence alignments were prepared for the full-length DNA-A, DNA-B and for the CP and Rep coding sequences using MUSCLE (Edgar, 2004). Phylogenetic trees were constructed using Bayesian inference with MrBayes v. 3.0b4 (Ronquist and Huelsenbeck, 2003), with the model selected by MrModeltest v. 2.2 (Nylander, 2004) by the Akaike Information Criterion (AIC). The analyses were carried out running 20,000,000 generations and excluding the first 2,000,000 generations as burn-in. Trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

### Virus taxonomy

Demarcation of viral species was based on DNA-A sequence comparisons, using the 89% identity threshold as determined by the *Geminiviridae* Study Group of the ICTV (Brown *et al.*, 2012). Pairwise comparisons were performed with the program SDT v. 1.0 (Muhire *et al.*, 2013), using the MUSCLE algorithm. The dataset included all DNA-A sequences obtained, plus the ones from the begomoviruses with greatest similarity based on the BLAST*n* analysis.

## Variability indices

The main descriptors of molecular variability were estimated for each specie/strain: total number of mutations (Eta), nucleotide diversity ($\pi$), mutation frequency, number of haplotypes (h), haplotype diversity (Hd). Variability indices were calculated using DnaSP v. 5.10 (Rozas *et al.*, 2003).

## Analysis of recombination

Recombination analysis was performed using the rdp, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods as implemented in Recombination Detection Program (RDP) version 3.44 (Martin *et al.*, 2010). Each of the two datasets (DNA-A and -B) comprised all SAMV sequences generated for this work. Alignments were scanned with default settings for the different methods. Statistical significance was inferred by *P*-values lower than a Bonferroni-corrected cut-off of 0.05. Only recombination events detected by at least three of the methods available in the program were considered reliable.

## Detection of positive and negative selection at amino acid sites

To detect sites under positive and negative selection, we analyzed all ORFs datasets using three maximum likelihood-based methods implemented in the DataMonkey server (www.datamonkey.org): SLAC, REL and PARRIS (Pond and Frost, 2005; Scheffler *et al.*, 2006). As recombinant sequences causes spurious selection results, we searched for breakpoints in sequences implicated as recombinant by GARD prior to running these analyses. Additionally, PARRIS allowed synonymous rate variation, topology and branch lengths to vary across recombination breakpoints. The SLAC method was also used to estimate the mean ratios of non-synonymous to synonymous substitutions (dN/dS) for all ORFs based on the inferred GARD phylogenetic trees. These methods were applied under the nucleotide substitution models determined in MODELTEST (Posada and Crandall, 1998). Bayes factors larger than 50 and *P*-values smaller than 0.1 were used as thresholds for the REL method.

## Detection of mixed infections and pseudorecombination

To verify the occurrence of mixed infections and possible pseudorecombination among SAMV strains, a restriction map was constructed using *Msp*I, a four-base cutter restriction enzyme which generates different restriction patterns for the DNA-A and DNA-B of each strain.

## Infectivity assay

Six *N. benthamiana* and ten *S. acuta* plants were biolistically inoculated with infectious clones corresponding to strains 1 and 2 of SAMV. Cognate DNA-A and -B components were cloned from the same sample: clones VIC07D1C plus VIC07D2B representing strain 1 DNA-A and -B, respectively, and VIC03D1C plus VIC03D2B representing strain 2 DNA-A and -B, respectively. Thirty micrograms of each DNA component were excised from the vector (using *Cla*I for VIC07D1C and VIC03D1C and *Bam*HI for VIC07D2B and VIC03D2B), and religated prior to inoculation. The inoculation pressure was adjusted to 50 psi, tungsten particles alone were used as negative control, and an infectious clone of *Tomato yellow spot virus* (ToYSV; Andrade *et al.*, 2006) was used as a positive control. The plants were evaluated at 14 days post inoculation (dpi). Infection was confirmed by visual observation of symptoms and RCA-RFLP analysis.

## Results

### A new begomovirus species (*Sida acuta mosaic virus*), subdivided into three strains, detected in *Sida acuta*

A total of 50 samples were collected, and 47 were positive for the presence of a begomovirus based on the detection of a 2,600 nt band after digestion of the RCA products with restriction enzymes (data not shown). A dataset comprising 65 complete sequences (33 DNA-A and 32 DNA-B; Table 1) was assembled for the analyses.

Pairwise sequence comparisons of the cloned genome sequences with those deposited in GenBank indicated that all clones corresponded to a single species, with <89% identity with other begomoviruses. Therefore this is a new species, for which the name *Sida acuta* mosaic virus (SAMV) is proposed. Three distinct groups of sequences were observed. Sequences within each group were >96% identical, whereas identities between sequences from different groups were of aprox. 94%. Thus, each group corresponds to a distinct strain of SAMV. Pairwise sequence identities (full length DNA-A) with the most closely related viruses are displayed in Figure 1.

No evidence of recombination was detected by the RDP3 program, either among the three strains or within each strain.

### Three distinct clades, representing three different strains from SAMV, observed in phylogenetic trees

The phylogenetic tree based on the nucleotide sequence of the full-length DNA-A (Figure 2) separated the isolates into three clades, each one corresponding to one of the three strains. However, only two clusters were observed in the DNA-B tree, with the two isolates from sample 03D (representing strain 2) clustering with strain 3 isolates.

Furthermore, isolates VIC16D3C, VIC16D2C1, VIC15D1C and VIC40D1C (strain 1 based on DNA-A sequence comparisons) clustered with strain 3 isolates.

The CP tree (Figure 3) had the same topology from the tree of the full length DNA-A, with the clades representing strains 1 and 2 forming a monophyletic branch. In the Rep tree (Figure 3) a slightly different topology was observed, with the strain 2 clade being monophyletic with the strain 3 clade instead of the strain 1 clade. The branches separating the three clades are longer in the Rep tree compared to the CP tree, reflecting the higher degree of variability in the Rep protein compared to the coat protein (see below). Regardless, all branches in both trees showed strong statistical support (Bayes posterior probabilities >0.84).

### Non-synonymous mutations are present in the CP and Rep ORFs, differing among the three strains

Non-synonymous mutations have occurred in the nucleotide sequences of both the CP and Rep proteins, and these are strongly correlated with the three strains of the virus. The CP sequences are very similar, with a single non-synonymous mutation at nucleotide position 617 separating strains 1 and 2 (tyrosine in the deduced amino acid sequence) from strain 3 (phenylalanine) (Table 2). Thus, the mutation caused a change between two amino acids from different classes, hydrophilic (tyrosine) and hydrophobic (phenylalanine).

The Rep sequence was much more variable, with 23 non-synonymous substitutions (Figure 4). Strain 2 isolates were the most variable, with 11 positions in which their amino acid sequence differend from that of strains 1 and 3 isolates. Strain 1 isolates differed from strains 2 and 3 isolates at seven positions, and strain 3 isolates different from strains 1 and 2 isolates at four positions. In one position, different amino acids were present in each of the three strains (Figure 4). Thirteen amino acid changes were between amino acids from the same class, but 10 changes were between amino acids from different classes (hydrophobic to hydrophilic, or vice-versa) (Figure 4). Three of the mutations occurred in the conserved domains involved in rolling-circle replication, but the invariable amino acids were always maintained, with only the variable ones in the motifs being changed (Figure 4).

## No positive selection detected in SAMV isolates

Only negatively selected amino acid sites were detected by the SLAC, REL and PARRIS methods in the sequences of all proteins encoded by the DNA-A and DNA-B of SiMlMV isolates (data not shown). This could be due to purifying selection, with deleterious mutations being purged from the population. Accordingly, dN/dS values calculated with SLAC were <1 for all ORFs (Table 4).

## High molecular variability of the SAMV population

Descriptors of molecular variability were calculated for the entire population and for the subpopulations representing strains 1 and 3 (Table 3). The small size of the strain 2 subpopulation (only two DNA-A sequences) precluded such analysis for this strain.

In general, the results indicate a high degree of variability for the entire population, and a lower degree (although still high) when each subpopulation was considered separately (Table 3). For the entire population, mutation rates for the full length DNA-A and for the CP and Rep proteins are in the order of $10^{-3}$. The values were one order of magnitude lower when each subpopulation was considered separately (Table 3). Likewise, nucleotide diversity values for each subpopulation were approximately 20-fold lower than the ones calculated for the entire population.

### Detection of mixed infection and pseudorecombination between SAMV strains

The restriction patterns obtained with *Msp*I indicated the occurrence of both mixed infections and pseudorecombination (Table 6). Thirty-four samples had a single infection (23 samples with strain 1, one with strain 2 and ten with strain 3). Three samples had mixed infections, and ten samples were infected by a pseudorecombinant with the DNA-A from strain 1 and the DNA-B from strain 3.

### Possible phenotypic difference among strains 1 and 2

Results of biolistic inoculation of *N. benthamiana* plants differed between the two strains. No plants were infected upon inoculation with strain 1, whereas five out of six inoculated plants were infected with strain 2 at 14 dpi (Table 5; Figure 6). All six plants inoculated with infectious clones of *Tomato yellow spot virus* (ToYSV) were infected at 14 dpi (Table 5). Similar results were obtained for *Sida acuta* plants: no plants were infected with strain 1, whereas two out of six inoculated plants were infected with strain 2 (Table 5; Figure 6).

## Discussion

We have characterized a population of a new begomovirus, *Sida acuta* mosaic virus (SAMV), infecting the non-cultivated host *Sida acuta*. Despite being sampled in a very small area, the population is divided into three different strains. Support for the existence of three strains was obtained from sequence comparisons, phylogenetic analysis and from the analysis of non-synonymous substitutions in the CP and Rep sequences.

Interestingly, the phylogenetic analysis based on the DNA-B component indicated the subdivision of SAMV isolates into only two clades, with isolates from sample 03D (whose cognate DNA-As were classified as strain 2) clustering with strain 3. Furthermore, some isolates whose cognate DNA-As were classified as strain 1 also clustered with strain 3 isolates. These results suggest that pseudorecombination between strains could be occurring, and this hypothesis is supported by the restriction analysis of the *Msp*I-digested RCA product which demonstrated infection of several samples by the DNA-A from strain 1 and the DNA-B from strain 3.

The phylogenetic trees based on the CP and Rep amino acid sequences (both proteins being encoded by the DNA-A component) had slightly different topologies, with isolates of strain 2 being closer to strain 1 isolates in the CP tree, and to strain 3 isolates in the Rep tree. This observation suggested that a recombination event could have occurred between strain 1 and strain 3 isolates. However, no recombination event was detected by RDP analysis. This was actually a surprising result, considering the high frequency of recombination among begomoviruses (Albuquerque *et al.*, 2012a; Padidam *et al.*, 1999; Tiendrebeogo *et al.*, 2012). The lack of a detectable recombination signal among SAMV isolates indicates that strain 2 isolates arose by the accumulation of mutations rather than by recombination between strains 1 and 3. This is supported by the higher frequency of non-synonymous mutations observed in the Rep gene of strain 2 isolates compared to strains 1 and 3. However, it must be taken into account that recombination between highly homologous sequences (>90% identity, which is the case among SAMV strains 1, 2 and 3) is difficult to be detected using the RDP program. Therefore, and considering the occurrence of mixed infection between strains 1 and 3 (for example in sample 04D), the possibility of recombination between these strains should not be completely discarded.

The Rep protein is the only begomovirus-encoded protein which is essential for genome replication. Mutations in the Rep protein could lead to changes in replication and maybe in the virus fitness, which is in constant evolution. The N-terminal region of Rep contains conserved motifs that are characteristic of many rolling-circle initiator proteins

(Ilyina and Koonin, 1992; Koonin and Ilyina, 1992). Motif I (FLTY) is required for specific dsDNA binding, while motif II (HLH) is a metal-binding site that may be involved in protein conformation and DNA cleavage (Arguello-Astorga and Ruiz-Medrano, 2001; Fontes *et al.*, 1992; Orozco and Hanley-Bowdoin, 1998). Motif III (YxxKD/E) is the catalytic site for DNA cleavage, with the hydroxyl group of the Y residue forming a covalent bond with the 5' phosphoryl group of the cleaved DNA strand (Laufs *et al.*, 1995; Orozco and Hanley-Bowdoin, 1996). The GRS motif is required for infection and viral genome replication (Nash *et al.*, 2011). Modeling studies suggested that some GRS motif residues contribute to the structural integrity of the Rep protein but do not obviate other potential functions of this motif (Nash *et al.*, 2011). The C-terminal domain contains a NTP binding motif (GxxxxGKT/S), specifying the phosphate binding fold (P-loop). Alteration of this motif led to loss of the ATPase and DNA helicase activities of Rep (Campos-Olivas *et al.*, 2002).

Three of the non-synonymous substitutions detected in strain 2 isolates were located in conserved motifs of the Rep protein: $E_{318}D$ located in motif III, $G_{670}S$ located in the Walker A (NTP-binding) motif, and $V_{781}I$ located in the Walker B motif. Although these three changes involve amino acids of the same class, it is not unreasonable to assume that strain 2 isolates could have differences in fitness compared to isolates of strains 1 and 3. However, results of the infectivity assay with strains 1 and 2 isolates were inconclusive, since no infected plants were obtained upon biolistic inoculation with the strain 1 isolate. Although this could be an indication of the poor fitness of this isolate, it could also mean simply that the DNA-A and/or DNA-B clones were not infectious due to cloning artifacts. Additional experiments must be conducted with other strain 1 clones to properly assess any fitness differences among SAMV isolates belonging to different strains.

Based on the non-synonymous substitutions detected in the Rep and CP sequences, we expected to detect evidence of positive selection favoring the most fit genomes. However, no such evidence was detected. So what is keeping these strains subdivided in the same host, specially when mixed infections and pseudorecombination are occurring ? Maybe we detected an initial sympatric speciation event, where the strains are diverging to generate new species. In these cases, the new species emerges from the same host due to the isolates infecting different host cells (Holmes, 2009; Pagan *et al.*, 2010). Continuing sampling at the same location over the course of many years could confirm this hypothesis.

## References

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

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

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

Andrade, E. C., Manhani, G. G., Alfenas, P. F., Calegario, R. F., Fontes, E. P. B., and Zerbini, F. M. (2006). *Tomato yellow spot virus*, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida* sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*. *Journal of General Virology* **87,** 3687-3696.

Arguello-Astorga, G. R., and 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,** 1465-85.

Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, F. M., Moriones, E., and Navas-Castillo, J. (2012). Family *Geminiviridae*. *In* "Virus Taxonomy. 9th Report of the International Committee on Taxonomy of Viruses" (A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz, Eds.), pp. 351-373. Elsevier Academic Press, London, UK.

Calegario, R. F., Ferreira, S. S., Andrade, E. C., and Zerbini, F. M. (2007). Characterization of *Tomato yellow spot virus*, (ToYSV), a novel tomato-infecting begomovirus from Brazil. *Pesquisa Agropecuária Brasileira* **42,** 1335-1343.

Campos-Olivas, R., Louis, J. M., Clerot, D., Gronenborn, B., and Gronenborn, A. M. (2002). 1H, 13C, and 15N assignment of the N-terminal, catalytic domain of the replication initiation protein from the geminivirus TYLCV. *Journal of Biomolecular NMR* **24,** 73-4.

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

Costa, A. S. (1955). Studies on *Abutilon* mosaic in Brazil. *Phytopathologische Zeitschrift* **24,** 97-112.

Costa, A. S., and Bennett, C. W. (1950). Whitefly transmitted mosaic of *Euphorbia prunifolia*. *Phytopathology* **40,** 266-283.

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

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 7160-4.

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

Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics* **9,** 267-276.

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

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

Fernandes, F. R., Cruz, A. R. R., Faria, J. C., Zerbini, F. M., and Aragão, F. J. L. (2009). Three distinct begomoviruses associated with soybean in central Brazil. *Archives of Virology* **154,** 1567-1570.

Fernandes, J. J., Carvalho, M. G., Andrade, E. C., Brommonschenkel, S. H., Fontes, E. P. B., and Zerbini, F. M. (2006). Biological and molecular properties of *Tomato rugose mosaic virus* (ToRMV), a new tomato-infecting begomovirus from Brazil. *Plant Pathology* **55,** 513-522.

Fontes, E. P. B., Luckow, V. A., and Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell* **4,** 597-608.

García-Arenal, F., Fraile, A., and Malpica, J. M. (2003). Variation and evolution of plant virus populations. *International Microbiology* **6,** 225-232.

Ge, L. M., Zhang, J. T., Zhou, X. P., and Li, H. Y. (2007). Genetic structure and population variability of tomato yellow leaf curl China virus. *Journal of Virology* **81,** 5902-5907.

Holmes, E. C. (2009). The evolutionary genetics of emerging viruses. *Annual Review of Ecology, Evolution and Systematics* **40,** 353-372.

Ilyina, T. V., and Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Research* **20,** 3279-85.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B., and Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *Journal of Virological Methods* **116,** 209-211.

Koonin, E. V., and Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokariotic plasmid rolling circle DNA replication initiator proteins. *Journal of General Virology* **73,** 2763-2766.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995). Identification of the nicking tyrosine of geminivirus Rep protein. *FEBS Letters* **377,** 258-262.

Lima, A. T. M., Sobrinho, R. R., Gonzalez-Aguilera, J., Rocha, C. S., Silva, S. J. C., Xavier, C. A. D., Silva, F. N., Duffy, S., and Zerbini, F. M. (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *Journal of General Virology* **94,** 418-431.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26,** 2462-2463.

McDonald, B. A., and Linde, C. (2002). The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica* **124,** 163-180.

Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Rivera-Bustamante, R., Malathi, V. G., Briddon, R. W., and 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.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibanez, J. T., and Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *Journal of Virology* **85,** 1182-1192.

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

Orozco, B. M., and Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *Journal of Virology* **70,** 148-58.

Orozco, B. M., and Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *Journal of Biological Chemistry* **273,** 24448-24456.

Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265,** 218-224.

Pagan, I., Fraile, A., Fernandez-Fueyo, E., Montes, N., Alonso-Blanco, C., and Garcia-Arenal, F. (2010). *Arabidopsis thaliana* as a model for the study of plant-virus co-evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365,** 1983-1995.

Paprotka, T., Metzler, V., and Jeske, H. (2010). The first DNA 1-like alpha satellites in association with New World begomoviruses in natural infections. *Virology* **404,** 148-157.

Pond, S. L., and Frost, S. D. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21,** 2531-3.

Posada, D., and Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14,** 817-8.

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

Ribeiro, S. G., Ambrozevicius, L. P., Ávila, A. C., Bezerra, I. C., Calegario, R. F., Fernandes, J. J., Lima, M. F., Mello, R. N., Rocha, H., and Zerbini, F. M. (2003). Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Archives of Virology* **148,** 281-295.

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

Ribeiro, S. G., Martin, D. P., Lacorte, C., Simões, I. C., Orlandini, D. R. S., and Inoue-Nagata, A. K. (2007). Molecular and biological characterization of *Tomato chlorotic mottle virus* suggests that recombination underlies the evolution and diversity of Brazilian tomato begomoviruses. *Phytopathology* **97,** 702-711.

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

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

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003). DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496-2497.

Ruiz-Ferrer, V., Goytia, E., Martinez-Garcia, B., Lopez-Abella, D., and Lopez-Moya, J. J. (2004). Expression of functionally active helper component protein of Tobacco etch potyvirus in the yeast *Pichia pastoris*. *Journal of General Virology* **85,** 241-249.

Scheffler, K., Martin, D. P., and Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22,** 2493-9.

Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Junior, B. T., Assunção, I. P., Lima, G. S. A., Pio-Ribeiro, G., Mizubuti, E. S. G., and Zerbini, F. M. (2012). Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathology* **61,** 457-467.

Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Júnior, B. T., Assunção, I. P., Lima, G. S. A., Pio-Ribeiro, G., Mizubuti, E. S. G., and Zerbini, F. M. (2011). High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in northeastern Brazil. *Archives of Virology* **156,** 2205-2213.

Tavares, S. S., Ramos-Sobrinho, R., Gonzalez-Aguilera, J., Lima, G. S. A., Assunção, I. P., and Zerbini, F. M. (2012). Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on *Sida* spp. *Planta Daninha* **30,** 305-315.

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

**Table 1.** Origin and strain classification of the *Sida acuta* mosaic virus (SAMV) clones obtained in this work.

| Sample | Clone | Component | Enzyme | Strain[a] |
|---|---|---|---|---|
| 01D | VIC01D5C | DNA-A | *Cla*I | 1 |
| 01D | VIC01D1B | DNA-B | *Bam*HI | 1 |
| 03D | VIC03D1C | DNA-A | *Cla*I | 2 |
| 03D | VIC03D3C | DNA-A | *Cla*I | 2 |
| 03D | VIC03D1B | DNA-B | *Bam*HI | 2 |
| 03D | VIC03D1B1 | DNA-B | *Bam*HI | 2 |
| 03D | VIC03D2B | DNA-B | *Bam*HI | 2 |
| 04D | VIC04D1C | DNA-A | *Cla*I | 3 |
| 04D | VIC04D1EV | DNA-A | *Eco*RV | 3 |
| 04D | VIC04D3C | DNA-A | *Cla*I | 1 |
| 05D | VIC05D1C | DNA-A | *Cla*I | 1 |
| 05D | VIC05D3B | DNA-B | *Bam*HI | 1 |
| 06D | VIC06D1C | DNA-A | *Cla*I | 1 |
| 06D | VIC06D2EV | DNA-A | *Eco*RV | 1 |
| 06D | VIC06D1B | DNA-B | *Bam*HI | 1 |
| 07D | VIC07D1C | DNA-A | *Cla*I | 1 |
| 07D | VIC07D1C1 | DNA-A | *Cla*I | 1 |
| 07D | VIC07D2C | DNA-A | *Cla*I | 1 |
| 07D | VIC07D1B | DNA-B | *Bam*HI | 1 |
| 07D | VIC07D2B | DNA-B | *Bam*HI | 1 |
| 09D | VIC09D1C | DNA-A | *Cla*I | 1 |
| 09D | VIC09D3EV | DNA-A | *Eco*RV | 1 |
| 09D | VIC09D1B | DNA-B | *Bam*HI | 1 |
| 10D | VIC10D3C | DNA-A | *Cla*I | 1 |
| 10D | VIC10D1B | DNA-B | *Bam*HI | 1 |
| 12D | VIC12D1C | DNA-A | *Cla*I | 1 |
| 12D | VIC12D2C | DNA-A | *Cla*I | 1 |
| 12D | VIC12D1B | DNA-B | *Bam*HI | 1 |
| 13D | VIC13D1B | DNA-B | *Bam*HI | n.d. |
| 13D | VIC13D1B1 | DNA-B | *Bam*HI | n.d. |
| 14D | VIC14D1C | DNA-A | *Cla*I | 1 |
| 14D | VIC14D2B | DNA-B | *Bam*HI | 1 |
| 15D | VIC15D2C | DNA-A | *Cla*I | 1 |
| 15D | VIC15D2C1 | DNA-A | *Cla*I | 1 |
| 15D | VIC15D5C | DNA-A | *Cla*I | 1 |
| 15D | VIC15D1B | DNA-B | *Bam*HI | 1 |
| 15D | VIC15D1C | DNA-B | *Cla*I | 1 |
| 16D | VIC16D2C | DNA-A | *Cla*I | 1 |
| 16D | VIC16D2C1 | DNA-B | *Cla*I | 1 |
| 16D | VIC16D3C | DNA-B | *Cla*I | 1 |
| 18D | VIC18D1EV | DNA-A | *Eco*RV | 3 |

| | | | | |
|---|---|---|---|---|
| 18D | VIC18D1C | DNA-B | *Cla*I | 3 |
| 19D | VIC19D1C | DNA-A | *Cla*I | 3 |
| 19D | VIC19D2C | DNA-A | *Cla*I | 3 |
| 19D | VIC19D1B | DNA-B | *Bam*HI | 3 |
| 20D | VIC20D1C | DNA-B | *Cla*I | n.d. |
| 23D | VIC23D4C | DNA-A | *Cla*I | 1 |
| 23D | VIC23D1B | DNA-B | *Bam*HI | 1 |
| 24D | VIC24D1C | DNA-A | *Cla*I | 3 |
| 27D | VIC27D1C | DNA-B | *Cla*I | n.d. |
| 32D | VIC32D2B | DNA-B | *Bam*HI | n.d. |
| 36D | VIC36D1C | DNA-A | *Cla*I | 3 |
| 40D | VIC40D2C | DNA-A | *Cla*I | 1 |
| 40D | VIC40D1C | DNA-B | *Cla*I | 1 |
| 41D | VIC41D3C | DNA-A | *Cla*I | 3 |
| 41D | VIC41D3C1 | DNA-B | *Cla*I | 3 |
| 42D | VIC42D1C | DNA-B | *Cla*I | n.d. |
| 42D | VIC42D2C | DNA-B | *Cla*I | n.d. |
| 44D | VIC44D1B | DNA-B | *Bam*HI | n.d. |
| 44D | VIC44D3B | DNA-B | *Bam*HI | n.d. |
| 45D | VIC45D1EV | DNA-A | *Eco*RV | 1 |
| 46D | VIC46D1B | DNA-B | *Bam*HI | n.d. |
| 46D | VIC46D2C | DNA-B | *Cla*I | n.d. |
| 47D | VIC47D1C | DNA-A | *Cla*I | 3 |
| 47D | VIC47D2C | DNA-A | *Cla*I | 3 |

[a] Strain classification is based on the DNA-A sequence. When a DNA-B component was cloned without a cognate DNA-A, the strain was not determined (n.d.).

**Table 2.** Alignment of the C-terminal region of the coat protein of *Sida acuta* mosaic virus (SAMV) isolates. The single difference in the amino acid sequence among the three strains is indicated in bold underline (tyrosine for strain 1 and 2 isolates, phenylalanine for strain 3 isolates).

| Sample | Isolate | C-Terminal region of CP | Strain |
|---|---|---|---|
| 01D | VIC01D5C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 04D | VIC04D3C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 05D | VIC05D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 06D | VIC06D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 06D | VIC06D2EV | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 07D | VIC07D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 07D | VIC07D1C1 | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 07D | VIC07D2C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 09D | VIC09D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 09D | VIC09D3EV | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 10D | VIC10D3C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 12D | VIC12D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 12D | VIC12D2C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 14D | VIC14D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 15D | VIC15D2C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 15D | VIC15D2C1 | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 15D | VIC15D5C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 16D | VIC16D2C | WKVNNHVV**Y**NHQEAGKYDNHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 23D | VIC23D4C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 40D | VIC40D2C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 45D | VIC45D1EV | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 1 |
| 03D | VIC03D1C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 2 |
| 03D | VIC03D3C | WKVNNHVV**Y**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 2 |
| 04D | VIC04D1C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 04D | VIC04D1EV | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 18D | VIC18D1EV | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 19D | VIC19D1C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 19D | VIC19D2C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 24D | VIC24D1C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 36D | VIC36D1C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 41D | VIC41D3C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 47D | VIC47D1C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |
| 47D | VIC47D2C | WKVNNHVV**F**NHQEAGKYENHTENALLLYMACTHASNPVYATLKIRIYFYDSITN* | 3 |

**Table 3.** Genetic structure of the *Sida acuta* mosaic virus (SAMV) population from Viçosa, MG, Brazil.

| **Sequences** | **Strains** | **No. of sequences** | **Total number of sites**[b] | **h**[c] | **Hd** | **ETA** | $\pi$ | **Mutation frequency** |
|---|---|---|---|---|---|---|---|---|
| Full lenght DNA-A | All[a] | 33 | 2646 | 27 | 0.981 ± 0.015 | 192 | 0.02378 ± 0.00268 | 2.20E-03 |
| | 1 | 21 | 2649 | 16 | 0.957 ± 0.032 | 28 | 0.00185 ± 0.00023 | 5.03E-04 |
| | 3 | 10 | 2656 | 9 | 0.978 ± 0.054 | 13 | 0.00123 ± 0.00026 | 4.89E-04 |
| CP | All | 33 | 756 | 15 | 0.854 ± 0.053 | 54 | 0.02254 ± 0.00260 | 2.16E-03 |
| | 1 | 21 | 756 | 9 | 0.681 ± 0.113 | 9 | 0.00179 ± 0.00044 | 5.67E-04 |
| | 3 | 10 | 756 | 5 | 0.800 ± 0.100 | 5 | 0.00176 ± 0.00048 | 6.61E-04 |
| Rep | All | 33 | 1077 | 14 | 0.898 ± 0.031 | 76 | 0.02383 ± 0.00264 | 2.14E-03 |
| | 1 | 21 | 1077 | 8 | 0.805 ± 0.064 | 10 | 0.00149 ± 0.00027 | 4.42E-04 |
| | 3 | 10 | 1077 | 4 | 0.711 ± 0.117 | 3 | 0.00087 ± 0.00022 | 2.79E-04 |

[a] All tree strains from SAMV.

[b] Number of site excluding gaps.

[c] h, Number of haplotypes; Hd, Haplotype diversity; ETA, Total number of mutations; $\pi$, Nucleotide diversity.

**Table 4.** Ratio of non-synonymous to synonymous mutations (dN/dS) for each ORF from the DNA-A and DNA-B of *Sida acuta* mosaic virus (SAMV) isolates, calculated by the SLAC method.

| | **DNA-A** | | | | | **DNA-B** | |
|---|---|---|---|---|---|---|---|
| | **CP** | **Rep** | **C4** | **Trap** | **Ren** | **NSP** | **MP** |
| **dN/dS** | 0.10555 | 0.25768 | 0.79165 | 0.68363 | 0.32359 | 0.22854 | 0.09220 |

**Table 5.** Number of samples within each *Sida acuta* mosaic virus (SAMV) strain in single infection, mixed infection and pseudo-recombination based on RCA-RFLP analysis.

| **Strains** | **No. of samples** |
|---|---|
| Strain 1 | 23 |
| Strain 2 | 1 |
| Strain 3 | 10 |
| Mixed infection | 3 |
| DNA-A of strain 1 replicating DNA-B of strains 1 and 3 | 3 |
| DNA-A of strain 1 replicating DNA-B of strain 3 | 4 |
| DNA-A of strains 1 and 3 replicating DNA-B of strain 3 | 3 |

**Table 6.** Results of infectivity assay with isolates of strains 1 and 2 of *Sida acuta* mosaic virus (SAMV).

| **Treatment** | ***N. benthamiana*** | ***S. acuta*** |
|---|---|---|
| | **Inf./Inoc.** | **Inf./Inoc.** |
| SAMV strain 1 | 0/6 | 0/10 |
| SAMV strain 2 | 5/6 | 2/10 |
| ToYSV | 6/6 | 0/10 |
| Water | 0/6 | 0/10 |

## Figure legends

**Figure 1.** Pairwise sequence comparisons between *Sida acuta* mosaic virus (SAMV) clones obtained in this work plus the most closely related begomoviruses.

**Figure 2.** Bayesian phylogenetic tree based on the nucleotide sequences of the full length genome of *Sida acuta* mosaic virus (SAMV) isolates obtained in this work. In the DNA-A tree, three distinct clades, corresponding to the three different strains, can be observed. The DNA-B tree contains only two clades. Strain 1 isolates are indicated in blue, strain 2 isolates in green, and strain 3 isolates in red. DNA-B sequences not classified into strains due to the absence of a cognate DNA-A are indicated in yellow. Numbers on the branches represent Bayesian posterior probabilities.

**Figure 3.** Bayesian phylogenetic trees based on the deduced amino acid sequences of the CP and Rep proteins of *Sida acuta* mosaic virus (SAMV) isolates obtained in this work. Three distinct clades corresponding to the three different strains (blue for strain 1, green for strain 2 and red for strain 3) can be observed. The CP tree has the same topology of the DNA-A tree. In the Rep tree, strain 2 isolates clustered with strain 3 isolates. Numbers on the branches represent Bayesian posterior probabilities.

**Figure 4.** Schematic representation of the *Sida acuta* mosaic virus (SAMV) Rep protein, showing conserved motifs and non-synonymous mutations. The bold letters represent the amino acid code, with U representing any hydrophobic amino acid and X representing any amino acid. Green bars represent the position of conserved motifs in the coding region. Yellow lines represent non-synonymous mutations in which the change was between

amino acids in the same class. Red lines represent non-synonymous mutations in which the change was between amino acids from different classes. The letters above the yellow and red lines represent the amino acid present in isolates of each strain (strain 1/strain 2/strain 3). Numbers in parentheses represent the nucleotide position of each mutation.

**Figure 5. A.** *In silico*-predicted *Msp*I restriction patterns for the DNA-A and DNA-B of *Sida acuta* mosaic virus (SAMV) isolates belonging to strains 1 and 3. **B.** Results of *Msp*I digestion from four samples (4, 16, 23 and 24) representing the pattern for single infection by strains 1 (23) and strain 3 (24), mixed infection (4) and pseudorecombination between strain 1 DNA-A and strain 3 DNA-B (16).

**Figure 6.** Symptoms observed in *Nicotiana benthamiana* (**A-D**) and *Sida acuta* (**E-H**) plants biolistically inoculated with *Sida acuta* mosaic virus (SAMV), at 14 days post-inoculation. **A, E.** Plants inoculated with tungsten particles alone (negative control). **B, F.** Plant inoculated with a strain 1 isolate. **C, G.** Plants inoculated with a strain 2 isolate. **D, H.** Plants inoculated with *Tomato yellow spot virus* (ToTSV).

Figure 1

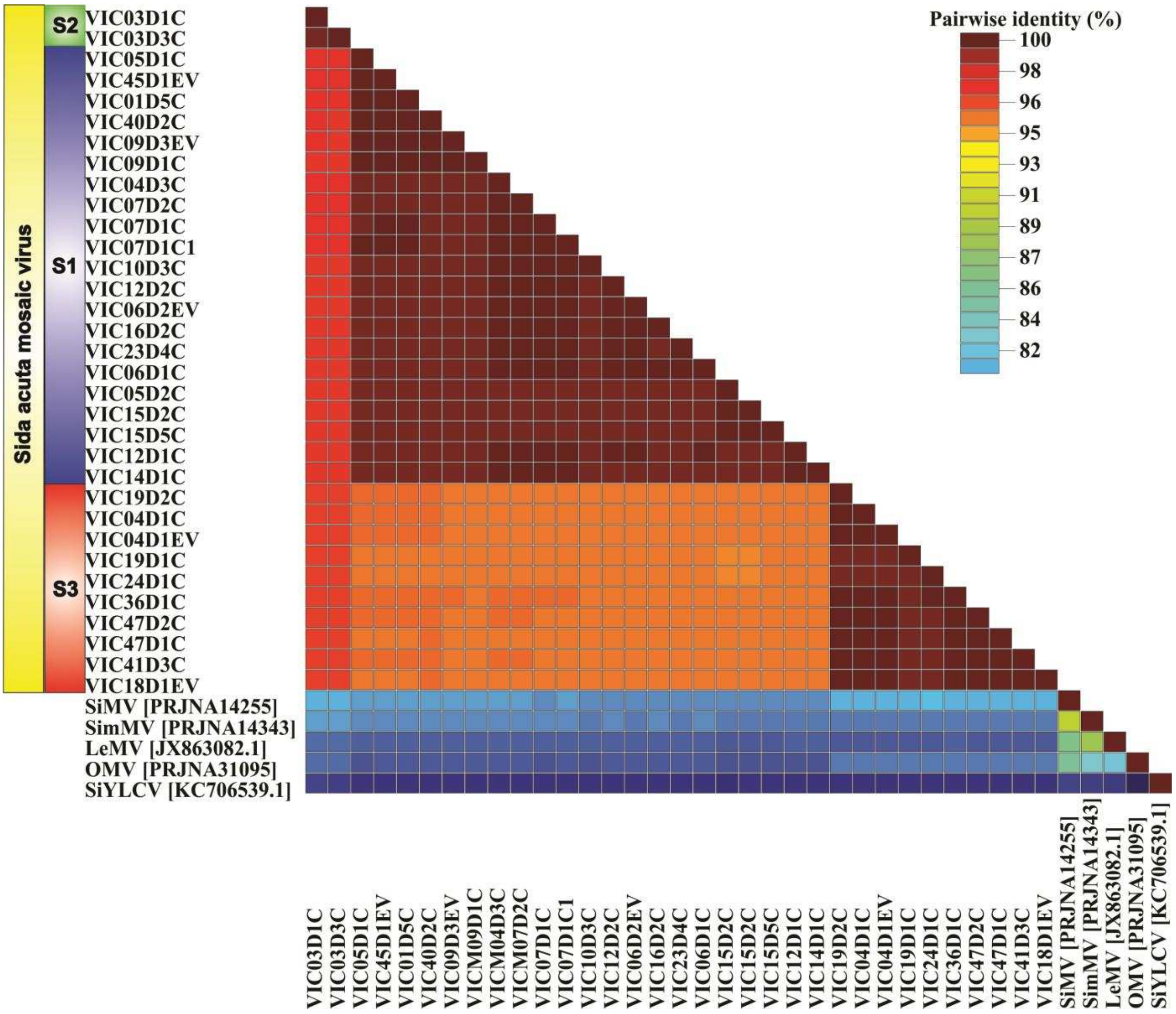

Sida acuta mosaic virus
S2
S1
S3
VIC03D1C
VIC03D3C
VIC05D1C
VIC45D1EV
VIC01D5C
VIC40D2C
VIC09D3EV
VIC09D1C
VIC04D3C
VIC07D2C
VIC07D1C
VIC07D1C1
VIC10D3C
VIC12D2C
VIC06D2EV
VIC16D2C
VIC23D4C
VIC06D1C
VIC05D2C
VIC15D2C
VIC15D5C
VIC12D1C
VIC14D1C
VIC19D2C
VIC04D1C
VIC04D1EV
VIC19D1C
VIC24D1C
VIC36D1C
VIC47D2C
VIC47D1C
VIC41D3C
VIC18D1EV
SiMV [PRJNA14255]
SimMV [PRJNA14343]
LeMV [JX863082.1]
OMV [PRJNA31095]
SiYLCV [KC706539.1]
Pairwise identity (%)
100
98
96
95
93
91
89
87
86
84
82
VIC03D1C
VIC03D3C
VIC05D1C
VIC45D1EV
VIC01D5C
VIC40D2C
VIC09D3EV
VICM09D1C
VICM04D3C
VICM07D2C
VIC07D1C
VIC07D1C1
VIC10D3C
VIC12D2C
VIC06D2EV
VIC16D2C
VIC23D4C
VIC06D1C
VIC15D2C
VIC15D2C
VIC15D5C
VIC12D1C
VIC14D1C
VIC19D2C
VIC04D1C
VIC04D1EV
VIC19D1C
VIC24D1C
VIC36D1C
VIC47D2C
VIC47D1C
VIC41D3C
VIC18D1EV
SiMV [PRJNA14255]
SimMV [PRJNA14343]
LeMV [JX863082.1]
OMV [PRJNA31095]
SiYLCV [KC706539.1]

**Figure 2**

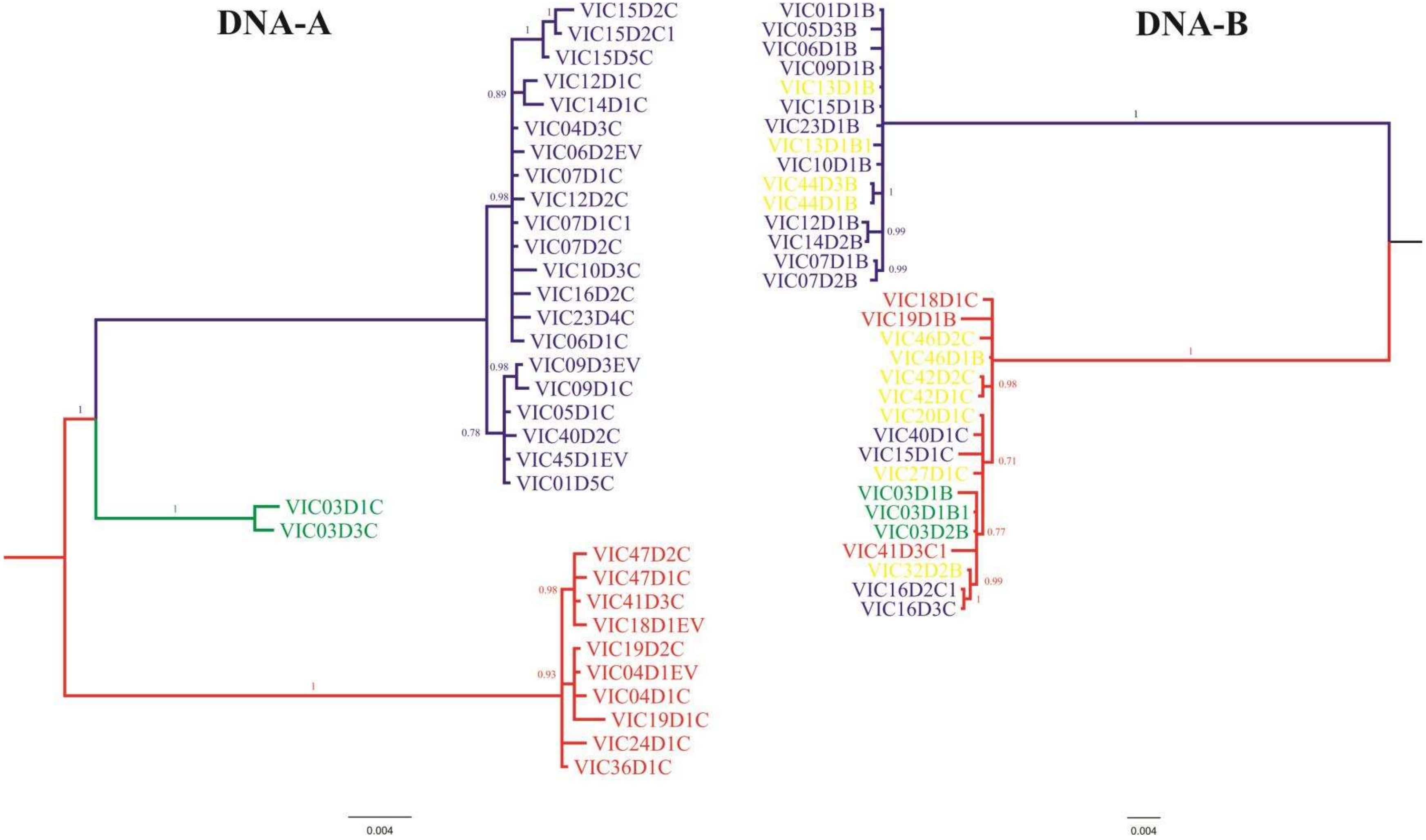

DNA-A
VIC15D2C
VIC15D2C1
VIC15D5C
VIC12D1C
VIC14D1C
VIC04D3C
VIC06D2EV
VIC07D1C
VIC12D2C
VIC07D1C1
VIC07D2C
VIC10D3C
VIC16D2C
VIC23D4C
VIC06D1C
VIC09D3EV
VIC09D1C
VIC05D1C
VIC40D2C
VIC45D1EV
VIC01D5C
VIC03D1C
VIC03D3C
VIC47D2C
VIC47D1C
VIC41D3C
VIC18D1EV
VIC19D2C
VIC04D1EV
VIC04D1C
VIC19D1C
VIC24D1C
VIC36D1C
1
1
0.89
0.98
0.98
0.78
1
1
1
0.98
0.93
0.004
DNA-B
VIC01D1B
VIC05D3B
VIC06D1B
VIC09D1B
VIC13D1B
VIC15D1B
VIC23D1B
VIC13D1B1
VIC10D1B
VIC44D3B
VIC44D1B
VIC12D1B
VIC14D2B
VIC07D1B
VIC07D2B
VIC18D1C
VIC19D1B
VIC46D2C
VIC46D1B
VIC42D2C
VIC42D1C
VIC20D1C
VIC40D1C
VIC15D1C
VIC27D1C
VIC03D1B
VIC03D1B1
VIC03D2B
VIC41D3C1
VIC32D2B
VIC16D2C1
VIC16D3C
1
1
0.99
0.99
1
0.98
0.71
0.77
0.99
1
0.004

**Figure 3**

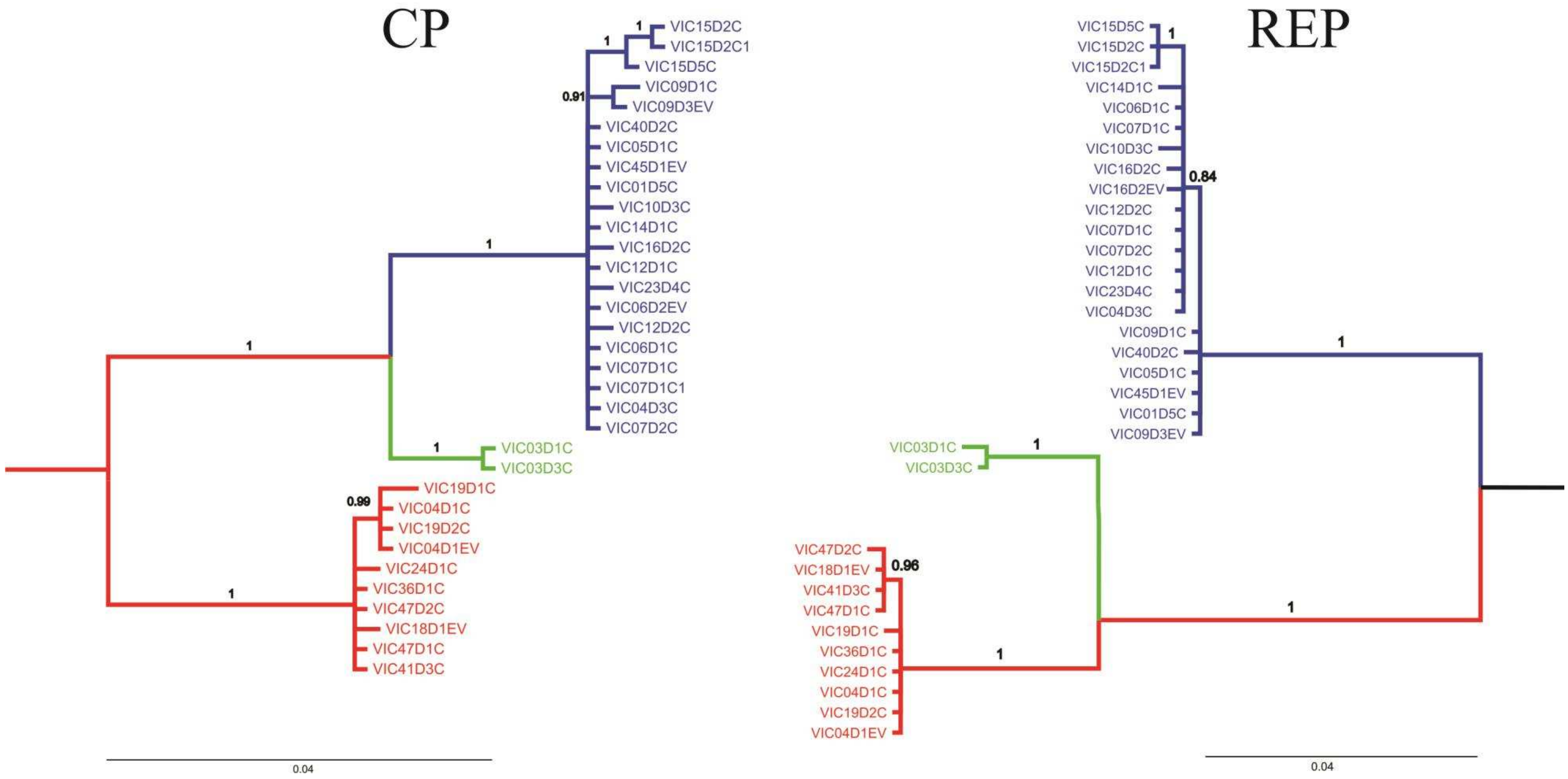

CP
1
1
VIC15D2C
VIC15D2C1
VIC15D5C
0.91
VIC09D1C
VIC09D3EV
VIC40D2C
VIC05D1C
VIC45D1EV
VIC01D5C
VIC10D3C
VIC14D1C
1
VIC16D2C
VIC12D1C
VIC23D4C
VIC06D2EV
VIC12D2C
1
VIC06D1C
VIC07D1C
VIC07D1C1
VIC04D3C
VIC07D2C
1
VIC03D1C
VIC03D3C
0.99
VIC19D1C
VIC04D1C
VIC19D2C
VIC04D1EV
VIC24D1C
1
VIC36D1C
VIC47D2C
VIC18D1EV
VIC47D1C
VIC41D3C
0.04
REP
VIC15D5C
1
VIC15D2C
VIC15D2C1
VIC14D1C
VIC06D1C
VIC07D1C
VIC10D3C
VIC16D2C
0.84
VIC16D2EV
VIC12D2C
VIC07D1C
VIC07D2C
VIC12D1C
VIC23D4C
VIC04D3C
VIC09D1C
1
VIC40D2C
VIC05D1C
VIC45D1EV
VIC01D5C
VIC09D3EV
VIC03D1C
1
VIC03D3C
VIC47D2C
0.96
VIC18D1EV
VIC41D3C
VIC47D1C
1
VIC19D1C
VIC36D1C
1
VIC24D1C
VIC04D1C
VIC19D2C
VIC04D1EV
0.04

**Figure 4**

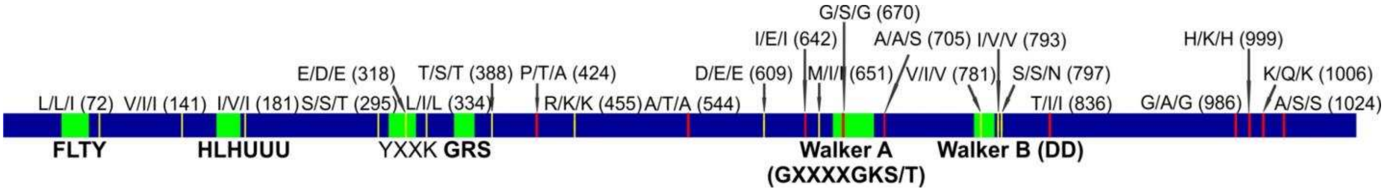

G/S/G (670)
I/E/I (642)
A/A/S (705)
I/V/V (793)
H/K/H (999)
E/D/E (318)
T/S/T (388)
P/T/A (424)
D/E/E (609)
M/I/I (651)
V/I/V (781)
S/S/N (797)
K/Q/K (1006)
L/L/I (72)
V/I/I (141)
I/V/I (181)
S/S/T (295)
L/I/L (334)
R/K/K (455)
A/T/A (544)
T/I/I (836)
G/A/G (986)
A/S/S (1024)
FLTY
HLHUUU
YXXK
GRS
Walker A
(GXXXXGKS/T)
Walker B (DD)

**Figure 5**

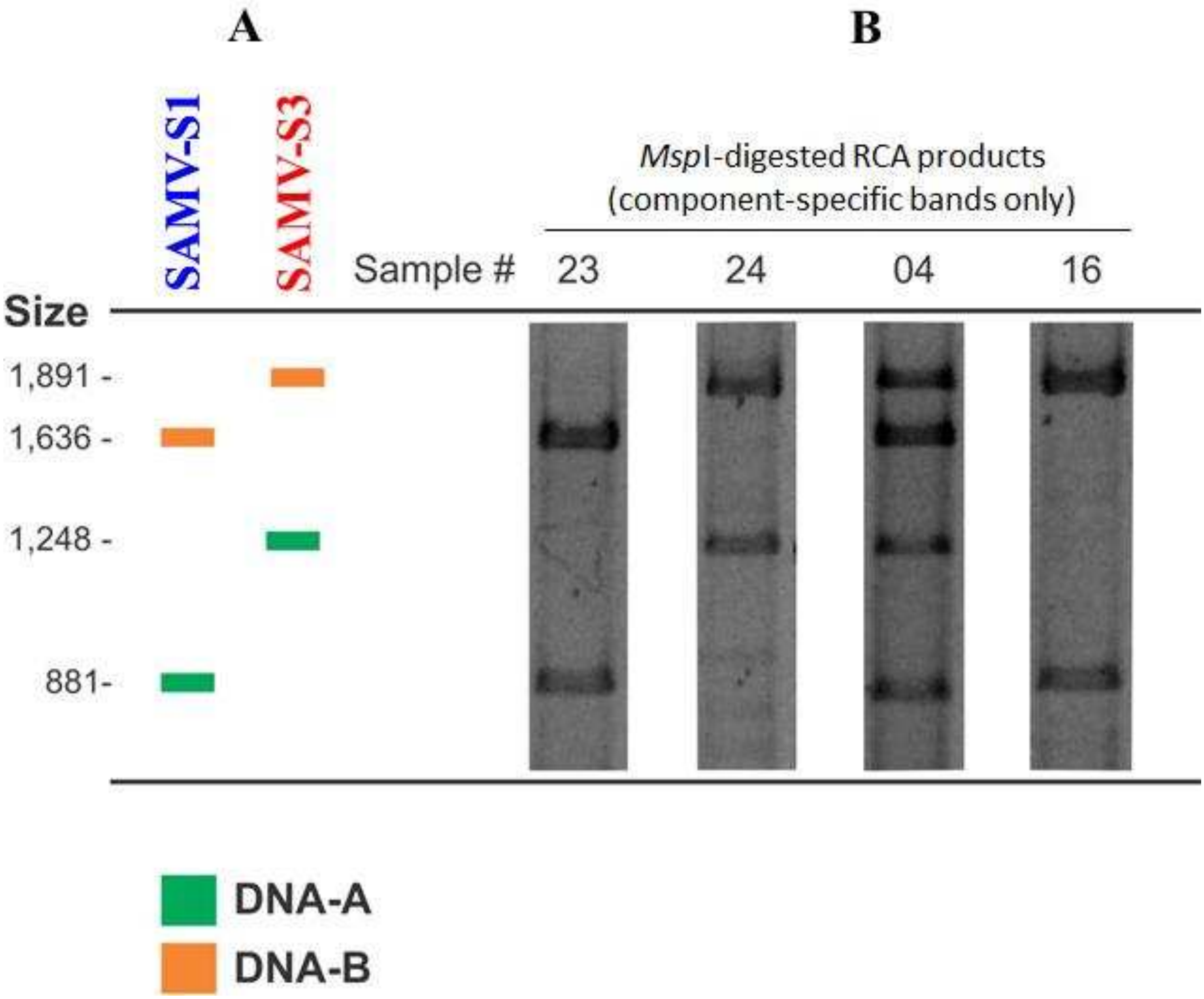

A
B
SAMV-S1
SAMV-S3
MspI-digested RCA products
(component-specific bands only)
Sample #
23
24
04
16
Size
1,891 -
1,636 -
1,248 -
881-
DNA-A
DNA-B

**Figure 6**