ROBERTO RAMOS SOBRINHO

BEGOMOVIRUS EVOLUTION IN THE NEW WORLD: GENETIC VARIABILITY OF POPULATIONS AND ITS EFFECT ON SPECIATION

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

R175b
2014

Ramos Sobrinho, Roberto, 1986-
Begomovirus evolution in the New World : genetic variability of populations and its effect on speciation / Roberto Ramos Sobrinho. - Viçosa, MG, 2014.
x, 116f. : il. (algumas color.) ; 29 cm.

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

1. Geminivirus. 2. Recombinação (Genética). 3. Evolução. 4. Mosca. 5. Vírus. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-gradução em Fitopatologia. II. Título.

CDD 22. ed. 595.7

ROBERTO RAMOS SOBRINHO

BEGOMOVIRUS EVOLUTION IN THE NEW WORLD: GENETIC VARIABILITY OF POPULATIONS AND ITS EFFECT ON SPECIATION

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: 26 de fevereiro de 2014

Prof. Siobain Marie Deirdre Duffy	Prof. Luis Cláudio Vieira da Cunha
Dr. Fábio Nascimento da Silva	Prof. Poliane Alfenas Zerbini

Prof. Francisco Murilo Zerbini Júnior
(Orientador)

AGRADECIMENTOS

À Deus pelo dom da vida e por me dar força e sustentação durante as adversidades encontradas durante a realização deste trabalho;

À minha mãe, Anilza Rocha, ao meu pai José Neto Ramos e ao meu padrasto Gildásio dos Santos pelo esforço e dedicação na minha criação, e pela contribuição para realização dos meus sonhos, sempre acreditando que eles eram possíveis;

À minha família e amigos que tanto contribuíram ao longo de toda a minha vida, ensinando com muita sabedoria, valores que manterei sempre comigo;

Ao Professor Francisco Murilo Zerbini, pelos ensinamentos transmitidos ao longo desses anos de convivência, pela orientação na condução desse trabalho, pela confiança e principalmente por ter sempre acreditado que tudo isso seria possível (minha eterna gratidão);

À Professora Siobain Duffy, por ter-me acolhido com tanto carinho e dedicação quanto estive junto ao seu grupo de pesquisa nos Estados Unidos (I wish to thank you for your attention and friendship);

Ao Professor Eduardo Seiti Gomide Mizubuti, pela orientação e ensinamentos transmitidos;

Aos Professores Gaus Silvestre Andrade de Lima e Iraíldes Pereira Assunção pela orientação e confiança;

Ao Professor Luiz Carlos Caetano pelo conhecimento imensurável transmitido, amizade, apoio e dedicação;

À Professora Poliane Alfenas Zerbini por ter acreditado em minha capacidade e contribuído ao longo desse processo para construção do conhecimento;

Aos Professores e funcionários do departamento de fitopatologia pela imensa contribuição durante o doutorado;

Aos meus avós, Manoel Vieira Ramos, Maria do Socorro Ramos, Antônio Rocha e Maria Helena Alves, por acreditarem e ensinarem sempre que eu podia ser mais;

Aos meus irmãos pelo apoio e por estarem presentes em minha vida;

Aos meus amigos do Laboratório de Virologia Vegetal Molecular da Universidade Federal de Viçosa, Jorge, Alison, Renan, Joaquim, Tatiana, Fábio, César, Larissa, Fernanda, Thalita, Camila, Vitor, João Paulo, Osvaldo, Adriana, André, Chaianne, Daniel, David, Diêgo, Diogo, Fábio Assad, Glória, Hermano, Ana Paula, Josiane, Lenin, Márcio, Marcos, Pedro e Silvia;

Aos amigos de doutorado que cursaram disciplinas comigo, pelo apoio e amizade, em especial aos amigos Jaime Honorato, Hugo Medeiros, Tadeu Silva, Alessandro Nicole, Rafael Alfenas, Danilo Pinho e Alessandro Fortunato;

Aos meus amigos do Duffy Lab da Rutgers University, Yee Mey, Preshita Gadkari, Lily Li, Daniel Cardinale, Shannon Loelius, Yuliya Olifer, Kendra Avinger, Christopher Lin, Allison Hicks e Mark Labrador (thanks so much to all of you);

Aos meus amigos nos Estados Unidos, Foday Kamara, Bre Da e John Rivera (thanks for your help and friendship);

Aos amigos do Laboratório de Cultura de Tecidos, em especial aos amigos Elyabe, Marcos, Diego e Prof. Wagner;

Aos meus amigos Robson, Ricardo, Antônio, Gian, Demetrius, Marcos, Cido, Anderson, Nem, Jefferson, Marcondes, Aurélio, Jane, Davidson, Júnior, Tony, Leandro e Maykon;

À Alicia Zem, por ser tão especial e por todo carinho e amizade;

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

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudo;

À CAPES, pela concessão da bolsa de estudo no exterior;

À Embrapa Arroz e Feijão, pelo apoio durante realizações das coletas de campo, em especial às Doutoras Raquel Mello e Flávia Rabelo;

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

BIOGRAFIA

ROBERTO RAMOS SOBRINHO, filho de José Neto Ramos e Anilza Rocha, nasceu em Cedro de São João, Sergipe, no dia 12 de Julho de 1986.

Em 2004 iniciou o curso de Ciências Biológicas na Universidade Federal de Alagoas (UFAL), Maceió, Alagoas, graduando-se em fevereiro de 2008. No mês seguinte, ingressou no curso de pós-graduação em Agronomia em nível de Mestrado na mesma universidade, submetendo-se à defesa de dissertação em fevereiro de 2010.

Em março de 2010, iniciou o curso de Doutorado em Fitopatologia na Universidade Federal de Viçosa (UFV), submetendo-se à defesa de tese em 26 de fevereiro de 2014.

CONTENTS

RESUMO

RAMOS-SOBRINHO, Roberto, D.Sc., Universidade Federal de Viçosa, fevereiro de 2014. **Evolução de begomovírus no Novo Mundo: Variabilidade genética de populações e seu efeito na especiação.** Orientador: Francisco Murilo Zerbini Júnior. Co-orientador: Eduardo Seiti Gomide Mizubuti.

Vírus pertencentes à família *Geminiviridae* possuem genoma circular de fita simples e infectam uma ampla gama de hospedeiros causando perdas significativas, principalmente em países tropicais e subtropicais. A família é dividida em sete gêneros (*Begomovirus*, *Becurtovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* e *Turncurtovirus*) de acordo com o tipo de inseto vetor, gama de hospedeiros, organização genômica e filogenia. Begomovírus são transmitidos por mosca-branca e causam epidemias em culturas economicamente importantes em todo o mundo. Plantas silvestres/não-cultivadas possuem um papel epidemiológico importante, agindo como reservatórios de begomovírus e como hospedeiros promíscuos onde recombinação pode ocorrer. Além disso, resultados anteriores sugerem que populações de begomovírus presentes em plantas não-cultivadas possuem maior variabilidade genética do que aquelas presentes em plantas cultivadas. A maioria dos begomovírus do Novo Mundo (NM) possuem dois componentes genômicos denominados DNA-A e DNA-B, e seu genoma tem a capacidade de evoluir rapidamente via mutação, pseudo-recombinação e recombinação. Neste contexto, este trabalho objetivou: (*i*) avaliar o efeito do hospedeiro na variabilidade genética de populações de begomovírus; e (*ii*) estudar o efeito de recombinação na evolução de begomovírus no Novo Mundo. Para o primeiro objetivo, duas plantas leguminosas cultivadas (feijão-comum, *Phaseolus vulgaris*, e fava, *P. lunatus*) e uma não-cultivada (*Macroptilium lathyroides*) foram intensamente amostradas em duas regiões no Brasil entre 2005 e 2012. Um total de 212 genomas (DNA-A) foram clonados e sequenciados, e populações dos begomovírus *Bean golden mosaic virus* (BGMV) e *Macroptilium yellow spot virus* (MaYSV) foram obtidas a partir dos três hospedeiros. Os resultados indicam que a variabilidade genética presumida dos hospedeiros não afetou a variabilidade viral. O MaYSV (N = 99) apresentou maior variabilidade genética que o BGMV (N = 147), com a população de BGMV (porém não de MaYSV) estruturada por hospedeiro e região geográfica. Para o segundo objetivo, conjuntos de dados incluindo todas as sequências-referência de DNA-A e DNA-B de begomovírus do NM foram obtidas do GenBank. As análises indicaram que begomovírus sul-americanos não exibem monofilia geográfica, com evidência de pelo menos dois eventos independentes de introdução nesta região.

Recombinação foi detectada como um importante mecanismo evolutivo agindo sobre a evolução do DNA-A. Encontrou-se evidência de elevada variabilidade genética em begomovírus presentes na América Central e Caribe, e em menor grau em begomovírus na América do Sul (AS). Novos eventos de introdução podem acelerar a evolução desses patógenos na AS, favorecendo a recombinação e aumentando a adaptabilidade e/ou a virulência dos begomovírus nativos.

ABSTRACT

RAMOS-SOBRINHO, Roberto, D.Sc., Universidade Federal de Viçosa, February 2014. **Begomovirus evolution in the New World: Genetic variability of populations and its effect on speciation.** Adviser: Francisco Murilo Zerbini Júnior. Committee member: Eduardo Seiti Gomide Mizubuti.

Viruses belonging to the family *Geminiviridae* have circular ssDNA genomes and infect a broad range of plant species causing devastating diseases, mainly in subtropical and tropical countries. The family is divided into seven genera (*Begomovirus*, *Becurtovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*) according to the type of insect vector, host range, genome organization and phylogeny. Begomoviruses are whitefly-transmitted and are among the most damaging pathogens causing epidemics in economically important crops worldwide. Wild/non-cultivated plants play a crucial epidemiological role, acting as begomovirus reservoirs and as "mixing vessels" where recombination can occur. Furthermore, previous results suggest that begomovirus populations in non-cultivated hosts are more genetically variable then those infecting cultivated hosts. Most New World (NW) begomoviruses have two genomic components designated as DNA-A and DNA-B, and their genomes have the capacity to evolve quickly via mutation, reassortment and recombination. In this context, this work aimed: (*i*) to assess the effect of the host on the standing genetic variability of begomovirus populations; and (*ii*) to study the effect of recombination on the evolution of New World begomoviruses. For the first objective, cultivated (common bean, *Phaseolus vulgaris*, and lima bean, *P. lunatus*) and non-cultivated (*Macroptilium lathyroides*) legume hosts were intensively sampled from two regions across Brazil between 2005 and 2012. A total of 212 full-length DNA-A components were cloned and sequenced, and populations of the begomoviruses *Bean golden mosaic virus* (BGMV) and *Macroptilium yellow spot virus* (MaYSV) were obtained from the three hosts. Our results indicate that the presumed genetic variability of the host did not affect viral variability. MaYSV (N = 99) showed higher genetic variability than BGMV (N = 147), with the BGMV (but not the MaYSV) population being structured based on both host and geography. For the second objective, datasets including all DNA-A and DNA-B reference sequences of NW begomoviruses were obtained from GenBank. Our analyses indicate that South American begomoviruses do not exhibit geographic monophyly, with evidence of at least two independent introduction events in this region. Recombination was detected as a very important evolutionary mechanism acting on DNA-A evolution. We found evidence of

greater genetic variability in begomoviruses from Central America and the Caribbean compared to South America (SA). Additional introduction events may impact the evolution of begomoviruses in SA by favoring recombination and introducing new virulence features to indigenous South American begomoviruses.

INTRODUCTION

Viruses belonging to the family *Geminiviridae* have circular ssDNA genomes and infect a broad range of plant species causing devastating diseases, mainly in subtropical and tropical countries (Legg & Fauquet, 2004; Morales, 2006). The family is divided into seven genera (*Begomovirus*, *Becurtovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*) according to the type of insect vector, host range, genome organization and phylogeny (Varsani *et al.*, 2014). Viruses classified within the genus *Begomovirus* are transmitted by the whitefly *Bemisia tabaci* and constitute one of the largest and most important groups of emerging plant viruses, growing in importance as the number of species characterized and their economic impact continues to increase (Brown *et al.*, 2012). Begomoviruses specifically limit production of tomatoes, peppers and legumes in the New World (the Americas) (Morales, 2006; Navas-Castillo *et al.*, 2011). In Brazil, *Bean golden mosaic virus* (BGMV) has been an important pathogen infecting beans (*Phaseolus* spp.) since the 1970s (Costa, 1976; Gilbertson *et al.*, 1991; Faria & Maxwell, 1999), causing yield losses between 40 to 100% (Morales, 2006). Begomoviruses are unique among plant viruses in their strong geographic segregation, with New World and Old World begomoviruses forming clearly distinct phylogenetic clades (Briddon *et al.*, 2010). Recently, phylogeographic segregation of tomato-infecting begomoviruses was observed in Brazil, where different viral species were prevalent in different tomato growing areas (Rocha *et al.*, 2013).

Begomoviruses from the Old World (Europe, Asia, Africa and Australia) have one or two genomic components (known as mono- and bipartite begomoviruses, respectively), and are often associated with circular ssDNA molecules designated as alphasatellites (previously DNA-1) and betasatellites (previously DNA β) (Briddon *et al.*, 2003; Briddon & Stanley, 2006). Alphasatellites have been recently identified in Brazil and Venezuela in association with the bipartite begomoviruses *Cleome leaf crumple virus* (ClLCrV), *Euphorbia mosaic*

virus (EuMV) and *Melon chlorotic mosaic virus* (MeCMV) (Paprotka *et al.*, 2010b; Romay *et al.*, 2010).

Most Begomoviruses from the New World have two genomic components known as DNA-A and DNA-B (Fauquet *et al.*, 2005). *Tomato leaf deformation virus* (ToLDeV), which has only a single, DNA-A-like component, was the first monopartite begomovirus reported to naturally occur in the New World (Melgarejo *et al.*, 2013). The two components of bipartite begomoviruses do not share significant sequence identity, except for a region with approximately 200 nt known as the common region (CR), which includes the origin of replication (Hanley-Bowdoin *et al.*, 1999). The genome of bipartite begomoviruses encodes six to eight proteins: the replication-associated protein (Rep), which is the initiator of rolling circle replication and has nucleic acid binding, endonuclease and ATPase activities (Fontes *et al.*, 1992; Orozco *et al.*, 1997); the trans-activating protein (TrAP), a transcriptional factor of the *cp* and *nsp* genes and also a suppressor of post-transcriptional gene silencing (Sunter & Bisaro, 1992; Voinnet *et al.*, 1999; Wang *et al.*, 2005); the replication-enhancer protein (Ren), an accessory factor required for optimal replication (Sunter *et al.*, 1990; Pedersen & Hanley-Bowdoin, 1994), and the coat protein (CP), responsible for encapsidation of the ssDNA genome and also essential for insect transmission (Briddon *et al.*, 1990; Hofer *et al.*, 1997). The V2 and AV2 proteins function as anti-defense proteins to inhibit post-transcriptional gene silencing. V2 also provides the movement function for monopartite viruses (Hanley-Bowdoin *et al.*, 2013). Some begomoviruses encode the AC4 protein, which is involved with gene silencing suppression (Hanley-Bowdoin *et al.*, 2013). Two proteins are encoded by the DNA-B: the nuclear shuttle protein (NSP) and the movement protein (MP), which are involved in intra- and intercellular viral movement, respectively (Noueiry *et al.*, 1994).

The sequence of the origin of replication (*ori*) is conserved among components of the same virus but is variable among species, except for a ~30 nt sequence conserved among all species (Davies *et al.*, 1987; Lazarowitz, 1992). This sequence comprises an inverted GC-rich

repeat which forms a hairpin structure carrying an invariant nonanucleotide sequence (5'-TAATATTAC-3') found in all begomoviruses (Heyraud-Nitschke *et al.*, 1995; Orozco & Hanley-Bowdoin, 1998). Cleavage of the nonanucleotide is essential to the initiation of replication and is performed by the Rep protein (Fontes *et al.*, 1994; Laufs *et al.*, 1995; Orozco *et al.*, 1998). The Rep catalytic domain (located at the N-terminal portion of the protein) includes three motifs conserved in proteins involved in rolling circle replication (motif I: FLTY, motif II: HxH and motif III: YxxxV) (Ilyina & Koonin, 1992) and the iteron-related domain (IRD) (Arguello-Astorga & Ruiz-Medrano, 2001).

Begomovirus populations exhibit high molecular variability (Ariyo *et al.*, 2005; Ge *et al.*, 2007; Silva *et al.*, 2011; Silva *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013). The main evolutionary processes shaping the molecular variability of plant virus populations are mutation and recombination (García-Arenal *et al.*, 2003; Seal *et al.*, 2006; Lima *et al.*, 2013). Viruses with divided genomes (including the bipartite begomoviruses) may also evolve by pseudorecombination (reassortment), in which whole genomic components are exchanged amongst distinct viruses without intermolecular recombination (García-Arenal *et al.*, 2001). Experimental evidence of the ability of begomoviruses to form viable reassortants has been obtained in laboratory settings (Gilbertson *et al.*, 1993; Sung & Coutts, 1995; Andrade *et al.*, 2006), although there are limited reports of their natural occurrence under field conditions (Paplomatas *et al.*, 1994; Pita *et al.*, 2001).

Mutation is the primary source of variability in plant virus populations (Roossinck, 1997; García-Arenal *et al.*, 2001; 2003; Lima *et al.*, 2013) and previous studies have shown that ssDNA viruses may evolve as quickly as RNA viruses which use an error-prone RNA-dependent RNA polymerase for replicating their genomes (Drake, 1991; Shackelton *et al.*, 2005; Shackelton & Holmes, 2006; Duffy *et al.*, 2008; Duffy & Holmes, 2009). Geminiviruses exhibit high levels of within-host molecular variability (Isnard *et al.*, 1998; Ge *et al.*, 2007; Van Der Walt *et al.*, 2008; Lima *et al.*, 2013; Rocha *et al.*, 2013) and there is

evidence that their rapid evolution might be, at least in part, driven by mutational processes acting specifically on ssDNA (Duffy *et al.*, 2008; Harkins *et al.*, 2009). Substitution rates for temporally sampled *Tomato yellow leaf curl virus* (TYLCV) genome sequences were estimated to be 2.88×10^{-4} subs/site/year for the full-length genome and 4.63×10^{-4} subs/site/year for the CP (Duffy & Holmes, 2008), values which are similar to those of RNA viruses (Holland *et al.*, 1982; Domingo & Holland, 1997; Roossinck, 2003). These high substitution rates were validated for the bipartite begomovirus *East African cassava mosaic virus* (EACMV) (Duffy & Holmes, 2009), suggesting that they are representative of multiple begomoviruses. The dN/dS ratios calculated for all coding sequences analyzed in these studies were lower than 1.0, indicating purifying selection and, consequently, that the substitutions rates were not overestimated due to adaptive selection, reflecting a rapid mutational dynamics for these viruses (Duffy & Holmes, 2008; 2009).

Recombination is the process by which a DNA or RNA segment becomes incorporated into a different strand during replication (Padidam *et al.*, 1999). Recombination is a common evolutionary process acting on geminivirus genomes (Padidam *et al.*, 1999), and seems to contribute heavily to the standing molecular variability of begomoviruses, increasing their evolutionary potential and local adaptation (Harrison & Robinson, 1999; Padidam *et al.*, 1999; Berrie *et al.*, 2001; Monci *et al.*, 2002; Lima *et al.*, 2013; Rocha *et al.*, 2013). The high recombination frequency in this group of viruses can be partly explained by recombination-dependent replication (RDR) (Jeske *et al.*, 2001; Preiss & Jeske, 2003) in addition to the well-documented rolling-circle replication (RCR) (Saunders *et al.*, 2001). Furthermore, the frequent occurrence of mixed infections (Torres-Pacheco *et al.*, 1996; Harrison *et al.*, 1997; Sanz *et al.*, 2000; Pita *et al.*, 2001; Ribeiro *et al.*, 2003; García-Andrés *et al.*, 2006; Davino *et al.*, 2009) in which more than one virus can replicate simultaneously in the same nucleus (Morilla *et al.*, 2004) also favors the occurrence of recombination.

Recombination events have been directly implicated in the emergence of new begomovirus diseases and epidemics in cultivated hosts (Zhou *et al.*, 1997; Pita *et al.*, 2001; Monci *et al.*, 2002). In fact, the devastating epidemics of cassava mosaic disease caused by EACMV in Uganda and neighboring countries (Zhou *et al.*, 1997; Pita *et al.*, 2001), the TYLCV epidemics in the western Mediterranean basin (Monci *et al.*, 2002; García-Andrés *et al.*, 2006; García-Andrés *et al.*, 2007), and the epidemics of *Cotton leaf curl virus* (CLCuV) in Pakistan (Zhou *et al.*, 1997; Idris & Brown, 2002) were all caused by a complex of begomoviruses including several recombinants.

A comparative analysis using genomic sequences of ssDNA viruses from several families revealed a conserved, non-random pattern of distribution of recombination breakpoints (Lefeuvre *et al.*, 2009). Although the mechanistic aspects of recombination in ssDNA viruses remain unknown (Padidam *et al.*, 1999), this non-random distribution of recombination breakpoints is conserved amongst mono- and bipartite viruses, with hot spots in the 5'-portion of the Rep gene and the 5'-end of the common region (Lefeuvre *et al.*, 2007a; Lefeuvre *et al.*, 2007b). Evidence suggests that the recombination breakpoints tend to occur outside or on the periphery of coding sequences. In agreement, a reduced number of breakpoints have been found within genes encoding structural proteins, such as the *cp* gene (Lefeuvre *et al.*, 2007a). The high nucleotide variability in the N-terminal portion of the Rep gene is accompanied by strong purifying selection that preserves the amino acid sequence (Lima *et al.*, 2013; Rocha *et al.*, 2013). These results suggest that natural selection acting against viruses expressing recombinant proteins is an important determinant of the non-random distribution of recombination breakpoints in most ssDNA viruses (Lefeuvre *et al.*, 2009). Furthermore, it has been shown that recombination events that preserve co-evolved intragenome interactions (protein-protein and/or protein-DNA) are also favored by selection (Martin *et al.*, 2011).

The advent of rolling-circle amplification using the phi29 DNA polymerase (Inoue-Nagata *et al.*, 2004) has greatly facilitated the characterization of novel begomoviruses infecting a number of cultivated and non-cultivated hosts (Castillo-Urquiza *et al.*, 2008; Varsani *et al.*, 2009; Paprotka *et al.*, 2010a; Wyant *et al.*, 2011; Silva *et al.*, 2012; Tavares *et al.*, 2012) as well as the rapid cloning of a large number of genomes, allowing studies of viral population genetics on a genomic scale (Haible *et al.*, 2006; Silva *et al.*, 2011; González-Aguilera *et al.*, 2012; Silva *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013). Such studies suggested that begomovirus populations infecting wild/non-cultivated hosts have a high degree of genetic variability (Silva *et al.*, 2011; Silva *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013), whereas, begomovirus populations infecting cultivated hosts seem to have lower molecular variation (Faria & Maxwell, 1999; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013). This higher genetic variability in non-cultivated hosts seems to be explained mostly by recombination events (Lima *et al.*, 2013). Therefore, we hypothesized that non-cultivated hosts, which are more genetically variable, would harbor more genetically variable virus sequences, whereas cultivated hosts would select for fewer haplotypes.

In this context, this work aimed: (*i*) to assess the effect of the host on the standing genetic variability of begomovirus populations; (*ii*) to study the effect of recombination on the evolution of New World begomoviruses.

Literature Cited

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.

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

ARIYO, O.A.; KOERBLER, M.; DIXON, A.G.O.; ATIRI, G.I.; WINTER, S. Molecular variability and distribution of Cassava mosaic begomoviruses in Nigeria. Journal of Phytopathology, v. 153, p. 226-231, 2005.

BERRIE, L.C.; RYBICKI, E.P.; REY, M.E.C. Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. Journal of General Virology, v. 82, p. 53-58., 2001.

BRIDDON, R.W.; BULL, S.E.; AMIN, I.; IDRIS, A.M.; MANSOOR, S.; BEDFORD, I.D.; DHAWAN, P.; RISHI, N.; SIWATCH, S.S.; ABDEL-SALAM, A.M.; BROWN, J.K.; ZAFAR, Y.; MARKHAM, P.G. Diversity of DNA beta, a satellite molecule associated with some monopartite begomoviruses. Virology, v. 312, p. 106-121, 2003.

BRIDDON, R.W.; PATIL, B.L.; BAGEWADI, B.; NAWAZ-UL-REHMAN, M.S.; FAUQUET, C.M. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. BMC Evolutionary Biology, v. 10, p. 97, 2010.

BRIDDON, R.W.; PINNER, M.S.; STANLEY, J.; MARKHAM, P.G. Geminivirus coat protein gene replacement alters insect specificity. Virology, v. 177, p. 85-94, 1990.

BRIDDON, R.W.; STANLEY, J. Subviral agents associated with plant single-stranded DNA viruses. Virology, v. 344, p. 198-210, 2006.

BROWN, J.K.; FAUQUET, C.M.; BRIDDON, R.W.; ZERBINI, F.M.; MORIONES, E.; NAVAS-CASTILLO, J. Family *Geminiviridae*. In: KING, A.M.Q.; ADAMS, M.J.; CARSTENS, E.B.; LEFKOWITZ, E.J. (Eds.) Virus Taxonomy. 9th Report of the International Committee on Taxonomy of Viruses. London, UK: Elsevier Academic Press, 2012. p. 351-373.

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. Whitefly-transmitted plant diseases. Annual Review of Phytopathology, v. 14, p. 429-440, 1976.

DAVIES, J.W.; STANLEY, J.; DONSON, J.; MULLINEAUX, P.M.; BOULTON, M.I. Structure and replication of geminivirus genomes. Journal of Cell Science, v. 7, p. 95-107, 1987.

DAVINO, S.; NAPOLI, C.; DELLACROCE, C.; MIOZZI, L.; NORIS, E.; DAVINO, M.; ACCOTTO, G.P. Two new natural begomovirus recombinants associated with the tomato yellow leaf curl disease co-exist with parental viruses in tomato epidemics in Italy. Virus Research, v. 143, p. 15-23, 2009.

DOMINGO, E.; HOLLAND, J.J. RNA virus mutations and fitness for survival. Annual Review of Microbiology, v. 51, p. 151-178, 1997.

DRAKE, J.W. A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of the National Academy of Sciences, USA, v. 88, p. 7160-7164, 1991.

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

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

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

FARIA, J.C.; MAXWELL, D.P. Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil. Phytopathology, v. 89, p. 262-268, 1999.

FAUQUET, C.M.; MAYO, M.A.; MANILOFF, J.; DESSELBERGER, U.; BALL, L.A. (Eds.) Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press. 1259 p. 2005.

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

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

GARCÍA-ANDRÉS, S.; MONCI, F.; NAVAS-CASTILLO, J.; MORIONES, E. Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. Virology, v. 350, p. 433-442, 2006.

GARCÍA-ANDRÉS, S.; TOMAS, D.M.; SANCHEZ-CAMPOS, S.; NAVAS-CASTILLO, J.; MORIONES, E. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. Virology, v. 365, p. 210-219, 2007.

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

GARCÍA-ARENAL, F.; 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.

GILBERTSON, R.L.; HIDAYAT, S.H.; MARTINEZ, R.T.; LEONG, S.A.; FARIA, J.C.; MORALES, F.J.; MAXWELL, D.P. Diferentiation of bean-infecting geminiviruses by nucleic acid hybridization probes and aspects of bean golden mosaic in Brazil. Plant Disease, v. 75, p. 336-342, 1991.

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

GONZÁLEZ-AGUILERA, J.; TAVARES, S.S.; SOBRINHO, R.R.; XAVIER, C.A.D.; DUEÑAS-HURTADO, F.; LARA-RODRIGUES, R.M.; SILVA, D.J.H.D.; ZERBINI, F.M. Genetic structure of a Brazilian population of the begomovirus *Tomato severe rugose virus* (ToSRV). Tropical Plant Pathology, v. 37, p. 346-353, 2012.

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.

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

HANLEY-BOWDOIN, L.; BEJARANO, E. R.; ROBERTSON, D.; MANSOOR S. Geminiviruses: masters at redirecting and reprogramming plant processes. Nature Reviews Microbiology, v. 11, p. 777-788, 2013.

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.

HARRISON, B.D.; ROBINSON, D.J. Natural genomic and antigenic variation in white-fly transmitted geminiviruses (begomoviruses). Annual Review of Phytopathology, v. 39, p. 369-398, 1999.

HARRISON, B.D.; ZHOU, X.; OTIM NAPE, G.W.; LIU, Y.; ROBINSON, D.J. Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. Annals of Applied Biology, v. 131, p. 437-448, 1997.

HEYRAUD-NITSCHKE, F.; SCHUMACHER, S.; LAUFS, J.; SCHAEFER, S.; SCHELL, J.; GRONENBORN, B. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. Nucleic Acids Research, v. 23, p. 910-916, 1995.

HOFER, P.; BEDFORD, I.D.; MARKHAM, P.G.; JESKE, H.; FRISCHMUTH, T. Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. Virology, v. 236, p. 288-295, 1997.

HOLLAND, J.; SPINDLER, K.; HORODYSKI, F.; GRABAU, E.; NICHOL, S.; VANDEPOL, S. Rapid evolution of RNA genomes. Science, v. 215, p. 1577-1585, 1982.

IDRIS, A.M.; BROWN, J.K. Molecular analysis of Cotton leaf curl virus-Sudan reveals an evolutionary history of recombination. Virus Genes, v. 24, p. 249-256, 2002.

ILYINA, T.V.; KOONIN, E.V. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Research, v. 20, p. 3279-3285, 1992.

INOUE-NAGATA, A.K.; ALBUQUERQUE, L.C.; ROCHA, W.B.; NAGATA, T. A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. Journal of Virological Methods, v. 116, p. 209-211, 2004.

ISNARD, M.; GRANIER, M.; FRUTOS, R.; REYNAUD, B.; PETERSCHMITT, M. Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. Journal of General Virology, v. 79, p. 3091-3099, 1998.

JESKE, H.; LUTGEMEIER, M.; PREISS, W. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. EMBO Journal, v. 20, p. 6158-6167, 2001.

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

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

LEFEUVRE, P.; LETT, J.M.; REYNAUD, B.; MARTIN, D.P. Avoidance of protein fold disruption in natural virus recombinants. PLoS Pathogens, v. 3, p. e181, 2007a.

LEFEUVRE, P.; LETT, J.M.; VARSANI, A.; MARTIN, D.P. Widely conserved recombination patterns among single-stranded DNA viruses. Journal of Virology, v. 83, p. 2697-2707, 2009.

LEFEUVRE, P.; MARTIN, D.P.; HOAREAU, M.; NAZE, F.; DELATTE, H.; THIERRY, M.; VARSANI, A.; BECKER, N.; REYNAUD, B.; LETT, J.M. Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. Journal of General Virology, v. 88, p. 3458-3468, 2007b.

LEGG, J.; FAUQUET, C. Cassava mosaic geminiviruses in Africa. Plant Molecular Biology, v. 56, p. 585-599, 2004.

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 higher genetic variability in begomovirus populations infecting non-cultivated hosts. Journal of General Virology, v. 94, p. 418-431, 2013.

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

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.

MORALES, F.J. History and current distribution of begomoviruses in Latin America. Advances in Virus Research, v. 67, p. 127-162, 2006.

MORILLA, G.; KRENZ, B.; JESKE, H.; BEJARANO, E.R.; WEGE, C. Tête-à-tête of tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single nuclei. Journal of Virology, v. 78, p. 10715-10723, 2004.

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

NOUEIRY, A.O.; LUCAS, W.J.; GILBERTSON, R.L. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell, v. 76, p. 925-932, 1994.

OROZCO, B.M.; GLADFELTER, H.J.; SETTLAGE, S.B.; EAGLE, P.A.; GENTRY, R.N.; HANLEY-BOWDOIN, L. Multiple *cis* elements contribute to geminivirus origin function. Virology, v. 242, p. 346-356, 1998.

OROZCO, B.M.; HANLEY-BOWDOIN, L. Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. Journal of Biological Chemistry, v. 273, p. 24448-24456, 1998.

OROZCO, B.M.; MILLER, A.B.; SETTLAGE, S.B.; HANLEY-BOWDOIN, L. Functional domains of a geminivirus replication protein. Journal of Biological Chemistry, v. 272, p. 9840-9846, 1997.

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

PAPLOMATAS, E.J.; PATEL, V.P.; HOU, Y.M.; NOUEIRY, A.O.; GILBERTSON, R.L. Molecular characterization of a new sap-transmissible bipartite genome geminivirus infecting tomatoes in Mexico. Phytopathology, v. 84, p. 1215-1224, 1994.

PAPROTKA, T.; BOITEUX, L.S.; FONSECA, M.E.N.; RESENDE, R.O.; JESKE, H.; FARIA, J.C.; RIBEIRO, S.G. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. Virus Research, v. 149, p. 224-233, 2010a.

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, 2010b.

PEDERSEN, T.J.; HANLEY-BOWDOIN. Molecular characterization of the AL3 protein encoded by a bipartite geminivirus. Virology, v. 202, p. 1070-1075, 1994.

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.

PREISS, W.; JESKE, H. Multitasking in replication is common among geminiviruses. Journal of Virology, v. 77, p. 2972-2980, 2003.

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.

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.

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.

ROOSSINCK, M.J. Mechanisms of plant virus evolution. Annual Review of Phytopathology, v. 35, p. 191-209, 1997.

ROOSSINCK, M.J. Plant RNA virus evolution. Current Opinion in Microbiology, v. 6, p. 406-409, 2003.

SANZ, A.I.; FRAILE, A.; GARCÍA-ARENAL, F.; ZHOU, X.; ROBINSON, D.J.; KHALID, S.; BUTT, T.; HARRISON, B.D. Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. Journal of General Virology, v. 81, p. 1839-1849, 2000.

SAUNDERS, K.; BEDFORD, I.D.; STANLEY, J. Pathogenicity of a natural recombinant associated with ageratum yellow vein disease: implications for geminivirus evolution and disease aetiology. Virology, v. 282, p. 38-47, 2001.

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

SHACKELTON, L.A.; HOLMES, E.C. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. Journal of Virology, v. 80, p. 3666-3669, 2006.

SHACKELTON, L.A.; PARRISH, C.R.; TRUYEN, U.; HOLMES, E.C. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proceedings of the National Academy of Sciences, USA, v. 102, p. 379-384, 2005.

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.

SUNG, Y.K.; COUTTS, R.H. Pseudorecombination and complementation between potato yellow mosaic geminivirus and tomato golden mosaic geminivirus. Journal of General Virology, v. 76, p. 2809-2815., 1995.

SUNTER, G.; BISARO, D.M. Transactivation of geminivirus AR1 and BR2 gene expression by the viral AL2 gene product occurs at the level of transcription. Plant Cell, v. 4, p. 1321-1331, 1992.

SUNTER, G.; HARTITZ, M.D.; HORMUZDI, S.G.; BROUGH, C.L.; BISARO, D.M. Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for eficient DNA replication. Virology, v. 179, p. 69-77, 1990.

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.

TORRES-PACHECO, I.; GARZÓN-TIZNADO, J.A.; BROWN, J.K.; BECERRA-FLORA, A.; RIVERA-BUSTAMANTE, R. Detection and distribution of geminiviruses in Mexico and the Southern United States. Phytopathology, v. 86, p. 1186-1192, 1996.

VAN DER WALT, E.; MARTIN, D.P.; VARSANI, A.; POLSTON, J.E.; RYBICKI, E.P. Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. Virology Journal, v. 5, p. 104, 2008.

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.; DENT, K.; MONJANE, A.L.; RYBICKI, E.P.; MARTIN, D.P. A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. Virology Journal, v. 6, p. 36, 2009.

VOINNET, O.; PINTO, Y.M.; BAULCOMBE, D.C. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proceedings of the National Academy of Sciences, USA, v. 96, p. 14147-14152, 1999.

WANG, H.; BUCKLEY, K.J.; YANG, X.; BUCHMANN, R.C.; BISARO, D.M. Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. Journal of Virology, v. 79, p. 7410-7418, 2005.

WYANT, P.S.; GOTTHARDT, D.; SCHAFER, B.; KRENZ, B.; JESKE, H. The genomes of four novel begomoviruses and a new *Sida micrantha mosaic virus* strain from Bolivian weeds. Archives of Virology, v. 156, p. 347-352, 2011.

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.

CHAPTER 1

CONTRASTING GENETIC STRUCTURE BETWEEN TWO BEGOMOVIRUSES INFECTING THE SAME LEGUMINOUS HOSTS

Ramos-Sobrinho, R., Xavier, C.A.D., Pereira, H.M.B., Lima, G.S.A., Assunção, I.P., Mizubuti, E.S.G., Duffy, S., Zerbini, F.M. Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. Journal of General Virology, *submitted.*

Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts

Roberto Ramos Sobrinho[a], César Augusto Diniz Xavier[a], Hermano Monteiro de Barros Pereira[a], Gaus Silvestre de Andrade Lima[b], Iraíldes Pereira Assunção[b], Eduardo Seiti Gomide Mizubuti[a], Siobain Duffy[c#], Francisco Murilo Zerbini[a#]

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

[b]Departamento de Fitossanidade/CECA, Universidade Federal de Alagoas, Rio Largo, AL, 57100-000, Brazil

[c]Department of Ecology, Evolution and Natural Resources, Rutgers, The State University Of New Jersey, New Brunswick, NJ, 08901, USA

[#]Corresponding author: Siobain Duffy

Phone: (+1-848) 932-6299; E-mail: duffy@aesop.rutgers.edu

[#]Corresponding author: Francisco Murilo Zerbini

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

Summary – Begomoviruses are whitefly-transmitted, single-stranded DNA plant viruses and are among the most damaging pathogens causing epidemics in economically important crops worldwide. Wild/non-cultivated plants play a crucial epidemiological role, acting as begomovirus reservoirs and as 'mixing vessels' where recombination can occur. Previous results suggest a higher degree of genetic variability in begomovirus populations from non-cultivated hosts compared to cultivated hosts. To assess this supposed host effect on the genetic variability of begomovirus populations, cultivated (common bean, *Phaseolus vulgaris*, and lima bean, *P. lunatus*) and non-cultivated (*Macroptilium lathyroides*) legume hosts were intensively sampled from two regions across Brazil. A total of 212 full-length DNA-A genome segments were sequenced from samples collected between 2005 and 2012, and populations of the begomoviruses *Bean golden mosaic virus* (BGMV) and *Macroptilium yellow spot virus* (MaYSV) were obtained. We found, for each begomovirus species, similar genetic variation between populations infecting cultivated and non-cultivated hosts, indicating that the presumed genetic variability of the host did not *a priori* affect viral variability. The MaYSV population (N = 99) was more variable than the BGMV population (N = 147), which was explained by numerous recombination events in MaYSV. MaYSV and BGMV showed distinct distributions of genetic variation, with the BGMV population (but not MaYSV) being structured by both host and geography.

Introduction

Viruses belonging to the family *Geminiviridae* have circular ssDNA genomes and are widely distributed in tropical and subtropical regions, infecting several economically important crop species (Legg & Fauquet, 2004; Morales, 2006). The family is divided into seven genera (*Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*) according to the type of insect vector, host range, genome organization and phylogeny (Varsani *et al.*, 2014). Viruses classified within the genus *Begomovirus* are transmitted by the whitefly *Bemisia tabaci* (Brown *et al.*, 2012). Begomoviruses usually found in the New World have two genomic components known as DNA-A and DNA-B, but only the DNA-A is used to properly assign taxonomy (Brown *et al.*, 2012). The DNA-A contains genes involved in replication and encapsidation of the viral progeny, while the DNA-B contains genes responsible for intra- and intercellular movement (Brown *et al.*, 2012). Begomoviruses are among the most damaging pathogens infecting cultivated plants worldwide (Legg & Fauquet, 2004; Morales & Anderson, 2001; Navas-Castillo *et al.*, 2011; Varma & Malathi, 2003), and specifically limit production of tomatoes, peppers and legumes in the New World (Morales, 2006; Navas-Castillo *et al.*, 2011). In Brazil, *Bean golden mosaic virus* (BGMV) has been an important pathogen infecting beans (*Phaseolus* spp.) since the 1970s (Costa, 1976; Faria & Maxwell, 1999; Gilbertson *et al.*, 1991b), causing yield losses between 40 to 100% (Morales, 2006). Despite its economic and social importance, few studies have focused on BGMV and other begomovirus populations infecting these crops (Faria & Maxwell, 1999; Lima *et al.*, 2013). Other begomovirus species, such as *Macroptilium yellow spot virus* (MaYSV) and *Sida micrantha mosaic virus* (SimMV), have been reported naturally infecting common bean and non-cultivated legume hosts, but there is no information about their epidemiological importance for bean crops in Brazil.

Wild/non-cultivated plants from different botanical families can sustain a high species diversity of begomoviruses (Castillo-Urquiza *et al.*, 2008; Fiallo-Olive *et al.*, 2012; García-

Andrés *et al.*, 2006; Silva *et al.*, 2012; Tavares *et al.*, 2012; Wyant *et al.*, 2011), and can play an important epidemiological role serving as alternate/reservoir hosts, preventing local extinctions of the virus when the cultivated host is absent (Alabi *et al.*, 2008; Barbosa *et al.*, 2009; Rocha *et al.*, 2013). In these cases, whiteflies transmitting begomoviruses between cultivated and non-cultivated hosts contribute to virus evolution and disease epidemics (Alabi *et al.*, 2008; Power, 2000). Additionally, mixed infections by different begomoviruses are common in wild hosts (Alabi *et al.*, 2008; García-Andrés *et al.*, 2006; Monde *et al.*, 2010), facilitating recombination among distantly related begomoviruses. Recombination is an important evolutionary mechanism in begomoviruses (Lefeuvre *et al.*, 2009; Lefeuvre *et al.*, 2007b; Lima *et al.*, 2013; Martin *et al.*, 2011; Martin *et al.*, 2005), and was associated with higher genetic variability found in begomoviruses populations infecting primarily non-cultivated hosts (Lima *et al.*, 2013).

We have carried out a large-scale study to obtain more information about the genetic structure and factors shaping genetic variability in begomovirus populations infecting legume hosts in Brazil. Foliar samples of common bean (*P. vulgaris*), lima bean (*P. lunatus*) and of the weed *Macroptilium lathyroides* (located near common bean or lima bean fields) were collected over an eight-year period from different regions where the begomoviruses MaYSV and BGMV were previously reported infecting cultivated and/or non-cultivated hosts (Gilbertson *et al.*, 1991a; Lima *et al.*, 2013; Silva *et al.*, 2012; Wyant *et al.*, 2012). We hypothesized that non-cultivated hosts, which are more genetically variable, would harbor more genetically variable virus sequences, whereas cultivated hosts would select for fewer haplotypes. Instead, we found that the presumed genetic variability of the host did not affect viral sequence variability. Our results corroborate previous studies indicating high genetic variability in MaYSV populations due to interspecific recombination (Lima *et al.*, 2013; Silva *et al.*, 2012) and low genetic variability in BGMV populations (Faria & Maxwell, 1999). Nevertheless, MaYSV and BGMV showed different biogeographical patterns, and the genetic

structure of BGMV populations was strongly shaped by geography and host, while that of MaYSV populations was not.

Results

A total of 515 plant samples (300 common bean, 115 lima bean and 100 *M. lathyroides*) were collected. From these samples, 212 full-length DNA-A components were cloned and sequenced (each component was sequenced from individual samples; Suppl. Table S1).

Using pairwise comparisons of the DNA-A sequences and the ≥89% nucleotide identity criterion established by *Geminiviridae* Study Group of the ICTV (Brown *et al.*, 2012), the 212 isolates were assigned to four different begomovirus species: BGMV, MaYSV, *Macroptilium yellow vein virus* (MaYVV) and *Soybean chlorotic spot virus* (SoCSV). The nucleotide identity percentage ranged from 89%-100% between isolates belonging to the same species and 79%-90% between isolates belonging to different species. The higher than expected identity of 90% between species was only found for certain isolates of MaYSV and SoCSV (data not shown), but SoCSV is a potential parent of a recombination event within MaYSV (see below).

A total of 152 BGMV and 99 MaYSV full-length DNA-A sequences were analyzed. The MaYSV dataset contained 55 sequences previously published (Lima *et al.*, 2013; Silva *et al.*, 2012) plus 44 new sequences described here, and the BGMV dataset contained five sequences previously published (Fernandes *et al.*, 2009; Gilbertson *et al.*, 1991a; Silva *et al.*, 2012) plus 147 new sequences described here (Suppl. Tables S1 and S2). Additional representative sequences of other begomovirus species from Brazil were retrieved from GenBank for comparative analysis (Suppl. Table S2).

The prevalence of begomoviruses in leguminous hosts in Brazil shifts temporally and spatially

MaYSV was cloned from the three host species, but was found only in the northeastern region (states of Alagoas, AL, and Sergipe, SE). MaYSV is known to naturally infect non-cultivated plants (*M. lathyroides*, *Calopogonium mucunoides* and *Canavalia* sp.) as well as common bean (Lima *et al.*, 2013; Silva *et al.*, 2012). Our results extend its natural host range to include the cultivated *P. lunatus*. BGMV was cloned from lima bean in AL, common bean in the central region (states of Minas Gerais, MG, Goiás, GO, and the Federal District, DF), and *M. lathyroides* in MG. SoCSV was found in *M. lathyroides* samples from MG and MaYVV was only isolated from *M. lathyroides* plants adjacent to a lima bean field in AL (Suppl. Table S1).

The diversity of begomoviruses infecting *P. lunatus* in the northeast was different between our sampling times. In 2005, only BGMV isolates were obtained from this host in AL and Pernambuco (PE), but in 2011, nearly a third of the clones obtained from *P. lunatus* were assigned as MaYSV. In two surveys in 2011 (this study and Lima *et al.*, 2013), MaYSV was the begomovirus most frequently cloned from legume hosts in AL (87 of 146 clones obtained), while BGMV and MaYVV were less prevalent (41 and 18 clones, respectively). These results highlight the rapid emergence of MaYSV in legume crops in AL, where it is now the prevalent begomovirus.

MaYSV populations are more variable than BGMV populations

The genetic variability of MaYSV was evenly distributed among isolates infecting the different hosts, and was similar to the variability calculated for all isolates (Table 1). Among the different BGMV subpopulations from each of the three hosts, isolates infecting lima bean were slightly more variable than those from other hosts, despite a smaller population size than isolates from common bean (Table 1). The genetic variability calculated for all BGMV

isolates was higher than that for each individual subpopulation (Table 1). The MaYSV Rep gene was four times more variable than its CP gene (Table 1), and the MaYSV population showed greater genetic variability than the BGMV populations, especially in the Rep gene (Table 1).

Recombination is very common in MaYSV but not in BGMV

To investigate putative recombination events, full-length DNA-A genomes of MaYSV and BGMV were analyzed using the RDP3 package (Table 2). For MaYSV, a complex pattern of recombination was found, with seven unique events. Six events had putative breakpoints in the Rep gene and the common region, and only one event was in the CP gene (Table 2). In contrast, only two putative recombination events were detected in BGMV (Table 2). These events were restricted to isolates infecting lima bean from AL, where almost all isolates differed from other BGMV isolates by a recombinant region with breakpoints in the Rep and CP genes, and where one isolate (BR:Ata2:05) has a recombinant region in its CP gene and the common region (Table 2).

The effect of recombination in the CP and Rep genes was visualized using neighbor-net analysis. The networks included MaYSV, BGMV and other begomoviruses that showed greatest nucleotide identity to putative recombinant regions detected in MaYSV and BGMV by RDP3 analysis (Table S2). In the MaYSV CP dataset, isolates BR:Bas1:09 and BR:Mac1:10 may also be recombinant, with another begomovirus donating sequence to their common ancestor (Fig. 1a). The network of MaYSV Rep sequences confirmed the complex pattern of recombination detected by RDP3 (both intra- and inter-species events; Fig. 1b). Isolates in Cluster I were closely related to *Sida yellow blotch virus* (SiYBV) and *Sida mosaic Alagoas virus* (SiMAlV, obtained from *Sida* sp.), isolates in Cluster II to SoCSV, and those in Cluster III to *Blainvillea yellow spot virus* (BlYSV, from *Blainvillea rhomboidea*) (Fig. 1b). Recombination was also suggested in the BGMV datasets, albeit at a lower degree (Fig. 2).

The *M. lathyroides*-infecting BGMV isolates BR:Car3:10 and BR:Car4:10 from AL were more closely placed to other begomovirus species in the CP network compared to the Rep network (Fig. 2). Similarly, the lima bean-infecting isolates BR:Ata2:05, BR:Rec1:05 and BR:Rec2:05 were located away from the main BGMV clusters in both the CP and Rep networks (Fig. 2).

Full-length DNA-A datasets including the begomovirus species used in the neighbor-net analysis were checked for recombination using RDP3. For MaYSV, the event involving isolates in Cluster I had as putative minor parents SiYBV and SiMAlV, and the single event involving the CP gene had as putative minor parents *Tomato interveinal chlorosis virus* (ToICV) and *Macroptilium yellow net virus* (MaYNV) (data not shown). For BGMV, both events represent intra-species recombination, with different BGMV isolates identified as the putative parents (data not shown).

BGMV but not MaYSV populations show genetic structuring by geography and host

The MaYSV CP and Rep ML phylogenetic trees are not congruent (Fig. 3). The Rep tree has four well-supported clades that are not resolved in the CP tree. In both trees there is no evidence for structuring by location or host species, as each clade includes isolates from different hosts and/or locations (Fig. 3). These results were confirmed by subdivision analysis based on full-length DNA-A sequences ($F_{st/host}$=0.0223; $F_{st/location}$=0.0794), which indicates that putative subpopulations have little genetic differentiation.

The BGMV CP and Rep ML phylogenetic trees are similar (Fig. 4), with strong evidence for population structuring both by location and host plant. Both trees showed three main clades, with isolates from AL (primarily from lima bean) being a separate population from MG, GO and DF (common bean) and MG (non-cultivated plants). Isolates from Florestal (MG) formed a sister subpopulation to isolates from Unaí (MG), GO and DF (Fig. 4). As Unaí is closer to fields located in the neighboring state of GO (*e.g.*, 85 km from

Cristalina, GO and 470 km from Florestal, MG), geographical proximity can explain the clustering observed in the phylogenetic trees. Two isolates (BR:Rec1:05 and BR:Rec2:05) from Recife (PE) clustered with MG isolates, suggesting that there has been migration between these states (Fig. 4). Based on F_{st} analysis for full-length DNA-A sequences, BGMV showed strong genetic differentiation between subpopulations isolated from different hosts/locations ($F_{st/host/location}$= 0.8667).

Purifying selection predominates in BGMV and MaYSV populations

To investigate the extent in which selection pressures have shaped the standing genetic variability in MaYSV and BGMV populations, CP and Rep datasets were analyzed using different maximum likelihood-based methods. Subpopulations were analyzed separately to minimize the impact of recombination on the results. Most datasets showed d_N/d_S mean ratios smaller than 1 (Table 3), indicating negative selection. Although neutral evolution may be occurring in the BGMV Rep infecting *Macroptilium* (d_N/d_S =1.0), this analysis contained only four distinct sequences, and a larger dataset needs to be analyzed before definitive conclusions can be drawn.

A higher number of sites were under statistically detectable negative selection than positive selection in both MaYSV and BGMV populations. For example, using the FEL method on the MaYSV CP dataset, only three sites (142, 163 and 250) were identified as under positive selection (with site 250 also detected by SLAC), while 72 sites were under negative selection (Suppl. Table S3). A few sites under positive selection were detected in MaYSV Rep datasets. In clusters I, II and III, one, six and two sites were detected by REL, REL and FEL, respectively (Suppl. Table S3). In the BGMV CP gene, seven sites under positive selection were identified in the population infecting *P. vulgaris* (Suppl. Table S3). A larger number of sites under positive selection were detected in the small (only four haplotypes) *Macroptilium*-infecting BGMV dataset: 28 sites were under positive selection,

with only two sites under negative selection (Suppl. Table S3). PARRIS did not identify any sites under positive selection in any of the datasets analyzed, emphasizing that negative selection is the most important selective force acting upon BGMV and MaYSV populations.

Discussion

The species diversity of begomoviruses has been surveyed extensively (Ala-Poikela *et al.*, 2005; Albuquerque *et al.*, 2012; Bull *et al.*, 2006; Castillo-Urquiza *et al.*, 2008; Fernandes *et al.*, 2009; García-Andrés *et al.*, 2006; Lozano *et al.*, 2009; Ndunguru *et al.*, 2005; Reddy *et al.*, 2005; Ribeiro *et al.*, 2003; Rothenstein *et al.*, 2006; Sserubombwe *et al.*, 2008; Tavares *et al.*, 2012), and some studies have investigated the genetic structure of viral populations in cultivated and non-cultivated hosts in different geographical regions (Ge *et al.*, 2007; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012; Silva *et al.*, 2011). Strikingly, in spite of its great economic importance as the most damaging viral pathogen of bean crops in South America (Morales, 2006), no studies on the genetic variability of BGMV have been conducted since Faria & Maxwell (1999) analyzed 20 partial DNA-A sequences. The need to properly assess such variability is highlighted by the recent approval of a genetically-modified BGMV-resistant common bean by the Brazilian National Biosafety Commission (http://www.ctnbio.gov.br/index.php/content/view/16526.html) in which resistance is due to RNA silencing (Bonfim *et al.*, 2007), a process which is highly sequence-specific (Raja *et al.*, 2010). Our study of 272 full-length DNA-A begomovirus sequences across Brazil found no effect of genetic background of the host species on molecular variability of begomovirus populations, confirmed the high genetic variability in MaYSV populations, and detected large differences between BGMV populations infecting each of three leguminous hosts.

BGMV was the predominant begomovirus species found infecting lima bean in northeastern Brazil in 2005, but comprised a minority of isolates sampled in this region in

2011. Instead, MaYSV was the main begomovirus infecting legumes in 2011 in AL, but was not present in the central region, where BGMV was still prevalent. MaYSV was reported for the first time infecting non-cultivated legumes in the northeast in 2010 (Silva *et al.*, 2012), common bean in AL in 2011 (Lima *et al.*, 2013) and now naturally infecting another cultivated host, *P. lunatus* (this work). The absence of MaYSV in samples from 2005 could be related to the cloning method used at that time, which was biased for previously known sequences. However, a study using cloning-independent (*i.e.*, unbiased) pyro-sequencing of cultivated and non-cultivated legume samples collected in 2003 and 2004 in three northeastern states (AL, PE and Bahia), also did not detect the presence of MaYSV (Wyant *et al.*, 2012). Our combined results provide strong evidence for the recent emergence of MaYSV as an important agricultural pathogen, for a rapid expansion of its host range and for its swift spread throughout northeastern Brazil. Interestingly, MaYSV seems to be replacing BGMV as the dominant begomovirus infecting the economically important common bean and lima bean crops in AL. If the trends in AL hold for other regions in Brazil, monitoring the spread of MaYSV would be critical to guarantee the success of BGMV-resistant transgenic common bean.

High genetic variability in begomovirus populations infecting different non-cultivated hosts has been reported (Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012; Silva *et al.*, 2011). Conversely, begomovirus populations infecting cultivated hosts seem to have lower variability (Faria & Maxwell, 1999; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013). Here, we found that viral sequence variability was similar in one non-cultivated and two cultivated hosts for both BGMV and MaYSV. Our results on BGMV, based on 152 full-length DNA-A sequences, are in agreement with Faria & Maxwell (1999). We found little genetic variability in common bean-infecting BGMV isolates, while the lima bean-infecting population, which showed evidence of recombination, was slightly more variable. The variability in MaYSV populations observed here was similar to previous reports (Lima *et*

al., 2013; Silva *et al.*, 2012), but when compared to other begomovirus species, MaYSV was 2.5 times more variable than the next most variable begomovirus reported in Brazil, BlYSV (González-Aguilera *et al.*, 2012; Rocha *et al.*, 2013; Silva *et al.*, 2011). We found additional begomovirus species infecting *M. lathyroides*, consistent with previously published reports where this host sustains a high species diversity of begomoviruses (Silva *et al.*, 2012).

Recombination is very common and important for the emergence of different geminiviruses, and occurs within species, between species and across genera in the family (García-Andrés *et al.*, 2007a; García-Andrés *et al.*, 2007b; Lefeuvre *et al.*, 2009; Lefeuvre *et al.*, 2007b; Monci *et al.*, 2002; Padidam *et al.*, 1999; Pita *et al.*, 2001; Schnippenkoetter *et al.*, 2001; Tiendrebeogo *et al.*, 2012; Umaharan *et al.*, 1998). Different studies have shown that it is also an important evolutionary mechanism acting within begomovirus populations in Brazil (Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012; Silva *et al.*, 2011), with detectable events of recombination occurring in regular, non-random regions in the genome (Lima *et al.*, 2013; Rocha *et al.*, 2013).

A complex recombination pattern was found in MaYSV. The more variable populations all showed evidence of recombination (*e.g.*, the frequently recombinant Rep gene in MaYSV), suggesting that recombination could, in part, explain the higher genetic variability found in MaYSV, consistent with previously published reports (Lima *et al.*, 2013). Although we are unable to provide definitive evidence that the parental sequences identified here are the actual parents, these results provide a strong indication that MaYSV evolved through recombination between begomoviruses infecting cultivated and non-cultivated hosts belonging to different botanical families. Similar results were observed for tomato-infecting begomoviruses in Brazil, which are thought to have arisen from interspecies recombination between begomoviruses infecting non-cultivated hosts (Ribeiro *et al.*, 2007; Rocha *et al.*, 2013). Although some studies indicate that a large number of recombinants arising from events between distantly related genomes are defective and probably would be removed from

the population by selection (Liu *et al.*, 1999; Martin *et al.*, 2005), our results continue to emphasize the importance of recombination for microevolution and macroevolution of agronomically important begomoviruses.

MaYSV isolates with detectable recombination in the Rep gene were found in all three hosts studied, but isolates with recombination in the CP region were only found twice, both in *M. lathyroide*s in 2010. The lack of spread of CP recombinants could be evidence that they suffer a greater fitness cost than Rep recombinants. Several studies have shown that selection pressure is an important factor determining which recombinants survive in viral populations (Lefeuvre *et al.*, 2007a; Martin *et al.*, 2011; Rokyta & Wichman, 2009; Simon-Loriere *et al.*, 2009), and purifying selection is strongest in the CP gene of begomoviruses (Lefeuvre *et al.*, 2009). Interestingly, recombination seems infrequent or selected against in BGMV populations, and the few putative events detected here were restricted to isolates located in AL.

The genetic structure of begomovirus populations is determined by mutation, recombination and the interplay between adaptation to host species and vector biotypes, and is influenced by the geographic distribution of the hosts, vectors and other begomoviruses (Lima *et al.*, 2013; Navas-Castillo *et al.*, 2011; Rocha *et al.*, 2013). Recently, phylogeographic segregation of begomoviruses infecting the same crop was observed in Brazil, where different species were prevalent in different tomato-growing areas (Rocha *et al.*, 2013). Here, similar results were observed for BGMV infecting cultivated and non-cultivated legume hosts.

MaYSV populations were not neatly subdivided according to host or sampled area, providing strong evidence of migration among fields. These phylogenetic results were confirmed by extremely low F_{st} values. In contrast, evidence of structuring by both host and geography was found for BGMV populations. However, related isolates were found infecting both lima bean and *M. lathyroides* in AL (Silva *et al.*, 2012), and infecting common bean and *M. lathyroides* in MG, which indicate that this segregation is not absolute. Furthermore, while

there appears to have been migration between fields located in PE and MG, BGMV isolates in northeastern and central Brazil cannot be considered to comprise one unified population.

The contrasting population structure between MaYSV and BGMV could be due to the distances among sampling sites. The MaYSV isolates are separated by a maximum of 240 km (Santana do Ipanema, AL and Barra de Santana, PB), but all but one isolate (BR:Bas1:10) were sampled no more than 105 km apart. BGMV was isolated within MG from fields 470 km apart (Unaí and Florestal), and the maximum distance between sampling sites was 1425 km (Cristalina, GO and Arapiraca, AL). Over this larger area, we observed clustering between isolates collected at closer locations, such as isolates from Unaí (MG) forming one population with isolates from GO and DF, whose sampling sites were no more than 85 km away. The observed lack of population structure for MaYSV may be due to sampling in a smaller region than BGMV, and that substructure could appear over larger distances.

Incidentally, our results could impact the implementation of BGMV-resistant transgenic common bean in Brazil. As this resistance is based on the highly sequence-specific mechanism of RNA silencing (Bonfim *et al.*, 2007), the effectiveness of transgenic cultivars must be assessed against the genetic variation in BGMV and other common bean-infecting begomoviruses present in Brazil. In particular, in AL where the highly variable MaYSV is widespread and efficiently infects common bean, the use of this cultivar may fail to control the losses due to begomoviruses. Fortunately, MaYSV seems to be currently restricted to a small part of northeastern Brazil, and efforts could be made to contain its spread into other regions. The far less variable BGMV found in the main areas growing common bean in central Brazil provides hope that the RNAi resistance may be durable under field conditions.

Purifying selection was the dominant force acting on BGMV and MaYSV CP and Rep genes, in agreement with most begomovirus studies (García-Andrés *et al.*, 2007a; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013; Sanz *et al.*, 1999; Silva *et al.*, 2012; Silva *et al.*, 2011). Positive selection was infrequently observed, and in both

species some selected sites in the CP are associated with insect transmission (*e.g.*, sites 123 for BGMV and 142 for MaYSV). The region between amino acids 123 and 152 is involved in virion assembly, systemic infection and insect transmissibility in monopartite begomoviruses (Caciagli *et al.*, 2009; Noris *et al.*, 1998) and with vector transmission in bipartite begomoviruses (Hohnle *et al.*, 2001; Kheyr-Pour *et al.*, 2000). Despite the long length of this transmission-determining region (compared to the concise motif conferring aphid transmission in potyviruses; James & Bryce, 2006), changes in only one (Kheyr-Pour *et al.*, 2000) or two (Hohnle *et al.*, 2001; Noris *et al.*, 1998) amino acids in this region will prevent whitefly transmission. Therefore, these positively-selected sites may be improving whitefly transmission in these begomovirus populations.

Although the N-terminal region in MaYSV Rep is known to be highly recombinant (and highly variable), the few sites we identified to be under positive selection were mostly located in the C-terminal region. Previously identified sites under positive selection in MaYSV Rep were hypothesized to be spurious results due to recombination events (Lima *et al.*, 2013), which is corroborated by our subpopulation-specific results. Interestingly, neutral evolution was observed in BGMV Rep (ω = 1.0) in the subpopulation infecting *M. lathyroides*. These results provide additional evidence that high nucleotide variability in the N-terminal portion of the Rep gene (*e.g.*, the MaYSV population) is accompanied by strong purifying selection that preserves the amino acid sequence. The Rep N-terminal region in geminiviruses includes conserved motifs essential for rolling-circle replication (Ilyina & Koonin, 1992; Koonin & Ilyina, 1992; Nash *et al.*, 2011). Conservation of the integrity of these elements is critical for successful infection cycles, despite the variation introduced by frequent recombination.

We have determined the genetic structure of two legume-infecting begomovirus populations. BGMV populations are less variable than MaYSV ones, mostly due to recombination acting upon MaYSV. BGMV populations are strongly structured by

geography/host, while MaYSV populations are not. These results suggest that genetic variability is an intrinsic viral property, rather than a malleable feature that could be affected by the host (also seen in Ge *et al.*, 2007). The emergence of the highly variable MaYSV in *Phaseolus* spp. could seriously complicate disease management in this important crop in Brazil and other countries in South America.

Methods

Sample collection and storage

Foliar samples with virus-like symptoms such as yellow mosaic, leaf curl and stunting were collected in common bean and lima bean fields in different states of Brazil. Lima bean (*P. lunatus*) landraces samples were collected in fields in PE in 2005 and AL states in 2005 and 2011 (detailed isolation information given in Suppl. Table S1). Common bean (*P. vulgaris*, carioca type) samples were collected in fields in AL in 2011 and in GO and MG states and in the DF in 2012 (Suppl. Table S1). Most common bean samples belonged to cultivar Pérola, except samples from AL, for which cultivar was not determined. Samples of non-cultivated *M. lathyroides* near common bean or lima bean fields were collected in AL and SE states in 2011 and MG in 2012 (Suppl. Table S1). For each sample collected in 2011 or 2012 the following information was recorded: plant species (and cultivar for the common bean samples), date of collection, GPS coordinates of the sampling location and a digital image of the sample at the time of collection. Incomplete information was available for samples collected in 2005. Samples were analyzed while fresh or were press dried at room temperature for storage as herbarium-like samples until analyzed.

DNA amplification and cloning

Total DNA was extracted from fresh tissue or herbarium-like samples as described by Doyle & Doyle (1987) and used as a template for rolling-circle amplification (RCA) of begomovirus genomes (Inoue-Nagata *et al.*, 2004). RCA products were individually cleaved with *Bam*HI, *Cla*I, *Hin*dIII or *Pst*I restriction endonucleases and ligated to the pBluescript KS+ plasmid vector (Stratagene), previously cleaved with the same enzyme. Different enzymes were used to ensure that no virus was left uncloned because it lacked one or more of the sites (only ten samples had undigested RCA products after incubation with all four restriction enzymes). Viral inserts were sequenced commercially by primer walking at Macrogen Inc. All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5'-TAATATT//AC-3').

Sequence comparisons

Full-length begomovirus genomes were assembled using CodonCode Aligner v. 4.1.1 (www.codoncode.com), and sequences were initially analyzed with the BLAST*n* algorithm (Altschul *et al.*, 1990) and the GenBank non-redundant nucleotide database (www.ncbi.nlm.nih.gov/genbank), to determine the viral species with which they shared greatest identity. These similar sequences from GenBank were used to assign proper taxonomy to the novel isolates using the program Species Demarcation Tool v. 1.0 (Muhire *et al.*, 2013).

Multiple sequence alignments and phylogenetic analysis

Multiple sequence alignments were prepared for the full-length DNA-A and for the CP and Rep coding sequences of each viral species using the MUSCLE algorithm (Edgar, 2004) and manually adjusted using Se-Al v. 2.0a11 (tree.bio.ed.ac.uk/software/seal/). Maximum Likelihood (ML) trees were inferred for CP and Rep sequences using RAxML

v. 7.0.3 (bioinformatics.oxfordjournals.org/content/22/21/2688.long), assuming a general time reversible nucleotide substitution model with a gamma model of rate heterogeneity. The CP and Rep genes were chosen for analysis because of their essential role for insect transmission and viral replication, therefore being subject to stricter variability constraints (Rojas *et al.*, 2005). Additionally, they encompass ~70% of the full-length DNA-A genome. The robustness of each individual branch was estimated from 2000 bootstrap replicates. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree) and Adobe Illustrator.

Genetic structure and variability indices

Partitioning of genetic variability and inferences about population structure were based on Wright's F fixation index (Wright, 1951) calculated using DnaSP v. 5.10 (Rozas *et al.*, 2003). Subpopulations were tested for structure according to host species and geographical location. The mean pairwise number of nucleotide differences per site (nucleotide diversity, π) was estimated for each population/subpopulation using DnaSP v. 5.10 (Rozas *et al.*, 2003).

Recombination analysis

Possible parental sequences and recombination breakpoints were determined using the rdp, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods implemented in Recombination Detection Program (RDP) v. 3.44 (Martin *et al.*, 2010). Alignments were scanned with default settings for the different methods and statistical significance was inferred by a *P*-value lower than a Bonferroni-corrected cut-off of 0.05. Only recombination events detected by at least three different methods were considered to be reliable. Putatively recombinant portions of genomes were BLASTed against the GenBank non-redundant nucleotide database to identify additional species that may have served as

parental sequences. These were added to the datasets for additional analysis (Suppl. Table S2). Evidence of non-tree-like evolution was assessed for CP and Rep datasets using the Neighbor-Net method (Bryant & Moulton, 2004) implemented in SplitsTree v. 4.10 (Huson & Bryant, 2006). Images of networks were edited using Adobe Illustrator.

Detection of positive and negative selection at amino acid sites

To detect sites under positive and negative selection the CP and Rep datasets were analyzed using four different ML-based methods available in the DataMonkey server (www.datamonkey.org): Single-Likelihood Ancestor Counting (SLAC), Fixed-Effect Likelihood (FEL), Random-Effect Likelihood (REL) and Partitioning for Robust Inference of Selection (PARRIS) (Kosakovsky-Pond & Frost, 2005; Scheffler *et al*., 2006). To avoid spurious selection results caused by recombination, different predominantly non-recombinant clusters of sequences were defined based on phylogenetic and recombination analyses. The SLAC method was also used to estimate the mean ratios of non-synonymous to synonymous substitutions (d_N/d_S) based on the inferred Genetic Algorithm Recombination Detection (GARD) (Kosakovsky-Pond & Frost, 2005) phylogenetic trees. A general time reversible nucleotide substitution model was assumed, and Bayes factors larger than 50 and a *P*-value smaller than 0.1 were used as significance thresholds for the REL and FEL methods.

Acknowledgements

The authors wish to thank Drs. Raquel Mello and Flávia Rabelo (Embrapa Arroz and Feijão) for their support during sample collection. R.R.S. was the recipient of a CNPq doctoral fellowship and a CAPES sandwich PhD fellowship. S.D. was supported by US NSF 1026095 and funds from the NJ Agricultural Experiment Station. F.M.Z. was supported by Fapemig (APQ-00949-09) and CNPq (484447/2011-4).

References

Ala-Poikela, M., Svensson, E., Rojas, A., Horko, T., Paulin, L., Valkonen, J. P. T. & Kvarnheden, A. (2005). Genetic diversity and mixed infections of begomoviruses infecting tomato, pepper and cucurbit crops in Nicaragua. *Plant Pathol* **54**, 448-459.

Alabi, O. J., Ogbe, F. O., Bandyopadhyay, R., Kumar, P. L., Dixon, A. G. O., Hughes, J. D. & Naidu, R. A. (2008). Alternate hosts of African cassava mosaic virus and East African cassava mosaic Cameroon virus in Nigeria. *Arch Virol* **153**, 1743-1747.

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

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Barbosa, J. C., Barreto, S. S., Inoue-Nagata, A. K., Reis, M. S., Firmino, A. C., Bergamin, A. & Rezende, J. A. M. (2009). Natural infection of *Nicandra physaloides* by *Tomato severe rugose virus* in Brazil. *J Gen Plant Pathol* **75**, 440-443.

Bonfim, K., Faria, J. C., Nogueira, E. O., Mendes, E. A. & Aragao, F. J. (2007). RNAi-mediated resistance to *Bean golden mosaic virus* in genetically engineered common bean (*Phaseolus vulgaris*). *Mol Plant-Microbe Interact* **20**, 717-726.

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

Bryant, D. & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255-265.

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G. & Stanley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *J Gen Virol* **87**, 3053-3065.

Caciagli, P., Piles, V. M., Marian, D., Vecchiati, M., Masenga, V., Mason, G., Falcioni, T. & Noris, E. (2009). Virion stability is important for the circulative transmission of tomato yellow leaf curl Sardinia virus by *Bemisia tabaci*, but virion access to salivary glands does not guarantee transmissibility. *J Virol* **83**, 5784-5795.

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

Costa, A. S. (1976). Whitefly-transmitted plant diseases. *Annu Rev Phytopath* **14**, 429-440.

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

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

Faria, J. C. & Maxwell, D. P. (1999). Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil. *Phytopathology* **89**, 262-268.

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

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*. *Arch Virol* **157**, 141-146.

García-Andrés, S., Accotto, G. P., Navas-Castillo, J. & Moriones, E. (2007a). Founder effect, plant host, and recombination shape the emergent population of begomoviruses

that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology* **359**, 302-312.

García-Andrés, S., Monci, F., Navas-Castillo, J. & Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology* **350**, 433-442.

García-Andrés, S., Tomas, D. M., Sanchez-Campos, S., Navas-Castillo, J. & Moriones, E. (2007b). Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. *Virology* **365**, 210-219.

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

Gilbertson, R. L., Faria, J. C., Hanson, S. F., Morales, F. J., Ahlquist, P. G., Maxwell, D. P. & Russell, D. R. (1991a). Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. *Phytopathology* **81**, 980-985.

Gilbertson, R. L., Hidayat, S. H., Martinez, R. T., Leong, S. A., Faria, J. C., Morales, F. J. & Maxwell, D. P. (1991b). Diferentiation of bean-infecting geminiviruses by nucleic acid hybridization probes and aspects of bean golden mosaic in Brazil. *Plant Dis* **75**, 336-342.

González-Aguilera, J., Tavares, S. S., Sobrinho, R. R., Xavier, C. A. D., Dueñas-Hurtado, F., Lara-Rodrigues, R. M., Silva, D. J. H. & Zerbini, F. M. (2012). Genetic structure of a Brazilian population of the begomovirus *Tomato severe rugose virus* (ToSRV). *Trop Plant Pathol* **37**, 346-353.

Hohnle, M., Hofer, P., Bedford, I. D., Briddon, R. W., Markham, P. G. & Frischmuth, T. (2001). Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible *Abutilon mosaic virus* isolate. *Virology* **290**, 164-171.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.

Ilyina, T. V. & 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 Res* **20**, 3279-3285.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B. & Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *J Virol Met* **116**, 209-211.

James, C. K. N. & Bryce, W. F. (2006). Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu Rev Phytopath* **44**, 183-212.

Kheyr-Pour, A., Bananej, K., Dafalla, G. A., Caciagli, P., Noris, E., Ahoonmanesh, A., Lecoq, H. & Gronenborn, B. (2000). *Watermelon chlorotic stunt virus* from the Sudan and Iran: sequence comparisons and identification of a whitefly-transmission determinant. *Phytopathology* **90**, 629-635.

Koonin, E. V. & Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokariotic plasmid rolling circle DNA replication initiator proteins. *J Gen Virol* **73**, 2763-2766.

Kosakovsky-Pond, S. L. & Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-1222.

Lefeuvre, P., Lett, J. M., Reynaud, B. & Martin, D. P. (2007a). Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.

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

Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B. & Lett, J. M. (2007b). Begomovirus 'melting pot' in the

south-west Indian Ocean islands: molecular diversity and evolution through recombination. *J Gen Virol* **88**, 3458-3468.

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

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. (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *J Gen Virol* **94**, 418-431.

Liu, L., Pinner, M. S., Davies, J. W. & Stanley, J. (1999). Adaptation of the geminivirus bean yellow dwarf virus to dicotyledonous hosts involves both virion-sense and complementary-sense genes. *J Gen Virol* **80**, 501-506.

Lozano, G., Trenado, H. P., Valverde, R. A. & Navas-Castillo, J. (2009). Novel begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and *Ipomoea indica*: taxonomic and phylogenetic implications. *J Gen Virol* **90**, 2550-2562.

Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B., Vincent, C., Reynaud, B. & Lett, J. M. (2011). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog* **7**, e1002203.

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

Martin, D. P., van der Walt, E., Posada, D. & Rybicki, E. P. (2005). The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* **1**, e51.

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**, 317-326.

Monde, G., Walangululu, J., Winter, S. & Bragard, C. (2010). Dual infection by cassava begomoviruses in two leguminous species (Fabaceae) in Yangambi, northeastern Democratic Republic of Congo. *Arch Virol* **155**, 1865-1869.

Morales, F. J. (2006). History and current distribution of begomoviruses in Latin America. *Adv Virus Res* **67**, 127-162.

Morales, F. J. & Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch Virol* **146**, 415-441.

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. & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). *Arch Virol* **158**, 1411-1424.

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

Navas-Castillo, J., Fiallo-Olivé, E. & Sánchez-Campos, S. (2011). Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol* **49**, 219-248.

Ndunguru, J., Legg, J., Aveling, T., Thompson, G. & Fauquet, C. (2005). Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virol J* **2**, 21.

Noris, E., Vaira, A. M., Caciagli, P., Masenga, V., Gronenborn, B. & Accotto, G. P. (1998). Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *J Virol* **72**, 10050-10057.

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

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* **82**, 655-665.

Power, A. G. (2000). Insect transmission of plant viruses: a constraint on virus variability. *Curr Opin Plant Biol* **3**, 336-340.

Raja, P., Wolf, J. N. & Bisaro, D. M. (2010). RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components. *Biochim Biophys Acta* **1799**, 337-351.

Reddy, R. V. C., Colvin, J., Muniyappa, V. & Seal, S. (2005). Diversity and distribution of begomoviruses infecting tomato in India. *Arch Virol* **150**, 845-867.

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. (2003). Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Arch Virol* **148**, 281-295.

Ribeiro, S. G., Martin, D. P., Lacorte, C., Simões, I. C., Orlandini, D. R. S. & 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. & Zerbini, F. M. (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J Virol* **87**, 5784-5799.

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

Rokyta, D. R. & Wichman, H. A. (2009). Genic incompatibilities in two hybrid bacteriophages. *Mol Biol Evol* **26**, 2831-2839.

Rothenstein, D., Haible, D., Dasgupta, I., Dutt, N., Patil, B. L. & Jeske, H. (2006). Biodiversity and recombination of cassava-infecting begomoviruses from southern India. *Arch Virol* **151**, 55-69.

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

Sanz, A. I., Fraile, A., Gallego, J. M., Malpica, J. M. & García-Arenal, F. (1999). Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. *J Mol Evol* **49**, 672-681.

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

Schnippenkoetter, W. H., Martin, D. P., Willment, J. A. & Rybicki, E. P. (2001). Forced recombination between distinct strains of *Maize streak virus*. *J Gen Virol* **82**, 3081-3090.

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. (2012). Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathol* **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. & Zerbini, F. M. (2011). High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in northeastern Brazil. *Arch Virol* **156**, 2205-2213.

Simon-Loriere, E., Galetto, R., Hamoudi, M., Archer, J., Lefeuvre, P., Martin, D. P., Robertson, D. L. & Negroni, M. (2009). Molecular mechanisms of recombination restriction in the envelope gene of the *Human immunodeficiency virus*. *PLoS Pathog* **5,** e1000418.

Sserubombwe, W. S., Briddon, R. W., Baguma, Y. K., Ssemakula, G. N., Bull, S. E., Bua, A., Alicai, T., Omongo, C., Otim-Nape, G. W. & Stanley, J. (2008). Diversity of begomoviruses associated with mosaic disease of cultivated cassava (*Manihot esculenta* Crantz) and its wild relative (*Manihot glaziovii* Mull. Arg.) in Uganda. *J Gen Virol* **89**, 1759-1769.

Tavares, S. S., Ramos-Sobrinho, R., Gonzalez-Aguilera, J., Lima, G. S. A., Assunção, I. P. & Zerbini, F. M. (2012). Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on *Sida* spp. *Planta Dan* **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. & Lett, J. M. (2012). Evolution of African cassava mosaic virus by recombination between bipartite and monopartite begomoviruses. *Virol J* **9**, 67.

Umaharan, P., Padidam, M., Phelps, R. H., Beachy, R. N. & Fauquet, C. M. (1998). Distribution and diversity of geminiviruses in Trinidad and Tobago. *Virology* **88**, 1262-1268.

Varma, A. & Malathi, V. G. (2003). Emerging geminivirus problems: a serious threat to crop production. *Ann Appl Biol* **142**, 145-164.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Zerbini, F. M. & Martin, D. P. (2014). Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Arch Virol* **159**, in press.

Wright, S. (1951). The genetical structure of populations. *Ann Eugen* **15**, 323-354.

Wyant, P. S., Gotthardt, D., Schafer, B., Krenz, B. & Jeske, H. (2011). The genomes of four novel begomoviruses and a new *Sida micrantha mosaic virus* strain from Bolivian weeds. *Arch Virol* **156**, 347-352.

Wyant, P. S., Strohmeier, S., Schafer, B., Krenz, B., Assuncao, I. P., Lima, G. S. D. & Jeske, H. (2012). Circular DNA genomics (circomics) exemplified for geminiviruses in bean crops and weeds of northeastern Brazil. *Virology* **427**, 151-157.

Table 1. Genetic variability of the begomoviruses BGMV and MaYSV infecting three distinct leguminous hosts.

Population	**No. of sequences**	**DNA-A π* (SD)†**	**CP π (SD)**	**Rep π (SD)**
BGMV (Total)	147	0.0489 (±0.0015)	0.0442 (±0.0012)	0.0371 (±0.0013)
BGMV (Common bean)	75	0.0067 (±0.0002)	0.0058 (±0.0002)	0.0043 (±0.0002)
BGMV (Lima bean)	59	0.0153 (±0.0039)	0.0117 (±0.0031)	0.0144 (±0.0036)
BGMV (*Macroptilium*)	13	0.0018 (±0.0005)	0.0020 (±0.0006)	0.0009 (±0.0002)
MaYSV (Total)	99	0.0627 (±0.0021)	0.0262 (±0.0020)	0.1055 (±0.0040)
MaYSV (Common bean)	50	0.0622 (±0.0032)	0.0224 (±0.0011)	0.1068 (±0.0065)
MaYSV (Lima bean)	21	0.0590 (±0.0062)	0.0212 (±0.0019)	0.0976 (±0.0119)
MaYSV (*Macroptilium*)	28	0.0662 (±0.0037)	0.0352 (±0.0056)	0.1037 (±0.0078)

* Pairwise, per-site nucleotide diversity.
† Standard deviation.

Table 2. Putative recombination events detected within the legume-infecting begomoviruses BGMV and MaYSV, based on full-length DNA-A sequences.

Event	Recombinant	Recombination breakpoints*		Parents		Method†	*P*-Value‡
		Begin	End	Minor	Major		
MaYSV							
1	Oaf12:11	2618	1660	Pdi1:10	Unknown	RGBMCS3	6.86×10^{-59}
	Crb3:11			Crb1:11	Unknown		
	Crb20:11, Crb4:11, Crb7:11, Crb8:11, Crb10:11, Crb13:11, Crb14:11, Crb16:11, Oaf1:11, Oaf2:11, Oaf3:11, Oaf4:11, Oaf5:11, Oaf7:11, Oaf8:11, Oaf10:11, Oaf11:11, Oaf26:11, Oaf29:11, Oaf14:11, Oaf16:11, Oaf28:11, Oaf17:11, Oaf20:11, Oaf21:11, Oaf22:11, Oaf27:11, Oaf23:11, Oaf24:11, Oaf25:11, Sti2:11, Sti3:11, Sti5:11, Sti35:11, Sti6:11, Sti7:11, Sti8:11, Sti9:11, Sti10:11, Sti11:11, Sti12:11, Sti14:11, Sti17:11, Sti18:11, Sti19:11, Sti20:11, Sti21:11, Sti22:11, Sti23:11, Sti24:11, Sti25:11, Sti26:11, Sti28:11, Sti31:11, Sti30:11, Sti32:11, Bat1:10, Pir1:10, Inp1:10			Crb2:11	Unknown		
2	Sti37:11	44	2039	Oaf6:11	Unknown	RBMCS3	5.03×10^{-28}
	Oaf21:11			Oaf9:11	Unknown		
3	Oaf19:11	(?)1747	2018	Sti11:11	Pdi1:10	RGBMCS3	4.46×10^{-23}
	Oaf6:11			Crb3:11	Crb1:11		
	Oaf9:11			Crb20:11	Crb2:11		
	Sti33:11			Crb4:11	Crb5:11		
4	Oaf11:11	1975	2157	Unknown	Oaf5:11	RGBMCS3	1.88×10^{-23}
	Oaf7:11			Unknown	Crb3:11		
	Oaf8:11			Unknown	Crb20:11		
	Oaf10:11			Unknown	Crb4:11		
	Oaf29:11			Unknown	Crb7:11		
	Oaf12:11			Unknown	Crb8:11		
	Oaf14:11			Unknown	Crb10:11		
	Oaf22:11			Unknown	Crb13:11		
	Oaf25:11			Unknown	Crb14:11		
	Sti2:11			Unknown	Crb16:11		
	Sti17:11			Unknown	Oaf1:11		

	Sti19:11			Unknown	Oaf2:11		
	Sti20:11			Unknown	Oaf3:11		
	Sti30:11			Unknown	Oaf4:11		
5	Oaf29:11	(?)1661	(?)1974	Crb10:11	Oaf25:11	RGBS3	$7.78x10^{-19}$
	Oaf7:11			Crb3:11	Sti17:11		
	Oaf8:11, Oaf10:11, Oaf11:11, Oaf12:11, Oaf14:11, Oaf22:11, Sti2:11, Sti20:11, Sti30:11			Crb20:11	Sti19:11		
6	Mac1:10	335	955	Unknown	Ced1:09	GBMCS3	$1.72x10^{-11}$
	Bas1:09			Unknown	Sti34:11		
7	Oaf28:11	2386	(?)2617	Unknown	Oaf11:11	GMCS	$2.99x10^{-20}$
	Crb3:11			Unknown	Oaf7:11		
	Crb20:11			Unknown	Oaf8:11		
	Crb4:11			Unknown	Oaf10:11		
	Crb7:11			Unknown	Oaf29:11		
	Crb8:11			Unknown	Oaf12:11		
	Crb10:11			Unknown	Oaf22:11		
	Crb13:11			Unknown	Oaf25:11		
	Crb14:11			Unknown	Sti2:11		
	Crb16:11			Unknown	Sti17:11		
	Oaf1:11			Unknown	Sti19:11		
	Oaf2:11, Oaf3:11, Oaf4:11, Oaf5:11, Oaf26:11, Oaf16:11, Oaf17:11, Oaf20:11, Oaf21:11, Oaf23:11, Oaf24:11, Oaf27:11, Sti3:11, Sti5:11, Sti35:11, Sti6:11, Sti7:11, Sti8:11, Sti9:11, Sti10:11, Sti11:11, Sti12:11, Sti14:11, Sti18:11, Sti21:11, Sti22:11, Sti37:11, Sti23:11, Sti24:11, Sti25:11, Sti26:11, Sti28:11, Sti31:11, Sti32:11, Bat1:10, Pir1:10, Inp1:10			Unknown	Sti20:11		
BGMV							
1	Ata2:05	504	52	Par:28:12	Ata1:05	RGBMCS3	$2.34x10^{-20}$

2	Vis9:11 Vis1:11, Vis3:11, Vis5:11, Vis6:11, Vis7:11, Vis8:11, Vis10:11, Vis11:11, Vis12:11, Vis13:11, Vis14:11, Pai1:11, Pai2:11, Pai3:11, Pai4:11, Pai5:11, Pai6:11, Pai7:11, Pai8:11, Pai9:11, Pai10:11, Pai11:11, Pai12:11, Pai13:11, Pai14:11, Pai15:11, Pai16:11, Pai17:11, Mur1:11, Mur6:11, Mur7:11, Mur8:11, Mur9:11, Mur10:11, Mur2:11, Mur3:11, Mur11:11, Mur12:11, Mur4:11, Mur5:11, Arp1:05, Eta1:05, Vic:05, Cno1:05, Cno2:05, Bel:05, Mrb:05, Mac:05, Mav:05, Pja:05, Ata1:05, Unp:05, Stm1:05, Chp:05, Rla:05, Mac2:10	410	1115	Rec1:05 Rec2:11	Car4:10 Car3:10	RGBMCS3	$1.87x10^{-13}$

* Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?), breakpoint could not be precisely pinpointed.
† R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimera; S, SisScan; 3, 3SEQ.
‡ The reported *P* values are for the method indicated in red, and they are the lowest *P* values calculated for the region in question.

Table 3. Mean values of non-synonymous to synonymous substitution ratios (d_N/d_S) for the subpopulation-specific CP and Rep genes of BGMV and MaYSV.

Datasets	d_N/d_S
BGMV CP (*P. vulgaris*)	0.0936
BGMV CP (*M. lathyroides*)	0.4844
BGMV CP (*P. lunatus*)	0.1207
BGMV Rep (*P. vulgaris*)	0.3541
BGMV Rep (*M. lathyroides*)	1.0000
BGMV Rep (*P. lunatus*)	0.2073
MaYSV CP	0.0919
MaYSV Rep CI	0.1757
MaYSV Rep CII	0.1725
MaYSV Rep CIII	0.2427

Figure legends

Fig. 1. Neighbor-Net network based on CP (**a**) and Rep (**b**) nucleotide sequences of MaYSV and selected begomoviruses from Brazil. Isolates from *Phaseolus vulgaris* are indicated in black, from *Phaseolus lunatus* in red, and from *Macroptilium lathyroides* in blue.

Fig. 2. Neighbor-Net network based on CP (**a**) and Rep (**b**) nucleotide sequences of BGMV and selected begomoviruses from Brazil. Isolates from *Phaseolus vulgaris* are indicated in black, from *Phaseolus lunatus* in red, and from *Macroptilium lathyroides* in blue.

Fig. 3. Midpoint-rooted maximum-likelihood trees based on CP (**a**) and Rep (**b**) nucleotide sequences of MaYSV isolates from *Phaseolus vulgaris* (indicated in black), *Phaseolus lunatus* (red) and *Macroptilium lathyroides* (blue). Nodes with bootstrap values higher than 50% and lower than 80% are indicated by empty circles, and those with values equal to or higher than 80% by filled circles. Isolates with identical names (first three letters) were collected at the same location.

Fig. 4. Midpoint-rooted maximum-likelihood trees based on CP (**a**) and Rep (**b**) nucleotide sequences of BGMV isolates from *Phaseolus vulgaris* (indicated in black), *Phaseolus lunatus* (red) and *Macroptilium lathyroides* (blue). An isolate infecting *Glycine max* is indicated in orange. Nodes with bootstrap values higher than 50% and lower than 80% are indicated by empty circles, and those with values equal to or higher than 80% by filled circles. Isolates with identical names (first three letters) were collected at the same location.

Figure 1

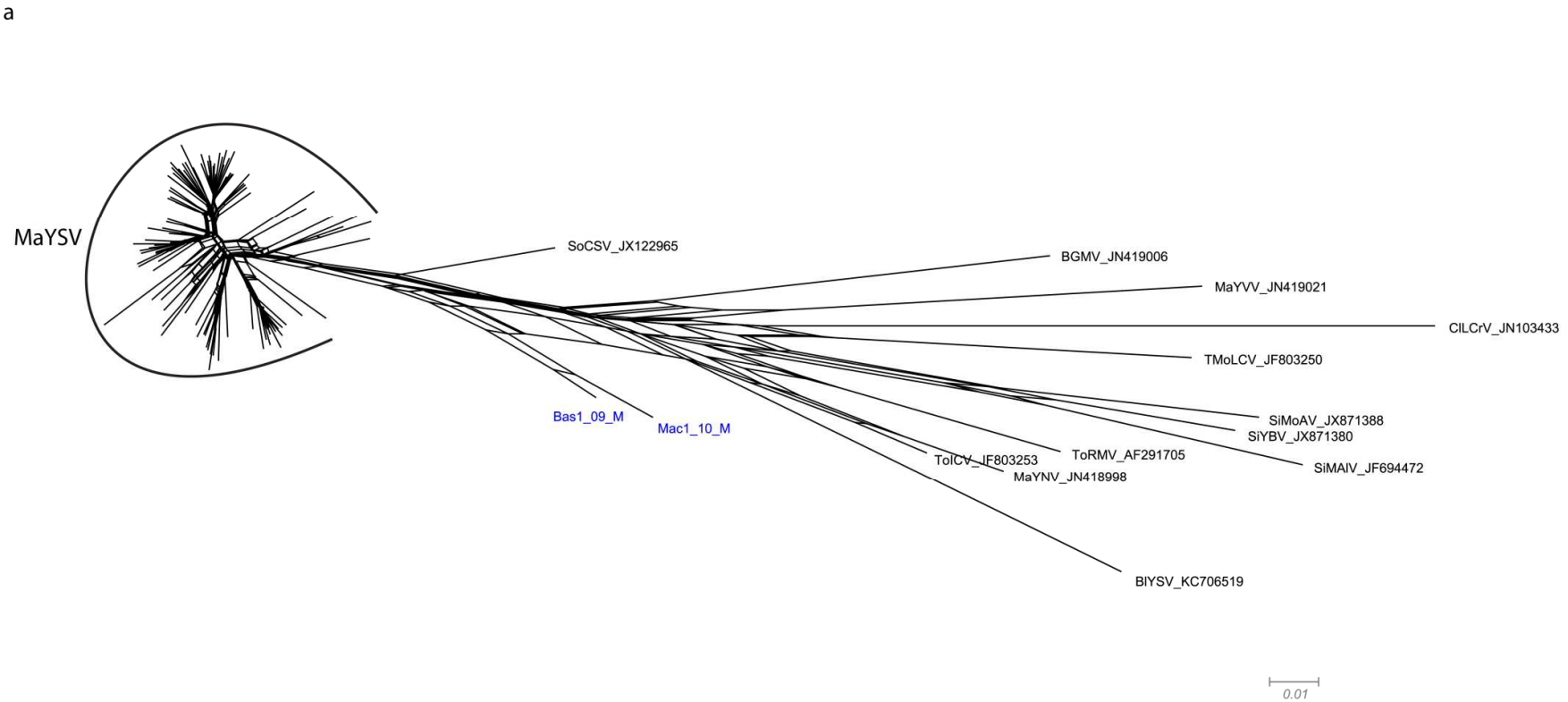

a
MaYSV
SoCSV_JX122965
BGMV_JN419006
MaYVV_JN419021
ClLCrV_JN103433
TMoLCV_JF803250
Bas1_09_M
Mac1_10_M
SiMoAV_JX871388
SiYBV_JX871380
ToICV_JF803253
ToRMV_AF291705
MaYNV_JN418998
SiMAIV_JF694472
BlYSV_KC706519
0.01

Figure 1

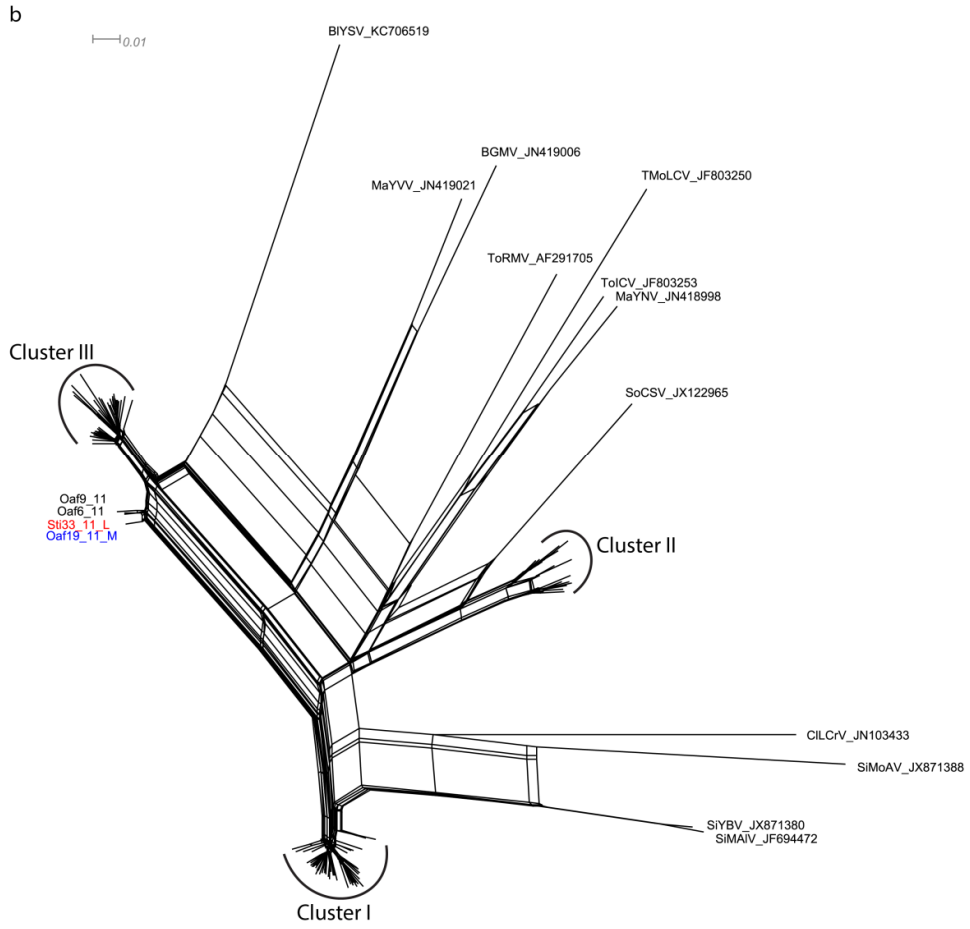

b
0.01
BIYSV_KC706519
BGMV_JN419006
MaYVV_JN419021
TMoLCV_JF803250
ToRMV_AF291705
ToICV_JF803253
MaYNV_JN418998
SoCSV_JX122965
Cluster III
Oaf9_11
Oaf6_11
Sti33_11_L
Oaf19_11_M
Cluster II
CILCrV_JN103433
SiMoAV_JX871388
SiYBV_JX871380
SiMAiV_JF694472
Cluster I

Figure 2

a

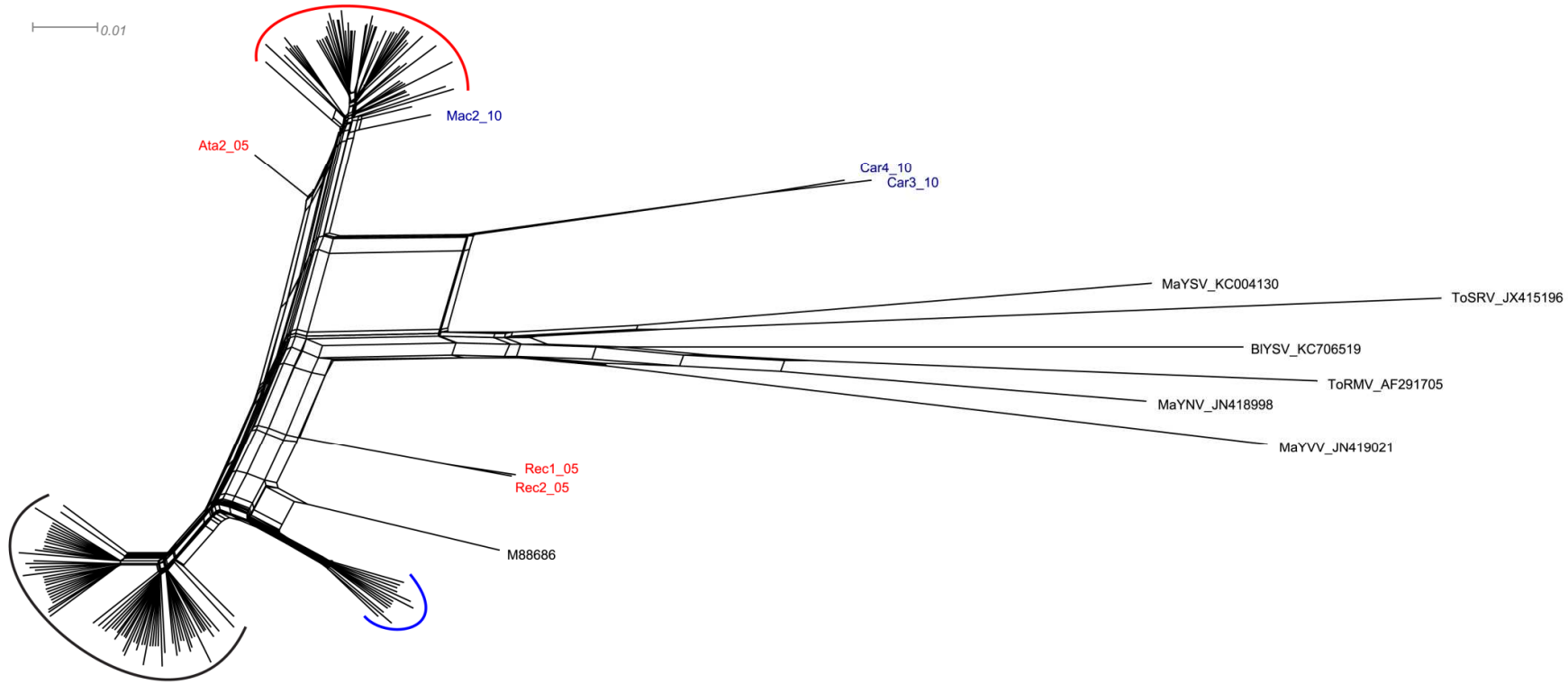

0.01
Mac2_10
Ata2_05
Car4_10
Car3_10
MaYSV_KC004130
ToSRV_JX415196
BlYSV_KC706519
ToRMV_AF291705
MaYNV_JN418998
MaYVV_JN419021
Rec1_05
Rec2_05
M88686

Figure 2

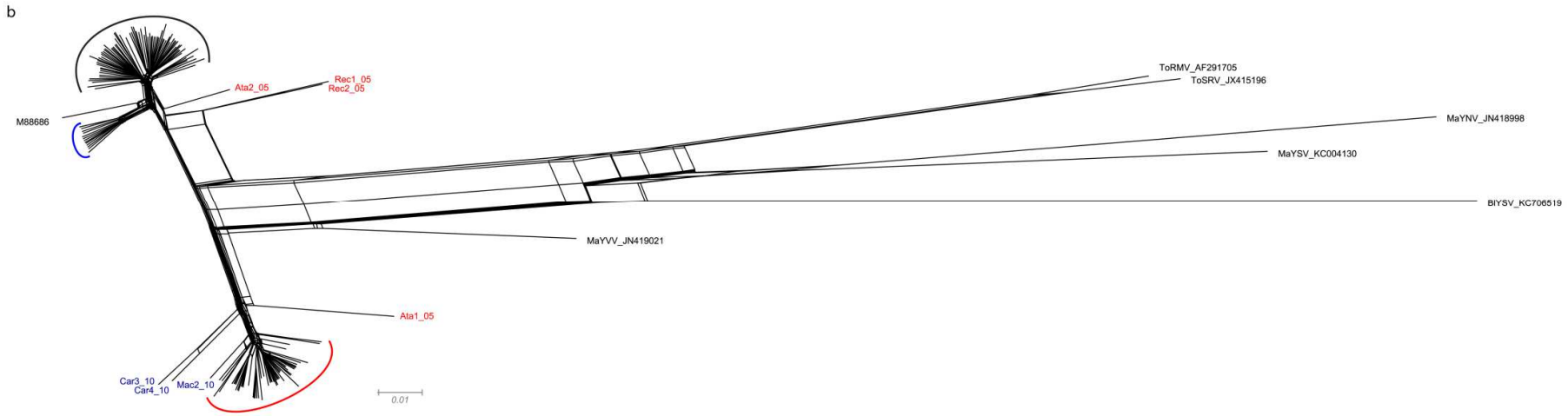

b
M88686
Ata2_05
Rec1_05
Rec2_05
ToRMV_AF291705
ToSRV_JX415196
MaYNV_JN418998
MaYSV_KC004130
BIYSV_KC706519
MaYVV_JN419021
Ata1_05
Car3_10
Car4_10
Mac2_10
0.01

Figure 3

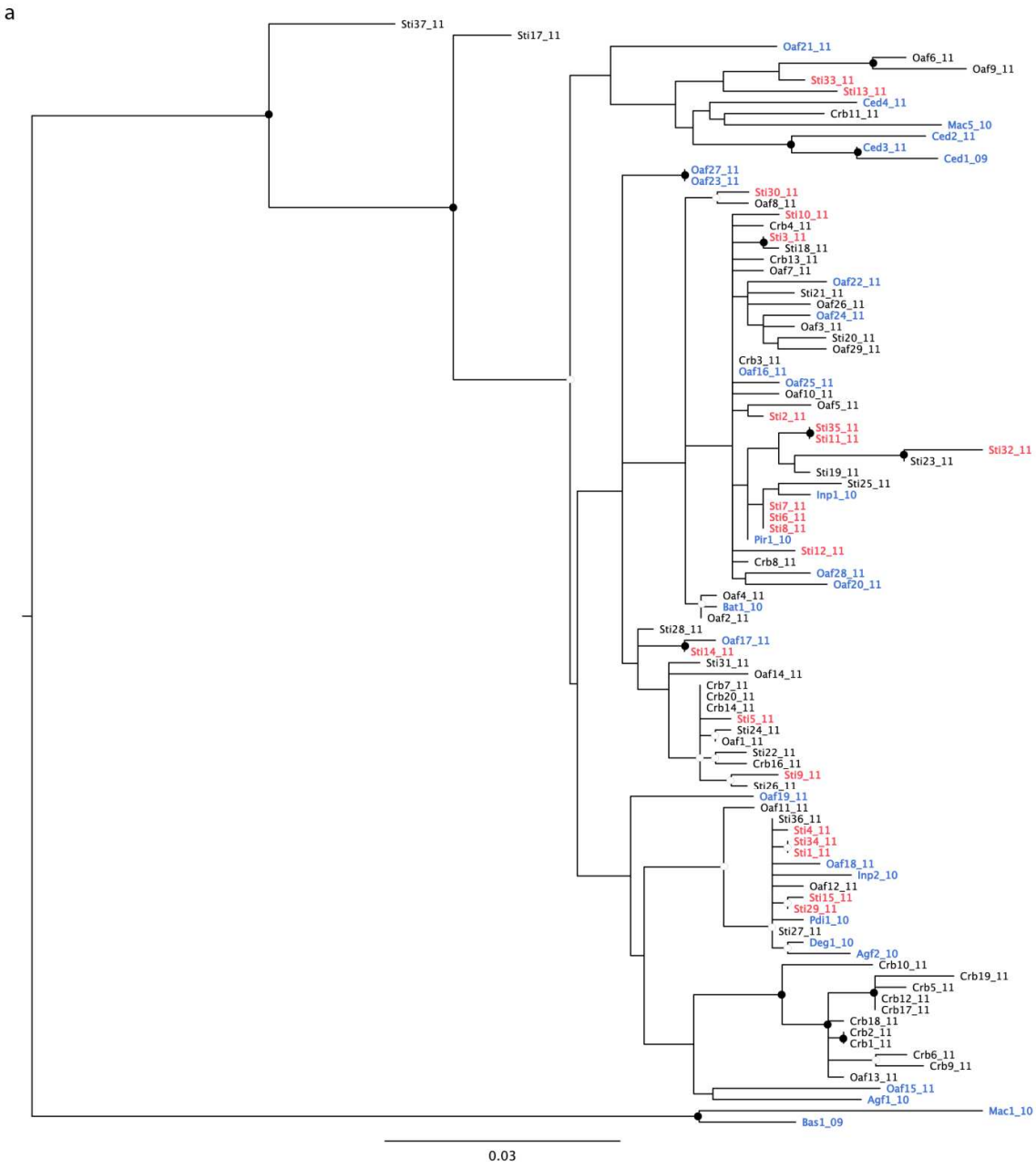

a
Sti37_11
Sti17_11
Oaf21_11
Oaf6_11
Oaf9_11
Sti33_11
Sti13_11
Ced4_11
Crb11_11
Mac5_10
Ced2_11
Ced3_11
Ced1_09
Oaf27_11
Oaf23_11
Sti30_11
Oaf8_11
Sti10_11
Crb4_11
Sti3_11
Sti18_11
Crb13_11
Oaf7_11
Oaf22_11
Sti21_11
Oaf26_11
Oaf24_11
Oaf3_11
Sti20_11
Oaf29_11
Crb3_11
Oaf16_11
Oaf25_11
Oaf10_11
Oaf5_11
Sti2_11
Sti35_11
Sti11_11
Sti32_11
Sti23_11
Sti19_11
Sti25_11
Inp1_10
Sti7_11
Sti6_11
Sti8_11
Pir1_10
Sti12_11
Crb8_11
Oaf28_11
Oaf20_11
Oaf4_11
Bat1_10
Oaf2_11
Sti28_11
Oaf17_11
Sti14_11
Sti31_11
Oaf14_11
Crb7_11
Crb20_11
Crb14_11
Sti5_11
Sti24_11
Oaf1_11
Sti22_11
Crb16_11
Sti9_11
Sti26_11
Oaf19_11
Oaf11_11
Sti36_11
Sti4_11
Sti34_11
Sti1_11
Oaf18_11
Inp2_10
Oaf12_11
Sti15_11
Sti29_11
Pdi1_10
Sti27_11
Deg1_10
Agf2_10
Crb10_11
Crb19_11
Crb5_11
Crb12_11
Crb17_11
Crb18_11
Crb2_11
Crb1_11
Crb6_11
Crb9_11
Oaf13_11
Oaf15_11
Agf1_10
Mac1_10
Bas1_09
0.03

Figure 3


b
Oaf24_11
Oaf3_11
Oaf26_11
Oaf20_11
Oaf5_11
Pir1_11
Oaf16_11
Sti21_11
Sti25_11
Sti12_11
Crb3_11
Crb4_11
Crb8_11
Crb13_11
Sti18_11
Sti3_11
Oaf28_11
Inp1_11
Sti10_11
Sti5_11
Crb14_11
Crb7_11
Sti8_11
Sti7_11
Sti6_11
Oaf17_11
Oaf27_11
Oaf23_11
Sti23_11
Sti14_11
Sti9_11
Sti26_11
Oaf4_11
Bat1_11
Sti22_11
Sti37_11
Crb16_11
Sti28_11
Sti24_11
Crb20_11
Oaf1_11
Sti35_11
Sti11_11
Oaf2_11
Sti32_11
Sti31_11
Oaf21_11
Crb10_11
Oaf11_11
Sti19_11
Sti17_11
Oaf25_11
Oaf29_11
Oaf10_11
Sti20_11
Oaf8_11
Sti2_11
Oaf12_11
Oaf22_11
Oaf14_11
Sti30_11
Oaf7_11
Oaf19_11
Oaf9_11
Sti33_11
Oaf6_11
Sti4_11
Sti1_11
Sti34_11
Deg1_11
Sti36_11
Oaf18_11
Sti15_11
Sti29_11
Sti27_11
Pdi1_10
Agf2_11
Inp2_11
Ced4_11
Agf1_10
Bas1_09
Mac1_10
Mac5_10
Crb11_11
Oaf13_11
Crb18_11
Crb2_11
Crb1_11
Crb17_11
Crb19_11
Crb12_11
Crb6_11
Crb5_11
Oaf15_11
Crb9_11
Sti13_11
Ced2_11
Ced3_11
Ced1_09
0.04

Figure 4

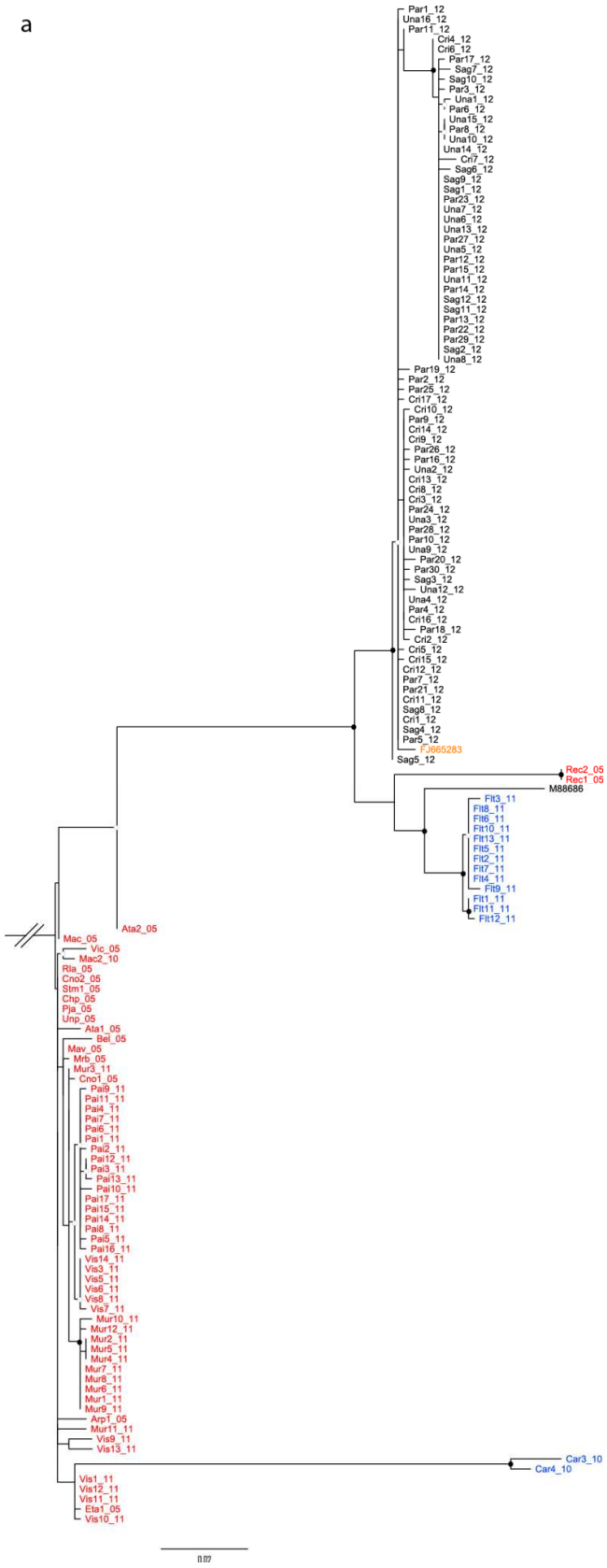
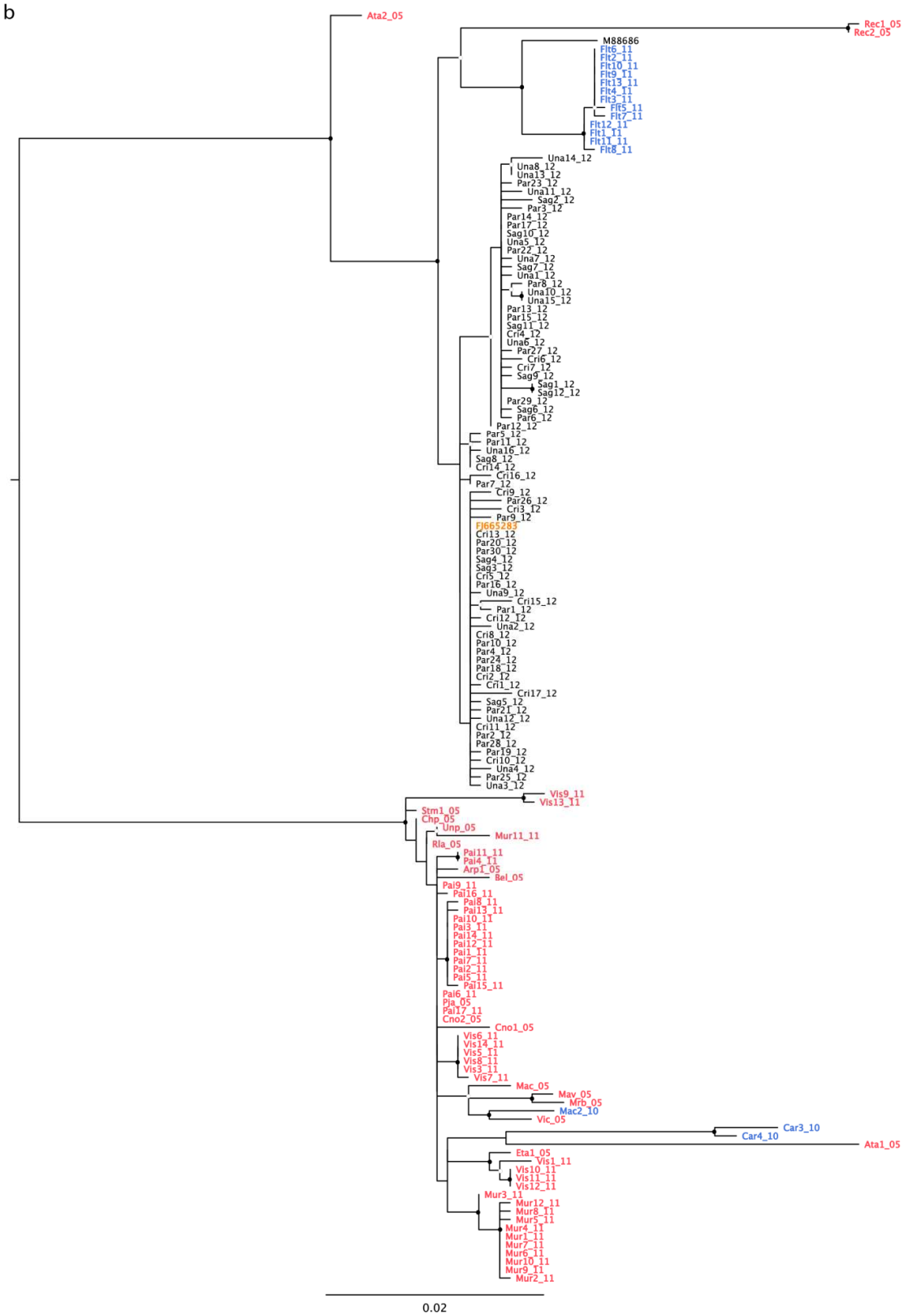

a
Par1_12
Una16_12
Par11_12
Cri4_12
Cri6_12
Par17_12
Sag7_12
Sag10_12
Par3_12
Una1_12
Par6_12
Una15_12
Par8_12
Una10_12
Una14_12
Cri7_12
Sag6_12
Sag9_12
Sag1_12
Par23_12
Una7_12
Una6_12
Una13_12
Par27_12
Una5_12
Par12_12
Par15_12
Una11_12
Par14_12
Sag12_12
Sag11_12
Par13_12
Par22_12
Par29_12
Sag2_12
Una8_12
Par19_12
Par2_12
Par25_12
Cri17_12
Cri10_12
Par9_12
Cri14_12
Cri9_12
Par26_12
Par16_12
Una2_12
Cri13_12
Cri8_12
Cri3_12
Par24_12
Una3_12
Par28_12
Par10_12
Una9_12
Par20_12
Par30_12
Sag3_12
Una12_12
Una4_12
Par4_12
Cri16_12
Par18_12
Cri2_12
Cri5_12
Cri15_12
Cri12_12
Par7_12
Par21_12
Cri11_12
Sag8_12
Cri1_12
Sag4_12
Par5_12
FJ665283
Sag5_12
Rec2_05
Rec1_05
M88686
Flt3_11
Flt8_11
Flt6_11
Flt10_11
Flt13_11
Flt5_11
Flt2_11
Flt7_11
Flt4_11
Flt9_11
Flt1_11
Flt11_11
Flt12_11
Ata2_05
Mac_05
Vic_05
Mac2_10
Rla_05
Cno2_05
Stm1_05
Chp_05
Pja_05
Unp_05
Ata1_05
Bel_05
Mav_05
Mrb_05
Mur3_11
Cno1_05
Pai9_11
Pai11_11
Pai4_11
Pai7_11
Pai6_11
Pai1_11
Pai2_11
Pai12_11
Pai3_11
Pai13_11
Pai10_11
Pai17_11
Pai15_11
Pai14_11
Pai8_11
Pai5_11
Pai16_11
Vis14_11
Vis3_11
Vis5_11
Vis6_11
Vis8_11
Vis7_11
Mur10_11
Mur12_11
Mur2_11
Mur5_11
Mur4_11
Mur7_11
Mur8_11
Mur6_11
Mur1_11
Mur9_11
Arp1_05
Mur11_11
Vis9_11
Vis13_11
Car3_10
Car4_10
Vis1_11
Vis12_11
Vis11_11
Eta1_05
Vis10_11

Figure 4


b
Ata2_05
Rec1_05
Rec2_05
M88686
Flt6_11
Flt2_11
Flt10_11
Flt9_11
Flt13_11
Flt4_11
Flt3_11
Flt5_11
Flt7_11
Flt12_11
Flt1_11
Flt11_11
Flt8_11
Una14_12
Una8_12
Una13_12
Par23_12
Una11_12
Sag2_12
Par3_12
Par14_12
Par17_12
Sag10_12
Una5_12
Par22_12
Una7_12
Sag7_12
Una1_12
Par8_12
Una10_12
Una15_12
Par13_12
Par15_12
Sag11_12
Cri4_12
Una6_12
Par27_12
Cri6_12
Cri7_12
Sag9_12
Sag1_12
Sag12_12
Par29_12
Sag6_12
Par6_12
Par12_12
Par5_12
Par11_12
Una16_12
Sag8_12
Cri14_12
Cri16_12
Par7_12
Cri9_12
Par26_12
Cri3_12
Par9_12
FJ665283
Cri13_12
Par20_12
Par30_12
Sag4_12
Sag3_12
Cri5_12
Par16_12
Una9_12
Cri15_12
Par1_12
Cri12_12
Una2_12
Cri8_12
Par10_12
Par4_12
Par24_12
Par18_12
Cri2_12
Cri1_12
Cri17_12
Sag5_12
Par21_12
Una12_12
Cri11_12
Par2_12
Par28_12
Par19_12
Cri10_12
Una4_12
Par25_12
Una3_12
Vis9_11
Vis13_11
Stm1_05
Chp_05
Unp_05
Mur11_11
Rla_05
Pai11_11
Pai4_11
Arp1_05
Bel_05
Pai9_11
Pai16_11
Pai8_11
Pai13_11
Pai10_11
Pai3_11
Pai14_11
Pai12_11
Pai1_11
Pai7_11
Pai2_11
Pai5_11
Pai15_11
Pai6_11
Pla_05
Pai17_11
Cno2_05
Cno1_05
Vis6_11
Vis14_11
Vis5_11
Vis8_11
Vis3_11
Vis7_11
Mac_05
Mav_05
Mrb_05
Mac2_10
Vic_05
Car3_10
Car4_10
Ata1_05
Eta1_05
Vis1_11
Vis10_11
Vis11_11
Vis12_11
Mur3_11
Mur12_11
Mur8_11
Mur5_11
Mur4_11
Mur1_11
Mur7_11
Mur6_11
Mur10_11
Mur9_11
Mur2_11
0.02

Suppl. Table S1. Begomovirus sequences reported in this work.

Sample code	Geographical coordinates		Viral species/Acronym	GenBank access #	Isolate	Host	Location	Date
282SE	S10°15'00"	W36°53'00"	*Macroptilium yellow spot virus*/MaYSV		BR:Ced2:11	*Macroptilium lathyroides*	Cedro de São João, SE	July, 2011
291SE	S10°15'00"	W36°53'00"	MaYSV		BR:Ced3:11	*Macroptilium lathyroides*	Cedro de São João, SE	July, 2011
304SE	S10°15'00"	W36°53'00"	MaYSV		BR:Ced4:11	*Macroptilium lathyroides*	Cedro de São João, SE	July, 2011
21AL	S09°40'39.3"	W036°46'38.8"	MaYSV		BR:Crb20:11	*Phaseolus vulgaris*	Craibas, AL	July, 2011
85AL	S09°32'28.8"	W037°17'28.3"	MaYSV		BR:Oaf26:11	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
61AL	S09°32'26.9"	W037°17'17.6"	MaYSV		BR:Oaf27:11	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
46AL	S09°32'27.8"	W037°17'22.1"	MaYSV		BR:Oaf28:11	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
87AL	S09°32'28.7"	W037°17'28.0"	MaYSV		BR:Oaf29:11	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
98AL	S09°23'24.6"	W037°12'47.8"	MaYSV		BR:Sti1:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
112AL	S09°23'24.9"	W037°12'47.6"	MaYSV		BR:Sti10:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
113AL	S09°23'24.6"	W037°12'47.5	MaYSV		BR:Sti11:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
114AL	S09°23'24.7"	W037°12'47.5"	MaYSV		BR:Sti12:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
114AL	S09°23'24.7"	W037°12'47.5"	MaYSV		BR:Sti13:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
115AL	S09°23'24.9"	W037°12'47.6"	MaYSV		BR:Sti14:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
116AL	S09°23'24.9"	W037°12'47.6"	MaYSV		BR:Sti15:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
117AL	S09°23'26.1"	W037°12'47.8"	MaYSV		BR:Sti17:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
119AL	S09°23'26.2"	W037°12'47.8"	MaYSV		BR:Sti18:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
119AL	S09°23'26.2"	W037°12'47.8"	MaYSV		BR:Sti19:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
99AL	S09°23'24.7"	W037°12'48.0"	MaYSV		BR:Sti2:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
120AL	S09°23'26.2"	W037°12'47.7"	MaYSV		BR:Sti20:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
121AL	S09°23'26.2"	W037°12'47.7"	MaYSV		BR:Sti21:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
122AL	S09°23'26.2"	W037°12'47.7"	MaYSV		BR:Sti22:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
124AL	S09°23'27.0"	W037°12'47.5"	MaYSV		BR:Sti23:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
125AL	S09°23'26.7"	W037°12'47.3"	MaYSV		BR:Sti24:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
127AL	S09°23'27.0"	W037°12'46.9"	MaYSV		BR:Sti25:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
128AL	S09°23'27.0"	W037°12'46.9"	MaYSV		BR:Sti26:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
129AL	S09°23'27.4"	W037°12'46.7"	MaYSV		BR:Sti27:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
134AL	S09°23'26.8"	W037°12'48.4"	MaYSV		BR:Sti28:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
152AL	S09°23'24.8"	W037°12'48.6"	MaYSV		BR:Sti29:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
100AL	S09°23'24.8"	W037°12'48.0"	MaYSV		BR:Sti3:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
154AL	S09°23'25.2"	W037°12'48.8"	MaYSV		BR:Sti30:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
135AL	S09°23'26.8"	W037°12'48.4"	MaYSV		BR:Sti31:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
147AL	S09°23'25.3"	W037°12'48.5"	MaYSV		BR:Sti32:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
156AL	S09°23'25.4"	W037°12'48.3"	MaYSV		BR:Sti33:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
97AL	S09°23'25.1"	W037°12'46.9"	MaYSV		BR:Sti34:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
105AL	S09°23'24.9"	W037°12'48.3"	MaYSV		BR:Sti35:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
118AL	S09°23'26.1"	W037°12'47.8"	MaYSV		BR:Sti36:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
123AL	S09°23'26.8"	W037°12'47.7"	MaYSV		BR:Sti37:11	*Phaseolus vulgaris*	Santana do Ipanema, AL	July, 2011
102AL	S09°23'24.9"	W037°12'48.0"	MaYSV		BR:Sti4:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
104AL	S09°23'24.8"	W037°12'48.2"	MaYSV		BR:Sti5:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
107AL	S09°23'25.1"	W037°12'48.3"	MaYSV		BR:Sti6:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
108AL	S09°23'25.2"	W037°12'48.4"	MaYSV		BR:Sti7:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
109AL	S09°23'25.0"	W037°12'48.3"	MaYSV		BR:Sti8:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011

111AL	S09°23’24.9”	W037°12’47.5”	MaYSV	BR:Sti9:11	*Phaseolus lunatus*	Santana do Ipanema, AL	July, 2011
1	-	-	*Bean golden mosaic virus*/BGMV	BR: Arp1:05	*Phaseolus lunatus*	Arapiraca, AL	2005
2	-	-	BGMV	BR: Eta1:05	*Phaseolus lunatus*	Estrela de Alagoas, AL	2005
3	-	-	BGMV	BR:Vic:05	*Phaseolus lunatus*	Viçosa, AL	2005
4	-	-	BGMV	BR:Rec1:05	*Phaseolus lunatus*	Recife, PE	2005
4	-	-	BGMV	BR:Rec2:05	*Phaseolus lunatus*	Recife, PE	2005
5	-	-	BGMV	BR:Cno1:05	*Phaseolus lunatus*	Coité do Noia, AL	2005
5	-	-	BGMV	BR:Cno2:05	*Phaseolus lunatus*	Coité do Noia, AL	2005
6	-	-	BGMV	BR:Bel:05	*Phaseolus lunatus*	Belém, AL	2005
7	-	-	BGMV	BR:Mrb:05	*Phaseolus lunatus*	Marimbondo, AL	2005
8	-	-	BGMV	BR:Mac:05	*Phaseolus lunatus*	Maceió, AL	2005
9	-	-	BGMV	BR:Mav:05	*Phaseolus lunatus*	Mar Vermelho, AL	2005
10	-	-	BGMV	BR:Pja:05	*Phaseolus lunatus*	Paulo Jacinto, AL	2005
11	-	-	BGMV	BR:Ata1:05	*Phaseolus lunatus*	Atalaia, AL	2005
11	-	-	BGMV	BR:Ata2:05	*Phaseolus lunatus*	Atalaia, AL	2005
12	-	-	BGMV	BR:Unp:05	*Phaseolus lunatus*	União dos Palmares, AL	2005
13	-	-	BGMV	BR:Stm1:05	*Phaseolus lunatus*	Santana do Mundaú, AL	2005
14	-	-	BGMV	BR:Chp:05	*Phaseolus lunatus*	Chã Preta, AL	2005
15	-	-	BGMV	BR:Rla:05	*Phaseolus lunatus*	Rio Largo, AL	2005
1AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis1:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
5AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis3:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
6AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis5:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
7AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis6:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
8AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis7:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
9AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis8:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
10AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis9:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
11AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis10:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
12AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis11:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
13AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis12:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
14AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis13:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
15AL	S09°19’22.4”	W036°19’57.6”	BGMV	BR:Vis14:11	*Phaseolus lunatus*	Viçosa, AL	July, 2011
158AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai1:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
159AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai2:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
161AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai3:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
162AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai4:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
163AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai5:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
164AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai6:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
165AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai7:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
166AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai8:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
167AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai9:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
168AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai10:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
169AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai11:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
170AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai12:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
173AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai13:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
174AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai14:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
176AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai15:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
177AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai16:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011
178AL	S09°27’11.3”	W036°36’34.3”	BGMV	BR:Pai17:11	*Phaseolus lunatus*	Palmeira dos Índios, AL	July, 2011

254AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur1:11	*Phaseolus lunatus*	Murici, AL	July, 2011
255AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur6:11	*Phaseolus lunatus*	Murici, AL	July, 2011
257AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur7:11	*Phaseolus lunatus*	Murici, AL	July, 2011
258AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur8:11	*Phaseolus lunatus*	Murici, AL	July, 2011
259AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur9:11	*Phaseolus lunatus*	Murici, AL	July, 2011
262AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur10:11	*Phaseolus lunatus*	Murici, AL	July, 2011
265AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur2:11	*Phaseolus lunatus*	Murici, AL	July, 2011
269AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur3:11	*Phaseolus lunatus*	Murici, AL	July, 2011
271AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur11:11	*Phaseolus lunatus*	Murici, AL	July, 2011
272AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur12:11	*Phaseolus lunatus*	Murici, AL	July, 2011
274AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur4:11	*Phaseolus lunatus*	Murici, AL	July, 2011
275AL	S09°17’10.0”	W035°57’43.3”	BGMV	BR:Mur5:11	*Phaseolus lunatus*	Murici, AL	July, 2011
310MG	-	-	BGMV	BR:Flt1:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
311MG	-	-	BGMV	BR:Flt2:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
323MG	-	-	BGMV	BR:Flt3:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
324MG	-	-	BGMV	BR:Flt4:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
325MG	-	-	BGMV	BR:Flt5:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
326MG	-	-	BGMV	BR:Flt6:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
329MG	-	-	BGMV	BR:Flt7:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
330MG	-	-	BGMV	BR:Flt8:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
332MG	-	-	BGMV	BR:Flt13:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
333MG	-	-	BGMV	BR:Flt9:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
334MG	-	-	BGMV	BR:Flt10:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
337MG	-	-	BGMV	BR:Flt11:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
338MG	-	-	BGMV	BR:Flt12:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
GO-2	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag11:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-3	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag1:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-4	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag2:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-7	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag3:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-9	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag7:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-11	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag4:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-15	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag5:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-22	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag6:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-25	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag12:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-26	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag8:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-27	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag9:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-29	S16°30’22.5”	W049°17’06.8”	BGMV	BR:Sag10:12	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	June, 2012
GO-63	S16°00’47”	W047°33’21.5”	BGMV	BR:Par14:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-64	S16°00’47”	W047°33’21.5”	BGMV	BR:Par1:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-65	S16°00’47”	W047°33’21.5”	BGMV	BR:Par2:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-66	S16°00’47”	W047°33’21.5”	BGMV	BR:Par15:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-68	S16°00’47”	W047°33’21.5”	BGMV	BR:Par16:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-70	S16°00’47”	W047°33’21.5”	BGMV	BR:Par3:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-71	S16°00’47”	W047°33’21.5”	BGMV	BR:Par17:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-72	S16°00’47”	W047°33’21.5”	BGMV	BR:Par4:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-73	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par5:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-75	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par6:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-76	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par18:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012

GO-77	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par7:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-78	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par19:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-79	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par8:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-80	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par20:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-81	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par21:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-82	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par22:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-85	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par23:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-88	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par9:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-89	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par10:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-91	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par24:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-92	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par25:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-93	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par26:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-96	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par27:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-97	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par11:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-102	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par28:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-103	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par29:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-104	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par30:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-106	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par12:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
GO-110	S15°57’57.4”	W047°31’35.8”	BGMV	BR:Par13:12	*Phaseolus vulgaris*	Paranoá, DF	June, 2012
MG-146	S16°43’02.2”	W046°45’14”	BGMV	BR:Una1:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-147	S16°43’02.2”	W046°45’14”	BGMV	BR:Una2:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-148	S16°43’02.2”	W046°45’14”	BGMV	BR:Una3:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-149	S16°43’02.2”	W046°45’14”	BGMV	BR:Una4:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-150	S16°43’02.2”	W046°45’14”	BGMV	BR:Una5:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-152	S16°43’02.2”	W046°45’14”	BGMV	BR:Una6:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-153	S16°43’02.2”	W046°45’14”	BGMV	BR:Una7:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-154	S16°43’02.2”	W046°45’14”	BGMV	BR:Una8:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-155	S16°43’02.2”	W046°45’14”	BGMV	BR:Una9:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-156	S16°43’02.2”	W046°45’14”	BGMV	BR:Una10:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-157	S16°43’02.2”	W046°45’14”	BGMV	BR:Una11:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-158	S16°43’02.2”	W046°45’14”	BGMV	BR:Una12:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-161	S16°43’02.2”	W046°45’14”	BGMV	BR:Una13:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-162	S16°43’02.2”	W046°45’14”	BGMV	BR:Una14:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-163	S16°43’02.2”	W046°45’14”	BGMV	BR:Una15:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
MG-165	S16°43’02.2”	W046°45’14”	BGMV	BR:Una16:12	*Phaseolus vulgaris*	Unaí, MG	June, 2012
GO-178	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri1:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-179	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri2:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-187	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri3:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-200	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri9:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-201	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri10:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-204	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri11:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-207	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri12:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-214	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri4:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-215	S17°05’29.2”	W047°37’01.4”	BGMV	BR:Cri5:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-221	S17°05’26.4”	W047°38’05.6”	BGMV	BR:Cri6:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-222	S17°05’26.4”	W047°38’05.6”	BGMV	BR:Cri7:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-223	S17°05’26.4”	W047°38’05.6”	BGMV	BR:Cri8:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-227	S17°05’26.4”	W047°38’05.6”	BGMV	BR:Cri13:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012

GO-228	S17°05'26.4"	W047°38'05.6"	BGMV	BR:Cri14:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-236	S17°05'26.4"	W047°38'05.6"	BGMV	BR:Cri15:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-238	S17°05'26.4"	W047°38'05.6"	BGMV	BR:Cri16:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
GO-241	S17°05'26.4"	W047°38'05.6"	BGMV	BR:Cri17:12	*Phaseolus vulgaris*	Cristalina, GO	June, 2012
181AL	S09°27'11.3"	W036°36'34.3"	*Macroptilium yellow vein virus*/MaYVV	BR:Pai18:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
182AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai19:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
183AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai20:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
184AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai21:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
185AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai22:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
187AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai23:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
188AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai24:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
189AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai25:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
190AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai26:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
191AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai27:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
192AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai28:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
194AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai29:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
195AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai30:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
196AL	S09°26'34.1"	W036°36'46.5"	MaYVV	BR:Pai31:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
197AL	S09°26'26.0"	W036°36'46.0"	MaYVV	BR:Pai32:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
198AL	S09°26'26.0"	W036°36'46.0"	MaYVV	BR:Pai33:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
199AL	S09°26'26.0"	W036°36'46.0"	MaYVV	BR:Pai34:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
200AL	S09°26'26.0"	W036°36'46.0"	MaYVV	BR:Pai35:11	*Macroptilium lathyroides*	Palmeira dos Índios, AL	July, 2011
321MG	-	-	*Soybean cholorotic spot virus*/SoCSV	BR:Flt13:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
328MG	-	-	SoCSV	BR:Flt14:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011
336MG	-	-	SoCSV	BR:Flt15:11	*Macroptilium lathyroides*	Florestal, MG	March, 2011

Suppl. Table S2. Begomovirus sequences retrieved from GenBank's non-redundant database.

Viral species/Isolate	GenBank access #	Host	Location	Date
Macroptilium yellow spot virus/MaYSV-[BR:Ced1:09]	JN419007	*Macroptilium lathyroides*	Cedro de São João, SE	2009
MaYSV-[BR:Bas1:09]	JN419005	*Macroptilium lathyroides*	Barra de Santana, PB	2009
MaYSV-[BR:Mac5:10]	JN419022	*Macroptilium lathyroides*	Maceió, AL	2010
MaYSV-[BR:Pdi1:10]	JN419020	*Macroptilium lathyroides*	Palmeira dos Índios, AL	2010
MaYSV-[BR:Inp2:10]	JN419019	*Macroptilium lathyroides*	Inhapi, AL	2010
MaYSV-[BR:Inp1:10]	JN419018	*Canavalia sp.*	Inhapi, AL	2010
MaYSV-[BR:Deg1:10]	JN419016	*Calopogonium mucunoides*	Delmiro Gouveia, AL	2010
MaYSV-[BR:Pir1:10]	JN419015	*Calopogonium mucunoides*	Piranhas, AL	2010
MaYSV-[BR:Agf2:10]	JN419014	*Macroptilium lathyroides*	Água das Flores, AL	2010
MaYSV-[BR:Agf1:10]	JN419013	*Macroptilium lathyroides*	Água das Flores, AL	2010
MaYSV-[BR:Bat1:10]	JN419012	*Macroptilium lathyroides*	Batalha, AL	2010
MaYSV-[BR:Mac1:10]	JN419009	*Macroptilium lathyroides*	Maceió, AL	2010
MaYSV-[BR:Crb7:11]	KC004134	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf25:11]	KC004133	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf24:11]	KC004132	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb6:11]	KC004131	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf14:11]	KC004130	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf13:11]	KC004129	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf12:11]	KC004128	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb5:11]	KC004127	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf11:11]	KC004126	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf10:11]	KC004125	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb4:11]	KC004124	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf9:11]	KC004123	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf6:11]	KC004122	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf8:11]	KC004121	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf5:11]	KC004120	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf4:11]	KC004119	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb3:11]	KC004118	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf3:11]	KC004117	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb2:11]	KC004116	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf7:11]	KC004115	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf2:11]	KC004114	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf1:11]	KC004113	*Phaseolus vulgaris*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf23:11]	KC004112	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb1:11]	KC004111	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Oaf22:11]	KC004110	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf21:11]	KC004109	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf20:11]	KC004108	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf19:11]	KC004107	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf18:11]	KC004106	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf17:11]	KC004105	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf16:11]	KC004104	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Oaf15:11]	KC004103	*Macroptilium lathyroides*	Olho D'água das Flores, AL	July, 2011
MaYSV-[BR:Crb19:11]	KC004102	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb18:11]	KC004101	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb17:11]	KC004100	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb16:11]	KC004099	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb14:11]	KC004097	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb13:11]	KC004096	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb12:11]	KC004095	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb11:11]	KC004094	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb10:11]	KC004093	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb9:11]	KC004092	*Phaseolus vulgaris*	Craibas, AL	July, 2011
MaYSV-[BR:Crb8:11]	KC004091	*Phaseolus vulgaris*	Craibas, AL	July, 2011
Bean golden mosaic virus/BGMV-[BR:Sag7:Soy:08]	FJ665283	*Glycine max*	Santo Antônio de Goiás, GO	2008
BGMV-[GO 87-1]	M88686	*Phaseolus vulgaris*	Santo Antônio de Goiás, GO	1991
BGMV-[BR:Car3:10]	JN419003	*Macroptilium lathyroides*	Caruaru, PE	2010
BGMV-[BR:Car4:10]	JN419004	*Macroptilium lathyroides*	Caruaru, PE	2010
BGMV-[BR:Mac2:10]	JN419006	*Macroptilium lathyroides*	Maceió, AL	2010
Blainvillea yellow spot virus-[BR:Vic09:10]	KC706519	*Blainvillea rhomboidea*	Viçosa, MG	2010
Macroptilium yellow vein virus-[BR:Mac4:10]	JN419021	*Macroptilium lathyroides*	Maceió, AL	2010
Tomato rugose mosaic virus-[Ube]	AF291705	*Solanum lycopersicum*	Uberlândia, MG	-
Tomato mottle leaf curl virus-[BR:Juaz2586:04]	JF803250	*Solanum lycopersicum*	Juazeiro, BA	2004
Tomato interveinal chlorosis virus-[BR:BSF2729:04]	JF803253	*Solanum lycopersicum*	Belém de S. Francisco, PE	2004
Macroptilium yellow net virus-[BR:Mur1:09]	JN418998	*Macroptilium lathyroides*	Murici, AL	2009
Soybean chlorotic spot virus-[BR:Jai9254.10]	JX122965	*Glycine max*	Jaíba, MG	2010
Cleome leaf crumple virus-[BR:AL-Ata1:07]	JN103433	*Cleome affinis*	Atalaia, AL	2007
Sida mottle Alagoas virus-[BR:Mar3:10]	JX871388	*Sida urens*	Maragogi, AL	2010
Sida mosaic Alagoas virus-[BgV02A.1.C61]	JF694472	*Sida* sp.	Rio Largo, AL	2003

Sida yellow blotch virus-[BR:Rla1:10]	JX871380	*Sida urens*	Rio Largo, AL	2010
Tomato severe rugose virus-[BR:780Tom3:08]	JX415196	*Solanum lycopersicum*	Luziânia, GO	2008

Suppl. Table S3. Positively and negatively selected sites in subpopulation-specific CP and Rep genes of BGMV and MaYSV.

Dataset	Number of haplotypes	Methods						
		SLAC		FEL		REL		PARRIS
		Positive	Negative	Positive	Negative	Positive	Negative	
MaYSV CP	86	**250***	**24**,**34**,**50**,53,56,**59**,**61**,**65**,**70**,**71**,**75**,**80**,**85**,**86**,**100**,**106**,**132**,**137**,**138**,**139**,**140**,**143**,**144**,**148**,**155**,**156**,**161**,**162**,**173**,**177**,**181**,**182**,**188**,**190**,**194**,**195**,**200**,**205**,**206**,**207**,**208**,**212**,**221**,**222**,**223**,**224**,**233**,**238**,**246**,**249**	142,163,**250**	12,**24**,**34**,37,**50**,52,**59**,60,**61**,**65**,66,69,**70**,**71**,73,74,**75**,**80**,**85**,**86**,**100**,**106**,109,112,122,127,**132**,134,**137**,**138**,**139**,**140**,**143**,**144**,**148**,149,**155**,**156**,**161**,**162**,**173**,**177**,**181**,**182**,185,187,**188**,**190**,191,192,**194**,**195**,**200**,204,**205**,**206**,**207**,**208**,209,211,**212**,218,**221**,**222**,**223**,**224**,229,**233**,237,**238**,**246**,**249**	ND†	ND	ND
MaYSV RePCI§	29	0	**27**,**172**,**221**,**223**,**257**,**260**,**291**,**316**	0	23,**27**,36,28,41,44,45,62,73,74,103,121,137,141,142,151,156,**172**,215,218,**221**,**223**,238,**257**,**260**,**261**,**291**,**316**	346	**27**,**172**,**291**,**316**	NS‡
MaYSV RePCII§	14	0	**114**,**154**,**155**,**172**,**178**,**202**,**213**,**258**,**261**,**273**	0	42,44,64,70,83,**114**,139,142,145,149,152,**154**,**155**,156,157,162,**172**,**177**,**178**,**179**,**202**,204,**213**,**258**,**260**,**261**,**273**,283,291,292,2**99**,**311**,321,330,**339**	158,181,197,250,335,346	**114**,**154**,**155**,**172**,**177**,**178**,**179**,**213**,**258**,**260**,**261**,**273**,**291**,**299**,**311**,**339**	NS
MaYSV RePCIII§	46	0	**71**,**110**,**213**,**217**,**228**,**256**,**258**,**261**,**290**,**310**,**334**	174,325	**10**,**12**,**21**,**38**,**44**,**48**,**71**,**73**,**94**,**99**,**110**,**111**,**178**,**179**,**213**,**217**,**219**,**228**,**256**,**258**,**261**,**290**,**291**,**298**,**299**,**310**,**334**,**339**	0	3,4,**10**,11,**12**,**21**,29,36,**38**,41,**44**,47,**48**,54,63,69,**71**,**73**,82,92,**94**,**99**,**110**,**111**,114,117,120,128,136,137,146,147,152,166,177,**178**,**179**,192,200,202,**213**,**217**,**219**,221,**228**,235,253,**256**,**258**,260,**261**,286,**290**,291,**298**,**299**,**310**,311,330,**334**,**339**	NS
BGMV CP (*P. vulgaris*)	33	0	0	0	3,114,122,139,177,242,247	11,27,48,79,123,179,227	0	NS
BGMV CP (*M. lathyroides*)	5	0	**219**	0	**219**	0	160,**219**	NS
BGMV CP (*P. lunatus*)	31	0	**154**	0	46,60,107,114,**154**,172,180,192,193,194,199,200,218,219	0	0	NS
BGMV ReP (*P. vulgaris*)	46	0	**106**,**150**,**163**	0	**14**,**25**,**62**,**74**,**89**,**94**,**106**,**150**,**151**,**156**,**163**,**200**,**208**,**268**	0	12,**14**,**25**,26,38,**62**,69,**74**,78,82,**89**,**94**,**106**,128,138,**150**,**151**,**156**,157,**163**,169,193,**200**,205,**208**,228,245,**268**,312,321	NS
BGMV ReP (*M. lathyroides*)	4	0	0	0	0	38,39,54,58,71,75,133,154,155,169,177,179,183,193,205,210,211,212,225,226,227,236,245,251,266,291,304,334	0	NS
BGMV ReP (*P. lunatus*)	33	0	**8**,**90**,**92**,**106**,**213**,**214**	0	**7**,**8**,**43**,**47**,**63**,**75**,**90**,**92**,**106**,**112**,**158**,**213**,**214**,**215**,**247**,**261**,**279**,**293**,**316**,**330**	0	**7**,**8**,17,38,39,**43**,**47**,50,**63**,**75**,87,**90**,**92**,105,**106**,**112**,127,135,138,142,152,154,155,**158**,169,175,179,**213**,**214**,**215**,218,219,**247**,250,**261**,**279**,**293**,**316**,328,**330**,331	NS

* Sites detected by more than one method are indicated in bold.

† Not determined.

‡ Not significant.

§ Clusters observed in Figure 1b.

CHAPTER 2

MIGRATION AND RECOMBINATION AFFECTING BEGOMOVIRUS EVOLUTION IN THE NEW WORLD

Ramos-Sobrinho, R., Seah, Y.M., Loelius, S., Silva, J.C.F., Zerbini, F.M., Duffy, S. Migration and recombination affecting begomovirus evolution in the New World. Journal of General Virology, *in preparation*.

Migration and recombination affecting begomovirus evolution in the New World

Roberto Ramos Sobrinho[a,b,c], Yee Mey Seah[a], Shannon Loelius[a], José Cleydson Ferreira da Silva[c], Francisco Murilo Zerbini[b,c#], Siobain Duffy[a#]

[a]Department of Ecology, Evolution and Natural Resources, Rutgers, The State University Of New Jersey, New Brunswick, NJ, 08901, USA

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

[c]National Research Institute on Plant-Pest Interactions, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil

[#]Corresponding author: Siobain Duffy

Phone: (+1-848) 932-6299; E-mail: duffy@aesop.rutgers.edu

[#]Corresponding author: Francisco Murilo Zerbini

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

Summary - Begomoviruses are whitefly-transmitted, single-stranded DNA plant viruses that constitute one of the largest and most important groups of emerging plant viruses, growing in importance as the number of species characterized and their economic impact continues to increase. Most New World (NW) begomoviruses have two genomic components designated as DNA-A and DNA-B, and their genomes have the capacity to evolve quickly via mutation, reassortment and recombination. Datasets including all DNA-A and DNA-B reference sequences of NW begomoviruses were obtained from GenBank and analyzed to understand begomovirus evolution in the NW. Recombination was confirmed as a very important evolutionary mechanism acting on DNA-A evolution. Phylogenetic relationships showed that South American begomoviruses do not exhibit geographic monophyly and that there were at least two independent introduction events in this region. Interestingly, we found evidence of a large pool of genetic variability not available to begomoviruses in South America (SA) due to geographical isolation. Additional introduction events may greatly affect begomovirus evolution in SA.

Introduction

Viruses belonging to the family *Geminiviridae* have circular ssDNA genomes and infect a broad range of plant species causing devastating diseases, mainly in subtropical and tropical countries (Legg & Fauquet, 2004; Morales, 2006). The family is divided into seven genera (*Begomovirus*, *Becurtovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*) according to the type of insect vector, host range, genome organization and phylogeny (Varsani *et al.*, 2014). Viruses classified within the genus *Begomovirus* are transmitted by the whitefly *Bemisia tabaci* and constitute one of the largest and most important groups of emerging plant viruses, growing in importance as the number of species characterized and their economic impact continues to increase (Brown *et al.*, 2012; Varma & Malathi, 2003). Most begomoviruses found in the New World (NW) have two genomic components designated as DNA-A and DNA-B, with the DNA-A alone being considered for taxonomy (Brown *et al.*, 2012). The DNA-A encodes proteins involved in viral replication, transcription, encapsidation and vector transmission, while the DNA-B contains genes responsible for viral movement within and between cells (Brown *et al.*, 2012). The DNA-A and DNA-B components of bipartite begomoviruses share a region of high nucleotide identity called the common region (CR). The CR contains *cis* elements that are implicated in viral replication and its regulation (Arguello-Astorga *et al.*, 1994; Eagle & Hanley-Bowdoin, 1997).

The begomovirus genome is able to evolve quickly through both mutation and recombination. DNA-A segments can be associated with different DNA-B segments (reassortment or pseudorecombination), or other new components and satellites (Briddon *et al.*, 2010; Rey *et al.*, 2012; Rojas *et al.*, 2005; Saunders *et al.*, 2002a; Seal *et al.*, 2006b). Very high mutation rates have been determined for begomoviruses (Duffy & Holmes, 2008; Duffy & Holmes, 2009) and recombination was found to be very common, with breakpoints

non-randomly distributed across the genome (Lefeuvre *et al.*, 2009; Lefeuvre *et al.*, 2007; Martin *et al.*, 2011; Martin *et al.*, 2005). As mixed infections by different begomoviruses are very common in non-cultivated hosts (Alabi *et al.*, 2008; García-Andrés *et al.*, 2006; Monde *et al.*, 2010; Silva *et al.*, 2012), these plants can play an important role serving as 'mixing vessels', facilitating recombination among distantly related begomoviruses and the generation of new viral genomes. Recombination of genomic components between existing begomoviruses can generate new begomovirus species, which can be better adapted than their progenitors (García-Andrés *et al.*, 2006). The expanded geographical range of *B. tabaci* populations (mostly the B biotype, which is more invasive and polyphagous) can facilitate the introduction of indigenous begomoviruses into crop plants and the exposure of potential host plants to begomoviruses (Gilbertson *et al.*, 2011; Seal *et al.*, 2006a). Therefore, the interplay between adaptation to host species and vector biotypes, as well as the geographic distribution of begomoviruses and of its insect vector, can influence the evolution of this virus group (Lima *et al.*, 2013; Navas-Castillo *et al.*, 2011; Rocha *et al.*, 2013).

We have carried out a study to obtain more information about the evolutionary mechanisms influencing the genetic diversity of begomoviruses in the NW. Datasets including all NW begomovirus sequences (DNA-A and DNA-B) were obtained from GenBank's non-redundant nucleotide database. We hypothesized that recombination affects NW begomoviruses evolution. Indeed, the network relationship among NW begomoviruses showed that recombination is a very important evolutionary mechanism affecting DNA-A evolution. We found that South American begomoviruses are not a monophyletic group, indicating at least two independent introductions in this region. Interestingly, one cluster is entirely comprised of viruses which are not yet in South America (SA), and could provide a significant amount of sequence variation to the existing clusters should it ever be introduced. Additional introductions of begomoviruses from elsewhere in the NW may greatly affect

begomovirus evolution in SA by increasing the probability of recombination and reassortment events which can introduce new virulence features into indigenous viruses.

Results

Begomovirus diversity in the New World

Datasets containing a total of 104 and 80 full-length DNA-A and DNA-B sequences were obtained, respectively. Each NW begomovirus species was represented by only one sequence (the reference sequence). Although most begomoviruses naturally occurring in the NW have two genomic components known as DNA-A and DNA-B (except for *Tomato leaf deformation virus* - ToLDeV, which has a single, DNA-A-like component; Melgarejo *et al.*, 2013) only the DNA-A is used to assign taxonomy (Brown *et al.*, 2012). Therefore, bipartite begomoviruses for which only the DNA-A sequence is available were included, leading to the higher number of DNA-A sequences retrieved. Using pairwise comparisons of the DNA-A sequences and the ≥89% nucleotide identity criterion established by *Geminiviridae* Study Group of the ICTV (Brown *et al.*, 2012), we confirmed that all 104 isolates belonged to different begomovirus species, with nucleotide identity percentages ranging from 68.9% to 88.7%, with most values between 70% and 81% (data not shown). For the DNA-B, nucleotide identity percentages ranged from 65.9% to 98.4%, with most values between 66% and 74% (data not shown).

Non-tree-like evolution

Network analysis of DNA-A sequences (Fig. 1a) revealed three main clusters (clusters I, II and III), which were also observed in the DNA-A phylogenetic tree (see below), with the remaining NW begomoviruses constituting an undefined group. A complex pattern of

recombination for NW begomoviruses was observed, with strong evidence of a recombinant origin for sequences in cluster I (Fig. 1a). Interestingly, we also found evidence for a recombinant origin for isolates in cluster III (Fig.1a), which contains isolates from the NW and OW.

The networks for the CP and Rep coding regions (Fig. 1b and 1c) are deeply incongruent. The Rep network shows a pattern very similar to the DNA-A, suggesting that this clustering is strongly influenced by recombination occurring mostly in the Rep gene. In the Rep network, the evidence of recombination in cluster III is stronger than for the DNA-A sequences (Fig. 1b).

Network analysis of DNA-B sequences indicated only a weak signal of recombination, with support for the cluster II observed in the DNA-A network (Fig. 1d). These results provide additional evidence that the viruses in cluster II represent an evolutionary lineage distinct from South American begomoviruses. In contrast to the DNA-A network, a cluster containing most of the South American isolates (except *Euphorbia yellow mosaic virus* - EuYMV) was observed, which was grouped separately from non-South American isolates (Fig. 1d).

The networks for the NSP and MP sequences (Fig. 1e and 1f) are incongruent, with the NSP network showing a pattern very similar to the DNA-B, except for *Cleome leaf crumple virus* (ClLCrV) which was grouped with non-South American isolates (Fig. 1e).

Recombination analysis

The RDP results confirmed the complex pattern of recombination for DNA-A genomes (Suppl. Table S2), as indicated by network analysis. At least 107 unique recombination events were detected, with breakpoints occurring mostly in the CR and Rep coding regions. All isolates in cluster I of the DNA-A tree shared a recombination event

(event 19; Suppl. Table S2) located in the Rep gene, but with the putative minor parent unknown. The same was observed for species in cluster III, which share a recombination event (event 53; Suppl. Table S2) with breakpoints located in the Ren coding region. Interestingly, there is a recombination event (event 38; Suppl. Table S2) shared amongst some NW begomoviruses in cluster III, as evidenced by network analysis, where *Corchorus golden mosaic virus* (CGMV, an OW begomovirus present in Vietnam that is closely related to NW begomoviruses) is the putative minor parent. For the DNA-B, only 26 unique recombination events were observed, with breakpoints occurring mostly in the CR (Suppl. Table S3).

Phylogenetic relationships

The DNA-A ML tree has three very well supported clusters and reflects the great species diversity among NW begomoviruses (Fig. 2). Cluster I contains isolates from all different geographic regions, whereas clusters II and III contain only non-South American isolates (Fig. 2). Phylogenetic relationships for other groups displayed very low bootstrap scores, indicating insufficient phylogenetic signal in the sequence dataset to give a statistically significant placement for all NW begomoviruses (Fig. 2). Sequences from Eastern South America (ESA) did not form a monophyletic clade apart from sequences in other NW regions. For example, in cluster I, a few Brazilian sequences grouped with begomoviruses from North and Central America and the Caribbean (Fig. 2). As recombination can affect our analysis, we redid the ML analysis without the recombinant region shared by all isolates in cluster I (event 19; Suppl. Table S1). EuYMV, a begomovirus that has so far been found only in Brazil, still clustered with non-South American begomoviruses (Suppl. Fig. S5). A similar result was observed for ToLDeV, a South

American begomovirus present in Ecuador and Peru (Melgarejo *et al.*, 2013), which clustered with begomoviruses from outside of this region.

For the DNA-B ML tree, the three main clusters observed in the case of the DNA-A were also present, but some topological incongruences could be discerned. For example, *Cotton leaf crumple virus* (CLCrV) and *Desmodium leaf distortion virus* (DeLDV), which were in the DNA-A cluster I, fell outside this cluster in the DNA-B tree (cluster IIa; Fig. 3). Similar results were observed for different isolates grouped in the DNA-A clusters I and II (Fig. 2 and 3). NW-like begomoviruses such as CGMV were placed, in both the DNA-A and DNA-B phylogenetic trees, basally located, in agreement with previously reported results (Ha *et al.*, 2006; 2008).

The CP and Rep ML trees are deeply incongruent (Suppl. Fig. S1 and S2), which is evidence of recombination. The Rep tree is very similar to the DNA-A tree and has two well-supported clades that are not resolved in the CP tree (clusters I and III; Suppl. Fig. S1 and S2). Cluster II is observed in both CP and Rep trees, but is not well-supported (Suppl. Fig. S1 and S2). Interestingly, the NSP and MP ML trees are not incongruent and both resemble the DNA-B tree (Suppl. Fig. S3 and S4). All clades observed for the DNA-B tree were also present in the NSP tree, but with lower bootstrap support. Although the clustering for MP sequences was very similar to NSP and the DNA-B, clades were not well-supported (Suppl. Fig. S3 and S4).

Phylogeography

Comparisons showed no difference among the models tested, so we opted for a strict molecular clock plus Bayesian Skyline demographic models to represent the phylogeographic relationships among DNA-A begomovirus sequences. The clustering observed was very similar to the DNA-A ML tree, with clusters I, II and III present. In cluster I, as in the DNA-

A tree, some Brazilian begomoviruses clustered with non-South American isolates. We found that the common ancestor for the viruses in cluster I was present in NA, with 0.99 posterior probability (Fig. 4). To confirm that EuYMV does not cluster with other Brazilian begomoviruses, we built a DNA-A ML tree forcing EuYMV to group apart from the non-South American isolates in cluster I and compared the likelihood between the two trees. The results showed that the tree where EuYMV clustered with non-South American begomoviruses was more likely ($p<0.0001$), confirming that EuYMV must be a migrant from outside SA. Interestingly, cluster II, which is composed by begomoviruses infecting different plant hosts in different regions, seems to represent an evolutionary lineage separated from South American begomoviruses. The common ancestor for isolates in this cluster was located in the Caribbean, with 1.00 posterior probability (Fig. 4). Although phylogeography indicates that cluster II has had an ancestor located in SA, it shows a low posterior probability (0.84; Fig. 4). Once again, NW-like begomoviruses were placed basally located (Fig. 4).

Discussion

Begomoviruses are unique among plant viruses in their strong geographic segregation, with NW and OW begomoviruses forming clearly distinct phylogenetic clades. Begomoviruses in the OW are much more diverse than those in the NW, and the DNA-A and DNA-B genomic components have different molecular evolutionary histories (Briddon *et al.*, 2010). Besides mutation and selection pressure, recombination is an important evolutionary force acting on begomovirus evolution (Lefeuvre *et al.*, 2009; Lefeuvre *et al.*, 2007; Martin *et al.*, 2011; Martin *et al.*, 2005). While intraspecific and interspecific recombination generally plays a minor role in the evolution of begomovirus populations (Lima *et al.*, 2013), recombination can be an important mechanism for begomovirus macroevolution. Here, we decided to assess the effect of recombination in NW begomoviruses evolution. Our results

confirmed the high species diversity of NW begomoviruses, with recombination detected as a very important mechanism affecting DNA-A evolution. Migration among the different biogeohraphic regions was detected to affect begomovirus evolution in the NW.

Pairwise comparisons showed a high degree of variability amongst begomoviruses infecting different plant hosts throughout the Americas, with the DNA-B displaying a higher range of nucleotide identities. Molecular variability in the DNA-A has been associated with recombination events occurring mostly in the Rep coding region (Lima *et al.*, 2013). We observed also that this region is highly recombinant, and a much lower number of recombination events were observed in the DNA-B dataset. Therefore, the greater molecular variability of DNA-B components observed here cannot be explained by recombination. A higher genetic variability of DNA-B components has been previously reported (Briddon *et al.*, 2010) and it was suggested that this could be explained by the absence of overlapping genes in the DNA-B and the different functions of the proteins encoded by these components, with greater room for variability in the DNA-B. As the DNA-A has more *cis*- and *trans*-interactions (*e.g.*, transactivation, Rep-iteron recognition, Rep-REn interaction and CP-vector interaction), genetic variation constraints should be stronger for the gene products of this component.

Moreover, the loss of the AV2 gene (coding an accessory movement protein) by NW begomoviruses may have increased selection pressure on the remaining movement protein BC1, leading for example to selection of a putative tyrosine phosphorylation site within BC1 that is much more conserved in NW begomoviruses and is possibly responsible for more efficient viral movement to compensate for the loss of the AV2 gene (Ho *et al.*, *unpublished results*). Studies with RNA plant viruses have shown that phosphorylation of MP regulates its localization and may account for virus movement between cells (Modena *et al.*, 2008).

Most NW begomoviruses have a bipartite genome and cognate components share a common region (CR), which ensures that replication of both components can be initiate by the DNA-A-encoded Rep protein, maintaining the integrity of the divided genomes (Arguello-Astorga *et al.*, 1994; Eagle & Hanley-Bowdoin, 1997; Fontes *et al.*, 1994). In most cases, the DNA-A can donate its CR, through recombination, to a non-cognate DNA-B and this event can facilitate the occurrence of reassortment (Hou & Gilbertson, 1996; Saunders *et al.*, 2002b). Studies have shown that begomoviruses in the SqLCV clade (here, cluster I) can yield viable reassortants, and these reassortment events can extend the host range of the begomovirus *Bean calico mosaic virus* (BCaMV) to cucurbit species. These begomoviruses share a highly similar Rep amino-terminal sequence and identical Rep-binding iterons, which can enhance their potential for reassortment (Brown *et al.*, 1999; Idris *et al.*, 2008).

Previous phylogenetic analyses indicated that *Cotton leaf crumple virus* has a complex evolutionary history probably involving both recombination between begomoviruses belonging to different clades and reassortment (Idris & Brown, 2004). Here, we found that most recombination events in the DNA-A are located in the CR and the Rep gene. Although these regions are known to be recombination hot spots (Lefeuvre *et al.*, 2009; Lefeuvre *et al.*, 2007), studies have provided evidence that the high nucleotide variability in the N-terminal portion of the Rep gene is accompanied by strong purifying selection that preserves the amino acid sequence (Lima *et al.*, 2013; Rocha *et al.*, 2013). The Rep N-terminal region in geminiviruses includes conserved motifs essential for rolling-circle replication (Ilyina & Koonin, 1992; Koonin & Ilyina, 1992; Nash *et al.*, 2011). Conservation of the integrity of these elements is critical for successful infection cycles, despite the variation introduced by frequent recombination. Several of the recombination events detected here involved the donation of the Rep N-terminal region and most of the CR, and this new genetic variability inserted by recombination may contribute to DNA-A diversification.

Noticeably, EuYMV (or its ancestor), identified here as a migrant from outside of SA, is perhaps the only Brazilian begomovirus that coexisted with non-South American begomoviruses in cluster I. Therefore, we believe that EuYMV received this genomic fragment before the migration event and then spread it, via recombination, to begomoviruses present in SA. These results highlight the importance of recombination for begomovirus evolution in the NW and are additional evidence that migration of begomoviruses present in cluster II can represent serious problems for crop protection in SA, as these 'migrants' could insert new molecular features into South American begomoviruses.

Network relationships of the DNA-A showed evidence for recombinant origin for NW and OW begomoviruses in cluster III. *Macroptilium golden mosaic virus* (MaGMV), a begomovirus naturally infecting *Macroptilium,* is also able to infect solanaceous hosts such as tomato and pepper and is closely related to *Potato yellow mosaic virus* (PYMV). Interestingly, although MaGMV shows features of NW begomoviruses, its REn protein is most similar to OW begomoviruses such as *Corchorus yellow vein virus* (CoYVV) (Collins *et al.*, 2010). These results can be evidence that some migration fo begomoviruses from NW to OW may have occurred.

Phylogenetic trees showed three major groups of NW begomoviruses (clusters I, II and III). This clustering is very similar to those observed in other published phylogenies of NW begomoviruses (Briddon *et al.*, 2010; Rocha *et al.*, 2013). Although begomoviruses from SA (as defined here, *i.e.*, excluding Colombia and Venezuela) do not form a monophyletic group, most of them cluster together and separately from begomoviruses located in Central America (CA), North America (NA) and the Caribbean (CB), which show a closer relationship to each other than to sequences from SA. ToLDeV, a Western South American begomovirus, clustered with other non-Eastern South American begomoviruses. These results suggest that geographic features (*e.g.*, the Amazon rainforest and the Andes mountains) are

efficient natural barriers to begomovirus spread by whiteflies and that begomoviruses are evolving independently in each area.

Phylogenetic relationships of *Bean yellow chlorosis virus* (BYCV), a Venezuelan begomovirus isolated from *Phaseolus vulgaris*, have shown that this isolate grouped in a distinct clade of NW begomoviruses, including non-South American begomoviruses (Fiallo-Olive *et al.*, 2013). The authors suggest that the absence of Brazilian begomoviruses, a neighbor country of Venezuela, is due to the presence of ecological barriers such as the Orinoco River and the Amazon rainforest, which can prevent the spread of *B. tabaci* between these countries. To reinforce these observations, two begomovirus species that are able to infect the same hosts (*Bean golden mosaic virus*, which is found from the eastern boundaries of the Amazon rainforest to Argentina, and *Bean golden yellow mosaic virus*, found from southern Mexico to Colombia) have been kept separated for a long time, probably due the effective natural barrier imposed by the Amazon rainforest, as this geographic feature separates the regions where these begomoviruses are indigenous (Costa, 1965; Morales, 2006; Morales & Anderson, 2001). Nevertheless, we found evidence that some long distance spread occurred between these areas (*i.e.*, between the northern and southern ends of the Amazon rainforest). EuYMV isolates were previously reported to be more closely related to Peruvian and North American begomoviruses than to Brazilian begomoviruses and it was hypothesized that EuYMV was introduced into Brazil, maybe brought with a vegetatively propagated ornamental relative of *Euphorbia heterophylla* such as *E. pulcherrima* or *E. milii* (Fernandes *et al.*, 2011). Our results confirmed that the Brazilian begomovirus EuYMV is a migrant from outside of SA.

Other South American begomoviruses were reported to be closely related to pepper- and *Euphorbia*-infecting begomoviruses from Central America, grouping in cluster I (Marquez-Martin *et al.*, 2012; Paprotka *et al.*, 2010), but we observed that this clustering is

affected by recombination events occurring in this clade. When we removed the recombinant region shared by all isolates in this cluster, only EuYMV clusters with non-South American begomoviruses. Together, these results represent strong evidence for at least two independent introductions of begomoviruses in ESA, showing that Eastern South American begomoviruses are not a simple paraphyletic group.

Cluster II contained only non-South American isolates, but they are not clustered by geographic location or plant host species. Although recombination seems to affect cluster I, no evidence of recombination events shared by all isolates in cluster II was found, which could explain this grouping. We believe it represents an evolutionary lineage separated from other NW begomoviruses, evolving in a separate way from other begomoviruses. Therefore, begomoviruses placed in cluster II may represent a source of genetic variability not available (due to geographical isolation) to begomoviruses in SA. Introduction of these viruses could greatly affect begomovirus evolution in SA as they can recombine and introduce new features into South American begomoviruses. Our results further indicate that migration is very frequent within the regions located at the northern and southern ends of the Amazon rainforest, as we can see that the main clusters observed here contain isolates from different regions, in some cases distant from these areas. These results show that migration followed by recombination may influence NW begomovirus evolution, highlighting long-distance spread as an important evolutionary event for begomovirus diversification.

The recent spread of the polyphagous *B. tabaci* biotype B (Morales, 2006) might have facilitated host switching of indigenous begomoviruses, providing new opportunities for recombination and reassortment to occur, leading to the emergence of novel viruses. Topological incongruences between DNA-A and DNA-B phylogenetic trees were observed here and are additional evidence of reassortment events between begomoviruses in the NW. *Bean golden yellow mosaic virus* (BGYMV) was reported to be able to form asymmetric

viable reassortants with a begomovirus species belonging to a different phylogenetic cluster (Garrido-Ramirez *et al.*, 2000). Incongruent placement for *Macroptilium mosaic Puerto Rico virus* (MaMPRV) suggest that its components may have evolved along different paths, with ancestors from the current begomovirus clades that contain PHYVV and BGYMV (Idris *et al.*, 2003). Together, these results represent additional evidence that genomic reassortment between different clusters may have occurred, showing the importance of reassortment in the evolution of NW begomoviruses.

Recombination was previously reported to affect the evolution of OW begomoviruses (Lefeuvre *et al.*, 2007). Here, we observed that, besides mutation, recombination seems pivotal for begomovirus evolution in the NW. Migration was detected to be very frequent into the different biogeographic regions, with its effect being highlighted by recombination, which was responsible for inserting new molecular features to indigenous begomoviruses and for mixing the genetic variability already present. Although migration and recombination deeply affect the evolution of NW begomoviruses, the maintenance of distinct clusters despite overlapping geographic regions and hosts was observed. These clusters represent sources of genetic variability, which can be transferred to begomoviruses in other clusters by migration and recombination.

Methods

Begomovirus datasets

All DNA-A and DNA-B reference sequences of begomoviruses from the NW were retrieved from GenBank's non-redundant nucleotide database (www.ncbi.nlm.nih.gov/genbank; accessed in September 2013). We also included sequences of some bipartite begomoviruses from the Old World (OW) because they are more similar to

NW than OW begomoviruses (Ha *et al.*, 2006; Ha, 2008). Sequences were classified according to sampling location for phylogeographic analysis (Suppl. Table S1). We used natural geographic features (*e.g.*, the Andes mountains and the Amazon rainforest) to divide the NW into five different geographic regions: Western South America (WSA) and Eastern South America (ESA), countries located western and eastern of the Andes mountains, respectively; Caribbean (CB), including Venezuela and Florida because they are physically close to countries located in the Caribbean Sea; Central America (CA), including Colombia; and North America (NA), including Mexico. Venezuela and Colombia were classified separately from South America due to their location above the Amazon rainforest, a natural barrier for whitefly dispersion.

Multiple sequence alignments

Multiple sequence alignments were prepared for the full-length DNA-A and DNA-B sequences and for the CP, Rep, NSP and MP coding sequences of each viral species using the MUSCLE algorithm (Edgar, 2004) and manually adjusted using Se-Al v. 2.0a11 (tree.bio.ed.ac.uk/software/seal/). To confirm assigned taxonomy, DNA-A sequences were analyzed using the program Species Demarcation Tool v. 1.0 (Muhire *et al.*, 2013).

Network and recombination analyses

Evidence of non-tree-like evolution was assessed for DNA-A, DNA-B, CP, Rep, NSP and MP datasets using the Neighbor-Net method (Bryant & Moulton, 2004) implemented in SplitsTree v. 4.10 (Huson & Bryant, 2006). Images of networks were edited using Adobe Illustrator.

Possible parental sequences and recombination breakpoints were determined for full-length DNA-A and DNA-B genomes using the rdp, Geneconv, Bootscan, Maximum Chi

Square, Chimaera, SisterScan and 3Seq methods implemented in Recombination Detection Program (RDP) v. 3.44 (Martin *et al.*, 2010). Alignments were scanned with default settings for the different methods and statistical significance was inferred by a *P*-value lower than a Bonferroni-corrected cut-off of 0.05. Only recombination events detected by at least three different methods were considered to be reliable.

Phylogenetic analysis

Maximum likelihood (ML) trees were inferred for all datasets using RAxML v. 7.0.3 (Stamatakis, 2006), assuming a general time reversible nucleotide substitution model with a gamma model of rate heterogeneity. The robustness of each individual branch was estimated from 2000 bootstrap replicates. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree) and Adobe Illustrator.

Phylogeography analysis

Spatial diffusion for the full-length DNA-A was estimated using a Bayesian MCMC statistical framework implemented in BEAST v.1.7.5 (Drummond & Rambaut, 2007). A matrix of geographic locations was created based on the sampling locations for each sequence/species. A full model was used, where all possible reversible exchange rates between locations were equally likely. The dataset was analyzed under demographic models of constant size, exponential growth and Bayesian Skyline. For all models, we used both relaxed (uncorrelated Lognormal) and strict molecular clocks, assuming a general time reversible nucleotide substitution model. MCMC analyses were run for 2×10^8 generations to achieve the convergence of parameters, which was assessed by calculation of the effective sample size (ESS) using TRACER v.1.4 (beast.bio.ed.ac.uk/Tracer). All parameter estimates showed ESS values >200, and their uncertainty were reflected in the 95% highest posterior

probability (HPP) intervals. The best model was chosen by calculating the Bayes factor from the posterior distribution output of each model using TRACER v.1.4. The MCC tree was summarized using TreeAnnotator v.1.7.5. (beast.bio.ed.ac.uk/TreeAnnotator), visualized and edited using FigTree and Adobe Illustrator.

Acknowledgements

R.R.S. was the recipient of a CNPq doctoral fellowship and a CAPES sandwich PhD fellowship. S.D. was supported by US NSF 1026095 and funds from the NJ Agricultural Experiment Station. F.M.Z. was supported by Fapemig (APQ-00949-09) and CNPq (484447/2011-4).

References

Alabi, O. J., Ogbe, F. O., Bandyopadhyay, R., Kumar, P. L., Dixon, A. G. O., Hughes, J. D. & Naidu, R. A. (2008). Alternate hosts of African cassava mosaic virus and East African cassava mosaic Cameroon virus in Nigeria. *Arch Virol* **153**, 1743-1747.

Arguello-Astorga, G. R., Guevara-González, R. G., Herrera--Estrella, L. R. & Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group-specific organization of interative elements: a model for replication. *Virology* **203**, 90-100.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-ul-Rehman, M. S. & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evol Biol* **10**, 97.

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

Brown, J. K., Ostrow, K. M., Idris, A. M. & Stenger, D. C. (1999). Biotic, molecular, and phylogenetic characterization of bean calico mosaic virus, a distinct Begomovirus species with affiliation in the squash leaf curl virus cluster. *Phytopathology* **89**, 273-280.

Bryant, D. & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255-265.

Collins, A., Rehman, M. M., Chowda-Reddy, R. V., Wang, A. M., Fondong, V., Brown, J. & Roye, M. (2010). Molecular characterization and experimental host range of an isolate of *Macroptilium golden mosaic virus* that infects *Wissadula amplissima* in Jamaica. *Virus Res* **150**, 148-152.

Costa, A. S. (1965). Three whitefly-transmitted virus diseases of beans in São Paulo, Brazil. *Plant Prot Bull FAO* **13**, 2-12.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

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

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

Eagle, P. A. & Hanley-Bowdoin, L. (1997). *Cis*-elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. *J Virol* **71**, 6947-6955.

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

Fiallo-Olive, E., Marquez-Martin, B., Hassan, I., Chirinos, D. T., Geraud-Pouey, F., Navas-Castillo, J. & Moriones, E. (2013). Complete genome sequences of two novel begomoviruses infecting common bean in Venezuela. *Arch Virol* **158**, 723-727.

Fontes, E. P. B., Eagle, P. A., Sipe, P. S., Luckow, V. A. & Hanley-Bowdoin, L. (1994). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J Biol Chem* **269**, 8459-8465.

García-Andrés, S., Monci, F., Navas-Castillo, J. & Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology* **350**, 433-442.

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

Gilbertson, R. L., Rojas, M. R. & Natwick, E. T. (2011). Development of integrated pest management (IPM) strategies for whitefly (*Bemisia tabaci*)-transmissible geminiviruses. In *The whitefly, Bemisia tabaci (Homoptera: Aleyrodidae) interaction with geminivirus-infected host plants*, pp. 323-356. Edited by W. M. O. Thompson. Dordrecht, The Netherlands: Springer.

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. *J Gen Virol* **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. *J Gen Virol* **89**, 312-326.

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

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.

Idris, A. M. & Brown, J. K. (2004). Cotton leaf crumple virus is a distinct western hemisphere begomovirus species with complex evolutionary relationships indicative of recombination and reassortment. *Phytopathology* **94**, 1068-1074.

Idris, A. M., Hiebert, E., Bird, J. & Brown, J. K. (2003). Two newly described begomoviruses of *Macroptilium lathyroides* and common bean. *Phytopathology* **93**, 774-783.

Idris, A. M., Mills-Lujan, K., Martin, K. & Brown, J. K. (2008). Melon chlorotic leaf curl virus: characterization and differential reassortment with closest relatives reveal adaptive virulence in the *Squash leaf curl virus* clade and host shifting by the host-restricted *Bean calico mosaic virus*. *J Virol* **82**, 1959-1967.

Ilyina, T. V. & 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 Res* **20**, 3279-3285.

Koonin, E. V. & Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokariotic plasmid rolling circle DNA replication initiator proteins. *J Gen Virol* **73**, 2763-2766.

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

Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B. & Lett, J. M. (2007). Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. *J Gen Virol* **88**, 3458-3468.

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

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. (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *J Gen Virol* **94**, 418-431.

Marquez-Martin, B., Maeso, D., Martinez-Ayala, A., Bernal, R., Teresa Federici, M., Vincelli, P., Navas-Castillo, J. & Moriones, E. (2012). Diverse population of a new bipartite begomovirus infecting tomato crops in Uruguay. *Arch Virol* **157**, 1137-1142.

Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B., Vincent, C., Reynaud, B. & Lett, J. M. (2011). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog* **7**, e1002203.

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

Martin, D. P., van der Walt, E., Posada, D. & Rybicki, E. P. (2005). The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* **1**, e51.

Melgarejo, T. A., Kon, T., Rojas, M. R., Paz-Carrasco, L., Zerbini, F. M. & Gilbertson, R. L. (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. *J Virol* **87**, 5397-5413.

Modena, N. A., Zelada, A. M., Conte, F. & Mentaberry, A. (2008). Phosphorylation of the TGBp1 movement protein of *Potato virus* X by a *Nicotiana tabacum* CK2-like activity. *Virus Res* **137**, 16-23.

Monde, G., Walangululu, J., Winter, S. & Bragard, C. (2010). Dual infection by cassava begomoviruses in two leguminous species (Fabaceae) in Yangambi, Northeastern Democratic Republic of Congo. *Arch Virol* **155**, 1865-1869.

Morales, F. J. (2006). History and current distribution of begomoviruses in Latin America. *Adv Virus Res* **67**, 127-162.

Morales, F. J. & Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch Virol* **146**, 415-441.

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. & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). *Arch Virol* **158**, 1411-1424.

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

Navas-Castillo, J., Fiallo-Olivé, E. & Sánchez-Campos, S. (2011). Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol* **49**, 219-248.

Paprotka, T., Metzler, V. & Jeske, H. (2010). The complete nucleotide sequence of a new bipartite begomovirus from Brazil infecting *Abutilon*. *Arch Virol* **155**, 813-816.

Rey, M. E. C., Ndunguru, J., Berrie, L. C., Paximadis, M., Berry, S., Cossa, N., Nuaila, V. N., Mabasa, K. G., Abraham, N., Rybicki, E. P., Martin, D., Pietersen, G. & Esterhuizen, L. L. (2012). Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the South-West Indian Ocean islands. *Viruses* **4**, 1753-1791.

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

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

Saunders, K., Bedford, I. D. & Stanley, J. (2002a). Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with ageratum yellow vein disease. *J Gen Virol* **83**, 907-913.

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

Seal, S. E., Jeger, M. J. & Van den Bosch, F. (2006a). Begomovirus evolution and disease management. *Adv Virus Res* **67**, 297-316.

Seal, S. E., Van den Bosch, F. & Jeger, M. J. (2006b). Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. *Crit Rev Plant Sci* **25**, 23-46.

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. (2012). Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathol* **61**, 457-467.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinf Appl Note* **22**, 2688-2690.

Varma, A. & Malathi, V. G. (2003). Emerging geminivirus problems: a serious threat to crop production. *Ann Appl Biol* **142**, 145-164.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Zerbini, F. M. & Martin, D. P. (2014). Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Arch Virol* **159**, DOI 10.1007/s00705-014-2050-2.

Figure legends

Fig. 1. Neighbor-net network based on full-length **(a)** DNA-A, **(b)** CP, **(c)** Rep, **(d)** DNA-B, **(e)** NSP and **(f)** MP nucleotide sequences of New World begomoviruses. Isolates from Eastern South America (ESA) are indicated in red, Western South America (WSA) in black, Central America (CA) in blue, North America (NA) in pink, the Caribbean (CB) in orange and Old World (OW) in green. A networked relationship instead of a bifurcating tree is suggestive of recombination.

Fig. 2. Midpoint-rooted maximum likelihood tree based on full-length DNA-A nucleotide sequences of New World begomoviruses. Isolates from Eastern South America (ESA) are indicated in red, Western South America (WSA) in black, Central America (CA) in blue, North America (NA) in pink, the Caribbean (CB) in orange and Old World (OW) in green. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal or higher than 85% by filled circles.

Fig. 3. Midpoint-rooted maximum likelihood tree based on full-length DNA-B nucleotide sequences of New World begomoviruses. Isolates from Eastern South America (ESA) are indicated in red, Western South America (WSA) in black, Central America (CA) in blue, North America (NA) in pink, the Caribbean (CB) in orange and Old World (OW) in green. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal or higher than 85% by filled circles.

Fig. 4. Phylogeographic clustering based on full-length DNA-A nucleotide sequences of New World begomoviruses. Isolates from Eastern South America (ESA) are indicated in red, Western South America (WSA) in black, Central America (CA) in blue, North America (NA)

in pink, the Caribbean (CB) in orange and Old World (OW) in green. Nodes with posterior probabilities equal to or higher than 0.99 are indicated by filled circles.

Fig. 1a

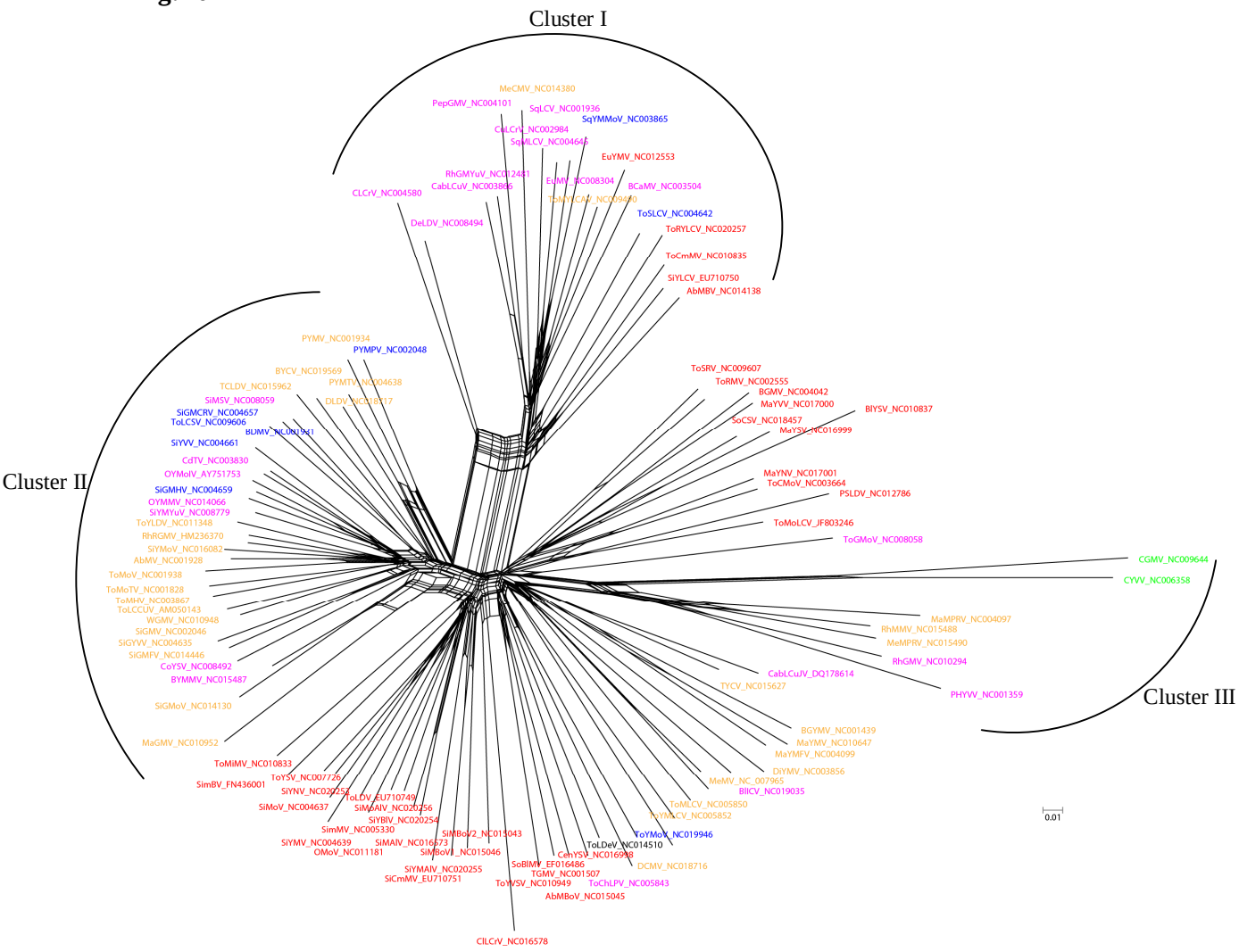

Cluster I
Cluster II
Cluster III
MeCMV_NC014380
PepGMV_NC004101
SqLCV_NC001936
SqYMMoV_NC003865
CaLCrV_NC002984
SqMLCV_NC004645
EuYMV_NC012553
RhGMYuV_NC012461
CabLCuV_NC003865
EuMV_NC008304
BCaMV_NC003504
TbMYCAV_NC009640
CLCrV_NC004580
DeLDV_NC008494
ToSLCV_NC004642
ToRYLCV_NC020257
ToCmMV_NC010835
SiYLCV_EU710750
AbMBV_NC014138
PYMV_NC001934
PYMPV_NC002048
BYCV_NC019569
PYMTV_NC004638
TCLDV_NC015962
DLDV_NC018617
SiMSV_NC008059
SiGMCRV_NC004657
ToLCSV_NC009606
BDMV_NC001941
SiYVV_NC004661
CdTV_NC003830
OYMoIV_AY751753
SiGMHV_NC004659
OYMMV_NC014066
SiYMYuV_NC008779
ToYLDV_NC011348
RhRGMV_HM236370
SiYMoV_NC016082
AbMV_NC001928
ToMoV_NC001938
ToMoTV_NC001828
ToMMV_NC003867
ToLCCUV_AM050143
WGMV_NC010948
SiGMV_NC002046
SiGYVV_NC004635
SiGMFV_NC014446
CoYSV_NC008492
BYMMV_NC015487
SiGMoV_NC014130
MaGMV_NC010952
ToMiMV_NC010833
ToYSV_NC007726
SimBV_FN436001
SiYNV_NC020257
ToLDV_EU710749
SiMoV_NC004637
SiMoAIV_NC020256
SiYBIV_NC020254
SimMV_NC005330
SiMBoV2_NC015043
SiYMV_NC004639
SiMAIV_NC016673
OMoV_NC011181
SiMBoV1_NC015046
SiYMAIV_NC020255
SiCmMV_EU710751
SoBIMV_EF016486
TGMV_NC001507
ToYVSV_NC010949
ToChLPV_NC005843
AbMBoV_NC015045
CILCrV_NC016578
ToLDeV_NC014510
CenYSV_NC016998
DCMV_NC018716
ToYMoV_NC019946
ToMLCV_NC005850
ToYMLCV_NC005852
MeMV_NC_007965
BIICV_NC019035
DiYMV_NC003856
MaYMFV_NC004099
MaYMV_NC010647
BGYMV_NC001439
PHYVV_NC001359
TYCV_NC015627
CabLCuJV_DQ178614
RhGMV_NC010294
MeMPRV_NC015490
RhMMV_NC015488
MaMPRV_NC004097
CYVV_NC006358
CGMV_NC009644
ToGMoV_NC008058
ToMoLCV_JF803246
PSLDV_NC012786
ToCMoV_NC003664
MaYNV_NC017001
MaYSV_NC016999
SoCSV_NC018457
BIYSV_NC010837
MaYMV_NC017000
BGMV_NC004042
ToRMV_NC002555
ToSRV_NC009607
0.01

Fig. 1b

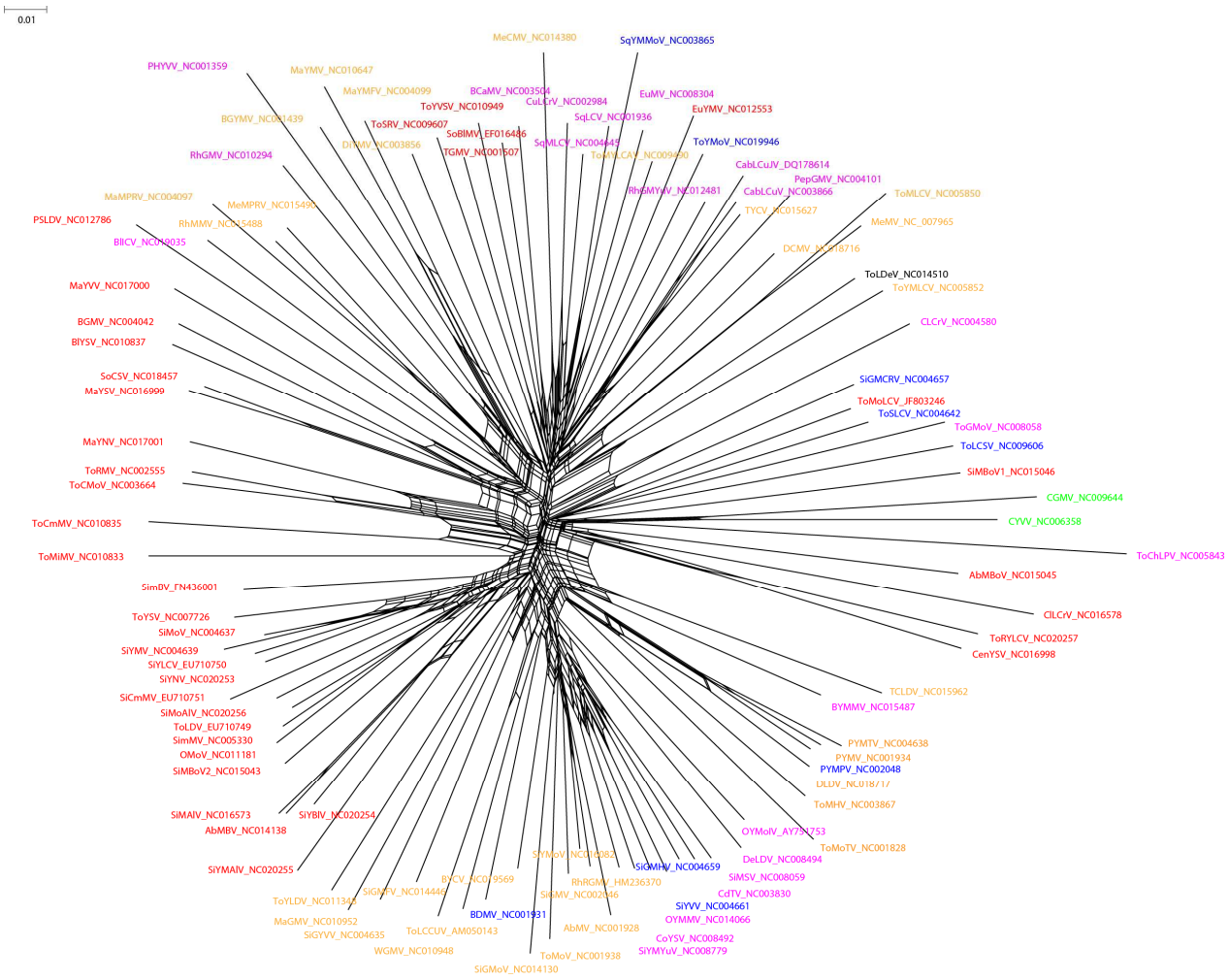

0.01
MeCMV_NC014380
SqYMMoV_NC003865
PHYVV_NC001359
MaYMV_NC010647
MaYMFV_NC004099
BCaMV_NC003504
CuLCrV_NC002984
EuMV_NC008304
ToYVSV_NC010949
EuYMV_NC012553
BGYMV_NC001439
SqLCV_NC001936
ToSRV_NC009607
SoBIMV_EF016446
ToYMoV_NC019946
DiMMV_NC003856
SqMLCV_NC004445
TGMV_NC001507
RhGMV_NC010294
CabLCuJV_DQ178614
PepGMV_NC004101
MaMPRV_NC004097
RhGMYV_NC012481
CabLCuV_NC003866
ToMLCV_NC005850
MeMPRV_NC015490
TYCV_NC015627
PSLDV_NC012786
RhMMV_NC015488
MeMV_NC_007965
BlCV_NC018035
DCMV_NC018716
ToLDeV_NC014510
MaYVV_NC017000
ToYMLCV_NC005852
BGMV_NC004042
CLCrV_NC004580
BlYSV_NC010837
SoCSV_NC018457
SiGMCRV_NC004657
MaYSV_NC016999
ToMoLCV_JF803246
ToSLCV_NC004642
ToGMoV_NC008058
MaYNV_NC017001
ToLCSV_NC009606
ToRMV_NC002555
SiMBoV1_NC015046
ToCMoV_NC003664
CGMV_NC009644
ToCmMV_NC010835
CYVV_NC006358
ToMiMV_NC010833
ToChLPV_NC005843
AbMBoV_NC015045
SimDV_FN436001
ClLCrV_NC016578
ToYSV_NC007726
SiMoV_NC004637
ToRYLCV_NC020257
SiYMV_NC004639
CenYSV_NC016998
SiYLCV_EU710750
SiYNV_NC020253
TCLDV_NC015962
SiCmMV_EU710751
BYMMV_NC015487
SiMoAIV_NC020256
ToLDV_EU710749
PYMTV_NC004638
SimMV_NC005330
PYMV_NC001934
OMoV_NC011181
PYMPV_NC002048
SiMBoV2_NC015043
ToMHV_NC003867
SiMAIV_NC016573
SiYBIV_NC020254
OYMoIV_AY750753
AbMBV_NC014138
ToMoTV_NC001828
DeLDV_NC008494
SiYMAIV_NC020255
SiGMHV_NC004659
SiMSV_NC008059
RhRGMV_HM236370
CdTV_NC003830
SiYVV_NC004661
OYMMV_NC014066
BDMV_NC001931
AbMV_NC001928
CoYSV_NC008492
SiYMYuV_NC008779
ToMoV_NC001938
SiGMoV_NC014130

Fig. 1c

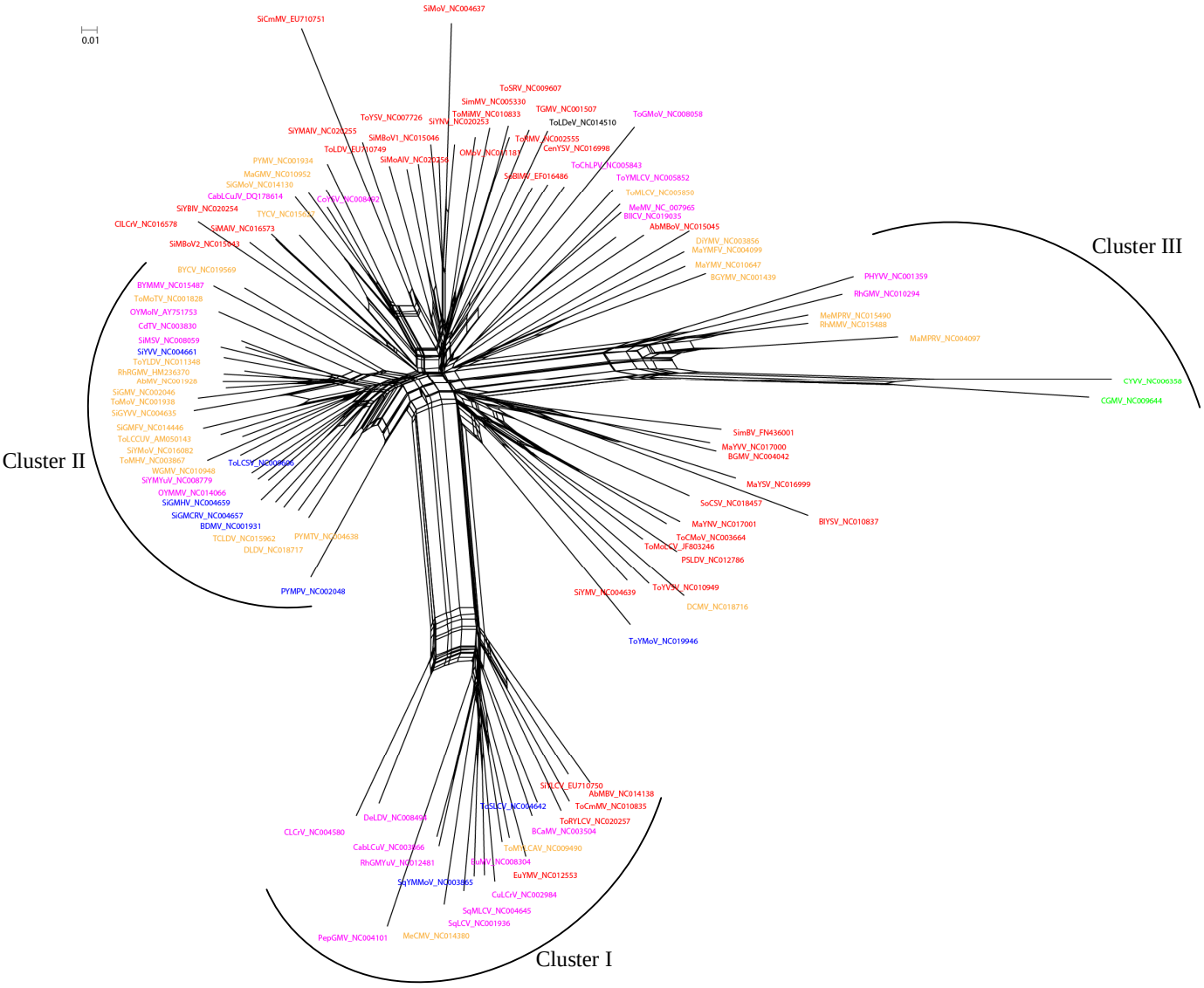

0.01
SiCmMV_EU710751
SiMoV_NC004637
ToSRV_NC009607
SimMV_NC005330
TGMV_NC001507
ToGMoV_NC008058
ToYSV_NC007726
SiYNV_NC020253
ToLDeV_NC014510
SiYMAIV_NC020255
SiMBoV1_NC015046
ToMiMV_NC010833
ToMMV_NC002555
CenYSV_NC016998
ToLDV_EU710749
OMoV_NC001181
PYMV_NC001934
SiMoAIV_NC020256
ToChLPV_NC005843
MaGMV_NC010952
SiBIMV_EF016486
ToYMLCV_NC005852
SiGMoV_NC014130
ToMLCV_NC005850
CabLCuJV_DQ178614
CoTYV_NC008462
MeMV_NC_007965
SiYBIV_NC020254
TYCV_NC015622
BlCV_NC019035
ClLCrV_NC016578
SiMAIV_NC016573
AbMBoV_NC015045
SiMBoV2_NC015643
DiYMV_NC003856
MaYMFV_NC004099
BYCV_NC019569
MaYMV_NC010647
BGYMV_NC001439
BYMMV_NC015487
PHYVV_NC001359
ToMoTV_NC001828
RhGMV_NC010294
OYMoIV_AY751753
MeMPRV_NC015490
CdTV_NC003830
RhMMV_NC015488
SiMSV_NC008059
MaMPRV_NC004097
SiYVV_NC004661
ToYLDV_NC011348
RhRGMV_HM236370
AbMV_NC001928
CYVV_NC006358
SiGMV_NC002046
ToMoV_NC001938
CGMV_NC009644
SiGYVV_NC004635
SiGMFV_NC014446
SimBV_FN436001
ToLCCUV_AM050143
MaYVV_NC017000
SiYMoV_NC016082
BGMV_NC004042
ToMHV_NC003867
ToLCSV_NC004686
MaYSV_NC016999
WGMV_NC010948
SiYMYuV_NC008779
SoCSV_NC018457
OYMMV_NC014066
BlYSV_NC010837
SiGMHV_NC004659
MaYNV_NC017001
SiGMCRV_NC004657
ToCMoV_NC003664
BDMV_NC001931
ToMoCV_JF803246
TCLDV_NC015962
PYMTV_NC004638
PSLDV_NC012786
DLDV_NC018717
ToYVV_NC010949
SiYMV_NC004639
PYMPV_NC002048
DCMV_NC018716
ToYMoV_NC019946
SiYLCV_EU710750
AbMBV_NC014138
ToSLCV_NC004642
ToCmMV_NC010835
DeLDV_NC008465
ToRYLCV_NC020257
CLCrV_NC004580
BCaMV_NC003504
CabLCuV_NC003366
ToMYLCAV_NC009490
RhGMYuV_NC012481
EuMV_NC008304
SiYMMoV_NC003865
EuYMV_NC012553
CuLCrV_NC002984
SqMLCV_NC004645
SqLCV_NC001936
PepGMV_NC004101
MeCMV_NC014380
Cluster I
Cluster II
Cluster III

Fig. 1d

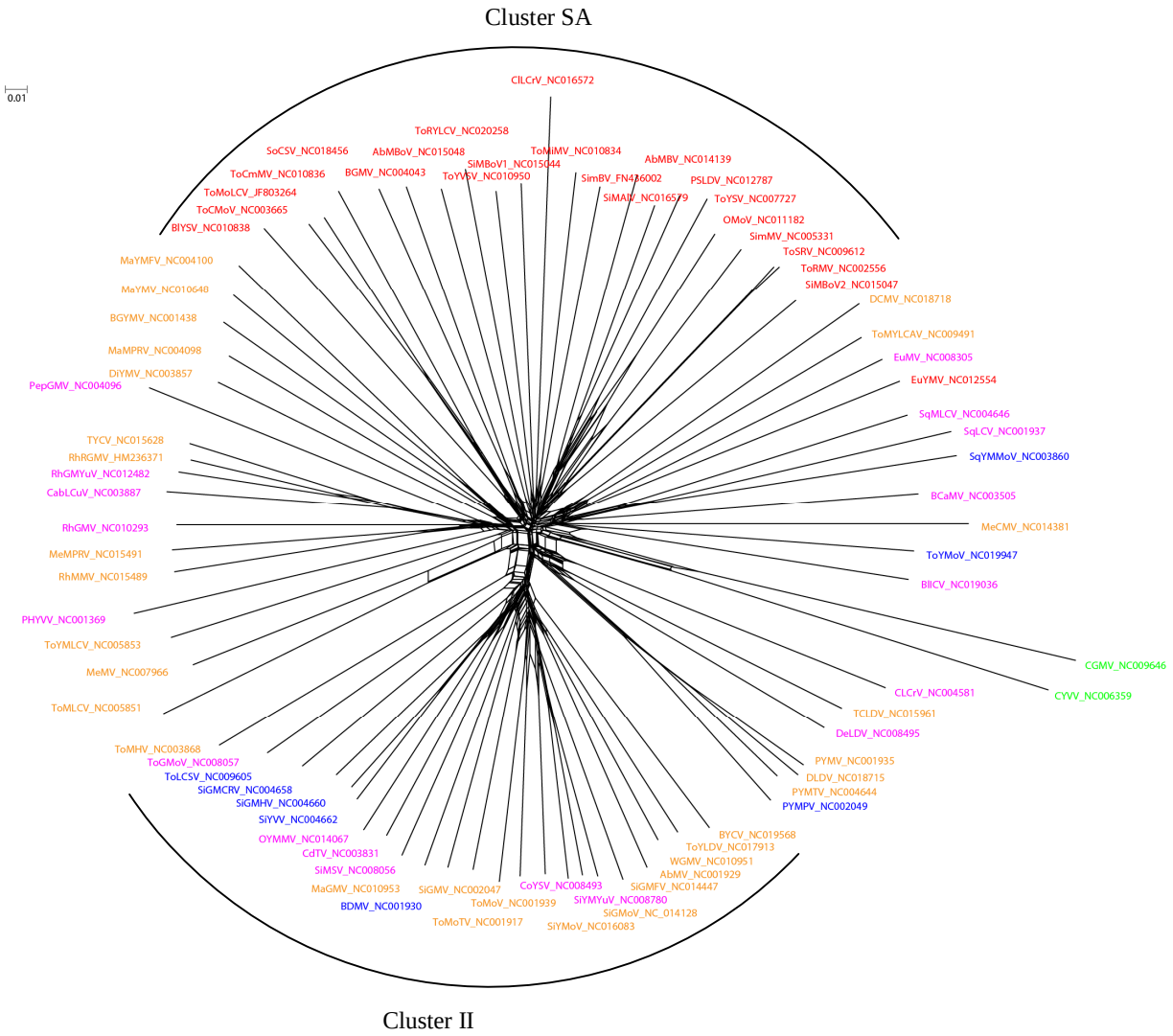

Cluster SA
0.01
ClLCrV_NC016572
ToRYLCV_NC020258
SoCSV_NC018456
AbMBoV_NC015048
ToNIMV_NC010834
AbMBV_NC014139
SiMBoV1_NC015044
ToCmMV_NC010836
BGMV_NC004043
ToYVSV_NC010950
SimBV_FN426002
PSLDV_NC012787
ToMoLCV_JF803264
SiMAIV_NC016579
ToYSV_NC007727
ToCMoV_NC003665
OMoV_NC011182
BlYSV_NC010838
SimMV_NC005331
ToSRV_NC009612
MaYMFV_NC004100
ToRMV_NC002556
MaYMV_NC010640
SiMBoV2_NC015047
DCMV_NC018718
BGYMV_NC001438
ToMYLCAV_NC009491
MaMPRV_NC004098
EuMV_NC008305
DiYMV_NC003857
EuYMV_NC012554
PepGMV_NC004096
SqMLCV_NC004646
SqLCV_NC001937
TYCV_NC015628
SqYMMoV_NC003860
RhRGMV_HM236371
RhGMYuV_NC012482
BCaMV_NC003505
CabLCuV_NC003887
RhGMV_NC010293
MeCMV_NC014381
MeMPRV_NC015491
ToYMoV_NC019947
RhMMV_NC015489
BllCV_NC019036
PHYVV_NC001369
ToYMLCV_NC005853
CGMV_NC009646
MeMV_NC007966
CLCrV_NC004581
CYVV_NC006359
ToMLCV_NC005851
TCLDV_NC015961
DeLDV_NC008495
ToMHV_NC003868
PYMV_NC001935
ToGMoV_NC008057
DLDV_NC018715
ToLCSV_NC009605
PYMTV_NC004644
SiGMCRV_NC004658
PYMPV_NC002049
SiGMHV_NC004660
SiYVV_NC004662
BYCV_NC019568
OYMMV_NC014067
ToYLDV_NC017913
CdTV_NC003831
WGMV_NC010951
SiMSV_NC008056
AbMV_NC001929
MaGMV_NC010953
SiGMV_NC002047
CoYSV_NC008493
SiGMFV_NC014447
BDMV_NC001930
ToMoV_NC001939
SiYMYuV_NC008780
SiGMoV_NC_014128
ToMoTV_NC001917
SiYMoV_NC016083
Cluster II

Fig. 1e

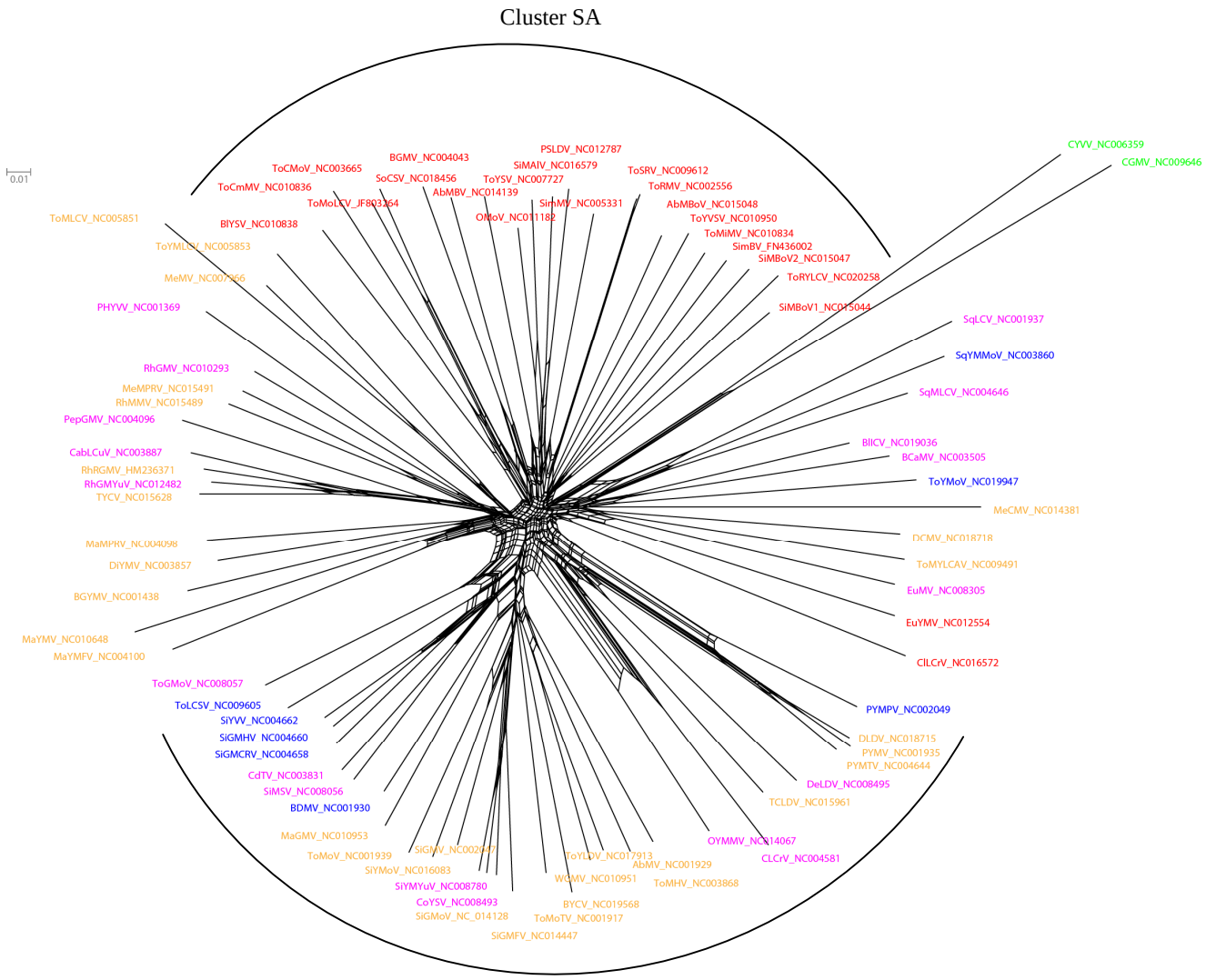

Cluster SA
0.01
PSLDV_NC012787
BGMV_NC004043
SiMAlV_NC016579
ToCMoV_NC003665
ToSRV_NC009612
SoCSV_NC018456
ToYSV_NC007727
ToCmMV_NC010836
AbMBV_NC014139
ToRMV_NC002556
ToMoLCV_JF803264
SimMV_NC005331
AbMBoV_NC015048
ToMLCV_NC005851
OMoV_NC011182
ToYVSV_NC010950
BIYSV_NC010838
ToMiMV_NC010834
SimBV_FN436002
ToYMLCV_NC005853
SiMBoV2_NC015047
MeMV_NC007366
ToRYLCV_NC020258
PHYVV_NC001369
SiMBoV1_NC015044
CYVV_NC006359
CGMV_NC009646
SqLCV_NC001937
SqYMMoV_NC003860
RhGMV_NC010293
MeMPRV_NC015491
SqMLCV_NC004646
RhMMV_NC015489
PepGMV_NC004096
BliCV_NC019036
CabLCuV_NC003887
BCaMV_NC003505
RhRGMV_HM236371
ToYMoV_NC019947
RhGMYuV_NC012482
TYCV_NC015628
MeCMV_NC014381
MaMPRV_NC004098
DCMV_NC018718
DiYMV_NC003857
ToMYLCAV_NC009491
EuMV_NC008305
BGYMV_NC001438
EuYMV_NC012554
MaYMV_NC010648
ClLCrV_NC016572
MaYMFV_NC004100
ToGMoV_NC008057
PYMPV_NC002049
ToLCSV_NC009605
SiYVV_NC004662
DLDV_NC018715
SiGMHV_NC004660
PYMV_NC001935
SiGMCRV_NC004658
PYMTV_NC004644
CdTV_NC003831
DeLDV_NC008495
SiMSV_NC008056
TCLDV_NC015961
BDMV_NC001930
MaGMV_NC010953
OYMMV_NC014067
ToMoV_NC001939
SiGMV_NC002047
ToYLDV_NC017913
CLCrV_NC004581
SiYMoV_NC016083
AbMV_NC001929
SiYMYuV_NC008780
WDMV_NC010951
ToMHV_NC003868
CoYSV_NC008493
BYCV_NC019568
SiGMoV_NC_014128
ToMoTV_NC001917
SiGMFV_NC014447
Cluster II

Fig. 1f

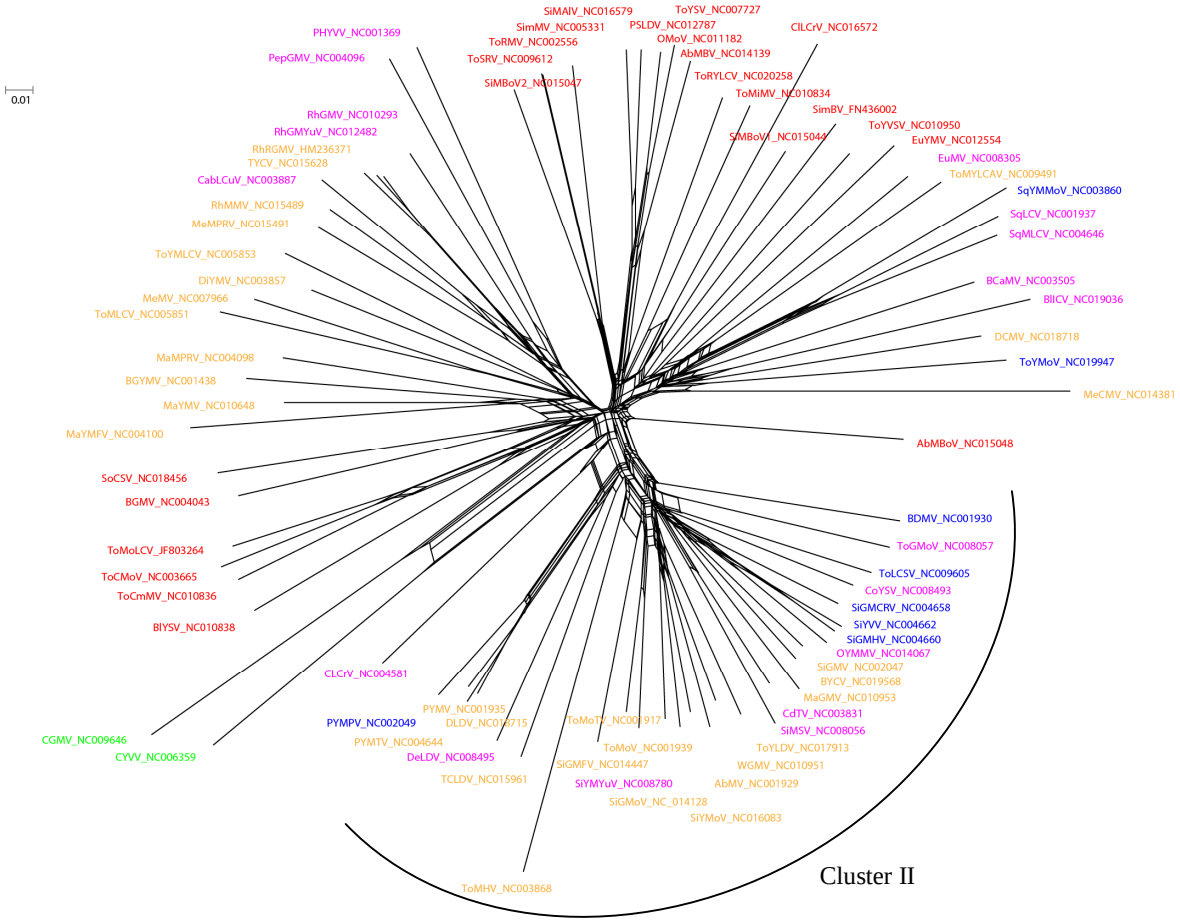

0.01
SiMAIV_NC016579
ToYSV_NC007727
SimMV_NC005331
PSLDV_NC012787
ClLCrV_NC016572
PHYVV_NC001369
ToRMV_NC002556
OMoV_NC011182
PepGMV_NC004096
ToSRV_NC009612
AbMBV_NC014139
SiMBoV2_NC015047
ToRYLCV_NC020258
ToMiMV_NC010834
RhGMV_NC010293
SimBV_FN436002
RhGMYuV_NC012482
ToYVSV_NC010950
SiMBoV1_NC015048
EuYMV_NC012554
RhRGMV_HM236371
EuMV_NC008305
TYCV_NC015628
ToMYLCAV_NC009491
CabLCuV_NC003887
SqYMMoV_NC003860
RhMMV_NC015489
SqLCV_NC001937
MeMPRV_NC015491
SqMLCV_NC004646
ToYMLCV_NC005853
BCaMV_NC003505
DiYMV_NC003857
BllCV_NC019036
MeMV_NC007966
ToMLCV_NC005851
DCMV_NC018718
MaMPRV_NC004098
ToYMoV_NC019947
BGYMV_NC001438
MeCMV_NC014381
MaYMV_NC010648
MaYMFV_NC004100
AbMBoV_NC015048
SoCSV_NC018456
BGMV_NC004043
BDMV_NC001930
ToGMoV_NC008057
ToMoLCV_JF803264
ToLCSV_NC009605
ToCMoV_NC003665
CoYSV_NC008493
ToCmMV_NC010836
SiGMCRV_NC004658
SiYVV_NC004662
BlYSV_NC010838
SiGMHV_NC004660
OYMMV_NC014067
CLCrV_NC004581
SiGMV_NC002047
BYCV_NC019568
PYMV_NC001935
MaGMV_NC010953
PYMPV_NC002049
DLDV_NC003715
ToMoTV_NC001917
CdTV_NC003831
SiMSV_NC008056
CGMV_NC009646
PYMTV_NC004644
ToMoV_NC001939
ToYLDV_NC017913
CYVV_NC006359
DeLDV_NC008495
SiGMFV_NC014447
WGMV_NC010951
TCLDV_NC015961
SiYMYuV_NC008780
AbMV_NC001929
SiGMoV_NC_014128
SiYMoV_NC016083
ToMHV_NC003868
Cluster II

Fig. 2


BlYSV_NC010837
MaYSV_NC016999
SoCSV_NC018457
TcBRV_NC009607
ToRMV_NC002555
PSLDV_NC012786
ToMoLCV_JF803246
ToCMoV_NC003664
MaYNV_NC017061
SlMBoV2_NC015043
CiLCrV_NC016578
TQMV_NC001507
SoBIMV_EF016986
ToYVSV_NC010949
TbYVSV_KC706643
CamYSV_NC016998
SlMBoV1_NC015046
SlYMAlV_NC020255
AbMBoV_NC015045
ToMlMV_NC010833
SlMoAlV_NC020256
SlmMV_NC005330
OMoV_NC011181
SlYNT_NC020253
SlMoV_NC004637
SlYMV_NC004639
ToYSV_NC007726
ToLDV_EU710749
SlCmMV_EU710751
BGMV_NC004042
MaYVV_NC017000
SimBV_FN436001
SlMAlV_NC016573
SlYBlV_NC020254
AbMBV_NC014138
ToCmMV_NC010835
SlYLCV_EU710750
DeLDV_NC008494
CLCrV_NC004580
ToSLCV_NC004642
BCaMV_NC003504
EuMV_NC008304
ToMYLCAV_NC009490
EuYMV_NC012553
SqLCV_NC001936
SlYMMoV_NC003865
SqMLCV_NC004545
CuLCrV_NC002984
MeCMV_NC014380
CabLCuV_NC003866
RhGMYuV_NC012481
PepGMV_NC004101
ToRYLCV_NC020257
Cluster I
BYCV_NC019569
PYMPV_NC002048
PYMV_NC001934
PYMTV_NC004538
DLDV_NC018717
TbLuV_NC015962
RhRGMV_HM236370
ToYLDV_NC011348
ToMoV_NC001938
SlGMV_NC002046
AbMV_NC001928
SlYMoV_NC016082
ToMHV_NC003867
ToLCGJV_AM050143
WGMV_NC010948
SlGMFV_NC014446
SlCYVV_NC004635
BYMMV_NC015462
ToMoTV_NC001828
SlMSV_NC008059
CdTV_NC003830
OYMoV_AY751753
SlYVV_NC004691
SlGMHV_NC004659
OYMMV_NC014066
ToLCSV_NC009606
SlYMYuV_NC008779
CoYSV_NC008492
MaGMV_NC010952
SlGMoV_NC014130
BDMV_NC001931
SlGMCRV_NC004657
Cluster II
ToYMLCV_NC005852
MeMV_NC007965
ToMLCV_NC005850
ToGMoV_NC008058
ToYMoV_NC019946
TbChLPV_NC005843
BlCV_NC019035
ToLDeV_NC019032
CabLCuJV_DQ178614
TYCV_NC015627
DCMV_NC018716
DlYMV_NC003856
MaYMV_NC010647
MaYMFV_NC004099
BGYMV_NC001439
PHYVV_NC001359
RhGMV_NC010294
MeMPRV_NC015490
MaMPRV_NC004097
RhMMV_NC015488
Cluster III
CGMV_NC009644
CYVV_NC006358
ToLCNDV_NC004611
0.2

Fig. 3

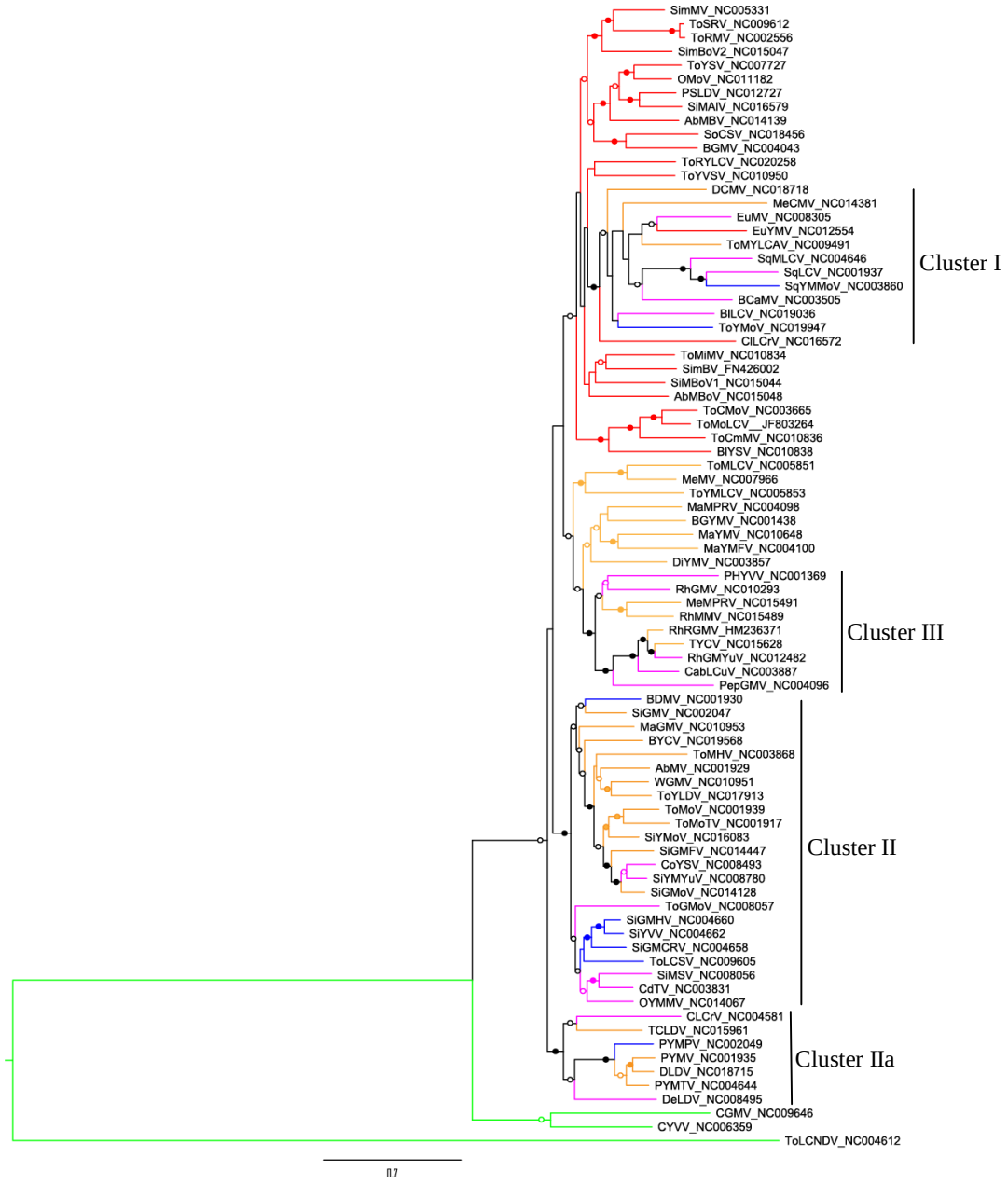
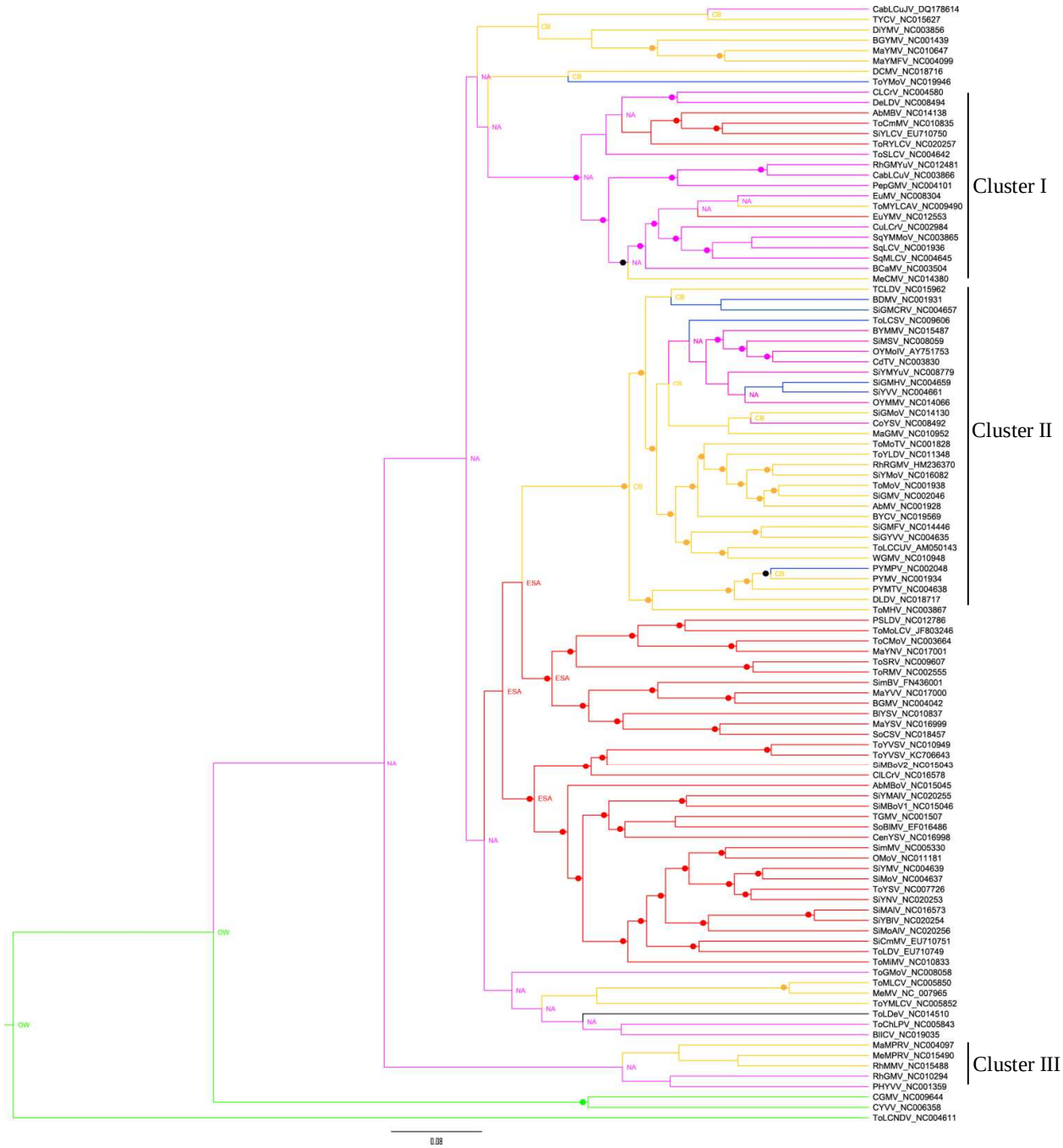

SimMV_NC005331
ToSRV_NC009612
ToRMV_NC002556
SimBoV2_NC015047
ToYSV_NC007727
OMoV_NC011182
PSLDV_NC012727
SiMAlV_NC016579
AbMBV_NC014139
SoCSV_NC018456
BGMV_NC004043
ToRYLCV_NC020258
ToYVSV_NC010950
DCMV_NC018718
MeCMV_NC014381
EuMV_NC008305
EuYMV_NC012554
ToMYLCAV_NC009491
SqMLCV_NC004646
SqLCV_NC001937
SqYMMoV_NC003860
BCaMV_NC003505
BILCV_NC019036
ToYMoV_NC019947
ClLCrV_NC016572
Cluster I
ToMiMV_NC010834
SimBV_FN426002
SiMBoV1_NC015044
AbMBoV_NC015048
ToCMoV_NC003665
ToMoLCV__JF803264
ToCmMV_NC010836
BIYSV_NC010838
ToMLCV_NC005851
MeMV_NC007966
ToYMLCV_NC005853
MaMPRV_NC004098
BGYMV_NC001438
MaYMV_NC010648
MaYMFV_NC004100
DiYMV_NC003857
PHYVV_NC001369
RhGMV_NC010293
MeMPRV_NC015491
RhMMV_NC015489
RhRGMV_HM236371
TYCV_NC015628
RhGMYuV_NC012482
CabLCuV_NC003887
PepGMV_NC004096
Cluster III
BDMV_NC001930
SiGMV_NC002047
MaGMV_NC010953
BYCV_NC019568
ToMHV_NC003868
AbMV_NC001929
WGMV_NC010951
ToYLDV_NC017913
ToMoV_NC001939
ToMoTV_NC001917
SiYMoV_NC016083
SiGMFV_NC014447
CoYSV_NC008493
SiYMYuV_NC008780
SiGMoV_NC014128
ToGMoV_NC008057
SiGMHV_NC004660
SiYVV_NC004662
SiGMCRV_NC004658
ToLCSV_NC009605
SiMSV_NC008056
CdTV_NC003831
OYMMV_NC014067
Cluster II
CLCrV_NC004581
TCLDV_NC015961
PYMPV_NC002049
PYMV_NC001935
DLDV_NC018715
PYMTV_NC004644
DeLDV_NC008495
Cluster IIa
CGMV_NC009646
CYVV_NC006359
ToLCNDV_NC004612
0.7

Fig. 4


CabLCuJV_DQ178614
TYCV_NC015627
DiYMV_NC003856
BGYMV_NC001439
MaYMV_NC010647
MaYMFV_NC004099
DCMV_NC018716
ToYMoV_NC019946
CLCrV_NC004580
DeLDV_NC008494
AbMBV_NC014138
ToCmMV_NC010835
SiYLCV_EU710750
ToRYLCV_NC020257
ToSLCV_NC004642
RhGMYuV_NC012481
CabLCuV_NC003866
PepGMV_NC004101
EuMV_NC008304
ToMYLCAV_NC009490
EuYMV_NC012553
CuLCrV_NC002984
SqYMMoV_NC003865
SqLCV_NC001936
SqMLCV_NC004645
BCaMV_NC003504
MeCMV_NC014380
TCLDV_NC015962
BDMV_NC001931
SiGMCRV_NC004657
ToLCSV_NC009606
BYMMV_NC015487
SiMSV_NC008059
OYMoIV_AY751753
CdTV_NC003830
SiYMYuV_NC008779
SiGMHV_NC004659
SiYVV_NC004661
OYMMV_NC014066
SiGMoV_NC014130
CoYSV_NC008492
MaGMV_NC010952
ToMoTV_NC001828
ToYLDV_NC011348
RhRGMV_HM236370
SiYMoV_NC016082
ToMoV_NC001938
SiGMV_NC002046
AbMV_NC001928
BYCV_NC019569
SiGMFV_NC014446
SiGYVV_NC004635
ToLCCUV_AM050143
WGMV_NC010948
PYMPV_NC002048
PYMV_NC001934
PYMTV_NC004638
DLDV_NC018717
ToMHV_NC003867
PSLDV_NC012786
ToMoLCV_JF803246
ToCMoV_NC003664
MaYNV_NC017001
ToSRV_NC009607
ToRMV_NC002555
SimBV_FN436001
MaYVV_NC017000
BGMV_NC004042
BlYSV_NC010837
MaYSV_NC016999
SoCSV_NC018457
ToYVSV_NC010949
ToYVSV_KC706643
SiMBoV2_NC015043
ClLCrV_NC016578
AbMBoV_NC015045
SiYMAlV_NC020255
SiMBoV1_NC015046
TGMV_NC001507
SoBIMV_EF016486
CenYSV_NC016998
SimMV_NC005330
OMoV_NC011181
SiYMV_NC004639
SiMoV_NC004637
ToYSV_NC007726
SiYNV_NC020253
SiMAlV_NC016573
SiYBlV_NC020254
SiMoAlV_NC020256
SiCmMV_EU710751
ToLDV_EU710749
ToMiMV_NC010833
ToGMoV_NC008058
ToMLCV_NC005850
MeMV_NC_007965
ToYMLCV_NC005852
ToLDeV_NC014510
ToChLPV_NC005843
BllCV_NC019035
MaMPRV_NC004097
MeMPRV_NC015490
RhMMV_NC015488
RhGMV_NC010294
PHYVV_NC001359
CGMV_NC009644
CYVV_NC006358
ToLCNDV_NC004611
Cluster I
Cluster II
Cluster III
CB
NA
ESA
OW
0.08

Suppl. Table S1. New World begomovirus species retrieved from GenBank's non-redundant nucleotide database.

Species	Acession #		Acronym	Location	Host
	DNA-A	DNA-B			
South America East					
Abutilon Brazil virus	NC014138	NC014139	AbMBV	Bahia, Brazil	*Abutilon sp.*
Abutilon mosaic Bolivia virus	NC015045	NC015048	AbMBoV	Santa Cruz, Bolivia	*Abutilon sp*
Bean golden mosaic virus	NC004042	NC004043	BGMV	Brazil	*Phaseolus vulgaris*
Blainvillea yellow spot virus	NC010837	NC010838	BlYSV	Brazil	*Blainvillea rhomboidea*
Centrosema yellow spot virus	NC016998	-	CenYSV	Brazil	*Centrosema brasilianum*
Cleome leaf crumple virus	NC016578	NC016572	ClLCrV	Brazil	*Cleome affinis*
Euphorbia yellow mosaic virus	NC012553	NC012554	EuYMV	Brazil	*Euphorbia heterophylla*
Macroptilium yellow net virus	NC017001	-	MaYNV	Brazil	*Macroptilium lathyroides*
Macroptilium yellow spot virus	NC016999	-	MaYSV	Brazil	*M. lathyroides*
Macroptilium yellow vein virus	NC017000	-	MaYVV	Brazil	*M. lathyroides*
Okra mottle virus	NC011181	NC011182	OMoV	Brazil	*Abelmoschus esculentus*
Passionfruit severe leaf distortion virus	NC012786	NC012787	PSLDV	Brazil	*Passiflora edulis*
Sida micrantha mosaic virus	NC005330	NC005331	SimMV	Brazil	*Sida micrantha*
Sida mosaic Alagoas virus	NC016573	NC016579	SiMAlV	Brazil	*Sida sp.*
Sida mosaic Bolivia virus 1	NC015046	NC015044	SiMBoV1	Bolivia	*S. micrantha*
Sida mosaic Bolivia virus 2	NC015043	NC015047	SiMBoV2	Bolivia	*S. micrantha*
Sida mottle Alagoas virus	NC020256	-	SiMoAlV	Brazil	*S. urens*
Sida mottle virus	NC004637	-	SiMoV	Brazil	*Sida sp.*
Sida yellow blotch virus	NC020254	-	SYBlV	Brazil	*S. urens*
Sida yellow mosaic Alagoas virus	NC020255	-	SiYMAlV	Brazil	*S. urens*
Sida yellow mosaic virus	NC004639	-	SiYMV	Brazil	*Sida sp.*
Sida yellow net virus	NC020253	-	SiYNV	Brazil	*S. micrantha*

Species	Acession # DNA-A	DNA-B	Acronym	Local	Host
South America East					
Soybean chlorotic spot virus	NC018457	NC018456	SoCSV	Brazil	*Glycine max*
Tomato chlorotic mottle virus	NC003664	NC003665	ToCMoV	Brazil	*Solanum lycopersicum*
Tomato common mosaic virus	NC010835	NC010836	ToCmMV	Brazil	*S. lycopersicum*
Tomato golden mosaic virus	NC001507	-	TGMV	Brazil	*S. lycopersicum*
Tomato mild mosaic virus	NC010833	NC010834	ToMiMV	Brazil	*S. lycopersicum*
Tomato rugose mosaic virus	NC002555	NC002556	ToRMV	Brazil	*S. lycopersicum*
Tomato rugose yellow leaf curl virus	NC020257	NC020258	ToRYLCV	Salto Grande, Uruguay	*S. lycopersicum*
Tomato severe rugose virus	NC009607	NC009612	ToSRV	Brazil	*Capsicum annuum*
Tomato yellow spot virus	NC007726	NC007727	ToYSV	Brazil	*S. lycopersicum*
Tomato yellow vein streak virus	NC010949	NC010950	ToYVSV	Brazil	*Solanum tuberosum*
Tomato mottle leaf curl virus	JF803246	JF803264	ToMoLCV	Distrito Federal, Brazil	*S. lycopersicum*
Tomato leaf distortion virus	EU710749	-	ToLDV	Brazil	*S. lycopersicum*
Sida common mosaic virus	EU710751	-	SiCmMV	Brazil	*S. rhombifolia*
Sida Brazil virus	FN436001	FN436002	SimBV	Mato Grosso do Sul, Brazil	*Sida sp.*
Sida yellow leaf curl virus	EU710750	-	SiYLCV	Brazil	*S. rhombifolia*
Soybean blistering mosaic virus	EF016486	-	SoBlMV	Argentina	*G. max*
South America West					
Tomato leaf deformation virus	NC014510	-	ToLDeV	Ecuador	*S. lycopersicum*
Caribbean					
Abutilon mosaic virus	NC001928	NC001929	AbMV	-	*Abutilon sellovianum*
Bean yellow chlorosis virus	NC019569	NC019568	BYCV	Barinas, Venezuela	*P. vulgaris*
Bean golden yellow mosaic virus	NC001439	NC001438	BGYMV	Puerto Rico	*P. vulgaris*

Species	Acession #		Acronym	Local	Host
	DNA-A	DNA-B			
Dalechampia chlorotic mosaic virus	NC018716	NC018718	DCMV	Venezuela	*Dalechampia sp.*
Datura leaf distortion virus	NC018717	NC018715	DLDV	Venezuela	*Datura stramonium*
Dicliptera yellow mottle virus	NC003856	NC003857	DiYMV	Florida, USA	*Dicliptera sexangularis*
Macroptilium golden mosaic virus	NC010952	NC010953	MaGMV	Jamaica	*Wissadula amplissima*
Macroptilium yellow mosaic Florida virus	NC004099	NC004100	MaYMFV	Florida, USA	*M. lathyroides*
Macroptilium yellow mosaic virus	NC010647	NC010648	MaYMV	Jamaica	*M. lathyroides*
Macroptilium mosaic Puerto Rico virus	NC004097	NC004098	MaMPRV	Puerto Rico	*M. lathyroides*
Melon chlorotic mosaic virus	NC014380	NC014381	MeCMV	Venezuela	*Citrullus lanatus*
Merremia mosaic Puerto Rico virus	NC015490	NC015491	MeMPRV	Puerto Rico	*Merremia quinquefolia*
Merremia mosaic virus	NC007965	NC007966	MeMV	Puerto Rico	*Merremia* sp.
Potato yellow mosaic Trinidad virus	NC004638	NC004644	PYMTV	Trinidad and Tobago	*S. lycopersicum*
Potato yellow mosaic virus	NC001934	NC001935	PYMV	Venezuela	*Solanum tuberosum*
Rhynchosai mild mosaic virus	NC015488	NC015489	RhMMV	Puerto Rico	*Rhynchosia minima*
Sida golden mosaic Florida virus	NC014446	NC014447	SiGMFV	Florida, USA	*Sida* sp.
Sida golden mosaic virus	NC002046	NC002047	SiGMV	Florida, USA	*S. santaremensis*
Sida golden mottle virus	NC014130	NC014128	SiGMoV	Florida, USA	*S. santaremensis*
Sida golden yellow vein virus	NC004635	-	SiGYVV	Jamaica	*Sida spp*
Sida yellow mottle virus	NC016082	NC016083	SiYMoV	Cuba	*M. coromandelianum*
Tobacco yellow crinkle virus	NC015627	NC015628	TYCV	Cuba	*Nicotiana tabacum*
Tomato chlorotic leaf distortion	NC015962	NC015961	TCLDV	Venezuela	*S. lycopersicum*
Tomato mild yellow leaf curl Aragua virus	NC009490	NC009491	ToMYLCAV	Venezuela	*S. lycopersicum*
Tomato mosaic Havana virus	NC003867	NC003868	ToMHV	Cuba	*S. lycopersicum*
Tomato mosaic leaf curl virus	NC005850	NC005851	ToMLCV	Trujillo, Venezuela	*S. lycopersicum*
Tomato mottle Taino virus	NC001828	NC001917	ToMoTV	Cuba	*S. lycopersicum*
Tomato mottle virus	NC001938	NC001939	ToMoV	Florida	*S. lycopersicum*

Species	Acession # DNA-A	DNA-B	Acronym	Local	Host
Tomato yellow leaf distortion virus	NC011348	NC017913	ToYLDV	Cuba	*S. lycopersicum*
Tomato yellow margin leaf curl virus	NC005852	NC005853	ToYMLCV	Merida, Venezuela	*S. lycopersicum*
Wissadula golden mosaic virus	NC010948	NC010951	WGMV	Saint Thomas, Jamaica	*W. amplissima*
Tobacco leaf curl Cuba virus	AM050143	-	ToLCCUV	Cuba	*N. tabacum*
Rhynchosia rugose golden mosaic virus	HM236370	HM236371	RhRGMV	Cuba	*R. minima*
Central America					
Bean dwarf mosaic virus	NC001931	NC001930	BDMV	Valle del Cauca, Colombia	*P. vulgaris*
Potato yellow mosaic Panama virus	NC002048	NC002049	PYMPV	Panama	*S. lycopersicum*
Sida golden mosaic Costa Rica virus	NC004657	NC004658	SiGMCRV	Costa Rica	*S. rhombifolia*
Sida golden mosaic Honduras virus	NC004659	NC004660	SiGMHV	Honduras	*S. rhombifolia*
Sida yellow vein virus	NC004661	NC004662	SiYVV	Honduras	*Sida sp.*
Squash yellow mild mottle virus	NC003865	NC003860	SqYMMoV	Costa Rica	*Cucurbita sp.*
Tomato leaf curl Sinaloa virus	NC009606	NC009605	ToLCSV	Santa Lucia, Nicaragua	*S. lycopersicum*
Tomato severe leaf curl virus	NC004642	-	ToSLCV	Guatemala	*S. lycopersicum*
Tomato yellow mottle virus	NC019946	NC019947	ToYMoV	Costa Rica	*S. lycopersicum*
North America					
Bean calico mosaic virus	NC003504	NC003505	BCaMV	Sonora, Mexico	*P. vulgaris*
Bean yellow mosaic Mexico virus	NC015487	-	BYMMV	Chiapas, Mexico	*P. vulgaris*
Blechum interveinal chlorosis virus	NC019035	NC019036	BlICV	Campeche, Mexico	*Blechum pyramidatum*
Cabbage leaf curl virus	NC003866	NC003887	CabLCuV	-	*Brassica oleracea*
Chino del tomate virus	NC003830	NC003831	CdTV	-	*S. lycopersicum*
Corchorus yellow spot virus	NC008492	NC008493	CoYSV	Yucatan, Mexico	*Corchorus siliquosus*

Species	Acession # DNA-A	DNA-B	Acronym	Local	Host
Cotton leaf crumple virus	NC004580	NC004581	CLCrV	Sonora, Mexico	*Gossypium sp.*
Cucurbit leaf crumple virus	NC002984	-	CuLCrV	USA	*Curcubita pepo*
Desmodium leaf distortion virus	NC008494	NC008495	DeLDV	Yucatan, Mexico	*Desmodium glabrum*
Euphorbia mosaic virus	NC008304	NC008305	EuMV	Yucatan, Mexico	*E. heterophylla*
Okra yellow mosaic Mexico virus	NC014066	NC014067	OYMMV	Morelos, Mexico	*A. esculentus*
Pepper golden mosaic virus	NC004101	NC004096	PepGMV	Tamaulipas, Mexico	*C. annuum*
Pepper huasteco yellow vein virus	NC001359	NC001369	PHYVV	Mexico	*C. annuum*
Rhynchosia golden mosaic virus	NC010294	NC010293	RhGMV	Sinaloa, Mexico	*G. max*
Rhynchosia golden mosaic Yucatan virus	NC012481	NC012482	RhGMYuV	Mexico	*R. minima*
Sida mosaic Sinaloa virus	NC008059	NC008056	SiMSV	Mexico	*Sida sp.*
Squash leaf curl virus	NC001936	NC001937	SqLCV	-	*Cucurbit sp.*
Squash mild leaf curl virus	NC004645	NC004646	SqMLCV	California, USA	*Cucurbit sp.*
Sida yellow mosaic Yucatan virus	NC008779	NC008780	SiYMYuV	Yucatan, Mexico	*S. acuta*
Tomato chino La Paz virus	NC005843	-	ToChLPV	La paz, Mexico	*S. lycopersicum*
Tomato golden mottle virus	NC008058	NC008057	ToGMoV	Sinaloa, Mexico	*S. lycopersicum*
Okra yellow mottle Iguala virus	AY751753	-	OYMoIV	Mexico	*A. esculentus*

Suppl. Table S2. Putative recombination events detected among begomoviruses from the New World, based on the full-length DNA-A.

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
1	1923	7	**Rep/IR**	CabLCuJV_DQ178614	CoYSV_NC008492	CabLCuV_NC003866	RGBMCS**3**	8.91E-93
2	2589	1925	**IR/Rep**	TYCV_NC015627	CabLCuV_NC003866	Unknown	RGBMCS**3**	4.34E-60
3	23	1520	**IR/Rep**	ToRMV_NC002555	ToCMoV_NC003664	ToSRV_NC009607	RGBMCS**3**	3.99E-58
4	1634	2649	**Rep/IR**	SiYMV_NC004639	Unknown	SiMoV_NC004637	RGBMCS**3**	6.87E-55
5	1993	2583	**Rep/IR**	AbMBV_NC014138	CabLCuV_NC003866	SiMAlV_NC016573	RGBMCS**3**	7.56E-53
6	2031	29	**Rep/IR**	SiYLCV_EU710750	EuYMV_NC012553	SiMoV_NC004637	RGBMCS**3**	4.29E-47
7	33	1878	**IR/Rep**	DeLDV_NC008494	CoYSV_NC008492	CabLCuV_NC003866	RGBMCS**3**	2.91E-38
8	45	1814	**IR/Rep**	PYMPV_NC002048	PYMV_NC001934	Unknown	RGBMCS**3**	1.00E-35
9	138	1930	**IR/Rep**	DLDV_NC018717	PYMV_NC001934	TCLDV_NC015962	RGBMCS**3**	8.41E-34
10	870	1944	**CP/Rep**	ToSLCV_NC004642	ToChLPV_NC005843	RhGMYuV_NC012481	RGBMCS**3**	1.25E-29
11	18	2064	**IR/Rep**	CLCrV_NC004580	OYMMV_NC014066	Unknown	RBMC**S**3	8.00E-39
12	79	2068	**IR/Rep**	ToLCNDV_NC004611	Unknown	ToYSV_NC007726	RBMC**S**3	2.18E-36
13	1873	45	**Rep/IR**	PYMV_NC001934	MaGMV_NC010952	PYMTV_NC004638	RGBMCS**3**	2.41E-27
14	2003	2488	**Rep**	ToRYLCV_NC020257	RhGMYuV_NC012481	CenYSV_NC016998	**R**GBMCS3	5.86E-26
15	1848	2479	**Rep**	MaYSV_NC016999	BlYSV_NC010837	SoCSV_NC018457	RGBMCS**3**	7.99E-26
16	91	1983	**IR/Rep**	PHYVV_NC001359	Unknown	SiMAlV_NC016573	RBMCS**3**	1.38E-25

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
17	1953	2525	**Rep**	ToCmMV_NC010835	CabLCuV_NC003866	SoBlMV_EF016486	RGBMCS3	4.83E-26
18	1937	2417	**Rep**	RhGMYuV_NC012481 BCaMV_NC003504 SqMLCV_NC004645 CabLCuV_NC003866	Unknown	RhGMV_NC010294	RGBMCS3	8.52E-24
19	1910	2403	**Rep**	EuMV_NC008304 ToSLCV_NC004642 MeCMV_NC014380 CuLCrV_NC002984 SqLCV_NC001936 SqYMMoV_NC003865 EuYMV_NC012553 ToMYLCAV_NC009490 CLCrV_NC004580 DeLDV_NC008494 ToCmMV_NC010835 SiYLCV_EU710750 AbMBV_NC014138	Unknown Unknown Unknown	ToMLCV_NC005850 MeMV_NC_007965 MaYMFV_NC004099	RGBMCS3	4.38E-26
20	182	1847	**IR/Rep**	CoYSV_NC008492 OYMMV_NC014066 SiMSV_NC008059 SiGMHV_NC004659 SiYMYuV_NC008779	SiGMV_NC002046	MaGMV_NC010952	RBMCS3	9.64E-23
21	1468	26	**Rep/IR**	SiCmMV_EU710751	MaGMV_NC010952	SiMoV_NC004637	RBMCS3	1.65E-20
22	1354	1138	**Trap/Ren**	SiMAlV_NC016573	SiYBlV_NC020254	Unknown	RGBMCS3	3.56E-20
23	1990	2416	**Rep**	PepGMV_NC004101	Unknown	RhMMV_NC015488	RGBMCS	3.53E-23

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
24	2667	1246 (?)	**IR/Ren/Trap**	SiMAlV_NC016573	AbMBV_NC014138	SiYBlV_NC020254	RG**3**	1.20E-19
25	1933	2201	**Rep**	MaYNV_NC017001 ToCMoV_NC003664	Unknown Unknown	SimMV_NC005330 ToMiMV_NC010833	R**G**BMCS3	4.90E-19
26	1936	2146	**Rep**	ToSRV_NC009607 ToRMV_NC002555	SimMV_NC005330 ToMiMV_NC010833	Unknown Unknown	RG**B**MCS3	5.60E-20
27	1950	168	**Rep/IR**	MaMPRV_NC004097	Unknown	RhMMV_NC015488	RB**M**C3	4.43E-17
28	131	296	**IR/CP**	OYMoIV_AY751753	SqLCV_NC001936	CdTV_NC003830	**R**GBMCS3	1.99E-16
29	2125	2384	**Rep**	MaYSV_NC016999 BlYSV_NC010837	Unknown Unknown	SoBlMV_EF016486 CabLCuJV_DQ178614	RGBMCS**3**	3.41E-16
30	1934	2145	**Rep**	SiYMoV_NC016082	Unknown	SiGMV_NC002046	**R**GBMCS3	7.03E-16
31	2028	2610	**Rep/IR**	SimBV_FN436001	MaYVV_NC017000	SiMoV_NC004637	RGBMC**S**3	1.33E-17
32	2211	32	**Rep/IR**	SiMoV_NC004637	Unknown	SiYNV_NC020253	**R**GBMCS3	1.96E-14
33	1916	114	**Rep/IR**	SiGMoV_NC014130	MaGMV_NC010952	OYMoIV_AY751753	RBM**C**S3	2.73E-12
34	1057	1965	**Ren/Rep**	RhGMV_NC010294 MeMPRV_NC015490	Unknown	ToYVSV_NC010949	RGBMC**S**	1.24E-14
35	1889	2619	**Rep/IR**	ToMoLCV_JF803246	PSLDV_NC012786	Unknown	RGBMC**S**	6.63E-20
36	864	1743	**CP/Rep**	BYCV_NC019569	Unknown	CdTV_NC003830	RGBMC**S**	3.59E-14
37	2558	1659	**IR/Rep**	BYMMV_NC015487	CdTV_NC003830	Unknown	MCS**3**	2.53E-13

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
38	1346	1925	**Ren/Trap/Rep**	RhMMV_NC015488 MaMPRV_NC004097 MeMPRV_NC015490	CGMV_NC009644	BDMV_NC001931	RBM**C**S3	1.58E-12
39	997	2370	**Ren/Rep**	CGMV_NC009644 CYVV_NC006358	Unknown	ToChLPV_NC005843	RBMC**S**3	1.13E-34
40	2321	2634	**Rep/IR**	OMoV_NC011181	Unknown	SiYNV_NC020253	R**G**BMCS3	1.20E-11
41	195	1649	**IR/Rep**	SimMV_NC005330	OMoV_NC011181	Unknown	RGMS**3**	8.61E-13
42	958	1255	**CP/Ren**	MaGMV_NC010952	Unknown	BDMV_NC001931	RG**B**MCS3	4.28E-12
43	2176	2469	**Rep**	SimBV_FN436001 BGMV_NC004042 MaYVV_NC017000	Unknown Unknown Unknown	CoYSV_NC008492 CabLCuJV_DQ178614 MaGMV_NC010952	RGMC**3**	2.20E-11
44 44	2014	2250(?)	**Rep**	CYVV_NC006358 CGMV_NC009644	Unknown Unknown	SiYMV_NC004639 ToYMoV_NC019946	**M**C3	1.47E-10
45	1869	2079(?)	**Rep**	ToMLCV_NC005850 MeMV_NC_007965	TCLDV_NC015962 PYMTV_NC004638	Unknown Unknown	**R**GBMCS	2.06E-10
46	2503	2644	**Rep/IR**	ToYSV_NC007726	Unknown	SiYNV_NC020253	RGB**S**3	2.60E-10
47	1873(?)	2342	**Rep**	ToYMoV_NC019946	Unknown	MeMV_NC_007965	**R**GBCS	9.62E-10
48	1897	2135	**Rep**	ToYVSV_NC010949 DCMV_NC018716 PSLDV_NC012786	Unknown	AbMBoV_NC015045	**R**GBMCS	1.26E-09
49	1656	2155	**Rep**	OMoV_NC011181	ToYSV_NC007726	SiYBlV_NC020254	RGBMC**S**	1.74E-15

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
50	23	919	**IR/CP**	ToYLDV_NC011348	Unknown	SiYMoV_NC016082	RMCS**3**	2.27E-09
51	2020	2153(?)	**Rep**	SiYMV_NC004639	ClLCrV_NC016578	OMoV_NC011181	RGMC3	4.85E-09
52	1925(?)	2159	**Rep**	CoYSV_NC008492 MaGMV_NC010952 PYMV_NC001934 SiGMoV_NC014130	Unknown	ToMoTV_NC001828	**R**GBMCS3	5.10E-09
53	1057(?)	1167	**Ren**	RhGMV_NC010294 PHYVV_NC001359 MaMPRV_NC004097 MeMPRV_NC015490 RhMMV_NC015488	TYCV_NC015627 RhGMYuV_NC012481 CabLCuV_NC003866 CabLCuJV_DQ178614	Unknown Unknown Unknown Unknown	**R**GMC3	5.87E-09
54	2047	2293	**Rep**	ToMiMV_NC010833	SiMBoV1_NC015046	SiYVV_NC004661	**R**GBMS	1.00E-08
55	2533	2674	**Rep/IR**	ToRYLCV_NC020257	SqMLCV_NC004645	CenYSV_NC016998	R**G**BMCS	1.80E-08
56	366	986	**CP**	SiMBoV1_NC015046	Unknown	SiYMAlV_NC020255	R**B**MCS3	3.25E-09
57	2487	2631	**Rep/IR**	SiMBoV1_NC015046 SiYMAlV_NC020255	ToMoTV_NC001828	Unknown	**R**BMC	2.88E-08
58	352	928	**CP**	SiYMAlV_NC020255	SiYBlV_NC020254	CenYSV_NC016998	**R**BMCS	1.47E-08
59	1874	2286	**Rep**	SiGMV_NC002046 ToYLDV_NC011348 AbMV_NC001928	CdTV_NC003830	ToMHV_NC003867	RGB**M**CS	1.39E-08
60	2028	2267	**Rep**	AbMBoV_NC015045	CenYSV_NC016998	SiMBoV2_NC015043	**R**GBMCS	3.74E-08

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
61	120(?)	387	**IR/CP**	ToSLCV_NC004642	Unknown	SiGMoV_NC014130	**M**C3	9.81E-08
62	993(?)	2534	**CP/Rep**	SiGMHV_NC004659 OYMMV_NC014066	Unknown	OYMoIV_AY751753	B**S**3	6.68E-10
63	2151	2518(?)	**Rep**	CabLCuJV_DQ178614 MaGMV_NC010952 SiCmMV_EU710751	ToMoTV_NC001828	ToChLPV_NC005843 CenYSV_NC016998	RMS**3**	2.92E-07
64	1983	2087	**Rep**	ToLDV_EU710749	SiCmMV_EU710751	SiMBoV2_NC015043	R**G**MC3	9.49E-09
65	2444	2571	**Rep**	SiYMYuV_NC008779	Unknown	SiGMoV_NC014130	RMC**3**	3.13E-07
66	292(?)	938	**CP**	MaYVV_NC017000	Unknown	SqMLCV_NC004645	M**C**3	3.90E-07
67	1470(?)	1965	**Trap/Rep**	SiCmMV_EU710751	BGMV_NC004042	SiMBoV2_NC015043	RGM**3**	6.16E-07
68	1884	2190(?)	**Rep**	SiYBlV_NC020254 SiMAlV_NC016573	SiGMFV_NC014446 SiGYVV_NC004635	SiMoV_NC004637 SiYNV_NC020253	R**B**MCS3	5.14E-07
69	2156(?)	2428	**Rep**	SiYMV_NC004639	Unknown	SiYNV_NC020253	RM**C**S3	9.34E-08
70	267	978	**CP**	ToLCSV_NC009606	SiYNV_NC020253	BYMMV_NC015487	RM**C**3	9.80E-07
71	2486(?)	318	**Rep/CP**	SqYMMoV_NC003865 SqLCV_NC001936 EuMV_NC008304 ToMYLCAV_NC009490	CabLCuV_NC003866	DiYMV_NC003856	RB**C**S	1.60E-06
72	1458	1955(?)	**Rep/Trap**	PHYVV_NC001359	Unknown	ToChLPV_NC005843	RM**C**	1.89E-06
73	1949	2440	**Rep**	ToYMLCV_NC005852	PYMTV_NC004638	BlICV_NC019035	RB**M**CS	2.38E-06

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
74	955	1251	**CP/Ren/Trap**	RhGMYuV_NC012481 TYCV_NC015627	Unknown	ToMYLCAV_NC009490	M**C**S	2.44E-06
75	2362	829	**Rep/CP**	ToChLPV_NC005843	Unknown	ToLDeV_NC014510	M**C**S	2.29E-06
76	1946	2082	**Rep**	SiGYVV_NC004635	CdTV_NC003830	SiGMFV_NC014446	**R**BMS	2.53E-06
77	955	2257(?)	**CP/Rep**	DCMV_NC018716	Unknown	ToYVSV_NC010949	MC**S**	4.91E-10
78	1378	1500	**Trap/Rep**	SiYLCV_EU710750 AbMBoV_NC015045 SiMBoV2_NC015043 SiMBoV1_NC015046 SiYMAlV_NC020255 SiCmMV_EU710751 ToYSV_NC007726 SiYNV_NC020253 SiMoV_NC004637 SiYMV_NC004639 CenYSV_NC016998 TGMV_NC001507	RhGMV_NC010294	ToLDeV_NC014510	R**M**C3	5.28E-06
79	2003(?)	2033(?)	**Rep**	AbMBV_NC014138	ToMoV_NC001938	Unknown	**R**G3	2.02E-06
80	1962	2028	**Rep**	CLCrV_NC004580	Unknown	SiGMCRV_NC004657	R**G**M	4.11E-06
81	342	1219	**CP/Ren/Trap**	CabLCuV_NC003866 PepGMV_NC004101 TYCV_NC015627 CabLCuJV_DQ178614	DiYMV_NC003856	SqMLCV_NC004645	M**C**S	6.20E-06
82	1993	2107	**Rep**	BYMMV_NC015487	ToGMoV_NC008058	Unknown	**R**GMC	6.50E-06

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
83	2438	9	**Rep/IR**	CoYSV_NC008492 SiMSV_NC008059 SiGMoV_NC014130	BlICV_NC019035	ToLCSV_NC009606	BM**C**3	1.09E-05
84	2367(?)	2567(?)	**Rep**	SiYMoV_NC016082 SiGMFV_NC014446 SiGYVV_NC004635 ToLCSV_NC009606 ToYLDV_NC011348 RhRGMV_HM236370 OYMoIV_AY751753	SiMAlV_NC016573 SiYBlV_NC020254	SiGMV_NC002046 ToMoV_NC001938	R**G**MC3	6.92E-04
85	2038	2186	**Rep**	ToYSV_NC007726 SiMBoV1_NC015046 OMoV_NC011181 SiMoAlV_NC020256 ToLDeV_NC014510 TGMV_NC001507 SoBlMV_EF016486	Unknown	BGYMV_NC001439	**R**BMCS3	8.85E-06
86	1393	1556	**Trap/Rep**	SiYBlV_NC020254 ToCmMV_NC010835 SiMAlV_NC016573	ToMLCV_NC005850 MeMV_NC_007965	BDMV_NC001931 SiGMFV_NC014446	**R**BC	3.07E-05
87	753(?)	1061	**CP/Ren**	SiGYVV_NC004635	Unknown	SiYMYuV_NC008779	**M**C3	2.21E-05
88	439	1111	**CP/Ren**	ToLDV_EU710749	OMoV_NC011181	SiCmMV_EU710751	MC**S**3	9.03E-06
89	2031(?)	2317	**Rep**	SimMV_NC005330	CenYSV_NC016998	OYMMV_NC014066	M**C**3	3.66E-05
90	2514	2610(?)	**Rep/IR**	WGMV_NC010948	ToMoLCV_JF803246	TCLDV_NC015962	**R**MC	3.95E-05
91	2031	2499	**Rep**	RhMMV_NC015488 MeMPRV_NC015490	BYCV_NC019569 CdTV_NC003830	Unknown	**M**CS	7.05E-05

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
92	2500	2653(?)	**Rep/IR**	SiYBlV_NC020254 SiMAlV_NC016573	PSLDV_NC012786	SiMoAlV_NC020256	**R**BCS	7.33E-05
93	2010	2106(?)	**Rep**	BlYSV_NC010837	SiGMFV_NC014446	Unknown	RGM**C**	2.89E-05
94	2432	25	**Rep/IR**	BDMV_NC001931 TCLDV_NC015962 ToMHV_NC003867 SiGMCRV_NC004657 ToMoV_NC001938 SiYVV_NC004661	Unknown	SiMoAlV_NC020256	**M**CS	3.60E-05
95	1981	2075	**Rep**	CdTV_NC003830 BYCV_NC019569 ToMoTV_NC001828 ToMoV_NC001938 RhRGMV_HM236370 SiMSV_NC008059 SiYVV_NC004661	Unknown	TCLDV_NC015962	**R**GBCS	7.77E-05
96	2427(?)	120	**Rep/IR**	BlYSV_NC010837	MaYMV_NC010647	MaYNV_NC017001	RBMC**S**	7.33E-05
97	2373	2510	**Rep**	MeMV_NC_007965 ToMLCV_NC005850	SiMoAlV_NC020256	Unknown	R**G**S	2.69E-04
98	2259(?)	2310	**Rep**	ToMLCV_NC005850 MeMV_NC_007965	SiGMV_NC002046 TCLDV_NC015962	ToMoLCV_JF803246	RMC**3**	5.81E-05
99	2424	2577(?)	**Rep**	TYCV_NC015627 MaYMFV_NC004099 DLDV_NC018717 CdTV_NC003830 SoCSV_NC018457	MaMPRV_NC004097	SiMoAlV_NC020256	M**C**3	3.20E-04

Events	Breakpoints		Genomic region	Recombinants	Putative parents		Methods	*P* value
	Begin	End			Minor	Major		
100	2474(?)	2635	**Rep/IR**	BYCV_NC019569	Unknown	SimMV_NC005330	RMC**S**3	7.80E-06
101	1251	1409	**Ren/Trap**	MaYMFV_NC004099 BGYMV_NC001439 MaYMV_NC010647	EuMV_NC008304	TCLDV_NC015962	**R**MC	2.38E-04
102	2060	2213	**Rep**	ToMoLCV_JF803246 PYMPV_NC002048 ToCMoV_NC003664 PSLDV_NC012786	Unknown	MaYMV_NC010647	RM**C**3	8.73E-06
103	1033(?)	1388	**Ren/Trap**	ToYVSV_NC010949 SoBlMV_EF016486	Unknown	SoCSV_NC018457	**M**C3	1.08E-04
104	754	1019	**CP/Ren**	ToMYLCAV_NC009490	Unknown	AbMBoV_NC015045	R**B**C	2.42E-03
105	27	131	**IR**	PYMTV_NC004638 PYMPV_NC002048 PYMV_NC001934	Unknown	SiGMV_NC002046	GM**S**	2.09E-03
106	1960(?)	2047	**Rep**	MaYMV_NC010647	MaYNV_NC017001	MaYMFV_NC004099	**M**C3	1.45E-04
107	166(?)	1062	**IR/Ren**	ToMoLCV_JF803246	ToCMoV_NC003664	Unknown	**M**CS	2.23E-03

[a] Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?), breakpoint could not be precisely pinpointed.
[b] R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimera; S, SisScan; 3, 3SEQ.
[c] The reported *P* values are for the methods indicated in red, and they are the lowest *P* values calculated for the region in question.

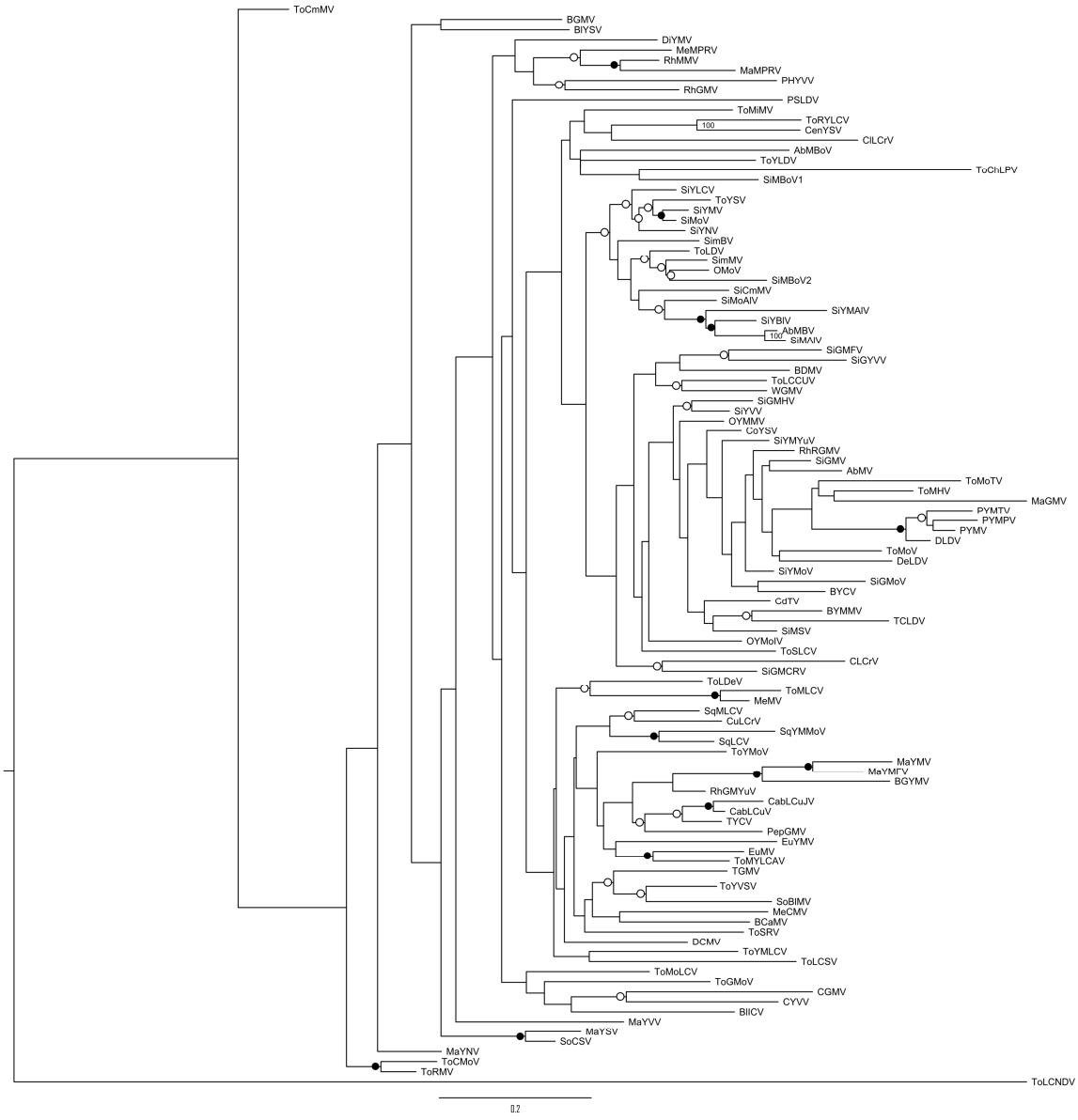
Suppl. Table S3. Putative recombination events detected among begomoviruses from the New World, based on the full-length DNA-B.

Events	Breakpoints		Genomic Region	Recombinants	Putative Parents		Methods	*P* value
	Begin	End			Minor	Major		
1	667	2361	**NSP/IR**	OMoV_NC011182	ToYSV_NC007727	SimMV_NC005331	RBMCS**3**	4.64E-52
2	592	1355	**NSP/MP**	OYMMV_NC014067	CLCrV_NC004581	BYCV_NC019568	**R**GBMCS3	1.75E-28
3	2465	2524	**IR**	PYMV_NC001935 DLDV_NC018715	CoYSV_NC008493	PYMTV_NC004644	R**G**BMCS	8.15E-17
4	2286	14	**IR**	CabLCuV_NC003887	Unknown	RhRGMV_HM236371	RG**B**MCS	3.70E-16
5	2482	64	**IR**	EuMV_NC008305	EuYMV_NC012554	BlICV_NC019036	RG**B**MCS	6.80E-12
6	1190	55(?)	**IR**	CabLCuV_NC003887	Unknown	RhRGMV_HM236371	RGB**M**CS	1.27E-09
7	84	1403	**IR**	ToMoV_NC001939	Unknown	ToMoTV_NC001917	M**C**S	8.21E-09
8	2509	2578(?)	**IR**	SiGMFV_NC014447 SiGMV_NC002047 TYCV_NC015628	ToMoTV_NC001917	Unknown	GM**S**	4.78E-09
9	2413	2586	**IR**	ToYLDV_NC017913 WGMV_NC010951 ToMoTV_NC001917 SiYMoV_NC016083	RhRGMV_HM236371	CLCrV_NC004581	RGBMC**S**	1.06E-09
10	2611	101	**IR**	SiMAlV_NC016579	SiMBoV1_NC015044	Unknown	R**G**BS	5.76E-08

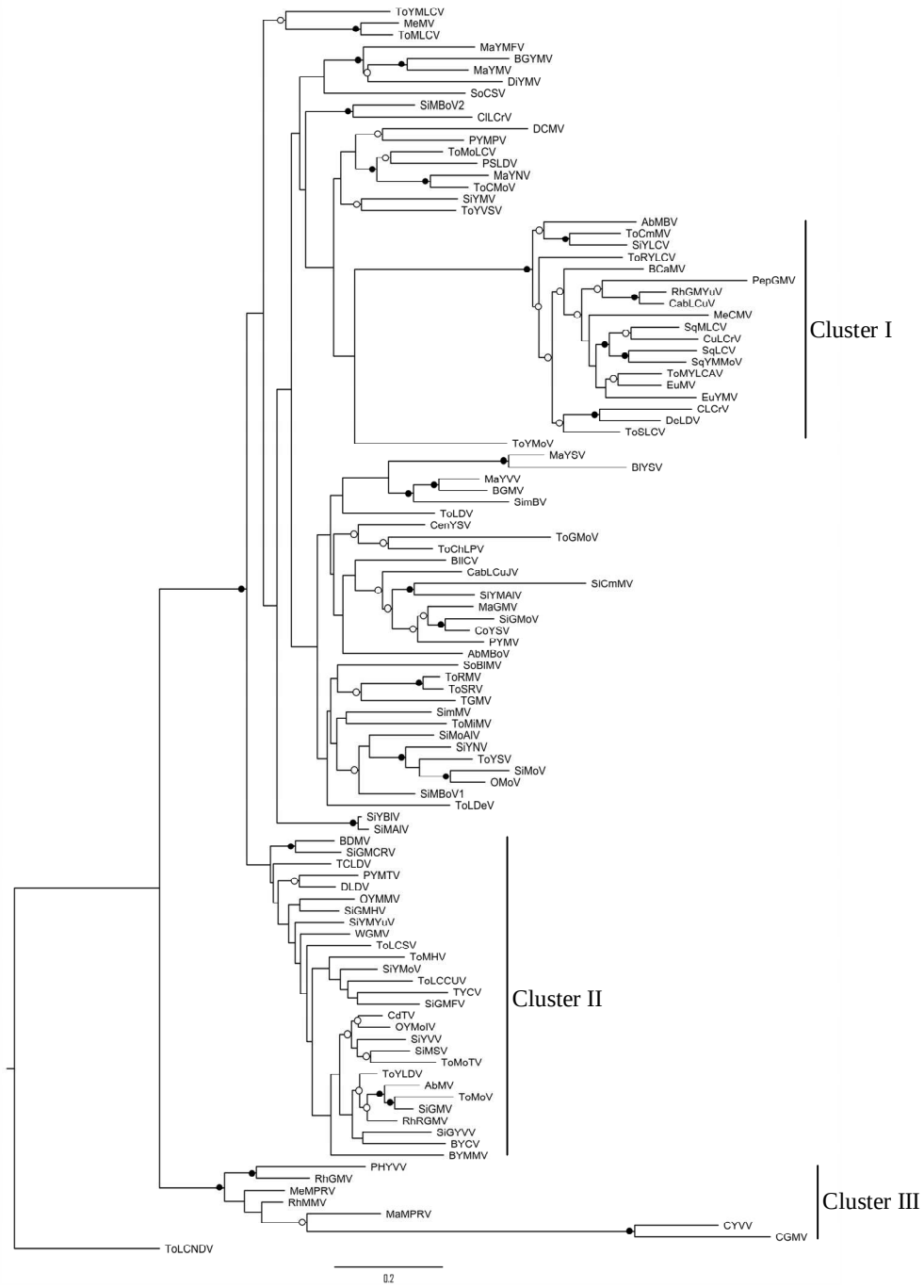
Events	Breakpoints		Genomic Region	Recombinants	Putative Parents		Methods	*P* value
	Begin	End			Minor	Major		
11	99	188	**IR**	PYMV_NC001935	Unknown	DLDV_NC018715	R**G**BM	1.36E-09
12	2448(?)	2515(?)	**IR**	SiMSV_NC008056	CoYSV_NC008493	OYMMV_NC014067	**G**BS	1.55E-07
13	401	743	**NSP**	SiGMV_NC002047	SiGMoV_NC_014128	Unknown	RGBC**S**	1.70E-09
14	90	146	**IR**	AbMBoV_NC015048	ToYVSV_NC010950	CdTV_NC003831	RGBMC**S**	4.53E-09
15	2430	2541	**IR**	SiGMoV_NC_014128	Unknown	SiYMYuV_NC008780	R**G**BMC	2.74E-06
16	2440	2540	**IR**	EuYMV_NC012554 ToMYLCAV_NC009491	SqMLCV_NC004646	ToYVSV_NC010950	RGBM**S**	2.15E-07
17	15	456	**IR/NSP**	DCMV_NC018718	CYVV_NC006359	BlICV_NC019036	MC**S**	7.47E-09
18	1392	1605	**MP**	DLDV_NC018715 PYMPV_NC002049 PYMV_NC001935	SiGMHV_NC004660 ToGMoV_NC008057	Unknown Unknown	**R**GBS	5.33E-05
19	1035	1235	**IR**	SiYMoV_NC016083	SimBV_FN436002	SiGMHV_NC004660 SiYVV_NC004662	**R**BMC	8.40E-05
20	2428	72	**IR**	SqMLCV_NC004646	BCaMV_NC003505	ToYVSV_NC010950	RMC**S**	3.17E-07
21	2408	31	**IR**	BCaMV_NC003505	SqYMMoV_NC003860	Unknown	RGMC**S**	4.00E-05
22	2200	60	**IR**	MeMPRV_NC015491	Unknown	MaYMFV_NC004100	RC**S**	3.02E-07

Events	Breakpoints		Genomic Region	Recombinants	Putative Parents		Methods	*P* value
	Begin	End			Minor	Major		
23	127	261(?)	**IR**	ToRYLCV_NC020258	SiMBoV2_NC015047	ToGMoV_NC008057	M**C**S	2.23E-04
24	1225	2119	**MP/IR**	RhMMV_NC015489	CabLCuV_NC003887	SiMBoV2_NC015047	MC**S**3	4.49E-15
25	2531	2597	**IR**	ToMoLCV_JF803264	SoCSV_NC018456	Unknown	**G**BS	2.27E-03
26	1722(?)	1834	**MP**	ToMHV_NC003868	PYMPV_NC002049	SiYMYuV_NC008780	R**M**C	1.61E-02

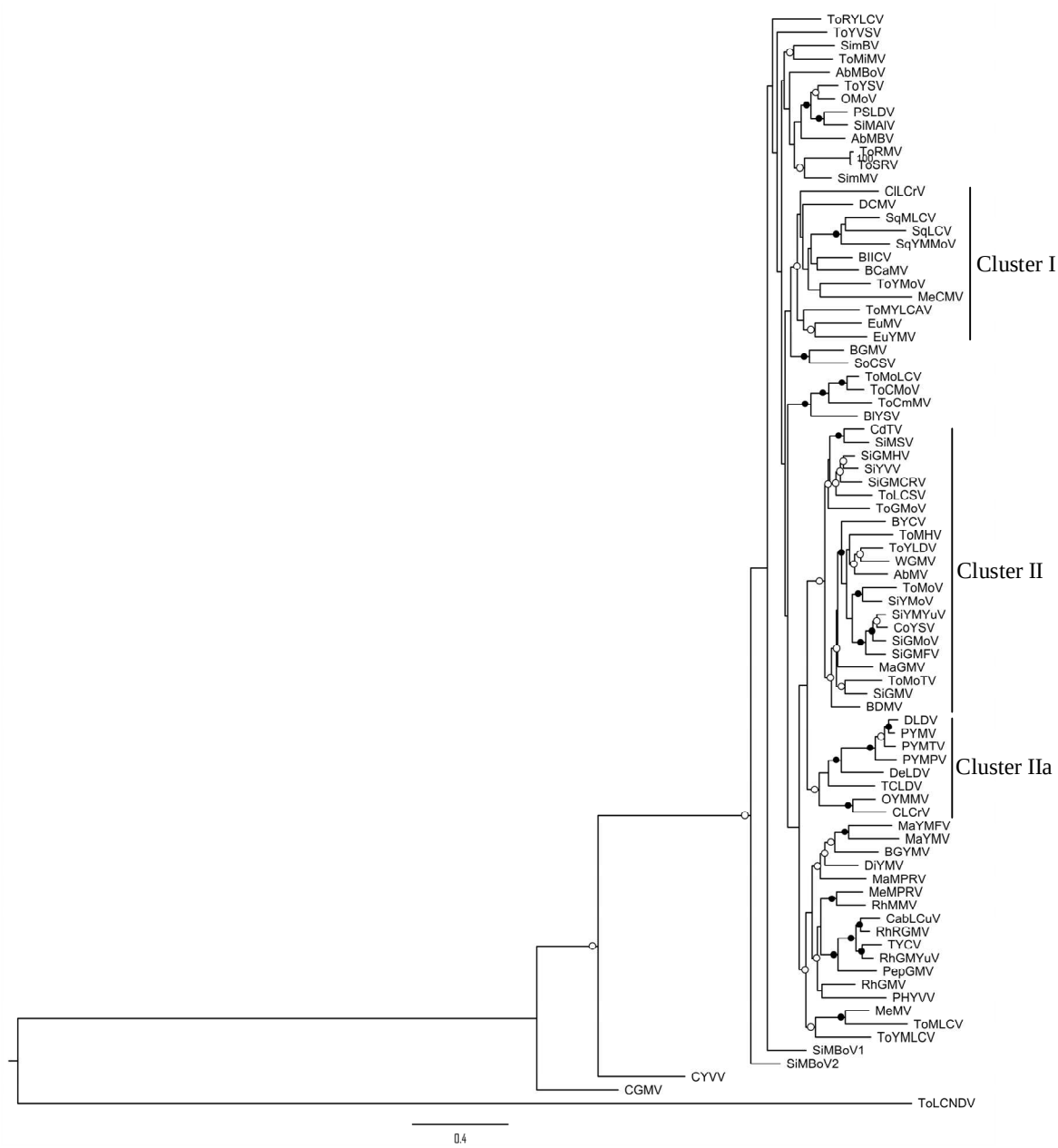
[a] Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?), breakpoint could not be precisely pinpointed.
[b] R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimera; S, SisScan; 3, 3SEQ.
[c] The reported *P* values are for the methods indicated in red, and they are the lowest *P* values calculated for the region in question.



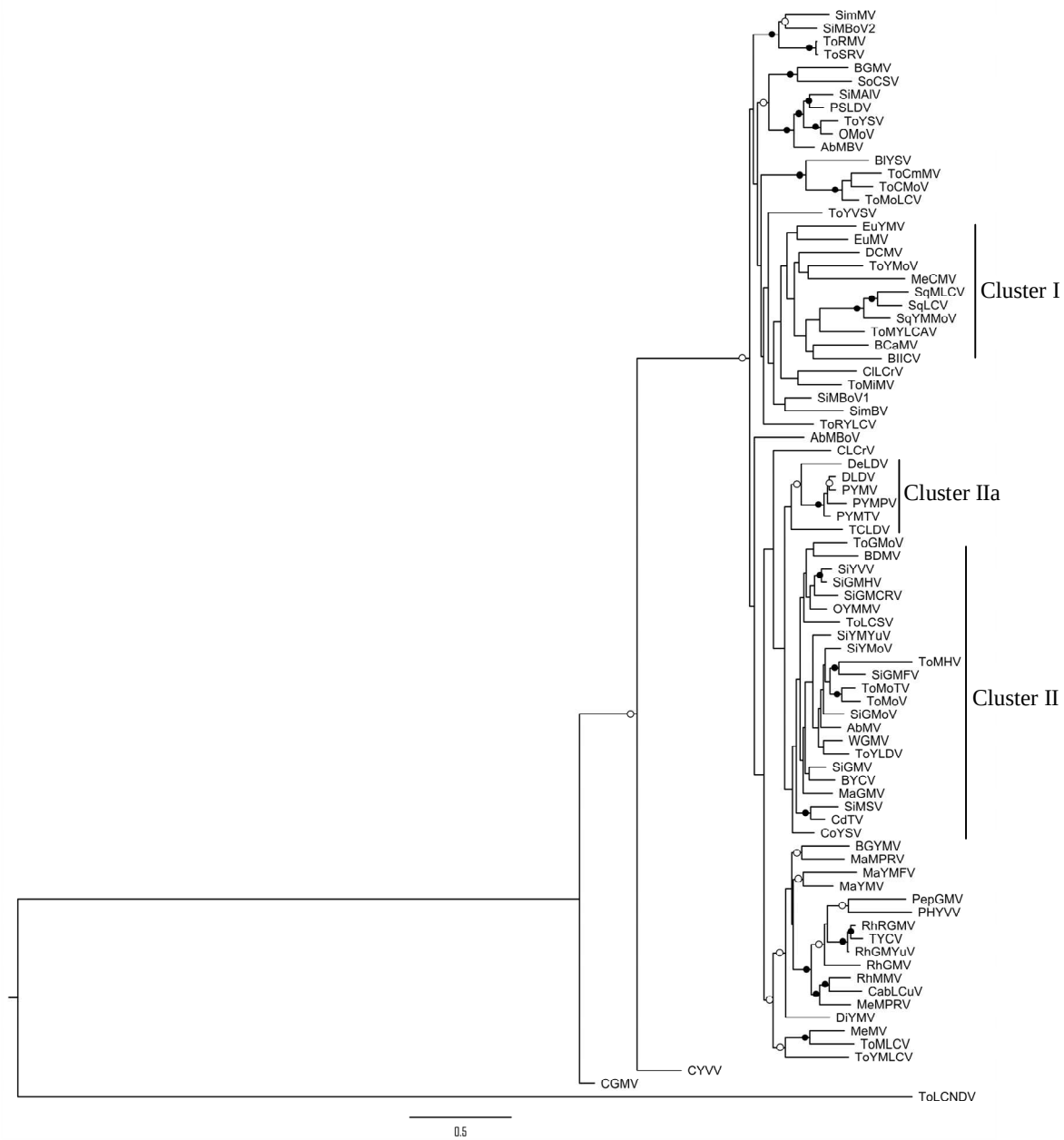
Suppl. Fig. S1. Midpoint-rooted maximum-likelihood tree based on CP nucleotide sequences of New World begomoviruses. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal to or higher than 85% by filled circles.



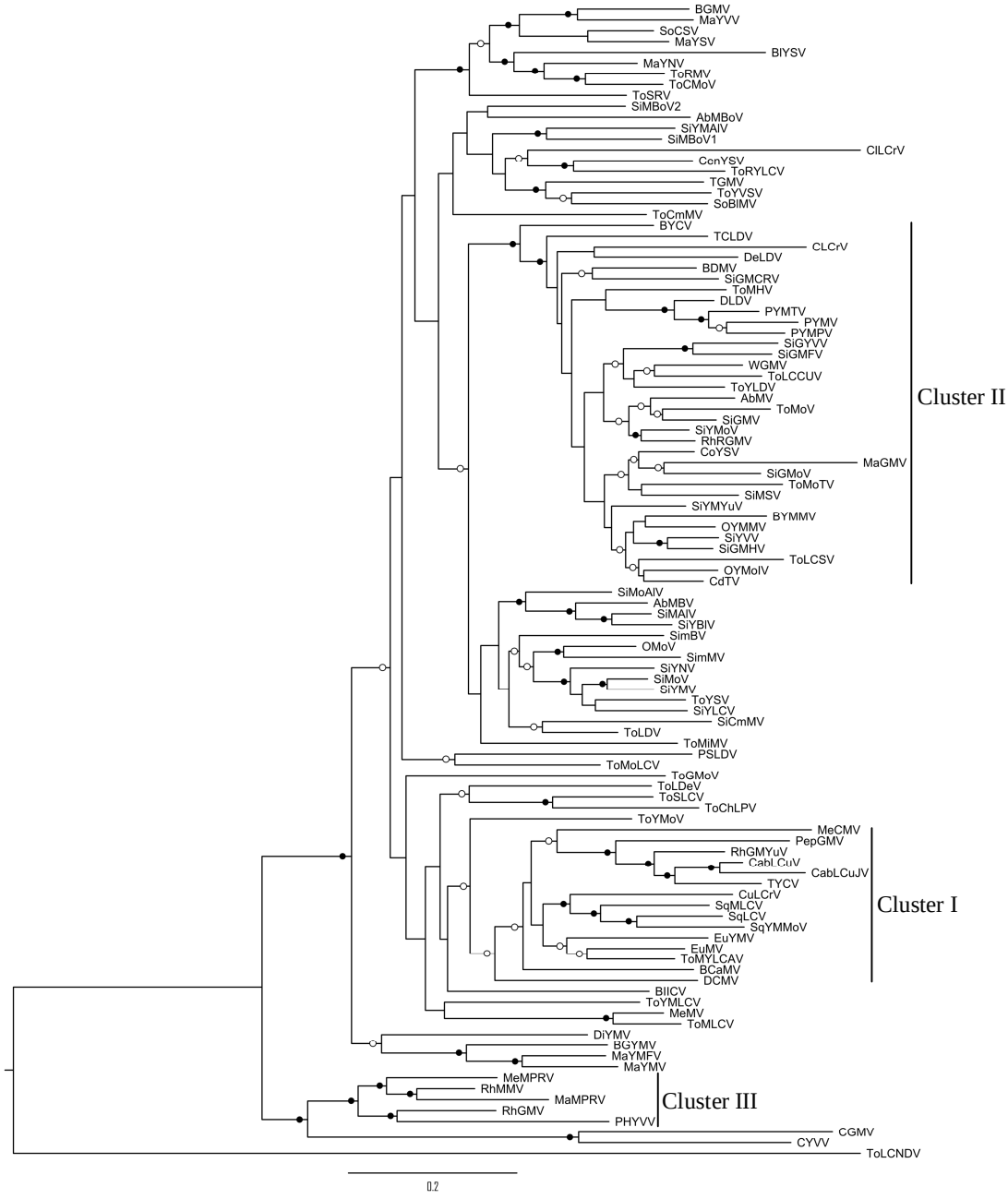
Suppl. Fig. S2. Midpoint-rooted maximum-likelihood tree based on Rep nucleotide sequences of New World begomoviruses. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal to or higher than 85% by filled circles.



Suppl. Fig. S3. Midpoint-rooted maximum-likelihood tree based on NSP nucleotide sequences of New World begomoviruses. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal to or higher than 85% by filled circles.



Suppl. Fig. S4. Midpoint-rooted maximum-likelihood tree based on MP nucleotide sequences of New World begomoviruses. Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal to or higher than 85% by filled circles.



Suppl. Fig. S5. Midpoint-rooted maximum-likelihood tree based on DNA-A nucleotide sequences without the recombinant region detected in event 19 (shared by all isolates in cluster I). Nodes with bootstrap values higher than 50% and lower than 85% are indicated by empty circles and those with values equal to or higher than 85% by filled circles.