ALISON TALIS MARTINS LIMA

EVOLUTIONARY DYNAMICS OF BEGOMOVIRUS POPULATIONS: VARIABILITY IN CULTIVATED AND NON-CULTIVATED HOSTS AND THEIR QUASISPECIES NATURE

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
2012

Ficha catalográfica preparada pela Seção de Catalogação e Classificação da Biblioteca Central da UFV

T
L732e
2012

Lima, Alison Talis Martins, 1982-
Evolutionary dynamics of begomovirus populations : variability in cultivated and non-cultivated hosts and their quasispecies nature / Alison Talis Martins Lima. – Viçosa, MG, 2012.
ix, 182f. : il. ; (algumas col.) ; 29cm.

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

1. Begomovírus. 2. Geminivírus. 3. Mutação (Biologia). 4. Recombinação (Genética). 5. Evolução (Biologia). I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22. ed. 571.9928

ALISON TALIS MARTINS LIMA

EVOLUTIONARY DYNAMICS OF BEGOMOVIRUS POPULATIONS: VARIABILITY IN CULTIVATED AND NON-CULTIVATED HOSTS AND THEIR QUASISPECIES NATURE

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: 28 de Agosto de 2012

Prof. Siobain Marie Deirdre Duffy	Prof. Karla Suemy Clemente Yotoko
Prof. Marisa Vieira de Queiroz	Prof. Eduardo Seiti Gomide Mizubuti

Prof. Francisco Murilo Zerbini Junior

(Orientador)

AGRADECIMENTOS

À Deus por direcionar meus caminhos permitindo a plena realização de meus sonhos;

Aos meus pais, por acreditarem nos meus sonhos e lutarem ao meu lado pela realização de cada um deles e por me ensinarem com muita sabedoria, valores que guardarei por toda a vida;

Ao meu irmão, Alexandre, pela amizade e confiança;

À minha noiva Gilcimara pela cumplicidade, carinho e atenção;

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;

Ao Professor Francisco Murilo Zerbini, pela oportunidade de crescimento profissional, pela valiosa orientação, dedicação e amizade durante a realização destes trabalhos;

À Professora Siobain Duffy, pela amizade, dedicação e por me acolher em seu grupo de trabalho,

Aos professores do DFP pela paciência, atenção e amizade;

Aos meus amigos do Laboratório de Virologia Vegetal Molecular: André, Álvaro, Amanda, Antônio Wilson, César, Danizinha, Diogo, Fernanda, Jorge, Joyce, Larissa, Marcelo, Márcio, Marcos, Renan, Roberto, Sarah, Sheila, Sílvia e Tathiana. Em especial: Carol, Dani, Fabio, Gloria, pela amizade e pelo agradável ambiente de trabalho;

Aos funcionários Joaquim, Fizinho, Délio e Sara pelo apoio para a realização deste trabalho e à amizade durante este tempo;

Aos meus avós, que apesar de ausentes, desejaram e sonharam ao meu lado;

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

BIOGRAFIA

ALISON TALIS MARTINS LIMA, filho de Joaquim Pereira Lima Filho e Terezinha Emiliana Martins Lima, nasceu em Ponte Nova, Minas Gerais, no dia 28 de dezembro de 1982.

Em 2001 iniciou o curso de Agronomia na Universidade Federal de Viçosa (UFV), Viçosa, Minas Gerais, graduando-se em março de 2007. No mesmo mês, ingressou no curso de pós-graduação em Fitopatologia em nível de mestrado na mesma universidade, submetendo-se à defesa de dissertação em julho de 2008.

Em agosto de 2008, iniciou o curso de Doutorado em Fitopatologia, submetendo-se a defesa de tese em 28 de agosto de 2012.

CONTENTS

Resumo vi
Abstract viii
Introduction 1
CHAPTER 1. M*alvaviscus yellow mosaic virus*, a recombinant weed-infecting begomovirus carrying a nanovirus-like nonanucleotide and a modified stem-loop structure. 16
  Abstract 18
  Introduction 19
  Material and Methods 21
  Results 24
  Discussion 29
  Acknowledgements 32
  References 32
CHAPTER 2. Synonymous site variation due to recombination explains higher begomovirus variability in non-cultivated hosts 46
  Summary 49
  Introduction 50
  Results 52
  Discussion 58
  Methods 64
  Acknowledgements 67
  References 68
CHAPTER 3. The relative contribution of recombination and mutation to the quasispecies nature of begomovirus populations 93
  Abstract 95
  Introduction 96
  Material and Methods 98
  Results 101
  Discussion 109
  References 116

RESUMO

LIMA, Alison Talis Martins, D.Sc., Universidade Federal de Viçosa, agosto de 2012. **Dinâmica evolutiva de populações de begomovírus infectando tomateiro e plantas daninhas.** Orientador: Francisco Murilo Zerbini Júnior. Coorientadora: Elizabeth Pacheco Batista Fontes.

A família *Geminiviridae* inclui vírus cujos genomas são compostos de uma ou duas moléculas de DNA fita simples circular, encapsidadas por uma única proteína estrutural em partículas icosaédricas geminadas. A família é composta pelos gêneros *Mastrevirus*, *Curtovirus*, *Topocuvirus* e *Begomovirus*, definidos com base no tipo de inseto vetor, gama de hospedeiros e organização genômica. Begomovírus (geminivírus transmitidos pela mosca-branca) são responsáveis por sérias ameaças à agricultura. No Brasil, um grande número de novas espécies de begomovírus associadas a hospedeiros cultivados e não cultivados tem sido caracterizado. Este trabalho teve como objetivos: (*i*) caracterizar molecularmente uma nova espécie de begomovírus infectando naturalmente plantas de *Malvaviscus arboreus*; (*ii*) estudar a dinâmica evolutiva de duas populações de begomovírus (*Tomato severe rugose virus* e *Macroptilium yellow spot virus*) infectando hospedeiros cultivados (*Solanum lycopersicum* e *Phaseolus vulgaris*) e hospedeiros não cultivados (*Sida* spp. e plantas daninhas leguminosas), utilizando um novo método de particionamento de variabilidade baseado em filogenia; e (*iii*) estudar a contribuição relativa da recombinação na dinâmica evolutiva de begomovírus de importância mundial. Para o primeiro objetivo, plantas de *Malvaviscus arboreus* exibindo mosaico amarelo, coletadas nos estados de São Paulo e Rio de Janeiro, foram submetidas à extração de DNA e o genoma viral foi amplificado por meio do mecanismo de amplificação por círculo rolante. A análise das sequências indicou que o vírus corresponde a uma nova espécie de begomovírus para o qual o nome *Malvaviscus* yellow mosaic virus (MalYMV) é proposto. Interessantemente, MalYMV tem características únicas dentro da família *Geminiviridae*: um nonanucleotídeo similar ao de nanovírus e alfassatélites (5'-TAGTATTAC-3') e uma sequência curta localizada imediatamente antes do nonanucleotídeo, capaz de formar uma pequena estrutura em forma de grampo embebida no grampo contendo o nonanucleotídeo. Para o segundo objetivo, áreas de tamanhos similares que conhecidamente abrigavam os begomovírus *Macroptilium yellow spot virus* (MaYSV) e *Tomato severe rugose virus* (ToSRV) foram amostradas. A população de MaYSV foi notavelmente mais variável do que a população do ToSRV, principalmente devido a um grande número de eventos de recombinação na região 5' do gene Rep. Por meio do

mapeamento dos eventos de recombinação em árvores de máxima verossimilhança baseadas nas sequências dos genes CP e Rep, foi possível distinguir as contribuições individuais associadas aos processos evolutivos de mutação e recombinação. Usando este novo método de particionamento, atribuiu-se a recombinação como fonte de 0 a 42% dos níveis de variabilidade dos genes Rep e CP do ToSRV, respectivamente, e 36 e 16% da variabilidade da Rep e CP do MaYSV, respectivamente. Para o terceiro objetivo, o uso desse novo método de particionamento de variabilidade baseado em filogenia foi expandido para conjuntos de dados de sequências de begomovírus coletados ao redor do mundo (disponíveis em bancos de dados públicos). Observou-se que a diversificação de populações de begomovírus é predominantemente dirigida pela dinâmica mutacional, mas a diferentes níveis. Os resultados indicam que a evolução de algumas populações pode ser significativamente dependente da natureza recombinante de seus genomas em adição à rápida dinâmica mutacional.

ABSTRACT

LIMA, Alison Talis Martins, D.Sc., Universidade Federal de Viçosa, August, 2012. **Evolutionary dynamics of begomovirus populations in cultivated and non-cultivaded hosts.** Adviser: Francisco Murilo Zerbini Júnior. Co-adviser: Elizabeth Pacheco Batista Fontes.

The *Geminiviridae* family includes viruses whose genomes are composed of one or two molecules of circular, single stranded DNA encapsided by a single structural protein into twinned icosahedral particles. The family comprises the genera *Mastrevirus*, *Curtovirus*, *Topocuvirus* and *Begomovirus*, based on the type of insect vector, host range and genome organization. Begomoviruses (whitefly-transmitted geminiviruses) are responsible for serious agricultural threats. In Brazil, a number of novel begomoviruses associated with cultivated and non-cultivated hosts have been characterized. In this context, this work aimed to: (*i*) molecularly characterize a novel begomovirus species naturally infecting *Malvaviscus arboreus*; (*ii*) study the evolutionary dynamics of two begomovirus populations (*Tomato severe rugose virus* and *Macroptilium yellow spot virus*) infecting cultivated (*Solanum lycopersicum* and *Phaseolus vulgaris*) and non-cultivated hosts (*Sida* spp. and leguminous weeds) using a novel phylogeny-based partitioning method; and (*iii*) study the relative contribution of recombination in the evolutionary dynamics of worldwide important begomoviruses. For the first objective, *Malvaviscus arboreus* plants showing a bright yellow mosaic, collected at the states of São Paulo and Rio de Janeiro, were submitted to DNA extraction and the viral genome was amplified by rolling-circle amplification using the DNA polymerase from phage phi29. Sequence analysis indicated that the virus corresponds to a novel begomovirus species for which the name *Malvaviscus* yellow mosaic virus (MalYMV) is proposed. Strikingly, MalYMV has unique properties within the family *Geminiviridae*: a nanovirus- and alphasatellite-like nonanucleotide (5'-TAGTATTAC-3') and a sequence located 5' of the nonanucleotide capable of forming a minor hairpin structure embedded in the major hairpin. For the second objective, crops and weeds in similarly sized areas known to harbor either *Macroptilium yellow spot virus* (MaYSV) or *Tomato severe rugose virus* (ToSRV) were intensively sampled. The MaYSV population was notably more variable than the ToSRV population, largely due to a number of recombination events in the 5'-end of its Rep gene. By mapping the recombination events onto maximum likelihood trees based on CP and Rep sequences, it was possible to distinguish the individual contributions associated with

the evolutionary processes of mutation and recombination. Using this novel partitioning method, recombination was assessed as the source of 0 and 42% of the standing molecular variability of the ToSRV Rep and CP, respectively, and 36 and 16% of the variability of MaYSV Rep and CP, respectively. For the third objective, the use of the novel phylogeny-based partitioning method of variability was expanded to begomovirus datasets collected from around the world (available in public databases). We observed that the diversification of begomovirus populations is predominantly driven by mutational dynamics, albeit at different extents. Recombination was assessed as the source of up to 50% of all inferred substitutions. These results indicate that the evolution of some begomovirus populations might be significantly dependent on the recombination-prone nature of their genomes in addition to their rapid mutational dynamics.

INTRODUCTION

Geminiviruses are circular single stranded DNA (ssDNA) plant viruses encapsidated in twinned icosahedral capsids (Zhang *et al.*, 2001). These pathogens infect a wide range of cultivated and non-cultivated hosts and are responsible for important worldwide epidemics in plants (Rojas *et al.*, 2005b). In Africa, cassava mosaic and mayze streak diseases are the major biotic constraints to cassava and corn production, respectively (Bosque-Perez, 2000; Legg e Thresh, 2000). A complex of begomoviruses (whitefly-transmitted geminiviruses) causing the tomato yellow leaf curl disease is responsible for heavy losses in tomato cultivation in the western Mediterranean basin (Picó *et al.*, 1996; Rojas *et al.*, 2005b). In the Americas, diseases caused by begomoviruses have significantly impacted tomato and bean production since the 1980's (Brown e Bird, 1992; Gilbertson *et al.*, 1993a; Blair *et al.*, 1995; Polston e Anderson, 1997; Morales e Jones, 2004).

The family *Geminiviridae* is divided into four genera: *Mastrevirus*, *Curtovirus*, *Topocuvirus* and *Begomovirus*(Brown *et al.*, 2012). The genus *Begomovirus* is the most economically important and harbors the largest number of species in the family (Fauquet *et al.*, 2005). Begomoviruses from the 'Old World' (Europe, Asia and Africa) have one or two genomic components (known as mono- and bipartite begomoviruses, respectively), and are often associated with circular ssDNA molecules designated alphasatellites (previously DNA-1) and betasatellites (previously DNA β) (Briddon *et al.*, 2003; Briddon e Stanley, 2006). Alphasatellite genomes are similar to the genomic component named DNA-R of the ssDNA nanoviruses, and contain an open reading frame (ORF) encoding a replication-associated protein (Rep), followed by an adenine rich sequence and a stem-loop structure which includes the origin of replication (Briddon e Stanley, 2006). Alphasatellites can replicate autonomously, but require the helper begomovirus for systemic infection and insect transmission (Saunders e Stanley, 1999; Saunders *et al.*, 2000; Saunders *et al.*, 2002).

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). Betasatellites are entirely dependent on their helper viruses for replication, movement and insect transmission. Betasatellite genomes contain an ORF, named betaC1, which encodes a protein involved in the induction of symptoms and suppression of post-transcriptional gene silencing (Cui *et al.*, 2004; Cui *et al.*, 2005; Briddon e Stanley, 2006).

Begomoviruses from the 'New World' (the Americas) have two genomic components known as DNA-A and DNA-B, each 2,500-2,600 nucleotides (nt) in size (Fauquet *et al.*, 2005). These components do not share significant sequence identity, except for an ~200 nt sequence known as the common region (CR) which includes the origin of replication (Hanley-Bowdoin *et al.*, 1999 142). The genome of bipartite begomoviruses encodes six to eight proteins: the replication-associated protein (Rep), which is the initiator of the rolling circle replication mechanism 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 *ns* genes and also a suppressor of post-transcriptional gene silencing (Sunter e Bisaro, 1992; Voinnet *et al.*, 1999; Wang *et al.*, 2005); the replication-enhancer protein (Ren), a viral factor required for optimal replication (Sunter *et al.*, 1990; Pedersen e 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). 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).

After the viral particles are inoculated into a host cell by the insect vector, the viral ssDNA spontaneously dissociates from the capsid (Lazarowitz, 1992). Then, the ssDNA genome is transported to the nucleus and converted into a double-stranded DNA (dsDNA)

intermediate, termed replicative form (RF). The exact mechanism of conversion from ssDNA to dsDNA remains unknown. The viral genome is replicated using the RF as template using a rolling circle mechanism, similar to the replication of the ssDNA bacteriophages φX174 and M13 (Stenger *et al.*, 1991; Stanley, 1995).

The origin of replication (*ori*) is located in the common region in both genomic components. The sequence of the *ori* is conserved among components of the same virus, but 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 geminiviruses (Heyraud-Nitschke *et al.*, 1995; Orozco e Hanley-Bowdoin, 1998). Cleavage of the nonanucleotide is essential to the initiation of the replication and is performed by the Rep protein (Fontes *et al.*, 1994; Laufs *et al.*, 1995; Orozco *et al.*, 1998). The sequence-specific binding of the Rep protein to the viral genome involves the recognition of two short direct repeat sequences and at least one inverted repeat known as iterons (Arguello-Astorga *et al.*, 1994). Following Rep binding, the dsDNA:Rep complex is stabilized by the Ren protein and host factors. Then, the Rep protein cleaves the nonanucleotide located in the hairpin, initiating rolling circle replication (Gutierrez, 1999). The Rep catalytic domain (located at the N-terminal portion of the protein) includes three conserved motifs in proteins involved in rolling circle replication (motif I: FLTY, motif II: HxH and motif III: YxxxV) (Ilyina e Koonin, 1992) and the iteron-related domain (Arguello-Astorga e Ruiz-Medrano, 2001 1966).

Movement of plant viruses within the host can be divided into two processes: cell-to-cell movement via plasmodesmata, and long distance movement in which the virus moves systemically through the plant's vascular system. Begomoviruses replicate in the cell nucleus and thus require an additional transport step from the nucleus to the cytoplasm, which is performed by NSP (Palmer e Rybicki, 1998). The MP associates with the cell membrane and

alters the siza exclusion limit of the plasmodesmata, mediating the transport of the viral genome (Noueiry *et al.*, 1994). The NSP and MP proteins act cooperatively to mediate intra- and intercellular viral DNA movement (Sanderfoot e Lazarowitz, 1995 1036), allowing the virus to infect the host systemically.

Two models have been proposed to explain the intracellular movement of begomoviruses (Levy e Tzfira, 2010). In the first model, known as 'couple-skating' (Kleinow *et al.*, 2008), NSP transports the viral ssDNA or dsDNA from the nucleus to the cell periphery, and MP acts in the plasmodesmata to facilitate cell-to-cell movement of the NSP-DNA complex (Sanderfoot e Lazarowitz, 1995; Frischmuth *et al.*, 2004; Frischmuth *et al.*, 2007; Kleinow *et al.*, 2008). In the second model, known as 'relay-race', NSP transports the dsDNA from the nucleus to the cytoplasm. Then, the dsDNA associates with MP, and the dsDNA-MP complex moves cell-to-cell through plasmodesmata (Noueiry *et al.*, 1994; Rojas *et al.*, 1998). The MP and NSP recognize the viral DNA in a size- and form-specific manner (Rojas *et al.*, 1998; Gilbertson *et al.*, 2003). The CP is required for long distance movement of monopartite, but not bipartite, begomoviruses (Rojas *et al.*, 2005a).

Begomovirus populations exhibit high molecular variability (Ariyo *et al.*, 2005; Ge *et al.*, 2007; Silva *et al.*, 2011; Silva *et al.*, 2012). 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). Viruses with divided genomes (including the bipartite begomoviruses) may also evolve by pseudorecombination (or 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 pseudorecombinants has been obtained in laboratory settings (Gilbertson *et al.*, 1993b; Sung e 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) 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 e Holmes, 2006; Duffy *et al.*, 2008). Geminiviruses exhibit high levels of within-host molecular variability (Isnard *et al.*, 1998; Ge *et al.*, 2007; Van Der Walt *et al.*, 2008) and there is evidence that the rapid evolution of geminiviruses might be, at least in part, driven by mutational processes acting specifically on ssDNA (Duffy *et al.*, 2008; Harkins *et al.*, 2009).

An experiment to assess the within-host molecular variability of the begomovirus *Tomato yellow leaf curl China virus* (TYLCCNV) under controlled conditions revealed an average mutation frequency of 3.5×10^{-4} and 5.3×10^{-4} after a 60-day infection period in *N. benthamiana* and tomato (*Solanum lycopersicon*), respectively (Ge *et al.*, 2007). Additionally, substitution rates were estimated for temporally sampled *Tomato yellow leaf curl virus* (Duffy e Holmes, 2008) genome sequences. Although unexpected, the substitution rates (2.88×10^{-4} subs/site/year for the full-length genome and 4.63×10^{-4} subs/site/year for the CP) were similar to those of RNA viruses (Holland *et al.*, 1982; Domingo e Holland, 1997; Roossinck, 2003). These high substitution rates were validated for the bipartite begomovirus *East African cassava mosaic virus* (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; Duffy & Holmes, 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). The high recombination frequency in this group of viruses can be partly explained by the existence of a putative strategy of 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 simultaneously replicate in the same nucleus (Morilla *et al.*, 2004), also favour 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 *East African cassva mosaic virus* (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). 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 9706). 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).

In addition to the already existing set of information about full-length genome sequences of begomoviruses (available in public databases), the advent of the rolling circle amplification technique using the phi29 DNA polymerase (Inoue-Nagata *et al.*, 2004) has provided new possibilities, such as 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) and viral population analysis on a genomic scale (Haible *et al.*, 2006). In this context, this work aimed to: (*i*) molecularly characterize a novel begomovirus species naturally infecting *Malvaviscus arboreus*, carrying unique properties within the family *Geminiviridae*: a nanovirus- and alphasatellite-like nonanucleotide (5'-TA**G**TATTAC-3') and a sequence located 5' of the nonanucleotide capable of forming a minor hairpin structure embedded in the major hairpin; (*ii*) study the evolutionary dynamics of two begomovirus populations (*Tomato severe rugose virus* and *Macroptilium yellow spot virus*) infecting cultivated (*Solanum lycopersicum* and *Phaseolus vulgaris*) and non-cultivated hosts (*Sida* spp. and leguminous weeds) in Brazil using a novel phylogeny-based partitioning method; and (*iii*) study the minimal relative contribution of recombination in the evolutionary dynamics of begomoviruses by expanding our partitioning method to worldwide important begomoviruses (sequences available in public databases).

References

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

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

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.

BLAIR, M.W.; BASSET, M.J.; ABOUZID, A.M.; HIEBERT, E.; POLSTON, J.E.; MCMILLAN, R.T.; GRAVES, W.; LAMBERTS, M. Ocurrence of bean golden mosaic virus in Florida. Plant Disease, v. 79, p. 529-533, 1995.

BOSQUE-PEREZ, N.A. Eight decades of maize streak virus research. Virus Research, v. 71, p. 107-121, 2000.

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.; 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.; BIRD, J. Whitefly-transmitted geminiviruses and associated disorders in the Americas and the Caribbean basin. Plant Disease, v. 76, p. 220-225, 1992.

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. (Ed.). *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, 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.

CUI, X.F.; LI, G.X.; WANG, D.W.; HU, D.W.; ZHOU, X.P. A begomovirus DNA beta-encoded protein binds DNA, functions as a suppressor of RNA silencing, and targets the cell nucleus. Journal of Virology, v. 79, p. 10764-10775, 2005.

CUI, X.F.; TAO, X.R.; XIE, Y.; FAUQUET, C.M.; ZHOU, X.P. A DNA beta associated with *Tomato yellow leaf curl China virus* is required for symptom induction. Journal of Virology, v. 78, p. 13966-13974, 2004.

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 of the United States of America, 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.

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, p.1259ed. 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.

FRISCHMUTH, S.; KLEINOW, T.; ABERLE, H.J.; WEGE, C.; HULSER, D.; JESKE, H. Yeast two-hybrid systems confirm the membrane association and oligomerization of BC1 but do not detect an interaction of the movement proteins BC1 and BV1 of Abutilon mosaic geminivirus. Archives of Virology, v. 149, p. 2349-2364, 2004.

FRISCHMUTH, S.; WEGE, C.; HULSER, D.; JESKE, H. The movement protein BC1 promotes redirection of the nuclear shuttle protein BV1 of Abutilon mosaic geminivirus to the plasma membrane in fission yeast. Protoplasma, v. 230, p. 117-123, 2007.

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.; FARIA, J.C.; AHLQUIST, P.; MAXWELL, D.P. Genetic diversity in geminiviruses causing bean golden mosaic disease: The nucleotide sequence of the infectious cloned DNA components of a Brazilian isolate of bean golden mosaic geminivirus. Phytopathology, v. 83, p. 709-715, 1993a.

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, 1993b.

GILBERTSON, R.L.; SUDARSHANA, M.; JIANG, H.; ROJAS, M.R.; LUCAS, W.J. Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: Insights into DNA trafficking. Plant Cell, v. 15, p. 2578-2591, 2003.

GUTIERREZ, C. Geminivirus DNA replication. Cellular and Molecular Life Sciences, v. 56, p. 313-329, 1999.

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.

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.

KLEINOW, T.; HOLEITER, G.; NISCHANG, M.; STEIN, M.; KARAYAVUZ, M.; WEGE, C.; JESKE, H. Post-translational modifications of *Abutilon mosaic virus* movement protein (BC1) in fission yeast. Virus Research, v. 131, p. 86-94, 2008.

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

LEVY, A.; TZFIRA, T. *Bean dwarf mosaic virus*: A model system for the study of viral movement. Molecular Plant Pathology, v. 11, p. 451-461, 2010.

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

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

MORALES, F.J.; JONES, P.G. The ecology and epidemiology of whitefly-transmitted viruses in Latin America. Virus Research, v. 100, p. 57-65, 2004.

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.

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.

PALMER, K.E.; RYBICKI, E.P. The molecular biology of mastreviruses. Advances in Virus Research, v. 50, p. 183-234, 1998.

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.

PICÓ, B.; DIEZ, M.J.; NUEZ, F. Viral diseases causing the greatest economic losses to the tomato crop. II. The tomato yellow leaf curl virus - a review. Scientia Horticulturae, v. 67, p. 151-196, 1996.

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.

POLSTON, J.E.; ANDERSON, P.K. The emergence of whitefly-transmitted geminiviruses in tomato in the western hemisphere. Plant Disease, v. 81, p. 1358-1369, 1997.

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.

ROJAS, A.; KVARNHEDEN, A.; MARCENARO, D.; VALKONEN, J.P.T. Sequence characterization of *Tomato leaf curl Sinaloa virus* and *Tomato severe leaf curl virus*: Phylogeny of New World begomoviruses and detection of recombination. Archives of Virology, v. 150, p. 1281-1299, 2005a.

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

ROJAS, M.R.; NOUEIRY, A.O.; LUCAS, W.J.; GILBERTSON, R.L. Bean dwarf mosaic geminivirus movement proteins recognize DNA in a form- and size-specific manner. Cell, v. 95, p. 105-113, 1998.

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.

SANDERFOOT, A.A.; LAZAROWITZ, S.G. Cooperation in viral movement: The geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. Plant Cell, v. 7, p. 1185-1194, 1995.

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.; BRIDDON, R.W.; MARKHAM, P.G.; WONG, S.M.; STANLEY, J. A unique virus complex causes *Ageratum yellow vein disease*. Proceedings of the National Academy of Sciences, USA, v. 97, p. 6890-6895, 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.

SAUNDERS, K.; BEDFORD, I.D.; STANLEY, J. Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with ageratum yellow vein disease. Journal of General Virology, v. 83, p. 907-913, 2002.

SAUNDERS, K.; STANLEY, J. A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: Evidence for interfamilial recombination between plant DNA viruses. Virology, v. 264, p. 142-152, 1999.

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.

STANLEY, J. Analysis of African cassava mosaic virus recombinants suggest strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle DNA replication. Virology, v. 206, p. 707-712, 1995.

STENGER, D.C.; REVINGTON, G.N.; STEVENSON, M.C.; BISARO, D.M. Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. Proceedings of the National Academy of Sciences, USA, v. 88, p. 8029-8033, 1991.

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.

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.; 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.

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

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

Malvaviscus yellow mosaic virus, A RECOMBINANT WEED-INFECTING BEGOMOVIRUS CARRYING A NANOVIRUS-LIKE NONANUCLEOTIDE AND A MODIFIED STEM-LOOP STRUCTURE

Lima, A.T.M., Almeida, M.S.S., Rocha, C.S., Barros, D.R., Castillo-Urquiza, G.P., Silva, F.N., Alfenas-Zerbini, P., Barbosa, J.C., Albuquerque, L.C., Inoue-Nagata, A.K., Kitajima, E.W. & Zerbini, F.M. *Malvaviscus yellow mosaic virus*, a recombinant weed-infecting begomovirus carrying a nanovirus-like nonanucleotide and a modified stem-loop structure. Virology Journal, *in preparation*.

Malvaviscus yellow mosaic virus, a recombinant weed-infecting begomovirus carrying a nanovirus-like nonanucleotide and a modified stem-loop structure

A.T.M. Lima[1], M. M. S. Almeida[2], C.S. Rocha[1], D.R. Barros[1*], G. P. Castillo-Urquiza[1], F.N. Silva[1], P. Alfenas-Zerbini[3], J. C. Barbosa[2], L. C. Albuquerque[2], A. K. Inoue-Nagata[2], E.W. Kitajima[4] & F. M. Zerbini[1]

[1]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, 36570-000, Brazil;

[2]Embrapa Hortaliças, Brasília, DF;

[3]Dep. de Microbiologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, 36570-000, Brazil;

[4]NAP/MEPA, ESALQ-USP, Piracicaba, SP, 13418-900, Brazil

[*]Current address: Dep. de Fitossanidade, Universidade Federal de Pelotas, Pelotas, RS, 96010-000, Brazil

Corresponding author: F.M. Zerbini

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

1 **Abstract**

2 Begomoviruses (family *Geminiviridae*) have a circular, ssDNA genome encapsidated in
3 twinned icosahedral particles. In Brazil, a number of begomoviruses infecting weeds have
4 been described, and evidence suggests that they have given rise to the viruses currently found
5 in crop plants. Here we describe a novel begomovirus infecting *Malvaviscus arboreus* plants
6 showing a bright yellow mosaic, collected at Campinas, São Paulo state in May 2005 and Rio
7 de Janeiro, Rio de Janeiro state in August 2009 and February 2011. Total DNA was extracted
8 and the viral genome was amplified by RCA, cloned and sequenced. Sequence analysis
9 indicated that the virus corresponds to novel species, for which the name Malvaviscus yellow
10 mosaic virus (MalYMV) is proposed. Strikingly, MalYMV has a nanovirus- and
11 alphasatellite-like nonanucleotide (TAGTATTAC). Moreover, a short sequence located 5' of
12 the nonanucleotide potentially forms a minor hairpin structure embedded in the major
13 hairpin. Intramolecular interactions involving the sequence of the atypical nonanucleotide
14 were predicted. Although MalYMV has been collected in Brazil, it is phylogenetically closer
15 to viruses from Central and North America. The *M. arboreus* plant at Campinas has been
16 displaying the observed yellow mosaic symptoms since at least the 1960's, which suggests
17 that MalYMV may be poorly transmitted (or not transmitted at all) by local whitefly
18 populations.

Introduction

The family *Geminiviridae* includes viruses whose genomes are comprised of one or two circular single-stranded DNA molecules encapsidated by a single structural protein in twinned icosahedral particles. The family consists of four genera, *Mastrevirus*, *Curtovirus*, *Topocuvirus* and *Begomovirus*, defined based on the type of insect vector, host range and genome organization. Begomoviruses have one or two genomic components, are transmitted by the whitefly *Bemisia tabaci* and infect dicot hosts [1]. The DNA-A of bipartite begomoviruses encodes viral proteins required for replication (Rep and Ren) [2-5], regulation of gene expression and suppression of RNA silencing (Trap) [6-8] and encapsidation of the viral progeny (CP) [9, 10], whereas the proteins necessary for inter- and intracellular trafficking (MP and NSP) are encoded by the DNA-B component [11, 12].

The replication origin is located in the common intergenic region of the two genomic components. The sequence of the origin is highly conserved between components of the same virus, but variable among species, except for a region of about 30 nucleotides conserved in all species [13, 14]. This region includes an inverted GC-rich repeat sequence, forming a potential conserved stem-loop structure with an invariant loop sequence (5'-TAATATTAC-3') found in all geminiviruses, which is the functional domain of the replication origin [15, 16]. An exception to this rule is the recently described curtovirus *Beet curly top Iran virus*, which carries the nonanucleotide sequence TAAGATT/CC [17].

During the last two decades begomoviruses have emerged as major plant pathogens, particularly in tropical and subtropical regions of the world, causing severe economic losses [18]. In Brazil, the most severely affected crops have been common bean and tomato [19, 20].

Native weed species acting as reservoirs can play an important role in plant virus epidemics [21]. Similarly to what is observed for begomoviruses in crops, the species

26 diversity of begomoviruses in weeds is very high, particularly in species of the Malvaceae
27 family [22-27].

28 The presence of several viruses in the field, all transmitted by the same insect vector,
29 facilitates the occurrence of mixed infections which in turn increase the likelihood of
30 recombination and pseudorecombination events leading to the emergence of novel species
31 [28-32]. Furthermore, the high evolution rates estimated for begomoviruses may provide a
32 rapid adaptation to new hosts [33, 34]. Therefore, the characterization of weed-infecting
33 begomoviruses can provide new insights into the evolutionary behavior of geminiviruses
34 [35].

35 In this study, we report the molecular characterization of a novel begomovirus, for
36 which the name Malvaviscus yellow mosaic virus (MalYMV) is proposed. MalYMV was
37 found naturally infecting *Malvaviscus arboreus* plants collected in São Paulo and Rio de
38 Janeiro states. Although MalYMV has a typical New World bipartite begomovirus genome, it
39 possesses unique characteristics within the *Geminiviridae* family: a nanovirus- and
40 alphasatellite-like nonanucleotide (5' TAGTATTAC 3') and a short sequence located 5' of the
41 nonanucleotide capable of forming a minor hairpin structure embedded in the major hairpin.

42 **Material and Methods**

43 *Sample collection and DNA extraction*

44 An approximately 50 year old *Malvaviscus arboreus* plant (family Malvaceae)
45 showing a bright yellow mosaic (Figure 1A) was collected at the experimental farm of the
46 Campinas Agronomical Institute (IAC), in Campinas, São Paulo state, Brazil, in May 2005.
47 The plant has since been vegetatively propagated in a greenhouse at the Universidade Federal
48 de Viçosa. Samples from six *M. arboreus* plants showing similar symptoms were also
49 collected from residential gardens at Rio de Janeiro, Rio de Janeiro state, in August 2009 and
50 February 2011. Total DNA was extracted from leaf discs according to Doyle & Doyle [36]
51 and preserved at -20°C.

52

53 *Cloning and sequencing of the full-length genomic components*

54 The genomic components were amplified by rolling-circle amplification (RCA) using
55 the DNA polymerase from phage phi29, according to a standard circular DNA cloning
56 method [37]. The concatamers were excised with *Hind* III or *Pst* I (which cleave the DNA-A
57 from all isolates at a single site) and *Spe* I or *Sac* I (which cleave the DNA-B from all isolates
58 at single site). Fragments of approximately 2,600 nucleotides, corresponding to one genomic
59 copy of each component, were ligated to the pBluescript KS+ plasmid vector (Stratagene).
60 The recombinant plasmids were used for *E. coli* transformation and the clones obtained were
61 completely sequenced by primer walking at Macrogen, Inc. (Seoul, South Korea).

62

63 *Plant inoculations and viral detection*

64 Approximately 2 µg of RCA-amplified DNA from the 50-year old *M. arboreus* plant
65 were used for biolistic inoculation of *Nicotiana benthamiana* plants [38]. Inoculated plants
66 were kept in a greenhouse with average daily temperatures of 26 ± 2°C. Two weeks after

67 inoculation, the plants were evaluated for the presence of symptoms, photographed and total
68 DNA was extracted from newly emerging leaves [39]. To confirm the viral infection, the total
69 DNA extract was used as a template for RCA, and the amplification products were submitted
70 to restriction analysis with *Hin*d III and *Spe* I. The DNA components obtained from the
71 inoculated *N. benthamiana* plants were cloned and completely sequenced.

72 In order to confirm the infectivity of the clones corresponding to the DNA-A and -B
73 from the BR:Ca01:05 isolate, the genomic components were excised from the plasmid vector,
74 re-ligated and biolistically inoculated onto *N. benthamiana* plants. Two weeks after
75 inoculation, total DNA was extracted from newly emerging leaves and the genomic
76 components were cloned and completely sequenced as described above.

77

78 *Sequence comparisons and phylogenetics analysis*

79 The complete DNA-A sequences were initially analyzed using the BLAST*n* algorithm
80 [40] to determine the viral species with the highest identity. Nucleotide and amino acid
81 sequences of the six isolates from *M. arboreus* plus additional representative begomoviruses
82 were used for pairwise comparisons. Percent nucleotide and amino acid identities for the
83 entire genome and for each viral ORF were calculated in EMBOSS Pairwise Alignment
84 Algorithms (http://www.ebi.ac.uk/Tools/emboss/align/ index.html), using the EMBOSS
85 needle (Global) with default settings.

86 A phylogenetic tree was constructed using Bayesian inference based on the full length
87 BR:Ca01:05 DNA-A and an additional 47 DNA-A and DNA-A-like sequences. A multiple
88 sequence alignment was constructed using Muscle [41] and manually edited in Mega 5.0
89 [42]. The Bayesian inference and Markov Chain Monte Carlo simulation were performed in
90 MrBayes 3.0 [43], with the evolutionary model selected by MrModeltest2.2 [44] in the
91 Akaike Information Criterion (AIC). Two runs with four Markov chains were conducted

92 simultaneously (each running 20,000,000 generations) starting from random initial trees.
93 Trees were sampled every 500 generations, resulting in 40,000 saved trees. Burn-in values
94 were set to the generation number after which the likelihood values were stationary. The
95 consensus phylogenetic tree was visualized using FigTree (tree.bio.ed.ac.uk/software/figtree).
96

97 *Prediction of the viral replication origin structure*

98 The putative secondary structures and the minimum free energy (ΔG) values at the
99 viral replication origin were predicted using Zuker's algorithm in the MFold web server for
100 single strand nucleic acid folding and hybridization prediction (version 3.2) with default
101 settings [45] (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form). Local free energy
102 with a high negative ΔG value reflects a stable internal structure.
103

104 *Recombination analysis*

105 Recombination analysis was performed for the full-length DNA-A sequence obtained
106 for the isolate BR:Ca01:05 plus New World begomoviruses from Americas and their isolates
107 using the rdp [46], geneconv [47], bootscan [48], maximum chi square [49], chimaera [50],
108 sister scan [51] and 3seq [52] methods as implemented in Recombination Detection Program
109 (RDP) version 3.0 [53]. Pairs of sequences sharing less than 70% identity were discarded.
110 Additionally, we estimated a threshold identity value between any sequence pair above which
111 no recombination event can be detected, as recommended in the RDP user manual. One
112 sequence of each pair sharing a value greater than the threshold was discarded. Alignments
113 were scanned with default settings for the different methods using a Bonferroni-corrected *p*
114 value cutoff of 0.05. Only recombination events detected by at least four of the analysis
115 methods available in the program and coupled with phylogenetic evidence were considered
116 reliable.

117 All sequences used in this work were obtained from the GenBank database under the
118 accession numbers given in Supplementary Table S1.
119
120 **Results**
121 *Characterization of a novel begomovirus infecting Malvaviscus arboreus*
122 The DNA-A of the begomovirus isolate obtained from the 50-year old *M. arboreus*
123 plant (named BR:Ca01:05) has 2,705 nucleotides and a typical New World bipartite
124 begomovirus genomic organization (Figure 1). The DNA-A contains five ORFs: *cp*, in the
125 virion sense, with 251 deduced amino acids, and *ren* (132 deduced amino acids), *trap* (129
126 deduced amino acids), *rep* (358 deduced amino acids) and *ac4* (87 deduced amino acids) in
127 the complementary sense.
128 The DNA-A nucleotide sequence of the BR:Ca01:05 isolate had the greatest identity
129 (75.7%) with *Tomato yellow spot virus* (ToYSV) (Supplementary Figure S1). Therefore,
130 based on the criteria established by the ICTV [54], this isolate represents a novel
131 begomovirus specie for which the name *Malvaviscus* yellow mosaic virus (MalYMV) is
132 proposed. We also cloned and completely sequenced five DNA-A components from six *M.*
133 *arboreus* samples collected at residential gardens in Rio de Janeiro (in August 2009 and
134 February 2011). These DNA-A components showed a similar genome organization and an
135 overall pairwise identity value of about 98% with the BR:Ca01:05 isolate (Supplementary
136 Figure S1), and therefore represent additional MalYMV isolates.
137 Attempts to clone the BR:Ca01:05 DNA-B component using the same approach were
138 unsuccessful. To overcome the problem, this isolate was inoculated onto *Nicotina*
139 *benthamiana* plants. *N. benthamiana* plants biolistically inoculated with RCA-amplified
140 DNA from *M. arboreus* displayed symptoms of viral infection two weeks after inoculation
141 (Figure 1). Viral infection was confirmed by PCR using degenerate primers for the DNA-A

142 [55]: a fragment of about 1,100 nt was amplified for four out of five inoculated *N.*
143 *benthamiana* plants (Figure 1). RCA-amplified DNA from the infected *N. benthamiana*
144 plants was cleaved with *Hin*d III, yielding unit-length DNA-A molecules. The sequences
145 obtained for the DNA-A components cloned from *N. benthamiana* plants were identical to
146 those previously obtained from *M. arboreus*. Then, based on a restriction analysis of the
147 DNA-A, we selected a restriction enzyme (*Spe* I) which does not cleave this component but
148 nevertheless generates a 2,600 nt fragment after digestion of the RCA-amplified DNA. We
149 cloned and completely sequenced the *Spe* I-generated 2,600 nt DNA fragment. The analysis
150 of the obtained sequences indicated that this component was a typical DNA-B of a New
151 World bipartite begomovirus (Figure 1). The common region comprised 206 nts and showed
152 an identity value of 95.6% with the corresponding BR:Ca01:05 DNA-A sequence, indicating
153 that both components were cognate DNAs.

154 The MalYMV DNA-B has 2,684 nts and two ORFs: *nsp* in the virion sense, with 256
155 deduced amino acids, and *mp* in the complementary sense, with 293 deduced amino acids.
156 Pairwise comparisons indicated a greatest identity value with ToYSV DNA-B (65.3%)
157 (Supplementary Figure S1). The sequence of the BR:RJ05:09 DNA-B was also determined
158 and showed an identity value of 96% with the BR:Ca01:05 DNA-B. The infection of *M.*
159 *arboreus* and *N. benthamiana* plants by a single begomovirus was confirmed by digestion of
160 the RCA-amplified DNA from each plant with *Msp* I (Figure 1A, 1B). The electrophoretic
161 profile was consistent with the presence of a single begomovirus, as the sum of the generated
162 fragments was approximately 5,200 base pairs.

163 The nt and deduced amino acid (aa) sequences from each MalYMV ORF were
164 compared to those of other begomoviruses. Pairwise comparisons indicated that the ORF
165 encoding the coat protein was the most conserved, with nucleotide identity values ranging
166 from 81 to 83.8%, and that the AC4 ORF was the least conserved, with values of 39-83%. In

general, the DNA-A ORFs displayed the greatest identity values with the corresponding ORFs from others malvaceous-infecting begomovirus and the tomato infecting begomoviruses *Tomato leaf distortion virus* and ToYSV (Supplementary Figure S1).

The replication origin of MalYMV

The hairpin motif located at the 5' intergenic region typically forms a loop containing two or three variable nucleotides plus an evolutionarily conserved nonanucleotide [56]. Strikingly, we observed an atypical nonanucleotide sequence (TAGTATTAC) as part of the stem-loop structure of the replication origin of all six MalYMV isolates (Figure 2A). The same nonanucleotide sequence is found in nanoviruses and alphasatellites [57]. Additionally, a short sequence located 5' of the nonanucleotide (5- AGGGCGAAGCCCT-3') potentially forms a minor hairpin structure embedded in the major hairpin (Figure 2B). We sequenced distinct clones for each viral isolate, and the replication origin sequences were identical in all of them.

To rule out the possibility that this atypical structure and nonanucleotide sequence were artifacts, the excised and re-ligated DNA-A and -B components from the recombinant plasmids (from which the sequences were determined) were used for biolistically inoculation of five *N. benthamiana* plants. Two weeks after inoculation, DNA-A and -B components were re-cloned and re-sequenced from these plants. The sequences were identical to those previously obtained for all isolates from *M. arboreus* (data not shown).

Characterization of MalYMV Rep protein and its specific binding sites in the viral genome

Although based on a small number of isolates, our results suggest that these atypical features present in the MalYMV replication origin have been evolutionarily conserved. In

192 this case, it is possible that they are recognized by the Rep protein since this protein acts as a
193 site-specific endonuclease with requirement of structure and sequence [16, 58, 59]. This
194 prompted us to further characterize the conserved motifs and domains present in the
195 MalYMV Rep protein and its specific binding sites in the viral genome.

196 The iteron-related domain (IRD) involved in the sequence-specific recognition of
197 iterons [60] was identified in the N-terminal MalYMV Rep as ${}_{5}\text{KPGFPIH}_{11}$. We also
198 identified the iterons in the cognate DNA components as ${}_{2620}\text{GGAAA}_{2624}/{}_{2627}\text{GGAAA}_{2631}$.
199 Based on a 254-sequence genome data set containing New and Old Word bipartite
200 begomovirus we were unable to find another begomovirus carrying the same iterons. Thus,
201 our results suggest that the MalYMV Rep binding sites are also unique among bipartite
202 begomoviruses.

203 The Rep protein is the initiator of rolling circle replication, with nucleic acid binding,
204 endonuclease and ATPase functions [2, 3]. Three conserved amino acids motifs [61] have
205 been identified in the MalYMV Rep protein catalytic domain: the rolling circle replication
206 motif I (${}_{15}\text{FLTYPQCS}_{22}$), motif II containing two histidines that form a putative metal ion
207 binding site (${}_{55}\text{NP}\underline{\text{H}}\text{L}\underline{\text{H}}\text{VLLQ}_{63}$), and motif III containing the active site tyrosine
208 (${}_{100}\text{VKD}\underline{\text{Y}}\text{VSKDGD}_{109}$).

209

210 *Phylogenetics analysis*

211 To determine the phylogenetic relationships of the atypical begomovirus characterized
212 in this study, a phylogenetic tree based on full-length DNA-A sequences was constructed
213 using Bayesian inference under the evolutionary model GTR+I+G. Begomoviruses clustered
214 according to their geographical origin, with five out of six clades being definitely supported
215 by posterior probability values greater than 0.95 (Figure 3). Clades I and II included

216 exclusively begomoviruses from South America, while clades IV, V and VI contained
217 begomoviruses reported in North and/or Central America.

218 It was notable that clade III includes begomovirus from North and Central America
219 plus four begomoviruses reported primarily in Brazil, including the MalYMV isolate
220 characterized in this work. The other three viruses were the recently reported *Sida yellow leaf*
221 *curl virus* (SiYLCV) and *Tomato common mosaic virus* (ToCmMV) [26] and *Abutilon Brazil*
222 *virus* (AbBV) [62]. We have demonstrated, however, that these three viruses share a similar
223 recombinant fragment spanning a large part of their Rep genes, derived from a begomovirus
224 from clade III. Our results support this unexpected clustering as a result of this recombination
225 event and not as an inadvertent introduction of these viruses from Central and/or North
226 America into Brazil (Rocha *et al.*, *manuscript in preparation*). Therefore, is possible that the
227 MalYMV genome has a similar recombinant origin. Alternatively, MalYMV could have been
228 introduced into Brazil through infected propagation material since *Malvaviscus arboreus* is
229 an ornamental species native from Central, North and western South America
230 (http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?105661).

231

232 *Recombination analysis*

233 The unexpected clustering of MalYMV with begomoviruses from Central and North
234 America prompted us to investigate the occurrence of recombination events involving
235 MalYMV isolates. Our recombination analysis using a 127-sequence New World
236 begomovirus data set (66 species) detected a total of 47 unique recombination events.

237 A recombination event in all MalYMV isolates within the Rep gene (nt position from
238 2,146 to 2,321) was confirmed by six detection methods contained in the RDP3 package (*P*-
239 values: rdp = 1.598×10^{-08}, Geneconv = 2.360×10^{-12}, Bootscan = 4.399×10^{-09}, Maximum χ^2
240 = 6.034×10^{-05}, Chimaera = 3.821×10^{-02}, Sister scan = 2.936×10^{-09}). This recombination

241 event has *Merremia mosaic virus* (a begomovirus primarily reported in Puerto Rico) as major
242 parent and an unknown virus as minor parent. However, this event is quite distinct from the
243 recombination event detected for AbBV, SiYLCVand ToCmMV.
244
245 **Discussion**
246 The advent of the rolling circle amplification technique using the phage phi-29 DNA
247 polymerase has allowed large-scale cloning of geminivirus genomes [63]. As a result, it is
248 now possible to performe exhaustive sampling of geminiviruses in cultivated and wild hosts.
249 Novel geminiviruses with distinct properties from most viruses in the family have been
250 identified and characterized recently [35]. Here we report the characterization of a
251 *Malvaviscus arboreus*-infecting begomovirus (MalYMV) containing a distinct
252 nonanucleotide and a stem-loop structure at the origin of replication unique in the family
253 *Geminiviridae*. In addition, our results imply that this begomovirus is phylogenetically closer
254 to begomoviruses from Central and North America than to those reported in Brazil. This
255 virus showed a maximum sequence identity of 75.7% with *Tomato yellow spot virus*, and
256 should therefore be considered a novel begomovirus species for which the name *Malvaviscus*
257 *yellow mosaic virus* (MalYMV) is proposed.
258 In Brazil, malvaceous-infecting begomovirus have been important contributors to the
259 genetic diversity of begomovirus infecting economically important crops, especially tomato.
260 A close phylogenetic relationship has been demonstrated among tomato-infecting
261 begomoviruses such as *Tomato yellow spot virus*, *Tomato leaf distortion virus* and *Tomato*
262 *mild mosaic virus* and *Sida*-infecting begomoviruses such as *Sida mottle virus* and *Sida*
263 *yellow mosaic virus* [26, 30]. Furthermore, recombinant begomoviruses that infect cultivated
264 hosts often have *Sida*-infecting begomovirus as parental viruses (*eg*, *Tomato severe rugose*
265 *virus*). Based on a extensive sampling in the main tomato-producing regions of the country,

266 *Sida*-infecting begomoviruses have been found (albeit sporadically) in tomato [26, 64]. These
267 results indicate that malvaceous-infecting begomovirus have either given rise to the
268 begomoviruses found in cultivated hosts (especially tomato) or have contributed with genetic
269 material via recombination.

270 Our phylogenetic analysis showed that MalYMV has a close phylogenetic relationship
271 with other begomoviruses from North and Central America. These results are unexpected but
272 not surprising since other recently reported begomoviruses in Brazil (AbBV, SiYLCV and
273 ToCmMV) also cluster in clade III. Although these results suggest initially a possible
274 introduction of these viruses into Brazil, we have demonstrated that they share a similar
275 recombination event with a parental virus from clade III. A small recombinant fragment
276 present in the MalYMV Rep gene was detected, but this event did not resemble the one
277 detected for AbBV, SiYLCV and ToCmMV. Therefore we hypothesize that the unexpected
278 inclusion of MalYMV in clade III is possibly due to the introduction of this virus into Brazil
279 through infected propagative material from Central, North or western South America (from
280 where *Malvaviscus arboreus* is native). This introduction must have been old since the *M.*
281 *arboreus* plant from which the BR:Ca01:05 isolate was obtained has been kept at Campinas
282 Agronomic Institute for at least 60 years.

283 Recently, two geminiviruses carrying an unusual nonanucleotide have been reported
284 from Iran [17] and South Africa [35]. Despite great differences in genomic organization, *Beet*
285 *curly top Iran virus* (BCTIV) and *Eragrostis curvula streak virus* (ECSV) have the same
286 unusual nonanucleotide (TAAGATTCC) as part of the stem-loop structures at the origin of
287 replication. All six MalYMV isolates analyzed in this work have a third type of
288 nonanucleotide, which is also present in alphasatellites and in nanoviruses (family
289 *Nanoviridae*) [57]. We are currently producing single nucleotide mutants (TA<u>A</u>TATTAC) to

290 assess the ability of the MalYMV Rep protein in replicating a viral genome with the wild-
291 type nonanucleotide.

292 To this date, no begomovirus carrying an atypical stem-loop structure has ever been
293 reported. Based on secondary structure predictions we detected a second hairpin embedded in
294 the major stem-loop, plus intramolecular interactions involving the nonanucleotide sequence
295 that significantly reduced the loop size. Mutations in the variable nucleotides adjacent to the
296 TGMV DNA-B nonanucleotide that altered loop sequence or structure significantly affected
297 its replication in the presence of wild-type DNA-A in tobacco protoplasts [59]. These results
298 indicate that the loop structure as well as its sequence are important for the functionality of
299 the origin of replication. Therefore, it is not unreasonable to assume that the intramolecular
300 interactions predicted in the MalYMV loop most likely do not occur *in vivo*. On the other
301 hand, the additional minor hairpin structure did not affect the loop size. This could imply in
302 the recognition by the Rep protein in a manner similar to that observed for other
303 begomoviruses carrying wild-type structures [59]. At the same time, the conservation of this
304 structure in all MalYMV isolates is intriguing and suggests a possible role in the recognition
305 of the MalYMV replication origin by the Rep protein. The conserved motifs I, II and III
306 found in rolling circle replication proteins [61] were all observed in the Rep protein primary
307 structure. However, we also detected deviations from the consensus sequence in some of
308 these motifs, for example, around the catalytic tyrosine in motif III (data not shown). The
309 possible role of these residues in the recognition of the atypical nonanucleotide and stem-loop
310 structure is being investigated.

311 The biological implications of these unusual features are unknown. However, if they
312 are recognized by Rep, and considering the iteron sequences which are unique among
313 bipartite begomoviruses, this could explain the apparent genetic isolation of this virus since it

would hinder, for example, the occurrence of pseudorecombination events with other viruses carrying wild-type structures.

Acknowledgements

We would like to thanks Dr. Arvind Varsani, School of Biological Sciences, University of Canterbury, for generating the similarity plots in the MatLab software. Part of this work was carried out using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation.

References

1. Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC: **Family *Geminiviridae*.** In *Virus Taxonomy Eighth Report of the International Committee on Taxonomy of Viruses.* Edited by Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA. San Diego: Elsevier Academic Press; 2005: 301-326
2. Fontes EPB, Luckow VA, Hanley-Bowdoin L: **A geminivirus replication protein is a sequence-specific DNA binding protein.** *Plant Cell* 1992, **4:**597-608.
3. Orozco BM, Miller AB, Settlage SB, Hanley-Bowdoin L: **Functional domains of a geminivirus replication protein.** *J Biol Chem* 1997, **272:**9840-9846.
4. Sunter G, Hartitz MD, Hormuzdi SG, Brough CL, Bisaro DM: **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* 1990, **179:**69-77.
5. Pedersen TJ, Hanley-Bowdoin: **Molecular characterization of the AL3 protein encoded by a bipartite geminivirus.** *Virology* 1994, **202:**1070-1075.
6. Sunter G, Bisaro DM: **Transactivation of geminivirus AR1 and BR2 gene expression by the viral AL2 gene product occurs at the level of transcription.** *Plant Cell* 1992, **4:**1321-1331.
7. Voinnet O, Pinto YM, Baulcombe DC: **Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants.** *Proc Natl Acad Sci U S A* 1999, **96:**14147-14152.
8. Wang H, Buckley KJ, Yang X, Buchmann RC, Bisaro DM: **Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins.** *Journal of Virology* 2005, **79:**7410-7418.
9. Briddon RW, Pinner MS, Stanley J, Markham PG: **Geminivirus coat protein gene replacement alters insect specificity.** *Virology* 1990, **177:**85-94.
10. Hofer P, Bedford ID, Markham PG, Jeske H, Frischmuth T: **Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate.** *Virology* 1997, **236:**288-295.
11. Noueiry AO, Lucas WJ, Gilbertson RL: **Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport.** *Cell* 1994, **76:**925-932.
12. Sanderfoot AA, Lazarowitz SG: **Cooperation in viral movement: The geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery.** *Plant Cell* 1995, **7:**1185-1194.
13. Davies JW, Stanley J, Donson J, Mullineaux PM, Boulton MI: **Structure and replication of geminivirus genomes.** *Journal of Cell Science* 1987, **7:**95-107.
14. Lazarowitz SG: **Geminiviruses: Genome structure and gene function.** *Critical Reviews in Plant Sciences* 1992, **11:**327-349.

358 15. Heyraud-Nitschke F, Schumacher S, Laufs J, Schaefer S, Schell J, Gronenborn B: **Determination of**
359 **the origin cleavage and joining domain of geminivirus Rep proteins.** *Nucleic Acids Research* 1995,
360 **23:**910-916.
361 16. Orozco BM, Hanley-Bowdoin L: **Conserved sequence and structural motifs contribute to the DNA**
362 **binding and cleavage activities of a geminivirus replication protein.** *Journal of Biological*
363 *Chemistry* 1998, **273:**24448-24456.
364 17. Yazdi HRB, Heydarnejad J, Massumi H: **Genome characterization and genetic diversity of beet**
365 **curly top Iran virus: a geminivirus with a novel nonanucleotide.** *Virus Genes* 2008, **36:**539-545.
366 18. Morales FJ: **History and current distribution of begomoviruses in Latin America.** *Advances in*
367 *Virus Research* 2006, **67:**127-162.
368 19. Faria JC, Maxwell DP: **Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil.**
369 *Phytopathology* 1999, **89:**262-268.
370 20. Zerbini FM, Andrade EC, Barros DR, Ferreira SS, Lima ATM, Alfenas PF, Mello RN: **Traditional**
371 **and novel strategies for geminivirus management in Brazil.** *Australasian Plant Pathology* 2005,
372 **34:**475-480.
373 21. Seal SE, Jeger MJ, Van den Bosch F: **Begomovirus evolution and disease management.** *Advances in*
374 *Virus Research* 2006, **67:**297-316.
375 22. Frischmuth T, Engel M, Lauster S, Jeske H: **Nucleotide sequence evidence for the occurrence of**
376 **three distinct whitefly-transmitted, *Sida*-infecting bipartite geminiviruses in Central America.** *J*
377 *Gen Virol* 1997, **78:**2675-2682.
378 23. Hofer P, Engel M, Jeske H, Frischmuth T: **Nucleotide sequence of a new bipartite geminivirus**
379 **isolated from the common weed *Sida rhombifolia* in Costa Rica.** *Virology* 1997, **78:**1785-1790.
380 24. Assunção IP, Listik AF, Barros MCS, Amorim EPR, Silva SJC, Izael OS, Ramalho-Neto CE, Lima
381 GSA: **Diversidade genética de begomovírus que infectam plantas invasoras na Região Nordeste.**
382 *Planta Daninha* 2006, **24:**239-244.
383 25. Guo XJ, Zhou XP: **Molecular characterization of a new begomovirus infecting *Sida cordifolia* and**
384 **its associated satellite DNA molecules.** *Virus Genes* 2006, **33:**279-285.
385 26. Castillo-Urquiza GP, Beserra Jr. JEA, Bruckner FP, Lima ATM, Varsani A, Alfenas-Zerbini P, Zerbini
386 FM: **Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil.**
387 *Archives of Virology* 2008, **153:**1985-1989.
388 27. Fiallo-Olive E, Martinez-Zubiaur Y, Moriones E, Navas-Castillo J: **Complete nucleotide sequence of**
389 ***Sida golden mosaic Florida virus* and phylogenetic relationships with other begomoviruses**
390 **infecting malvaceous weeds in the Caribbean.** *Archives of Virology* 2010, **155:**1535-1537.
391 28. Pita JS, Fondong VN, Sangare A, Otim-Nape GW, Ogwal S, Fauquet CM: **Recombination,**
392 **pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of**
393 **severe cassava mosaic disease in Uganda.** *J Gen Virol* 2001, **82:**655-665.
394 29. Monci F, Sanchez-Campos S, Navas-Castillo J, Moriones E: **A natural recombinant between the**
395 **geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibits a**
396 **novel pathogenic phenotype and is becoming prevalent in Spanish populations.** *Virology* 2002,
397 **303:**317-326.
398 30. Andrade EC, Manhani GG, Alfenas PF, Calegario RF, Fontes EPB, Zerbini FM: ***Tomato yellow spot***
399 ***virus*, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida***
400 **sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*.** *J Gen Virol*
401 2006, **87:**3687-3696.
402 31. Inoue-Nagata AK, Martin DP, Boiteux LS, Giordano LD, Bezerra IC, de Avila AC: **New species**
403 **emergence via recombination among isolates of the Brazilian tomato infecting Begomovirus**
404 **complex.** *Pesquisa Agropecuária Brasileira* 2006, **41:**1329-1332.
405 32. Ribeiro SG, Martin DP, Lacorte C, Simões IC, Orlandini DRS, Inoue-Nagata AK: **Molecular and**
406 **biological characterization of *Tomato chlorotic mottle virus* suggests that recombination underlies**
407 **the evolution and diversity of Brazilian tomato begomoviruses.** *Phytopathology* 2007, **97:**702-711.
408 33. Duffy S, Holmes EC: **Phylogenetic evidence for rapid rates of molecular evolution in the single-**
409 **stranded DNA begomovirus *Tomato yellow leaf curl virus*.** *Journal of Virology* 2008, **82:**957-965.
410 34. Duffy S, Holmes EC: **Validation of high rates of nucleotide substitution in geminiviruses:**
411 **Phylogenetic evidence from East African cassava mosaic viruses.** *J Gen Virol* 2009, **90:**1539-1547.
412 35. Varsani A, Shepherd DN, Dent K, Monjane AL, Rybicki EP, Martin DP: **A highly divergent South**
413 **African geminivirus species illuminates the ancient evolutionary history of this family.** *Virology*
414 *Journal* 2009, **6:**36.
415 36. Doyle JJ, Doyle JL: **A rapid DNA isolation procedure for small amounts of fresh leaf tissue.**
416 *Phytochemical Bulletin* 1987, **19:**11-15.

37. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T: **A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase.** *Journal of Virological Methods* 2004, **116:**209-211.

38. Aragão FJL, Barros LMG, Brasileiro ACM, Ribeiro SG, Smith FD, Sanford JC, Faria JC, Rech EL: **Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris L.*) co-transformed via particle bombardment.** *Theoretical and Applied Genetics* 1996, **93:**142-150.

39. Dellaporta SL, Wood J, Hicks JB: **A plant DNA minipreparation: Version II.** *Plant Molecular Biology Reporter* 1983, **1:**19-21.

40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of Molecular Biology* 1990, **215:**403-410.

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

42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Molecular Biology and Evolution* 2011**:**doi: 10.1093/molbev/msr1121. .

43. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.

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

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

46. Martin D, Rybicki EP: **RDP: detection of recombination amongst aligned sequences.** *Bioinformatics* 2000, **16:**562-563.

47. Padidam M, Sawyer S, Fauquet CM: **Possible emergence of new geminiviruses by frequent recombination.** *Virology* 1999, **265:**218-224.

48. Martin DP, Posada D, Crandall KA, Williamson C: **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.** *AIDS Res Hum Retroviruses* 2005, **21:**98-102.

49. Smith JM: **Analyzing the mosaic structure of genes.** *J Mol Evol* 1992, **34:**126-129.

50. Posada D, Crandall KA: **Evaluation of methods for detecting recombination from DNA sequences: computer simulations.** *Proc Natl Acad Sci U S A* 2001, **98:**13757-13762.

51. Gibbs MJ, Armstrong JS, Gibbs AJ: **Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences.** *Bioinformatics* 2000, **16:**573-582.

52. Boni MF, Posada D, Feldman MW: **An exact nonparametric method for inferring mosaic structure in sequence triplets.** *Genetics* 2007, **176:**1035-1047.

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

54. Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini FM, Zhou X: **Geminivirus strain demarcation and nomenclature.** *Archives of Virology* 2008, **153:**783-821.

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

56. Roberts S, Stanley J: **Lethal mutations within the conserved stem-loop of African cassava mosaic virus DNA are rapidly corrected by genomic recombination.** *J Gen Virol* 1994, **75:**3203-3209.

57. Gronenborn B: **Nanoviruses: genome organisation and protein function.** *Veterinary Microbiology* 2004, **98:**103-109.

58. Laufs J, Schumacher S, Geisler N, Jupin I, Gronenborn B: **Identification of the nicking tyrosine of geminivirus Rep protein.** *FEBS Lett* 1995, **377:**258-262.

59. Orozco BM, Hanley-Bowdoin L: **A DNA structure is required for geminivirus replication origin function.** *J Virol* 1996, **70:**148-158.

60. Arguello-Astorga GR, 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.** *Arch Virol* 2001, **146:**1465-1485.

61. Ilyina TV, Koonin EV: **Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria.** *Nucleic Acids Res* 1992, **20:**3279-3285.

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

63. Haible D, Kober S, Jeske H: **Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses.** *Journal of Virological Methods* 2006, **135:**9-16.

476 64. Cotrim MA, Krause-Sakate R, Narita N, Zerbini FM, Pavan MA: **Diversidade genética de**
477 **begomovírus em cultivos de tomateiro no Centro-Oeste Paulista.** *Summa Phytopathologica* 2007,
478 **33:**300-303.
479
480

Figure legends

Figure 1. (**A**) Symptoms of viral infection in a 50-year old *Malvaviscus arboreus* plant and (**B**) in a *N. benthamiana* plant biolistically inoculated with the RCA-amplified DNA from the *M. arboreus* plant. The leaf at the upper left is from a non-inoculated plant (negative control). The electrophoretic profile of the RCA-amplified DNA from *M. arboreus* and *N. benthamiana* plants digested with *Msp* I confirms infection by a single bipartite begomovirus. (**C**) Confirmation of viral infection of five *N. benthamiana* plants biolistically inoculated with the RCA-amplified DNA from *M. arboreus* (a *N. benthamiana* plant biolistically inoculated with *Tomato yellow spot virus* (ToYSV) was used as a positive control). (**D**) Genome organization of *Malvaviscus* yellow mosaic virus (MalYMV, BR:Ca01:05 isolate), showing the positions of virion sense ORFs (*cp* and *nsp*) and complementary sense ORFs (*ren*, *trap*, *rep*, *ac4* and *mp*).

Figure 2. Alignment of the sequence at the origin of replication of the MalYMV isolates and other bipartite and monopartite begomoviruses, indicating the hairpin motif (shaded in blue) and the atypical nonanucleotide (TAGTATTAC, shaded in light green). (**B**) Secondary structure of a typical begomovirus stem-loop structure (left) and the modified stem-loop of MalYMV (right). The ΔG values for each structure are indicated.

Figure 3. Bayesian 50% majority rule consensus tree based on the full-length DNA-A sequences of MalYMV (BR:Ca01:05 isolate) plus begomoviruses from the New and Old Worlds. Phylogenetic reconstruction was performed using MrBayes 3.1.2. [43] under the evolutionary model GTR+I+G and rooted on the curtovirus *Beet severe curly top virus* (BSCTV). Support for the nodes is presented as Bayesian posterior probability.

26

27 **Supplementary Figure S1.** **(A)** Percent sequence identities between the full-length DNA-A
28 (above of diagonal) and DNA-B (below of diagonal) of the BR:Ca01:05 isolate and the most
29 related begomoviruses from the GenBank database. **(B)** Percent sequence identities between
30 the nucleotides (above of diagonal) and deduced amino acid sequences (below of diagonal) of
31 the *cp* ORF, **(C)** *rep*, **(D)** *trap*, **(E)** *ren*, **(F)** *ac4*, **(G)** *mp*, and **(H)** *nsp*.
32

33

34

Figure 1

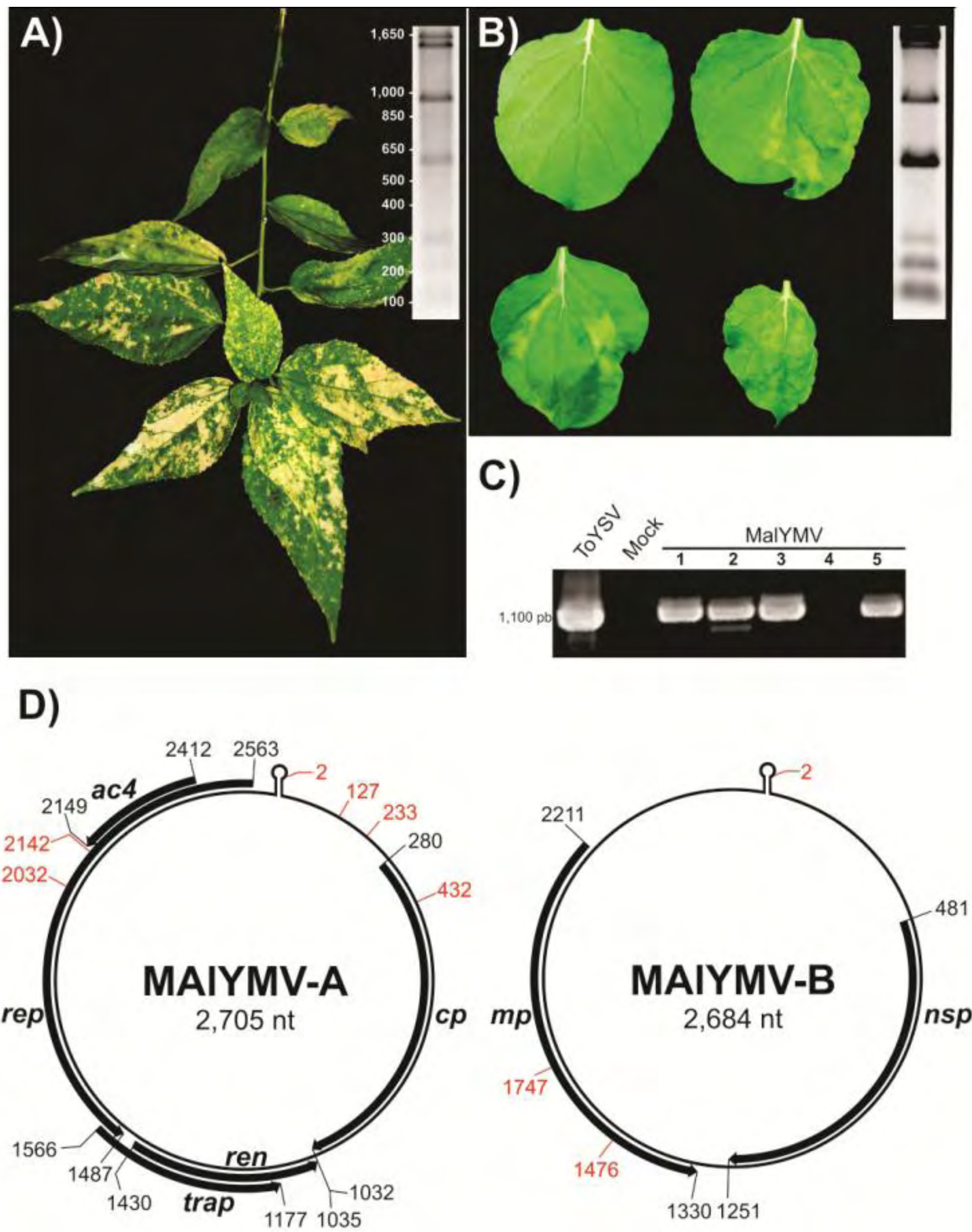


Figure 2

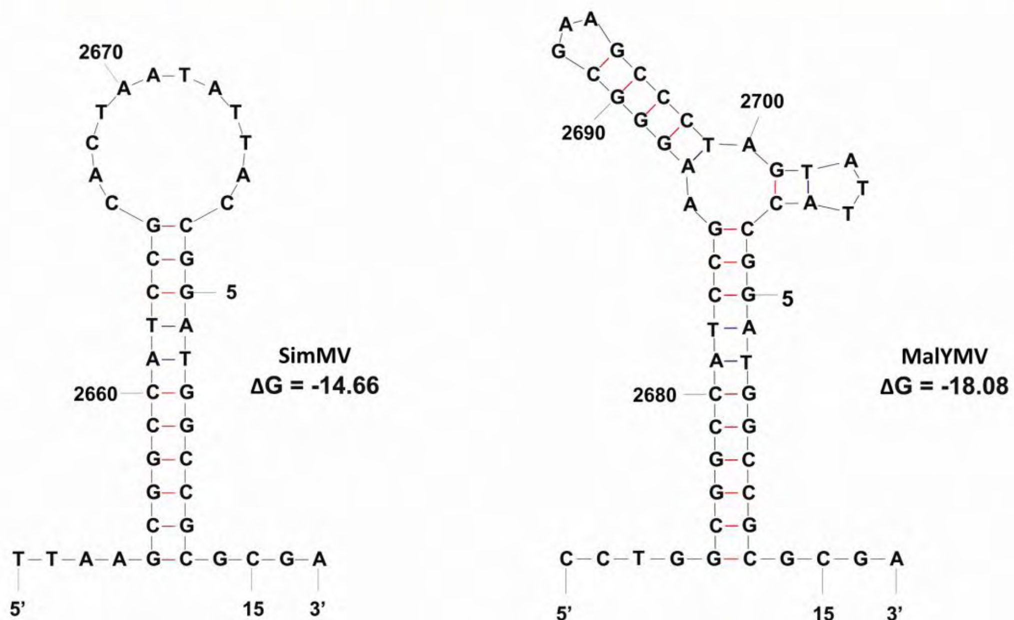
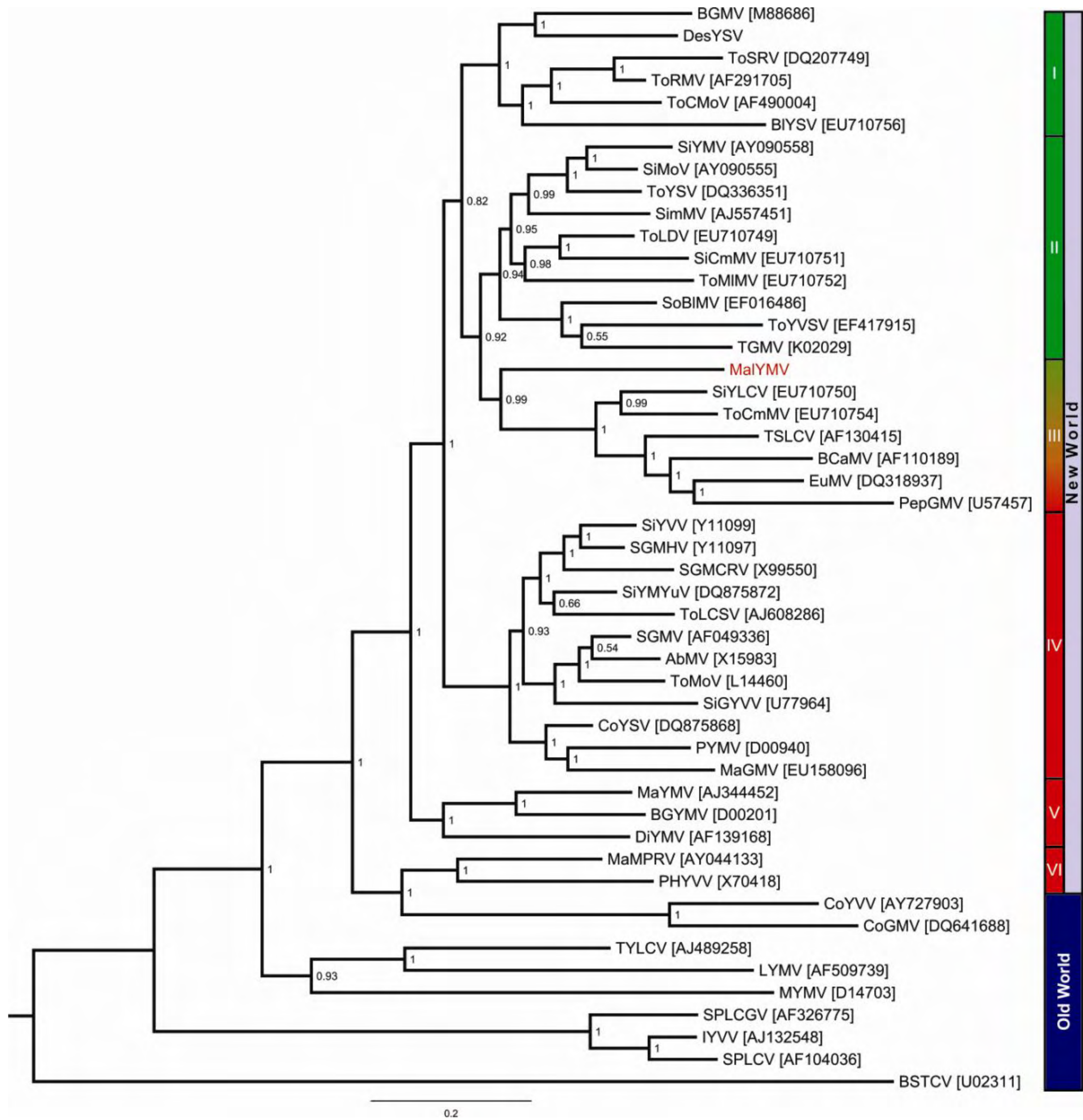
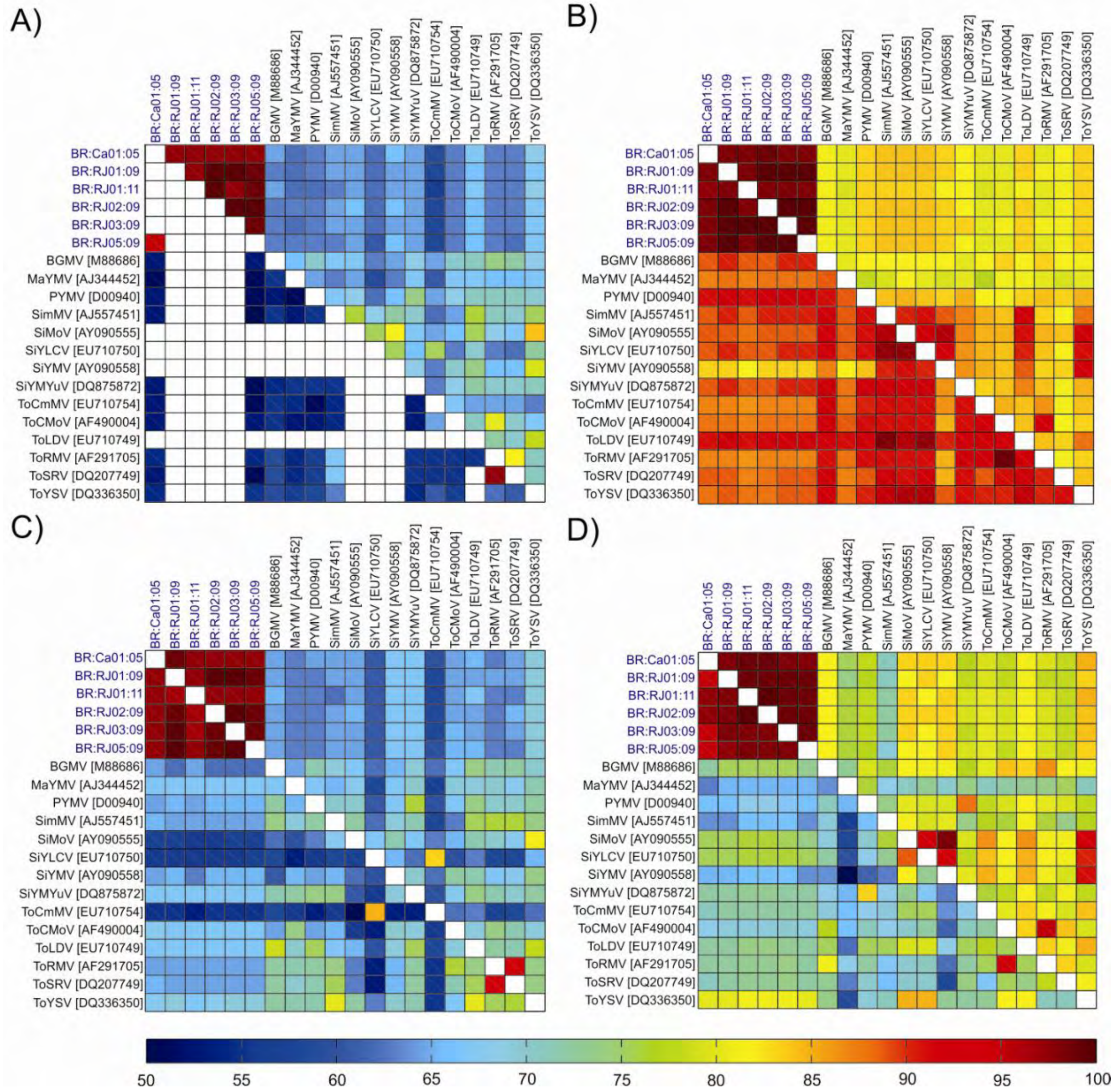


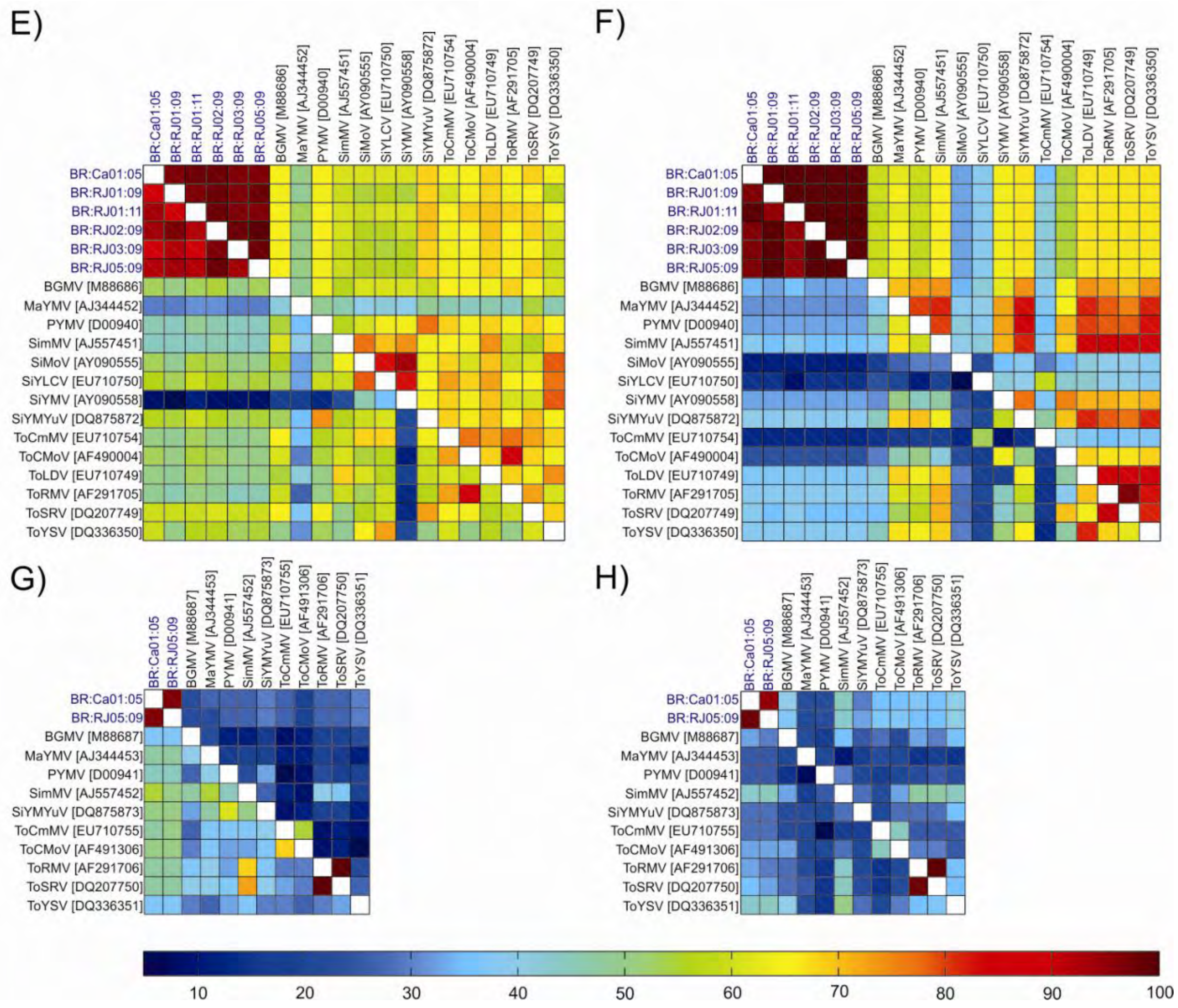
Figure 3



Supplementary Figure 1



Supplementary Figure 1 (cont.)



Supplementary Table S1. Begomovirus sequences used in this work.

Virus	Acronym	GenBank access # (DNA-A)
Abutilon Brazil virus	AbBV	FN434438
Abutilon mosaic virus	AbMV	X15983
Bean calico mosaic virus	BCaMV	AF110189
Bean dwarf mosaic virus	BDMV	M88179
Bean golden mosaic virus	BGMV	FJ665283
	BGMV	M88686
Bean golden yellow mosaic virus	BGYMV	AF173555
	BGYMV	AJ544531
	BGYMV	D00201
	BGYMV	DQ119824
	BGYMV	L01635
	BGYMV	M10070
	BGYMV	M91604
Blainvillea yellow spot virus	BlYSV	EU710756
Beet severe curly top virus	BSCTV	U02311
Cabbage leaf curl virus	CaLCuV	DQ178612
	CaLCuV	U65529
Chino del tomate virus	CdTV	DQ347945
	CdTV	DQ885456
Cotton leaf curl virus	CLCrV	AF480940
	CLCrV	AY083351
	CLCrV	AY742220
Cleome leaf crumple virus	ClLCrV	HM195184
Corchorus yellow spot virus	CoYSV	DQ875868
Corchorus yellow vein virus	CoYVV	AY727903
Corchorus golden mosaci virus	CoGMV	DQ641688
Curcubit leaf crumple virus	CuLCrV	AF224760
	CuLCrV	AF256200
Desmodium leaf distortion virus	DesLDV	DQ875870
Desmodium yellow spot virus	DesYSV	Unpublished
Dicliptera yellow mottle virus	DiYMV	AF139168
Euphorbia mosaic virus	EuMV	AF068642
	EuMV	DQ318937
	EuMV	DQ395342
Euphorbia yellow mosaic virus	EuYMV	JF756669
	EuYMV	JF756670
	EuYMV	JF756671
	EuYMV	JF756672
	EuYMV	JF756673
Ipomoea yellow vein virus	IYVV	AJ132548
Loofa yellow mosaic virus	LYMV	AF509739
Macroptillium golden mosaic virus	MaGMV	EF645647
	MaGMV	EU158096
Macroptillium mosaic Puerto Rico virus	MaMPRV	AF449192
	MaMPRV	AY044133
Macroptilium yellow mosaic Florida	MaYMFV	AY044135

virus		
Macroptilium yellow mosaic virus	MaYMV	AJ344452
Melon chlorotic leaf curl virus	MCLCuV	AF325497
Merremia mosaic virus	MeMV	DQ644557
Mungbean yellow mosaic virus	MYMV	D14703
Okra mottle virus	OMoV	EU914817
	OMoV	FJ686695
Okra yellow mosaic Mexico virus	OYMMV	DQ022611
	OYMMV	GU990614
	OYMMV	HM35059
	OYMMV	HQ020409
	OYMMV	HQ116414
Okra yellow mottle Iguala virus	OYMoIV	AY751753
Passionfruit severe leaf distortion virus	PSLDV	FJ972767
Pepper goden mosaic virus	PepGMV	U57457
Pepper huasteco yellow vein virus	PHYVV	X70418
Potato yellow mosaic Panama virus	PYMPV	Y15034
Potato yellow mosaic Trinidad virus	PYMTV	AF039031
Potato yellow mosaic virus	PYMV	AY120882
	PYMV	AY965897
	PYMV	D00940
Rhyncosia golden mosaic Sinaloa virus	RhGMSV	DQ406672
Rhynchosia golden mosaic virus	RhGMV	AF239671
	RhGMV	AF408199
Rhynchosia rugose golden mosaic virus	RhRGMV	HM236370
Sida golden mosaic Costa Rica virus	SGMCRV	X99550
Sida golden mosaic Honduras virus	SGMHV	Y11097
Sida golden mosaic virus	SGMV	AF049336
Sida common mosaic virus	SiCmMV	EU710751
Sida golden yellow vein virus	SiGYVV	AJ577395
	SiGYVV	U77964
Sida mosaic Brazil virus	SimBV	FN436001
Sida micrantha mosaic virus	SimMV	AJ557451
	SimMV	EU908733
	SimMV	FJ686693
	SimMV	FN436003
	SimMV	HM585431
	SimMV	HM585433
	SimMV	HM585439
Sida mottle virus	SiMoV	AY090555
Sida yellow leaf curl virus	SiYLCV	EU710750
Sida yellow mosaic virus	SiYMV	AY090558
Sida yellow mosaic Yucatan virus	SiYMYuV	DQ875872
Sida yellow vein virus	SiYVV	Y11099
Soybean blistering mosaic virus	SoBlMV	EF016486
Squash leaf curl virus	SqLCV	AF256203
	SqLCV	DQ285016
	SqLCV	M38183
Sweet potato leaf curl Georgia virus	SPLCGV	AF326775
Sweet potato leaf curl virus	SPLCV	AF104036

Tomato golden mosaic virus	TGMV	K02029
Tomato Chino La Paz virus	ToChLPV	AY339618
	ToChLPV	AY339619
	ToChLPV	DQ347948
	ToChLPV	DQ347949
	ToChLPV	HM459852
Tomato common mosaic virus	ToCmMV	EU710754
Tomato chlorotic mottle virus	ToCMoV	AF490004
	ToCMoV	AY090557
Tomato golden motlle virus	ToGMoV	AF132852
	ToGMoV	DQ520943
	ToGMoV	EF501976
Tobacco leaf curl Cuba virus	ToLCCUV	AM050143
Tomato leaf distortion virus	ToLDV	EU710749
Tomato mosaic Havana virus	ToMHV	EF088197
	ToMHV	Y14874
Tomato mild mosaic virus	ToMlMV	EU710752
Tomato mottle Taino virus	ToMoTV	AF012300
Tomato mottle virus	ToMoV	AY965900
	ToMoV	L14460
Tomato mild yellow leaf curl Aragua virus	ToMYLCAV	AY927277
	ToMYLCAV	NC009490
Tomato rugose mosaic virus	ToRMV	AF291705
Tomato severe leaf curl virus	ToSLCV	AF130415
	ToSLCV	AJ508784
	ToSLCV	AJ508785
	ToSLCV	DQ347946
	ToSLCV	DQ347947
Tomato severe rugose virus	ToSRV	DQ207749
Tomato yellow distortion leaf virus	ToYLDV	FJ174698
Tomato yellow leaf curl virus	TYLCV	AJ89258
Tomato yellow spot virus	ToYSV	DQ336350
	ToYSV	FJ538207
Tomato yellow vein streak virus	ToYVSV	EF417915
	ToYVSV	GQ387369
Tobacco yellow crinkle virus	TYCV	FJ213931
	TYCV	FJ222587
Tomato yellow margin leaf curl virus	TYMLCV	AY508993
Wissadula golden mosaic virus	WGMV	GQ355488

CHAPTER 2

SYNONYMOUS SITE VARIATION DUE TO RECOMBINATION EXPLAINS HIGHER GENETIC VARIABILITY IN BEGOMOVIRUS POPULATIONS INFECTING NON-CULTIVATED HOSTS

Lima, A.T.M., Ramos-Sobrinho, R., González-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, *submitted*.

Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts

Alison T.M. Lima[a], Roberto R. Sobrinho[a], Jorge González-Aguilera[a1], Carolina S. Rocha[a2], Sarah J.C. Silva[a3], César A.D. Xavier[a], Fábio N. Silva[a], Siobain Duffy[b#] and F. Murilo Zerbini[a#]

[a]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-000, Brazil

[b]Dep. 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

[1]Present address: Embrapa Trigo, Rodovia BR 285 km 294, CP 451, Passo Fundo, RS, 99001-970, Brazil

[2]Present address: Embrapa Soja, CP 231, Londrina, PR, 86001-970, Brazil

[3]Present address: Dep. de Fitossanidade/CECA, Universidade Federal de Alagoas, Rio Largo, AL, 57100-000, Brazil

Content category: *Plant viruses - DNA*

Running title: *Variability in weed begomoviruses due to recombination*

Words in summary: *191*

Words in main text: *5,377*

Summary

Begomoviruses are single-stranded DNA plant viruses which cause serious epidemics in economically important crops worldwide. Non-cultivated plants also harbor many begomoviruses, and it is believed that these hosts may act as reservoirs and as mixing vessels where recombination may occur. Begomoviruses are notoriously recombination-prone, and also display nucleotide substitution rates equivalent to those of RNA viruses. In Brazil, several indigenous begomoviruses have been described infecting tomatoes following the introduction of a new biotype of the whitefly vector in the mid-1990's. More recently, a number of viruses from non-cultivated hosts have also been described. Previous work has suggested that viruses infecting non-cultivated hosts have a higher degree of genetic variability compared to crop-infecting viruses. We intensively sampled cultivated and non-cultivated plants in similarly sized geographic areas known to harbor either the weed-infecting *Macroptilum yellow spot virus* (MaYSV) or the crop-infecting *Tomato severe rugose virus* (ToSRV), and compared the molecular evolution and population genetics of these two distantly-related begomoviruses. The results reinforce the assertion that infection of non-cultivated plant species leads to higher levels of standing genetic variability, and indicate that recombination, not adaptive selection, explains the higher begomovirus variability in non-cultivated hosts.

21 **Introduction**

22 Single-stranded DNA begomoviruses (whitefly-transmitted members of the
23 *Geminiviridae*) have become an important factor limiting crop production in tropical and
24 subtropical regions (Morales & Anderson, 2001; Rojas *et al.*, 2005; Seal *et al.*, 2006).
25 Begomoviruses causing cassava mosaic disease (CMD) are the major biotic constraint to
26 cassava cultivation in Africa (Legg & Fauquet, 2004; Legg & Thresh, 2000; Ndunguru *et al.*,
27 2005; Were *et al.*, 2004). A complex of at least six different begomovirus species is
28 responsible for the devastating tomato yellow leaf curl disease (TYLCD) (Moriones &
29 Navas-Castillo, 2000). In the Americas, diseases caused by begomoviruses have significantly
30 impacted tomato and bean production since the 1980's (Blair *et al.*, 1995; Brown & Bird,
31 1992; Gilbertson *et al.*, 1993; Morales & Jones, 2004; Polston & Anderson, 1997).

32 Previous studies have indicated a high intra- and interspecific diversity of
33 begomoviruses, which can facilitate adaptation to new climates and novel hosts (Monci *et al.*,
34 2002). Recombination is the most intensively studied population genetic process in
35 begomoviruses, and has been considered more significant than mutation by many researchers
36 (Lefeuvre *et al.*, 2007a; Lefeuvre *et al.*, 2009; Martin *et al.*, 2011; Martin *et al.*, 2005; Monci
37 *et al.*, 2002; Padidam *et al.*, 1999; Pita *et al.*, 2001). It appears to heavily contribute to
38 begomovirus genetic diversity, increasing the evolutionary potential and local adaptation of
39 strains (Berrie *et al.*, 2001; Graham *et al.*, 2010; Harrison & Robinson, 1999; Monci *et al.*,
40 2002; Padidam *et al.*, 1999). There is ample opportunity for recombination because multiple
41 begomovirus species are often found coinfecting the same plant (Davino *et al.*, 2009; García-
42 Andrés *et al.*, 2006; Harrison *et al.*, 1997; Pita *et al.*, 2001; Ribeiro *et al.*, 2003; Sanz *et al.*,
43 2000; Torres-Pacheco *et al.*, 1996), and more than one virus can simultaneously replicate in
44 the same nucleus (Morilla *et al.*, 2004). Additionally, the high recombination frequency
45 observed for begomoviruses may be explained by a theoretical recombination-dependent

46 replication mechanism (RDR) (Jeske *et al.*, 2001), in addition to the well-documented rolling
47 circle replication (RCR) (Saunders *et al.*, 2001). However, recent studies have also indicated
48 that begomoviruses can evolve by mutation alone as quickly as RNA viruses (Duffy &
49 Holmes, 2008; Duffy & Holmes, 2009), and positive selection – on mutations or the products
50 of recombination events – may also play a role in begomovirus evolutionary dynamics
51 (Monci *et al.*, 2002; Pita *et al.*, 2001; Zhou *et al.*, 1997).

52 Additionally, host use may play an important role in the standing genetic diversity of
53 begomovirus populations (Seal *et al.*, 2006). Several species of non-cultivated plants,
54 especially of the families Malvaceae, Euphorbiaceae, Fabaceae and Solanaceae, are known
55 hosts of begomoviruses (Morales & Anderson, 2001). These weed/wild hosts can serve as
56 reservoirs for infection of nearby crops (Alabi *et al.*, 2008; Barbosa *et al.*, 2009; Bedford *et*
57 *al.*, 1998; García-Andrés *et al.*, 2006), as overwintering refugia (Alabi *et al.*, 2007; Alabi *et*
58 *al.*, 2008; García-Andrés *et al.*, 2006) and as "mixing vessels" for interspecific coinfection
59 and recombination (García-Andrés *et al.*, 2006; Monde *et al.*, 2010; Silva *et al.*, 2012).
60 Increased host use and diminished bottlenecks would both potentially increase the effective
61 population size of begomovirus populations (Power, 2000; Seal *et al.*, 2006). Although there
62 is limited data on the variability of begomovirus populations in non-cultivated hosts, such
63 data suggest that it is higher that that observed in crop-infecting begomoviruses (Fiallo-Olive
64 *et al.*, 2012; Silva *et al.*, 2012; Silva *et al.*, 2011; Wyant *et al.*, 2011).

65 To gather data on the factors affecting genetic variability in begomovirus populations
66 and shed light on whether frequent infection of non-cultivated plants alters viral evolutionary
67 dynamics, we contrasted two populations of distantly related begomoviruses. We intensively
68 sampled crops and non-cultivated plants in similarly sized geographic areas known to harbor
69 either *Macroptilum yellow spot virus* or *Tomato severe rugose virus* (ToSRV). MaYSV is a
70 recently isolated species that was previously only reported in non-cultivated hosts in

71 northeastern Brazil (Silva *et al.*, 2012), whereas ToSRV is the most widespread tomato-
72 infecting begomovirus in Brazil (Fernandes *et al.*, 2008; Rocha, 2011; Zerbini *et al.*, 2005).
73 ToSRV can naturally infect other important crops such as chili pepper (Bezerra-Agasie *et al.*,
74 2006) and potato (Souza-Dias *et al.*, 2008) and is only rarely found in non-cultivated plants
75 such as *Nicandra physaloides* (family Solanaceae) (Barbosa *et al.*, 2009). Over a three year
76 period we obtained more than 50 full-length DNA-A sequences of each of these viruses
77 isolated from a mixture of crops and non-cultivated plants. We compared the molecular
78 evolution and population genetics of the mostly weed-infecting MaYSV to the predominately
79 tomato-infecting ToSRV. Our results bolster the assertion that infection of indigenous and
80 non-cultivated plant species leads to higher levels of standing genetic variability, apparently
81 driven by higher levels of detectable recombination.
82
83 **Results**
84 *Natural infection of the crop plant P. vulgaris by MaYSV, and of the wild host Sida spp. by*
85 *ToSRV*
86 MaYSV was originally described infecting leguminous non-cultivated hosts (*M. lathyroides,*
87 *Calopogonium mucunoides* and *Canavalia* sp.) (Silva *et al.*, 2012). However, in 2011
88 MaYSV was readily isolated from both *M. lathyroides* and a crop plant, *P. vulgaris* (common
89 bean), in Alagoas. Forty-four DNA-A components were cloned from these samples (Suppl.
90 Table S1). Pairwise comparisons of the DNA-A sequences (*data not shown*) revealed that
91 only one begomovirus was present, with 90-99% identity to MaYSV. Therefore, based on the
92 criteria established by the *Geminiviridae* Study Group of the ICTV (Brown *et al.*, 2012) these
93 genomic components represent additional isolates of this species. MaYSV appears to be the
94 prevalent begomovirus infecting common bean in this portion of Alagoas, since *Bean golden*
95 *mosaic virus* (BGMV) was not detected in *P. vulgaris* in our study. Overall, 56 whole

96 genomes were included in the MaYSV dataset, with 23 (41%) isolated from non-cultivated
97 hosts.
98 Despite sampling many non-cultivated species neighboring infected tomato plants, we
99 were only able to find ToSRV in two samples of the malvaceous non-cultivated host, *Sida* sp.
100 Therefore, 53 out of 55 whole ToSRV genomes were isolated from tomatoes, and only two
101 from non-cultivated hosts.
102
103 *The MaYSV population is more variable than the ToSRV population*
104 Although the ToSRV and MaYSV populations were sampled in similarly sized
105 geographical areas, they differed profoundly in their levels of standing genetic variability.
106 While the average pairwise number of nucleotide differences (π) for the full-length DNA-A
107 of the ToSRV population was 0.0084 (Table 1), this same statistic was about 8-fold higher
108 for the MaYSV population (π = 0.0658). Interestingly, the variability within the MaYSV
109 population was not evenly distributed throughout the genome. The MaYSV Rep gene was 3.8
110 times more variable than its CP gene, and more than 10-fold higher than the ToSRV Rep
111 (Table 1). Within the MaYSV Rep, the N-terminal half was much more variable than the C-
112 terminal half – again, a distribution of variation not shared by the ToSRV Rep (Fig. 1). The
113 nucleotide diversity of the ToSRV CP gene was similar to that of MaYSV.
114 We also compared the variability between MaYSV isolates sampled from crops and
115 non-cultivated hosts. Each MaYSV subset was still markedly more variable than the whole
116 ToSRV population, with isolates sampled from *P. vulgaris* and non-cultivated hosts showing
117 similar levels of standing genetic variability (π = 0.0626, S.D. = 0.003 and π = 0.0670, S.D. =
118 0.0043; respectively).
119
120

121 *Phylogenetic analysis*

122 The ToSRV CP and Rep ML phylogenetic trees are not congruent (Fig. 2), but the
123 topological differences between them were due to a strongly supported recombination event
124 in the CP gene of ToSRV-[BR:Vic20:10] (p=2.18×10^{-10}, Table 2). Without this isolate, the
125 clustering of isolates in the CP and Rep trees are identical (Suppl. Fig. S1), mimicking the
126 clustering found in the Rep tree on Fig. 2. Both ML trees showed strong support for isolates
127 from Florestal being a separate population from those from Carandaí, Jaíba and Viçosa (Fig.
128 2). There is rough clustering of isolates from each sampling site together, with Jaíba being
129 nested within isolates from Viçosa. However, the lack of bootstrap support for geographic
130 structure among these three sites (and the weakly supported clustering of Vic06 and Vic07
131 with isolates from Carandaí) suggests that migration may occur between these
132 subpopulations. Consistent with our phylogenetic analyses, Wright's fixation index F, based
133 on the ToSRV CP dataset indicated genetic differentiation amongst isolates sampled from
134 different geographic locations (F_{ST} = 0.51066).

135 The MaYSV CP and Rep trees were highly incongruent (Fig. 3). The MaYSV Rep
136 tree shows four clades with significant genetic distance between them (two with 100%
137 bootstrap support), and the CP shows fewer well-supported clades and less genetic variability
138 among isolates. Both trees showed little evidence of geographic structuring. In fact, Wright's
139 fixation index F based on the MaYSV CP dataset indicated less evidence of genetic
140 differentiation (F_{ST} = 0.11428) than the ToSRV population.

141

142 *MaYSV isolates have a mixture of different recombinant patterns*

143 While ToSRV showed little evidence that recombination has significantly contributed
144 to its evolution (Table 2), MaYSV showed both inter- and intra-specific recombination
145 events. Corroborating the conflicting MaYSV CP and Rep phylogenies, we identified a total

146 of six unique potential recombination events in the MaYSV population. Most events involved
147 breakpoints located inside the Rep gene and in the common region (events 1, 2, 3, 4 and 6;
148 Table 2), though one event had breakpoints within the CP gene (event 5; Table 2). Isolates
149 that showed evidence of a shared recombination event in the Rep sequence were readily
150 identified in well supported clades observed on the Rep ML phylogenetic tree (Fig. 3). There
151 were no strong geographic patterns associated with recombination events. For instance, four
152 out of five recombination events detected in MaYSV Rep (events 1, 2, 4 and 6) were detected
153 in isolates sampled in Olho d'Água das Flores, but two of these events (events 1 and 6) were
154 observed elsewhere.

155

156 *Adaptive selection does not explain the higher begomovirus variability in non-cultivated*
157 *hosts*

158 We investigated the extent that positive and negative selection at the amino acid level
159 had shaped the standing genetic variability in the CP and Rep data sets (composed by all
160 isolates obtained from cultivated and non-cultivated hosts) of both viruses. All data sets
161 showed dN/dS ratios (ω) lower than 1 (Table 3), indicating negative selection. However, the
162 wide variation of the values indicated that each gene/population might be under different
163 selective constraints. The ω value for the ToSRV CP dataset (0.446493) was considerably
164 higher than that observed for the MaYSV CP (0.0514088). The high ω of the ToSRV CP was
165 observed even with the elimination of the single recombinant isolate, Vic20 (ω = 0.27438).
166 The ω value for the ToSRV Rep (0.268859) was only slightly higher than that observed for
167 the MaYSV Rep dataset (0.208195).

168 Although Gard has been unable to detect the well-supported recombination event in
169 the ToSRV CP dataset, no sites were under statistically significant negative selection,
170 including or excluding Vic20 (Fig. 4 and Suppl. Fig. S2). Excluding the recombinant Vic20

171 from the data set, REL detected 20 sites under positive selection. However, the evidence for
172 selection on individual sites is weak, as no positively selected sites were detected using the
173 SLAC or PARRIS methods. Two negatively selected sites were detected by the SLAC
174 method in the ToSRV Rep dataset (Suppl. Fig. S2). In addition, nine positively selected sites
175 were detected by REL (Suppl. Fig. S1), with six of them located between motif III in the
176 catalytic domain and motif Walker A in the ATPase domain. PARRIS did not identify any
177 sites under positive selection.

178 Recombinant partitions were readily detected by Gard in the MaYSV CP and Rep
179 datasets. Despite, a higher number of sites were under detectable selection in both MaYSV
180 genes. In the CP data set, all codons were identified as being under negative selection by
181 REL. In the MaYSV Rep, 106 sites were shown to be under negative selection by SLAC or
182 REL, and 12 sites were under positive selection by REL. Curiously, most sites in the highly
183 variable Rep N-terminal showed high synonymous substitutions rates, although without
184 statistical evidence of negative selection (Fig. 4). Though GARD was able to detect
185 recombination events in this data set, none of the positively selected sites correlated with the
186 host from which the isolates were obtained (non-cultivated *vs*. cultivated), but were clearly
187 related to a given recombination event indicating that they most probably represent a spurious
188 selection signal due to recombination. Results obtained by PARRIS confirmed the absence of
189 sites under positive selection.

190 Together, our results indicate that adaptive selection at the amino acid level is not
191 driving the higher variability in the MaYSV population.

192

193 *Relative contribution of mutation and recombination to the genetic variability of ToSRV and*
194 *MaYSV populations*

195 We mapped all substitutions over the branches in our midpoint rooted ML trees, and
196 determined which branches were associated with well-supported recombination events. As
197 there were no detectable recombination events in the ToSRV Rep dataset, all of the
198 substitutions along the phylogeny are presumably due to mutation. Out of a total of 99
199 substitutions over the ToSRV CP phylogeny, the single recombination event detected
200 (spanning 50% of the CP gene) encompasses 42 substitutions. Therefore, 42.4% of the
201 standing genetic variability in the ToSRV CP could be attributed to recombination. In the
202 MaYSV CP tree, there were a total of 313 substitutions, and 49 substitutions were located
203 inside a putatively recombinant region (spanning about 75% of the CP gene) on a single
204 branch on the tree. In this considerably more variable dataset, recombination accounted for
205 only 16% of the substitutions.

206 As expected, a considerably higher number of substitutions was observed for the
207 MaYSV Rep ML tree: 553 substitutions. The unique recombination events 2, 4 and 6 were
208 readily assigned to the long branches (the brown, red and yellow branches, respectively, in
209 Fig. 3b), with another long branch leading to the clade containing events 4 and 6 (the orange
210 branch; Fig. 3b). For events 2 (shared by isolates Oaf6, Oaf9 and Oaf19) and 4 (shared by all
211 isolates in clade II), 22 and 36 substitutions were counted as associated to them, respectively.
212 Event 6 (shared by all 23 isolates in clade I) accounted for the largest individual contribution
213 amongst the unique recombination events: 37 substitutions. Additionally, 16 changes on the
214 branch leading to the clades associated with events 4 and 6 were either in the region of
215 overlap between events 4 and 6, or within the larger event 6, and were also assigned to η_r.
216 Event 3 (assigned to the purple terminal branch) was exclusively detected in the Inp1 isolate
217 and a total of 12 substitutions were counted in association with this event.

218 Although event 1 was well supported by the methods of analysis contained in RDP,
219 the exact location of the breakpoint was variable and frequently ended in the intergenic

220 region instead of the Rep gene. Only those substitutions in the terminal branches leading to
221 sequences whose end breakpoint of this event occurred inside the Rep gene (isolates Oaf3,
222 Oaf7, Oaf8, Oaf11, Oaf12, Oaf16, Oaf21, Oaf24, Oaf25, Crb10 and Inp1) were added to the
223 η_r. A total of 5 substitutions could be associated with this event.

224 Overall, 201 substitutions were counted as due to recombination over the MaYSV
225 Rep phylogeny, which represented a relative contribution of 36.3% to the standing genetic
226 variability.

227

228 **Discussion**

229 Most studies on the diversity of begomovirus populations have focused on species
230 diversity (Ala-Poikela *et al.*, 2005; Bull *et al.*, 2006; Fernandes *et al.*, 2008; García-Andrés *et*
231 *al.*, 2006; Lefeuvre *et al.*, 2007b; Lozano *et al.*, 2009; Ndunguru *et al.*, 2005; Reddy *et al.*,
232 2005; Ribeiro *et al.*, 2003; Rothenstein *et al.*, 2006; Sserubombwe *et al.*, 2008). In contrast,
233 little is known about the standing genetic variability within species, especially in
234 begomovirus populations in non-cultivated hosts. Here we present a study contrasting the
235 molecular variability of two begomovirus populations (of the mostly weed-infecting MaYSV
236 to the predominately tomato-infecting ToSRV) and discuss the relative contribution of
237 evolutionary processes on this diversity. Our results, based on 111 viral genome sequences
238 cloned from samples collected over a three-year period, support the hypothesis that although
239 the genetic structure of viral populations is modulated by the common processes of mutation,
240 recombination, selection and genetic drift (García-Arenal *et al.*, 2003; Roossinck, 2003), the
241 relative contribution of each process can vary widely among begomovirus populations.

242 The MaYSV population was notably more variable than the ToSRV population,
243 largely due to a number of recombination events in the Rep gene. Even MaYSV isolates
244 sampled from the crop *P. vulgaris* (collected from only 2 geographic locations: Craibas and

245 Olho d'Água das Flores) were considerably more variable than all ToSRV isolates sampled
246 from tomato plants (collected from 4 different geographic locations). Previous studies have
247 indicated that begomovirus populations infecting non-cultivated hosts are more variable than
248 crop-infecting ones (Rocha, 2011; Silva *et al.*, 2012). The standing genetic variabilities of
249 four tomato-infecting begomovirus populations were much lower than that of a begomovirus
250 population from a non-cultivated host (*Blainvillea yellow spot virus*, BlYSV) sampled in
251 southeastern Brazil (although these results are based on different sample sizes) (Rocha,
252 2011). The genetic structure was also determined for a population of *Cleome leaf crumple*
253 *virus* (ClLCrV) infecting an annual weed (*Cleome affinis*, family Capparaceae) often
254 associated with leguminous crops in Brazil (Silva *et al.*, 2011). Several recombination events
255 and a high molecular variability were estimated for this population.

256 Although an increasing body of evidence points to a higher genetic variability in
257 begomovirus populations infecting non-cultivated hosts, all previous studies have detected
258 whether various evolutionary processes have contributed to the population genetic variability
259 without attempting to determine their relative contribution to the total variability in these
260 populations. Our novel substitution counting method allowed us to assess the relative
261 contribution of recombination and mutation to the standing genetic variability of two
262 begomoviruses – a step towards disambiguating the effects of various evolutionary processes
263 on viral genetic variation.

264

265 *Recombination and the genetic structure of begomovirus populations*

266 Recombination is the most studied evolutionary process acting on geminivirus
267 populations and has greatly contributed to their evolution (Briddon *et al.*, 1996; García-
268 Andrés *et al.*, 2007a; García-Andrés *et al.*, 2007b; Padidam *et al.*, 1999; Pita *et al.*, 2001;
269 Schnippenkoetter *et al.*, 2001). In begomoviruses, the non-random location of recombination

270 breakpoints is conserved amongst mono- and bipartite genomes, with hot spots in the Rep N-
271 terminal portion and in the 5'-end of the common region (Lefeuvre *et al.*, 2007a; Lefeuvre *et*
272 *al.*, 2007b; Prasanna & Rai, 2007). Our analyses of the MaYSV population indicated that
273 most unique recombination events involved the Rep gene, with at least one breakpoint
274 located in the N-terminal portion and/or the 5'-end of the common region. In addition, we
275 also observed that the N-terminal portion accounted for the largest fraction of the total
276 variability in the Rep gene, suggesting that recombination may be the evolutionary process
277 responsible for this uneven distribution of polymorphisms.

278 It has been shown that recombination events that preserve co-evolved intragenome
279 interactions (protein-protein and/or protein-DNA) are favored by selection (Martin *et al.*,
280 2011). In fact, the linkage between the Rep N-terminal portion and the 5' common region was
281 not broken by recombination events detected in the MaYSV population. These portions were
282 exchanged as blocks that preserve them as originating from the same genetic background.
283 This could increase the frequency in which viable recombinants for the Rep sequence are
284 maintained by preserving the interaction between the catalytic domains (in the Rep N-
285 terminal portion) and their cognate iterative elements located in the 5' portion of the common
286 region (Martin *et al.*, 2011). Although a reduced number of breakpoints occur within the CP
287 encoding sequence, most of them are located in the central portion of the gene (Lefeuvre et
288 al., 2007b). In agreement, two unique events (the single event detected in the ToSRV
289 population and the MaYSV event 5) had initial breakpoints located in the central portion of
290 the gene and spanned its entire C-terminal portion.

291

292 *Purifying selection in viral genes*

293 Although previous studies indicate that the main type of selection acting on the CP
294 and Rep genes in begomovirus populations is purifying selection (García-Andrés *et al.*,

295 2007a; Sanz *et al.*, 1999; Silva *et al.*, 2012; Silva *et al.*, 2011), we also assessed the possible
296 contribution of adaptive selection in shaping the standing genetic variability in the ToSRV
297 and MaYSV populations, since a small fraction of codons could be under different selective
298 constraints. Interestingly, the dN/dS ratio estimated for the ToSRV CP was markedly high (ω
299 = 0.446), even after the exclusion of the recombinant isolate Vic20 (ω = 0.274). These values
300 are high even when compared to non-vector-borne ssRNA viruses such as *Prune dwarf virus*
301 (family *Bromoviridae*; ω = 0.222) and similar to *Pepper mild mottle virus* (genus
302 *Tobamovirus*; ω = 0.301) (Chare & Holmes, 2004). Despite the lower molecular variability,
303 REL detected stronger evidence of adaptive selection in the CP (two sites) and especially in
304 the ToSRV Rep (nine sites). Interestingly, out of nine positively selected sites in the ToSRV
305 Rep, six were located between motif III in the catalytic domain and motif Walker A in the
306 ATPase domain. This region is putatively involved in Rep oligomerization and interaction
307 with host factors (Hanley-Bowdoin *et al.*, 2004; Orozco *et al.*, 2000). Although REL is
308 considerably less conservative than SLAC in detecting positively selected sites, it is
309 important to note that the Rep ToSRV data set was free of any detectable evidence of
310 recombination that is known to affect the results of selection analyses.

311 In striking contrast, we did not detect any evidence of adaptive selection acting on
312 both MaYSV genes, indicating that adaptive selection did not contribute to the higher
313 molecular variability in the MaYSV population.

314 Although a higher variability was found in the Rep N-terminal portion (probably due
315 to recombination) it was composed mostly of synonymous substitutions. The Rep catalytic
316 domain is located in the N-terminal portion and includes three motifs which are conserved in
317 proteins involved in rolling circle replication (motif I: FLTY; motif II: HxH; and motif III:
318 YxxxV) (Ilyina & Koonin, 1992). Adjacent to these elements are the DNA-binding
319 specificity determinants involved in high affinity DNA-binding (Londono *et al.*, 2010). The

320 integrity of these elements in terms of amino acid sequence and structure is important for
321 viral replication and therefore suggests that although the nucleotide sequence in the 5' portion
322 of the Rep gene may harbor a high variability, most of these substitutions would tend to
323 preserve the amino acid sequence and consequently protein structure as well.

324

325 *Host preference and recombination*

326 The lower synonymous site variation in ToSRV population suggests that this
327 population may have been recently introduced in this geographical area and/or experienced a
328 recent genetic bottleneck. Low levels of molecular variability were also observed in
329 genetically differentiated subpopulations of begomoviruses causing tomato yellow leaf curl
330 disease (TYLCD) in Spain and Italy. This scenario was consistent with the foundation of
331 these subpopulations by few variants of limited genetic variability (García-Andrés et al.,
332 2007a). It is possible that the lack of alternative perennial hosts that would maintain high
333 ToSRV effective population sizes between cultivation periods results in a strong reduction of
334 variability. MaYSV is able to efficiently infect both P. vulgaris and perennial leguminous
335 non-cultivated hosts. These latter hosts could represent an important epidemiological
336 component for the MaYSV population for allowing that genetic variability is maintained
337 between cultivation periods. Furthermore, by infecting new hosts theses populations may be
338 subject to environments inaccessible to viruses that have a narrow host range (Power, 2000).
339 As a consequence, these viruses could experience high frequencies of interspecific
340 recombination resulting in a rapid evolution of these populations. Studies attempting to assess
341 the genetic variability of begomovirus populations on Solanum nigrum (an indigenous weed
342 in Spain) showed that this host is an efficient reservoir of all previously characterized
343 begomoviruses species and strains causing tomato yellow leaf curl disease (TYLCD) (García-
344 Andrés et al., 2006). Additionally, an uncharacterized recombinant isolate was detected in

345 this host. In fact, several new species of begomoviruses causing TYLCD are actually
346 interspecific recombinants (Monci et al., 2002; Navas-Castillo et al., 2000). Begomoviruses
347 causing cassava mosaic disease (CMD) are the major biotic constraint to cassava cultivation
348 in Africa (Legg & Fauquet, 2004; Legg & Thresh, 2000; Ndunguru et al., 2005; Were et al.,
349 2004), and multiple natural reservoirs for begomoviruses causing CMD have been described
350 (Alabi et al., 2007; Alabi et al., 2008). Associated with the ability to infect many alternative
351 hosts, these viruses appear to evolve largely by intra- and interspecific recombination (Berrie
352 et al., 2001; Fondong et al., 2000; Pita et al., 2001; Tiendrebeogo et al., 2012; Zhou et al.,
353 1997).
354
355 *Mutation accounts for most begomovirus genetic variability*
356 Our results suggest that recombination explains the higher molecular variability of the
357 MaYSV population compared to the ToSRV population. However, both populations were
358 dominated by mutational diversification. By mapping the recombination events onto ML
359 trees, we were able to distinguish the individual contributions associated with these two
360 mechanisms that create variability. Then, we took advantage of the additive character of η to
361 express the individual contributions of these process as fractions of the total variability (η_{total}
362 $= \eta_r + \eta_\mu$). This simple statistic is not without its flaws, chiefly that recombination is known
363 to reduce the accuracy of the phylogenetic trees upon which this method relies (Posada &
364 Crandall, 2002). However, phylogeny-independent measures of population variability, such
365 as pairwise comparisons, are non-additive, which complicates partitioning variability due to
366 recombination and mutation.
367 Some of the variation that our method attributes to mutation is undoubtedly due to
368 recombination that was not statistically detectable. Therefore, our estimates should be
369 interpreted as the minimal relative contribution of recombination. Furthermore, while our

370 method attributes 100% of changes on specified branches within indentified recombination
371 breakpoints to recombination, this is not a liberal assumption. The reconstructed sequence of
372 the recombinant region that finds each recombinant clade may not be the correct ancestral
373 state, but it is unlikely to contain dramatically more substitutions than the actual first
374 recombinant. All subsequent changes that occur in that region on branches to descendent
375 nodes are counted as mutations.

376 Using this novel partitioning, we assessed recombination as the source of up to 50%
377 of the variation in begomovirus populations. Importantly, our results indicate that its relative
378 contribution to the total variability does not necessarily correlate with the number of
379 detectable recombination events: the highest relative contribution of recombination was in the
380 ToSRV CP, which was the least variable dataset analyzed. On the other hand, results from the
381 MaYSV Rep indicate that in recombination hotspots, recombination may be at least as
382 important as mutation in generating variability.

383 We have assessed the molecular variability of two Brazilian begomoviruses
384 populations, and found that in certain portions of the genome recombination could be
385 responsible for a considerable fraction of the total variability in addition to mutation.
386 Additionally, we also found evidence that the distinct evolutionary patterns might, at least in
387 part, be related to the ability to infect non-cultivated, perennial hosts. Such ability would thus
388 be an important factor in the rapid evolution of these populations.

389

390 **Methods**

391 *Sequence data sets*

392 A total of 55 full-length DNA-A sequences corresponding to ToSRV and 56 corresponding to
393 MaYSV were analyzed (GenBank accession numbers JN419005, JN419007, JN419009,
394 JN419012-JN419016, JN419018-JN419020, JN419022 and KC004091-KC004134

395 (MaYSV), KC004068-KC004090 and JX865615-JX865650 (ToSRV)]. The ToSRV data set
396 comprised isolates obtained from tomato (53 isolates) and *Sida* sp. plants (2 isolates)
397 collected in three different municipalities in the state of Minas Gerais (MG) from 2008-2010.
398 Twelve of the MaYSV sequences were isolated from leguminous non-cultivated hosts
399 collected at nine municipalities of the states of Alagoas (AL), Sergipe (SE) and Paraiba (PB)
400 in 2009 and 2010 as previously described (Silva *et al.*, 2012), and 44 additional sequences
401 were obtained from *Macroptilium lathyroides* and the crop *Phaseolus vulgaris* (common
402 bean) samples showing typical symptoms of begomovirus infection collected in Olho d'Água
403 das Flores and Craibas (AL) in 2011 (Suppl. Table S1). The sampling areas in MG and
404 AL/SE/PB have similar sizes (aproximately 45,000 km^2).

405 Total DNA was extracted from fresh tissue or herbarium-like (pressed and dried)
406 samples as described by Doyle and Doyle (1987). Full-length viral genomes were amplified
407 by rolling-circle amplification, according to the standard circular DNA cloning method
408 (Inoue-Nagata *et al.*, 2004). Single genome-length fragments were excised from these
409 concatamers with *Bam*H I and ligated into the pBLUESCRIPT-KS+ (Stratagene) plasmid
410 vector, previously cleaved with the same enzyme. Viral inserts were commercially sequenced
411 (Macrogen Inc., Seoul, South Korea) by primer walking. All genome sequences were
412 organized to begin at the nicking site in the invariant nonanucleotide at the origin of
413 replication (5'-TAATATT//AC-3').

414

415 *Multiple sequence alignments and phylogenetic analysis*

416 Multiple sequence alignments were prepared for the full- length DNA-A and for the
417 CP and Rep coding sequences of each viral species (ToSRV and MaYSV) using Muscle
418 (Edgar, 2004). Maximum likelihood (ML) trees were inferred for CP and Rep sequences
419 using PAUP* v. 4.0 (Swofford, 2003). These genes were chosen due to their essential role for

420 insect transmission and viral replication, respectively. Besides, they encompass about 70% of
421 the full-length DNA-A genome. The best fit model of nucleotide substitution was determined
422 using Modeltest (Posada & Crandall, 1998) by the Akaike Information Criterion (AIC)
423 (Suppl. Table S2). The heuristic ML search was initiated with a neighbor-joining tree, with
424 optimization using the tree-bisection-reconnection algorithm. The robustness of each
425 individual branch was estimated from 2,000 nearest neighbor interchange bootstrap
426 replicates. Trees were visualized and edited using FigTree
427 (tree.bio.ed.ac.uk/software/figtree).
428
429 *Variability indices*
430 The haplotype diversity (H_d), the average pairwise number of nucleotide differences
431 per site (nucleotide diversity, π) and Wright's F fixation index were calculated using DnaSP
432 v. 5.10 (Rozas *et al.*, 2003). The π statistic was also calculated on a sliding window of 100
433 bases, with a step size of 10 bases in order to estimate the nucleotide diversity across the
434 length of the CP and Rep datasets.
435
436 *Recombination analysis*
437 Recombination analysis was performed using the RDP, Geneconv, Bootscan,
438 Maximum Chi Square, Chimaera, SisterScan and 3Seq methods as implemented in
439 Recombination Detection Program (RDP) version 3.0 (Martin *et al.*, 2010) (RDP project files
440 available from the authors upon request). Alignments were scanned with default settings for
441 the different methods. Statistical significance was inferred by *p*-values lower than a
442 Bonferroni-corrected cutoff. Only recombination events detected by at least three of the
443 analysis methods available in the program were considered reliable.
444

445 *Detection of positive and negative selection at amino acid sites*

446 To detect sites under positive and negative selection, we analyzed the CP and Rep
447 data sets using three ML-based methods implemented in DataMonkey
448 (www.datamonkey.org): Single Likelihood Ancestor Counting (SLAC), Random Effects
449 Likelihood (REL) and Partitioning for Robust Inference of Selection (PARRIS)
450 (Kosakovsky-Pond & Frost, 2005; Scheffler *et al.*, 2006). As recombinant sequences cause
451 spurious selection results, we searched for breakpoints in sequences implicated as
452 recombinant by GARD prior to running these analyses. Additionally, PARRIS allowed
453 synonymous rate variation, topology and branch lengths to vary across recombination
454 breakpoints. The SLAC method was also used to estimate the mean dN/dS ratios (ω) for the
455 CP and Rep datasets from both begomovirus populations based on the inferred GARD
456 phylogenetic trees. These methods were applied under the nucleotide substitution models
457 determined in Modeltest as described in Suppl. Table S2. Bayes factors larger than 50 and *p*
458 values smaller than 0.1 were used as thresholds for the REL method.

459

460 *Relative contribution of recombination and mutation in the genetic variability of* ToSRV *and*
461 MaYSV *populations*

462 A phylogeny-based approach was developed to determine the relative contribution of
463 recombination and mutation to the total variability observed in the ToSRV and MaYSV
464 populations. We identified groups of sequences descended from a shared recombination event
465 (based on RDP analysis), and frequently these formed clades on midpoint rooted CP and Rep
466 ML trees. The sequence of the ancestral node of each of these clades reflected the
467 recombination event (ancestral state reconstruction by marginal ML method in PAUP*), but
468 its parental node did not. Wherever possible, we assigned each recombination event to the
469 branch between these nodes. We did not consider the minor/major parent information from

470 RDP analyses to be definitive, and instead considered the portion of the sequence which
471 differed the most from the parental nodes on the ML tree to be the recombined, introduced
472 block. In the case of recombination events associated with only one sequence, the event was
473 assigned to the branch leading to the corresponding tip. Using PAUP*, we identified all
474 substitutions that occurred over each ML tree, and calculated η, the total number of
475 substitutions over the phylogeny (Fu & Li, 1993). We then looked at the location within the
476 gene of substitutions that occurred on branches associated with recombination. Substitutions
477 that occurred in the region likely introduced by recombination were added to $\eta_{recombination}$ (η_r),
478 the total number of substitutions on the phylogeny likely due to recombination. All other
479 substitutions, including those on branches associated with recombination but outside of the
480 region implicated in the recombination event, were added to $\eta_{mutation}$ (η_m). Thus, $\eta = \eta_r + \eta_m$.
481

482 **Acknowledgements**

483 The authors wish to thank Eduardo Mizubuti and Elizabeth Fontes for their continuous and
484 invaluable support. ATML was the recipient of a CNPq doctoral fellowship. SD was
485 supported by the NSF (DEB 1026095). FMZ is supported by CNPq (484447/2011-4) and
486 Fapemig (APQ-949-09).
487

488 **References**

489

490 **Ala-Poikela, M., Svensson, E., Rojas, A., Horko, T., Paulin, L., Valkonen, J. P. T. &
491 Kvarnheden, A. (2005).** Genetic diversity and mixed infections of begomoviruses
492 infecting tomato, pepper and cucurbit crops in Nicaragua. *Plant Pathol* **54**, 448-459.
493 **Alabi, O. J., Ogbe, F. O., Bandyopadhyay, R., Dixon, A. G., Hughes, J. & Naidu, R. A.
494 (2007).** The occurrence of African cassava mosaic virus and East African cassava
495 mosaic Cameroon virus in natural hosts other than cassava in Nigeria. *Phytopathology*
496 **97**, S3-S3.
497 **Alabi, O. J., Ogbe, F. O., Bandyopadhyay, R., Kumar, P. L., Dixon, A. G. O., Hughes, J.
498 D. & Naidu, R. A. (2008).** Alternate hosts of African cassava mosaic virus and East
499 African cassava mosaic Cameroon virus in Nigeria. *Arch Virol* **153**, 1743-1747.

500 **Barbosa, J. C., Barreto, S. S., Inoue-Nagata, A. K., Reis, M. S., Firmino, A. C.,**
501 **Bergamin, A. & Rezende, J. A. M. (2009).** Natural infection of *Nicandra*
502 *physaloides* by *Tomato severe rugose virus* in Brazil. *J Gen Plant Pathol* **75**, 440-443.
503 **Bedford, I. D., Kelly, A., Banks, G. K., Briddon, R. W., Cenis, J. L. & Markham, P. G.**
504 **(1998).** *Solanum nigrum*: an indigenous weed reservoir for a tomato yellow leaf curl
505 geminivirus in southern Spain. *Eur J Plant Pathol* **104**, 221-222.
506 **Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001).** Complete nucleotide sequence and
507 host range of South African cassava mosaic virus: further evidence for recombination
508 amongst begomoviruses. *J Gen Virol* **82**, 53-58.
509 **Bezerra-Agasie, I. C., Ferreira, G. B., Ávila, A. C. & Inoue-Nagata, A. K. (2006).** First
510 report of *Tomato severe rugose virus* in chili pepper in Brazil. *Plant Dis* **90**.
511 **Blair, M. W., Basset, M. J., Abouzid, A. M., Hiebert, E., Polston, J. E., McMillan, R. T.,**
512 **Graves, W. & Lamberts, M. (1995).** Ocurrence of bean golden mosaic virus in
513 Florida. *Plant Dis* **79**, 529-533.
514 **Briddon, R. W., Bedford, I. D., Tsai, J. H. & Markham, P. G. (1996).** Analysis of the
515 nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly
516 top virus, suggests a recombinant origin. *Virology* **219**, 387-394.
517 **Brown, J. K. & Bird, J. (1992).** Whitefly-transmitted geminiviruses and associated disorders
518 in the Americas and the Caribbean basin. *Plant Dis* **76**, 220-225.
519 **Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, F. M., Moriones, E. & Navas-**
520 **Castillo, J. (2012).** Family *Geminiviridae*. In *Virus Taxonomy 9th Report of the*
521 *International Committee on Taxonomy of Viruses*, pp. 351-373. Edited by A. M. Q.
522 King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. London, UK: Elsevier
523 Academic Press.
524 **Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G. & Stanley,**
525 **J. (2006).** Genetic diversity and phylogeography of cassava mosaic viruses in Kenya.
526 *J Gen Virol* **87**, 3053-3065.
527 **Chare, E. R. & Holmes, E. C. (2004).** Selection pressures in the capsid genes of plant RNA
528 viruses reflect mode of transmission. *J Gen Virol* **85**, 3149-3157.
529 **Davino, S., Napoli, C., Dellacroce, C., Miozzi, L., Noris, E., Davino, M. & Accotto, G. P.**
530 **(2009).** Two new natural begomovirus recombinants associated with the tomato
531 yellow leaf curl disease co-exist with parental viruses in tomato epidemics in Italy.
532 *Virus Res* **143**, 15-23.
533 **Doyle, J. J. & Doyle, J. L. (1987).** A rapid DNA isolation procedure for small amounts of
534 fresh leaf tissue. *Phytochem Bull* **19**, 11-15.
535 **Duffy, S. & Holmes, E. C. (2008).** Phylogenetic evidence for rapid rates of molecular
536 evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. *J*
537 *Virol* **82**, 957-965.
538 **Duffy, S. & Holmes, E. C. (2009).** Validation of high rates of nucleotide substitution in
539 geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *J*
540 *Gen Virol* **90**, 1539-1547.
541 **Edgar, R. C. (2004).** MUSCLE: A multiple sequence alignment method with reduced time
542 and space complexity. *BMC Bioinformatics* **5**, 1-19.
543 **Fernandes, F. R., Albuquerque, L. C., Giordano, L. B., Boiteux, L. S., Ávila, A. C. &**
544 **Inoue-Nagata, A. K. (2008).** Diversity and prevalence of Brazilian bipartite
545 begomovirus species associated to tomatoes. *Virus Genes* **36**, 251-258.
546 **Fiallo-Olive, E., Navas-Castillo, J., Moriones, E. & Martinez-Zubiaur, Y. (2012).**
547 Begomoviruses infecting weeds in Cuba: Increased host range and a novel virus
548 infecting *Sida rhombifolia*. *Arch Virol* **157**, 141-146.

549 **Fondong, V. N., Pita, J. S., Rey, M. E. C., Kochko, A., Beachy, R. N. & Fauquet, C. M.**
550 **(2000).** Evidence of synergism between African cassava mosaic virus and a new
551 double-recombinant geminivirus infecting cassava in Cameroon. *J Gen Virol* **81**, 287-
552 297.
553 **Fu, Y. & Li, W. H. (1993).** Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709.
554 **García-Andrés, S., Accotto, G. P., Navas-Castillo, J. & Moriones, E. (2007a).** Founder
555 effect, plant host, and recombination shape the emergent population of begomoviruses
556 that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology*
557 **359**, 302-312.
558 **García-Andrés, S., Monci, F., Navas-Castillo, J. & Moriones, E. (2006).** Begomovirus
559 genetic diversity in the native plant reservoir *Solanum nigrum*: Evidence for the
560 presence of a new virus species of recombinant nature. *Virology* **350**, 433-442.
561 **García-Andrés, S., Tomas, D. M., Sanchez-Campos, S., Navas-Castillo, J. & Moriones,**
562 **E. (2007b).** Frequent occurrence of recombinants in mixed infections of tomato
563 yellow leaf curl disease-associated begomoviruses. *Virology* **365**, 210-219.
564 **García-Arenal, F., Fraile, A. & Malpica, J. M. (2003).** Variation and evolution of plant
565 virus populations. *Int Microbiol* **6**, 225-232.
566 **Gilbertson, R. L., Faria, J. C., Ahlquist, P. & Maxwell, D. P. (1993).** Genetic diversity in
567 geminiviruses causing bean golden mosaic disease: the nucleotide sequence of the
568 infectious cloned DNA components of a Brazilian isolate of bean golden mosaic
569 geminivirus. *Phytopathology* **83**, 709-715.
570 **González-Aguilera, A., Tavares, S. S., Ramos-Sobrinho, R., Xavier, C. A. D., Dueñas-**
571 **Hurtado, F., Lara-Rodrigues, R. M., Silva, D. J. H. & Zerbini, F. M.** Genetic
572 structure of a Brazilian population of the begomovirus *Tomato severe rugose virus*
573 (ToSRV). *Trop Plant Pathol*, in press.
574 **Graham, A. P., Martin, D. P. & Roye, M. E. (2010).** Molecular characterization and
575 phylogeny of two begomoviruses infecting *Malvastrum americanum* in Jamaica:
576 Evidence of the contribution of inter-species recombination to the evolution of
577 malvaceous weed-associated begomoviruses from the Northern Caribbean. *Virus*
578 *Genes* **40**, 256-266.
579 **Hanley-Bowdoin, L., Settlage, S. B. & Robertson, D. (2004).** Reprogramming plant gene
580 expression: a prerequisite to geminivirus DNA replication. *Mol Plant Pathol* **5**, 149-
581 156.
582 **Harrison, B. D. & Robinson, D. J. (1999).** Natural genomic and antigenic variation in
583 white-fly transmitted geminiviruses (begomoviruses). *Annu Rev Phytopathol* **39**, 369-
584 398.
585 **Harrison, B. D., Zhou, X., Otim Nape, G. W., Liu, Y. & Robinson, D. J. (1997).** Role of a
586 novel type of double infection in the geminivirus-induced epidemic of severe cassava
587 mosaic in Uganda. *Ann Appl Biol* **131**, 437-448.
588 **Ilyina, T. V. & Koonin, E. V. (1992).** Conserved sequence motifs in the initiator proteins for
589 rolling circle DNA replication encoded by diverse replicons from eubacteria,
590 eucaryotes and archaebacteria. *Nucleic Acids Res* **20**, 3279-3285.
591 **Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B. & Nagata, T. (2004).** A simple
592 method for cloning the complete begomovirus genome using the bacteriophage phi 29
593 DNA polymerase. *J Virol Met* **116**, 209-211.
594 **Jeske, H., Lutgemeier, M. & Preiss, W. (2001).** DNA forms indicate rolling circle and
595 recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal* **20**,
596 6158-6167.

597 **Kosakovsky-Pond, S. L. & Frost, S. D. W. (2005).** Not so different after all: A comparison
598 of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-
599 1222.
600 **Lefeuvre, P., Lett, J. M., Reynaud, B. & Martin, D. P. (2007a).** Avoidance of protein fold
601 disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.
602 **Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009).** Widely conserved
603 recombination patterns among single-stranded DNA viruses. *J Virol* **83**, 2697-2707.
604 **Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A.,**
605 **Becker, N., Reynaud, B. & Lett, J. M. (2007b).** Begomovirus 'melting pot' in the
606 south-west Indian Ocean islands: Molecular diversity and evolution through
607 recombination. *J Gen Virol* **88**, 3458-3468.
608 **Legg, J. & Fauquet, C. (2004).** Cassava mosaic geminiviruses in Africa. *Plant Mol Biol* **56**,
609 585-599.
610 **Legg, J. P. & Thresh, J. M. (2000).** Cassava mosaic virus disease in East Africa: A dynamic
611 disease in a changing environment. *Virus Res* **71**, 135-149.
612 **Londono, A., Riego-Ruiz, L. & Arguello-Astorga, G. R. (2010).** DNA-binding specificity
613 determinants of replication proteins encoded by eukaryotic ssDNA viruses are
614 adjacent to widely separated RCR conserved motifs. *Arch Virol* **155**, 1033-1046.
615 **Lozano, G., Trenado, H. P., Valverde, R. A. & Navas-Castillo, J. (2009).** Novel
616 begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and
617 *Ipomoea indica*: Taxonomic and phylogenetic implications. *J Gen Virol* **90**, 2550-
618 2562.
619 **Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B.,**
620 **Vincent, C., Reynaud, B. & Lett, J. M. (2011).** Complex recombination patterns
621 arising during geminivirus coinfections preserve and demarcate biologically important
622 intra-genome interaction networks. *PLoS Pathog* **7**, e1002203.
623 **Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010).**
624 RDP3: A flexible and fast computer program for analyzing recombination.
625 *Bioinformatics* **26**, 2462-2463.
626 **Martin, D. P., van der Walt, E., Posada, D. & Rybicki, E. P. (2005).** The evolutionary
627 value of recombination is constrained by genome modularity. *PLoS Genet* **1**, e51.
628 **Monci, F., Sanchez-Campos, S., Navas-Castillo, J. & Moriones, E. (2002).** A natural
629 recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and
630 *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming
631 prevalent in Spanish populations. *Virology* **303**, 317-326.
632 **Monde, G., Walangululu, J., Winter, S. & Bragard, C. (2010).** Dual infection by cassava
633 begomoviruses in two leguminous species (Fabaceae) in Yangambi, Northeastern
634 Democratic Republic of Congo. *Arch Virol* **155**, 1865-1869.
635 **Morales, F. J. & Anderson, P. K. (2001).** The emergence and dissemination of whitefly-
636 transmitted geminiviruses in Latin America. *Arch Virol* **146**, 415-441.
637 **Morales, F. J. & Jones, P. G. (2004).** The ecology and epidemiology of whitefly-transmitted
638 viruses in Latin America. *Virus Res* **100**, 57-65.
639 **Morilla, G., Krenz, B., Jeske, H., Bejarano, E. R. & Wege, C. (2004).** Tête à tête of
640 tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single
641 nuclei. *J Virol* **78**, 10715-10723.
642 **Moriones, E. & Navas-Castillo, J. (2000).** *Tomato yellow leaf curl virus*, an emerging virus
643 complex causing epidemics worldwide. *Virus Res* **71**, 123-134.
644 **Navas-Castillo, J., Sanchez-Campos, S., Noris, E., Louro, D., Accotto, G. P. & Moriones,**
645 **E. (2000).** Natural recombination between Tomato yellow leaf curl virus-is and
646 Tomato leaf curl virus. *J Gen Virol* **81**, 2797-2801.

647 **Ndunguru, J., Legg, J., Aveling, T., Thompson, G. & Fauquet, C. (2005).** Molecular
648 biodiversity of cassava begomoviruses in Tanzania: Evolution of cassava
649 geminiviruses in Africa and evidence for East Africa being a center of diversity of
650 cassava geminiviruses. *Virol J* **2**, 21.
651 **Orozco, B. M., Kong, L. J., Batts, L. A., Elledge, S. & Hanley-Bowdoin, L. (2000).** The
652 multifunctional character of a geminivirus replication protein is reflected by its
653 complex oligomerization properties. *J Biol Chem* **275**, 6114-6122.
654 **Padidam, M., Sawyer, S. & Fauquet, C. M. (1999).** Possible emergence of new
655 geminiviruses by frequent recombination. *Virology* **265**, 218-224.
656 **Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M.**
657 **(2001).** Recombination, pseudorecombination and synergism of geminiviruses are
658 determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen*
659 *Virol* **82**, 655-665.
660 **Polston, J. E. & Anderson, P. K. (1997).** The emergence of whitefly-transmitted
661 geminiviruses in tomato in the western hemisphere. *Plant Dis* **81**, 1358-1369.
662 **Posada, D. & Crandall, K. A. (1998).** MODELTEST: Testing the model of DNA
663 substitution. *Bioinformatics* **14**, 817-818.
664 **Posada, D. & Crandall, K. A. (2002).** The effect of recombination on the accuracy of
665 phylogeny estimation. *J Mol Evol* **54**, 396-402.
666 **Power, A. G. (2000).** Insect transmission of plant viruses: A constraint on virus variability.
667 *Curr Opin Plant Biol* **3**, 336-340.
668 **Prasanna, H. C. & Rai, M. (2007).** Detection and frequency of recombination in tomato-
669 infecting begomoviruses of South and Southeast Asia. *Virol J* **4**, 111.
670 **Reddy, R. V. C., Colvin, J., Muniyappa, V. & Seal, S. (2005).** Diversity and distribution of
671 begomoviruses infecting tomato in India. *Arch Virol* **150**, 845-867.
672 **Ribeiro, S. G., Ambrozevicius, L. P., Ávila, A. C., Bezerra, I. C., Calegario, R. F.,**
673 **Fernandes, J. J., Lima, M. F., Mello, R. N., Rocha, H. & Zerbini, F. M. (2003).**
674 Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Arch*
675 *Virol* **148**, 281-295.
676 **Rocha, C. S. (2011).** Variability and genetic structure of begomovirus populations infecting
677 tomato and non-cultivated hosts in southwestern Brazil. DS Thesis, *Dep de*
678 *Fitopatologia*, 129 p. Viçosa, MG: Universidade Federal de Viçosa.
679 **Rojas, M. R., Hagen, C., Lucas, W. J. & Gilbertson, R. L. (2005).** Exploiting chinks in the
680 plant's armor: Evolution and emergence of geminiviruses. *Annu Rev Phytopathol* **43**,
681 361-394.
682 **Roossinck, M. J. (2003).** Plant RNA virus evolution. *Curr Opin Microbiol* **6**, 406-409.
683 **Rothenstein, D., Haible, D., Dasgupta, I., Dutt, N., Patil, B. L. & Jeske, H. (2006).**
684 Biodiversity and recombination of cassava-infecting begomoviruses from southern
685 India. *Arch Virol* **151**, 55-69.
686 **Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003).** DnaSP: DNA
687 polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**,
688 2496-2497.
689 **Sanz, A. I., Fraile, A., Gallego, J. M., Malpica, J. M. & García-Arenal, F. (1999).** Genetic
690 variability of natural populations of cotton leaf curl geminivirus, a single-stranded
691 DNA virus. *J Mol Evol* **49**, 672-681.
692 **Sanz, A. I., Fraile, A., García-Arenal, F., Zhou, X., Robinson, D. J., Khalid, S., Butt, T.**
693 **& Harrison, B. D. (2000).** Multiple infection, recombination and genome
694 relationships among begomovirus isolates found in cotton and other plants in
695 Pakistan. *J Gen Virol* **81**, 1839-1849.

696 **Saunders, K., Bedford, I. D. & Stanley, J. (2001).** Pathogenicity of a natural recombinant
697 associated with ageratum yellow vein disease: implications for geminivirus evolution
698 and disease aetiology. *Virology* **282**, 38-47.
699 **Scheffler, K., Martin, D. P. & Seoighe, C. (2006).** Robust inference of positive selection
700 from recombining coding sequences. *Bioinformatics* **22**, 2493-2499.
701 **Schnippenkoetter, W. H., Martin, D. P., Willment, J. A. & Rybicki, E. P. (2001).** Forced
702 recombination between distinct strains of *Maize streak virus*. *J Gen Virol* **82**, 3081-
703 3090.
704 **Seal, S. E., Jeger, M. J. & Van den Bosch, F. (2006).** Begomovirus evolution and disease
705 management. *Adv Virus Res* **67**, 297-316.
706 **Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Junior, B. T., Assunção, I. P., Lima, G. S.
707 A., Pio-Ribeiro, G., Mizubuti, E. S. G. & Zerbini, F. M. (2012).** Species diversity,
708 phylogeny and genetic variability of begomovirus populations infecting leguminous
709 weeds in northeastern Brazil. *Plant Pathol* **61**, 457-467.
710 **Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Júnior, B. T., Assunção, I. P., Lima, G. S.
711 A., Pio-Ribeiro, G., Mizubuti, E. S. G. & Zerbini, F. M. (2011).** High genetic
712 variability and recombination in a begomovirus population infecting the ubiquitous
713 weed *Cleome affinis* in northeastern Brazil. *Arch Virol* **156**, 2205-2213.
714 **Souza-Dias, J. A. C., Sawazaki, H. E., Pernambuco, P. C. A., Elias, L. M. & Maluf, H.
715 (2008).** Tomato severe rugose virus: Another begomovirus causing leaf deformation
716 and mosaic symptoms on potato in Brazil. *Plant Dis* **92**, 487-488.
717 **Sserubombwe, W. S., Briddon, R. W., Baguma, Y. K., Ssemakula, G. N., Bull, S. E.,
718 Bua, A., Alicai, T., Omongo, C., Otim-Nape, G. W. & Stanley, J. (2008).** Diversity
719 of begomoviruses associated with mosaic disease of cultivated cassava (*Manihot
720 esculenta* Crantz) and its wild relative (*Manihot glaziovii* Mull. Arg.) in Uganda. *J
721 Gen Virol* **89**, 1759-1769.
722 **Swofford, D. L. (2003).** PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
723 Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
724 **Tiendrebeogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A.,
725 Villemot, J., Traore, V. S., Konate, G., Traore, A. S., Barro, N., Reynaud, B.,
726 Traore, O. & Lett, J. M. (2012).** Evolution of *African cassava mosaic virus* by
727 recombination between bipartite and monopartite begomoviruses. *Virol J* **9**, 67.
728 **Torres-Pacheco, I., Garzón-Tiznado, J. A., Brown, J. K., Becerra-Flora, A. & Rivera-
729 Bustamante, R. (1996).** Detection and distribution of geminiviruses in Mexico and
730 the Southern United States. *Phytopathology* **86**, 1186-1192.
731 **Were, H. K., Winter, S. & Maiss, E. (2004).** Viruses infecting cassava in Kenya. *Plant Dis*
732 **88**, 17-22.
733 **Wyant, P. S., Gotthardt, D., Schafer, B., Krenz, B. & Jeske, H. (2011).** The genomes of
734 four novel begomoviruses and a new *Sida micrantha mosaic virus* strain from
735 Bolivian weeds. *Arch Virol* **156**, 347-352.
736 **Zerbini, F. M., Andrade, E. C., Barros, D. R., Ferreira, S. S., Lima, A. T. M., Alfenas, P.
737 F. & Mello, R. N. (2005).** Traditional and novel strategies for geminivirus
738 management in Brazil. *Australas Plant Pathol* **34**, 475-480.
739 **Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J. &
740 Harrison, B. D. (1997).** Evidence that DNA-A of a geminivirus associated with
741 severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J
742 Gen Virol* **78**, 2101-2111.
743
744
745

Table 1. Genetic variability of the begomoviruses *Macroptilium yellow spot virus* (MaYSV) and *Tomato severe rugose virus* (ToSRV).

Population	Number of sequences	DNA-A H_d	DNA-A π	CP π	Rep π
ToSRV (total)	55	0.999 (±0.004)	0.00844 (±0.00136)	0.00963 (±0.00238)	0.00743 (±0.00140)
MaYSV (total)	56	1.000 (±0.003)	0.06580 (±0.00231)	0.02852 (±0.00322)	0.10889 (±0.00456)
P. vulgaris	33	1.000 (±0.007)	0.06264 (±0.00300)	0.02189 (±0.00102)	0.11175 (±0.00599)
Non-cultivated hosts	23	1.000 (±0.013)	0.06702 (±0.00425)	0.03616 (±0.00670)	0.10547 (±0.00882)

H_d = Haplotype diversity
π = pairwise, per-site nucleotide diversity

Table 2. Recombination detected by RDP in the DNA-A of *Tomato severe rugose virus* (ToSRV) and *Macroptilium yellow spot virus* (MaYSV) populations.

Event	**Recombinant***	**Recombination breakpoints**		**Parents**		**Methods**†	***P*-value**‡
		Initial	**Final**	**Major**	**Minor**		
ToSRV							
1	Vic20:10	537	1011	Unknown	HV25:10	GBMCS**3**	2.18×10^{-10}
MaYSV							
1	Oaf12:11	2654	1661	Unknown	Pdi1:10	GBMCS**3**	6.07×10^{-60}
	Bat1:10			Unknown	Oaf1:10		
	Inp1:10			Unknown	Oaf2:10		
	Pir1:10			Unknown	Ced1:09		
	Crb8:11			Deg1:10	Ced1:09		
	Crb10:11			Inp2:10	Ced1:09		
	Crb13:11			Mac5:10	Ced1:09		
	Crb14:11			Crb9:11	Ced1:09		
	Crb15:11			Crb11:11	Ced1:09		
	Crb16:11			Crb12:11	Ced1:09		
	Oaf16:11			Crb17:11	Ced1:09		
	Oaf17:11			Crb18:11	Ced1:09		
	Oaf20:11			Crb19:11	Ced1:09		
	Oaf21:11			Oaf15:11	Ced1:09		
	Oaf22:11			Oaf18:11	Ced1:09		
	Oaf23:11			Oaf19:11	Ced1:09		
	Oaf1:11			Crb1:11	Ced1:09		
	Oaf2:11			Crb2:11	Ced1:09		
	Oaf7:11			Oaf6:11	Ced1:09		
	Oaf3:11			Oaf9:11	Ced1:09		
	Crb3:11			Crb5:11	Ced1:09		
	Oaf4:11,			Oaf13:11	Ced1:09		
	Oaf5:11, Oaf8:11, Crb4:11, Oaf10:11, Oaf11:11, Oaf14:11, Oaf24:11, Oaf25:11, Oaf25:11, Crb7:11			Crb6:11	Ced1:09		
2	Oaf19:11	(?) 1662	2039	Pdi1:10	Oaf24:11	GBMS**3**	8.50×10^{-23}
	Oaf6:11			Oaf1:10	Bat1:10		
	Oaf9:11			Oaf2:10	Inp1:10		
3	Inp1:10	1984	2648	Oaf14:11	Crb8:11	B**S**3	4.54×10^{-35}
4	Oaf25:11	1798	(?) 2653	Oaf24:11	Unknown	GBMC**S**3	3.51×10^{-22}
	Oaf22:11			Bat1:10	Unknown		
	Oaf7:11			Pir1:10	Unknown		
	Oaf8:11			Crb8:11	Unknown		
	Oaf10:11			Crb10:11	Unknown		
	Oaf11:11			Crb13:11	Unknown		
	Oaf12:11			Crb14:11	Unknown		
	Oaf14:11			Crb15:11	Unknown		
5	Mac1:10	415	1114	Oaf9:11	Unknown	GBM**3**	1.09×10^{-06}
	Bas1:09			Oaf2:10	Unknown		
6	Pir1:10	(?) 2159	(?) 2653	Oaf14:11	Unknown	MC**S**	1.39×10^{-20}
	Bat1:10			Oaf22:11	Unknown		
	Inp1:10			Oaf7:11	Unknown		
	Crb8:11			Oaf8:11	Unknown		
	Crb10:11			Oaf10:11	Unknown		
	Crb13:11			Oaf11:11	Unknown		
	Crb14:11			Oaf12:11	Unknown		
	Crb15:11, Crb16:11, Oaf16:11, Oaf17:11, Oaf20:11, Oaf21:11, Oaf23:11, Oaf1:11, Oaf2:11, Oaf3:11, Crb3:11, Oaf4:11, Oaf5:11, Crb4:11, Oaf24:11, Crb7:11			Oaf25:11	Unknown		

* Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?) Indicates that the breakpoint could not be precisely pinpointed.

† R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

‡The reported *P*-value is for the program in bold, underlined type and is the lowest *P*-value calculated for the event in question.

Table 3. Mean dN/dS values for the CP and Rep genes of *Tomato severe rugose virus* (ToSRV) and *Macroptilium yellow spot virus* (MaYSV).

Population	CP	Rep
ToSRV	0.446493/ 0.27438*	0.268859
MaYSV	0.0514088	0.208195

*Excluding the recombinant isolate Vic20:10

Table 4. Relative contribution of mutation and recombination to the standing genetic variability of *Tomato severe rugose virus* (ToSRV) and *Macroptilium yellow spot virus* (MaYSV) populations.

Population	CP			Rep		
	η_r	η_μ	η_{total}	η_r	η_μ	η_{total}
ToSRV	42 (42.4%)	57 (57.6%)	99	0	92 (100%)	92
MaYSV	49 (15.7%)	264 (84.3%)	313	201 (36.3%)	352 (63.7%)	553

Figure legends

Figure 1. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across the CP (A) and Rep (B) sequences from ToSRV (gray line) and MaYSV (black line) populations.

Figure 2. Midpoint-rooted maximum likelihood trees based on the CP **(A)** and Rep **(B)** nucleotide sequences of isolates from the ToSRV population. Nodes with bootstrap values equal or higher than 80% are indicated by filled circles, and those with values lower than 80% and higher 50% by empty circles. The contribution of parental sequences to the unique recombination event detected within the CP coding sequence (event 1, isolate Vic20; Table 2) is shown as a diagram above the branch where the substitutions due to recombination were mapped (in red color). Isolates sampled from non-cultivated hosts are shown in green.

Figure 3. Midpoint-rooted maximum likelihood trees based on the CP **(A)** and Rep **(B)** nucleotide sequences of isolates from the MaYSV population. Nodes with bootstrap values equal or higher than 80% are indicated by filled circles, and those with values lower than 80% and higher 50% by empty circles. The unique recombination events detected within the CP (Event 5) and Rep sequences (Events 1 – 4 and 6, Table 2) are shown as diagrams close to the branches where the substitutions due to each recombination event were mapped. Branches in blue, brown, purple, red, pink and yellow colors were assigned to the events 1-6, respectively. Additional substitutions due to the recombination events 4 and 6 were mapped over the orange branch. Isolates sampled from non-cultivated hosts are shown in green.

Figure 4. dN-dS values calculated across the codons of the CP **(A, C)** and Rep sequences **(B, D)** of isolates from the ToSRV **(A, B)** and MaYSV **(C, D)** populations using the Single Likelihood Ancestor Counting (SLAC) method. Sites with statistical evidence of negative selection are indicated by blue bars, while neutrally evolving sites are indicated by black bars.

Figure 1A

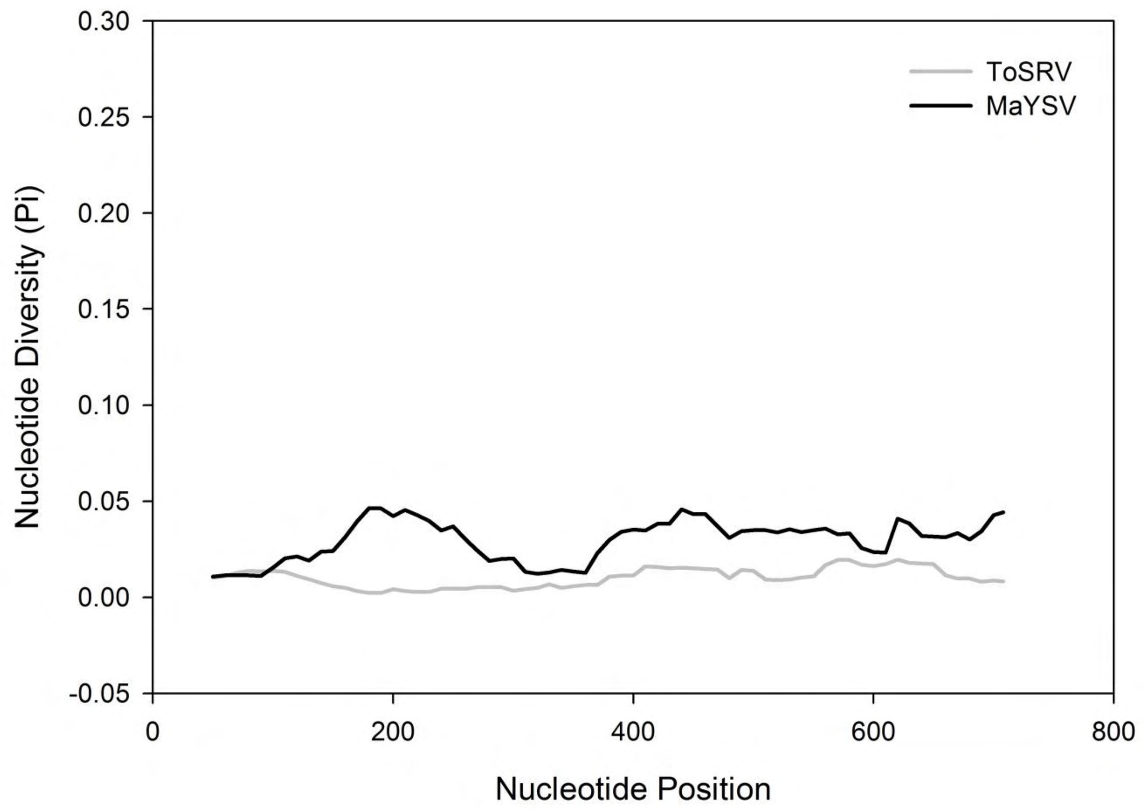


Figure 1B

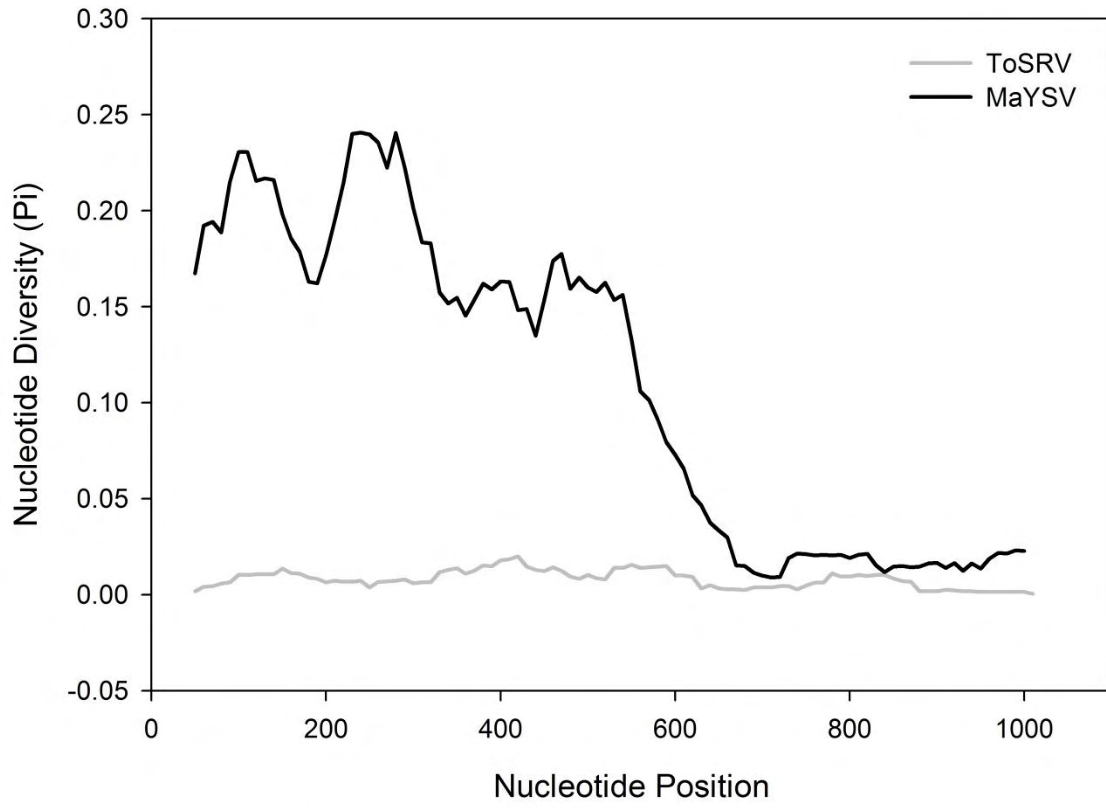
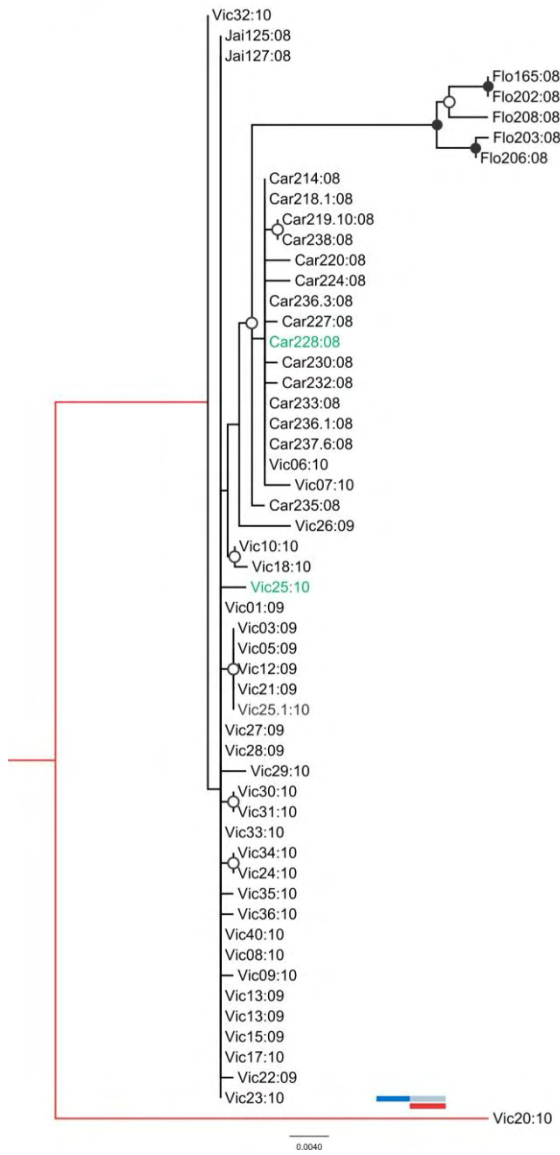


Figure 2



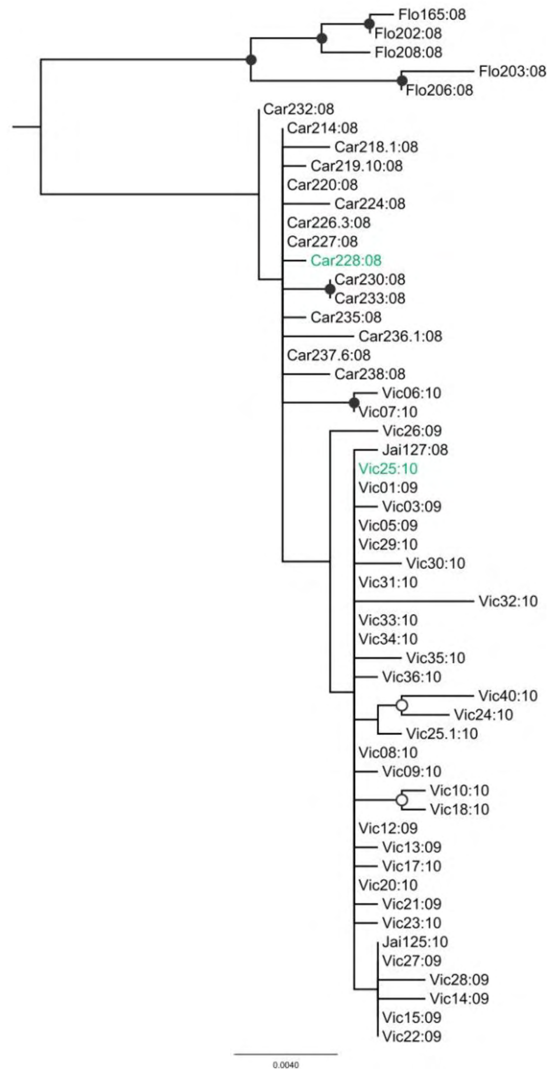


Figure 3

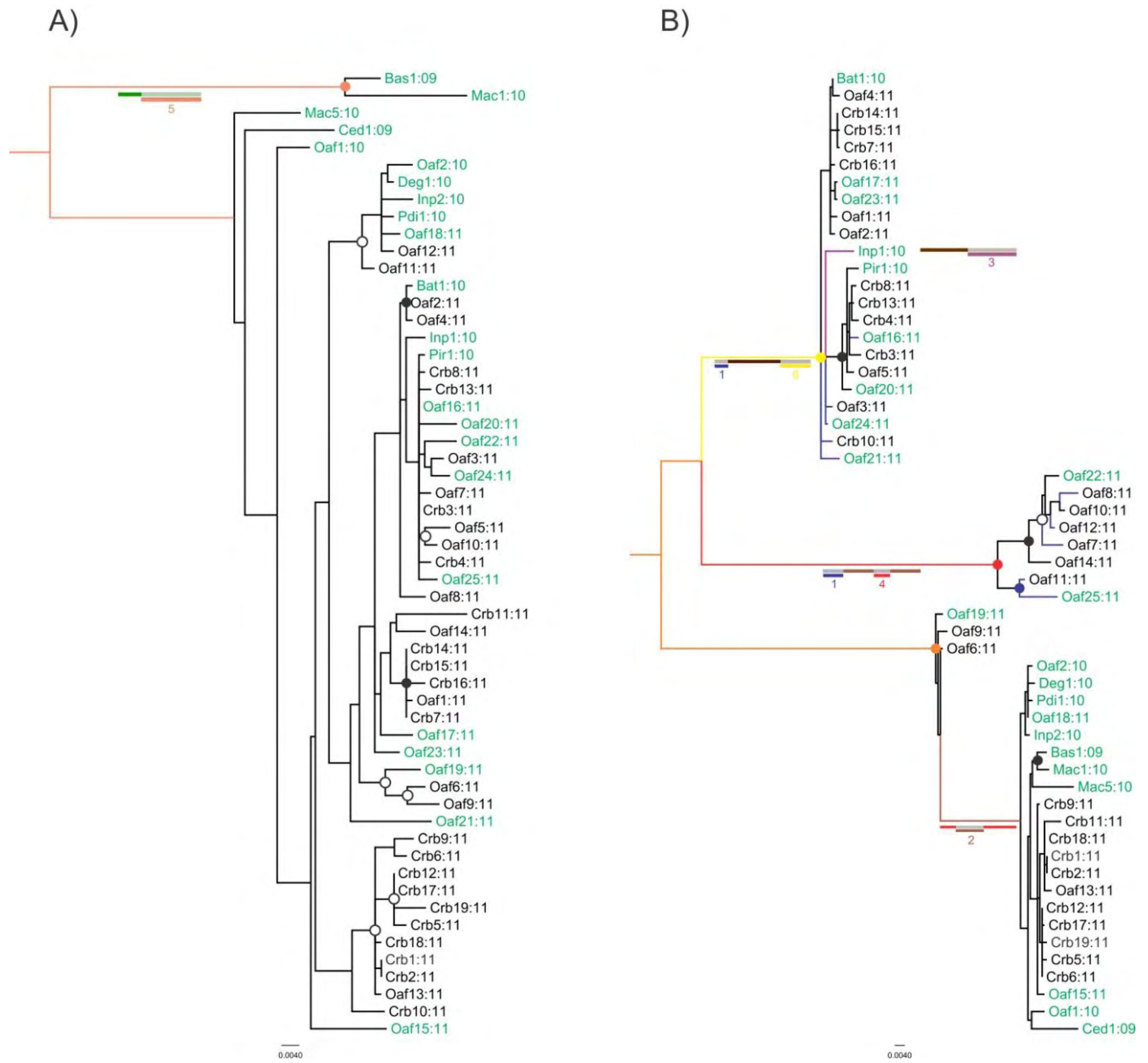
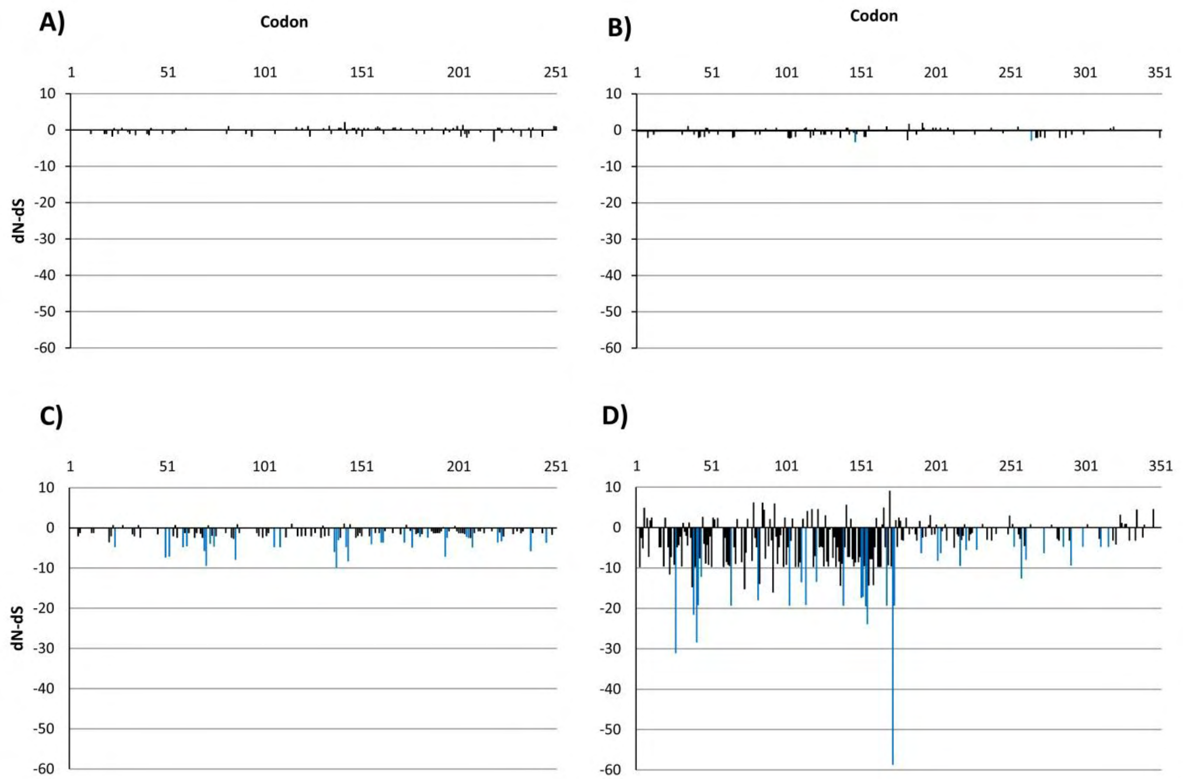


Figure 4



Supplementary Table S1. Begomovirus sequences used in this work.

Sample code	Location	Date	Host	Isolate	GenBank accession #	Reference
Tomato severe rugose virus (ToSRV)						
DV125	Jaíba	Jul 2008	*Solanum lycopersicum*	BR:Jai125:08		Rocha (2011)
DV127	Jaíba	Jul 2008	*Solanum lycopersicum*	BR:Jai127:08		Rocha (2011)
DV165	Florestal	Jul 2008	*Solanum lycopersicum*	BR:Flo165:08		Rocha (2011)
DV202	Florestal	Jul 2008	*Solanum lycopersicum*	BR:Flo202:08		Rocha (2011)
DV203	Florestal	Jul 2008	*Solanum lycopersicum*	BR:Flo203:08		Rocha (2011)
DV206	Florestal	Jul 2008	*Solanum lycopersicum*	BR:Flo206:08		Rocha (2011)
DV208	Florestal	Jul 2008	*Solanum lycopersicum*	BR:Flo208:08		Rocha (2011)
DV214	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car214:08		Rocha (2011)
DV218	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car218.1:08		Rocha (2011)
DV219	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car219.10:08		Rocha (2011)
DV220	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car220:08		Rocha (2011)
DV224	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car224:08		Rocha (2011)
DV226	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car226.3:08		Rocha (2011)
DV227	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car227:08		Rocha (2011)
DV228	Carandaí	Jul 2008	*Sida* sp.	BR:Car228:08		Rocha (2011)
DV230	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car230:08		Rocha (2011)
DV232	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car232:08		Rocha (2011)
DV233	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car233:08		Rocha (2011)
DV235	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car235:08		Rocha (2011)
DV236	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car236.1:08		Rocha (2011)
DV237	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car237.6:08		Rocha (2011)
DV238	Carandaí	Jul 2008	*Solanum lycopersicum*	BR:Car238:08		Rocha (2011)
HV25	Viçosa	May 2010	*Sida* sp.	BR:Vic25:10		González-Aguilera *et al.*, in press

J17-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic01:09	González-Aguilera *et al.*, in press
J20-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic03:09	González-Aguilera *et al.*, in press
J21-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic05:09	González-Aguilera *et al.*, in press
J28-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic06:10	González-Aguilera *et al.*, in press
J29-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic07:10	González-Aguilera *et al.*, in press
J30-2	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic08:10	González-Aguilera *et al.*, in press
J31-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic09:10	González-Aguilera *et al.*, in press
J32-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic10:10	González-Aguilera *et al.*, in press
J34-2	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic12:09	González-Aguilera *et al.*, in press
J35-7	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic13:09	González-Aguilera *et al.*, in press
J37-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic14:09	González-Aguilera *et al.*, in press
J38-2	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic15:09	González-Aguilera *et al.*, in press
J44-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic17:10	González-Aguilera *et al.*, in press
J46-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic18:10	González-Aguilera *et al.*, in press
J47-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic20:10	González-Aguilera *et al.*, in press
J48-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic21:09	González-Aguilera *et al.*, in press
J80-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic22:09	González-Aguilera *et al.*, in press
J85-3	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic23:10	González-Aguilera *et al.*, in press
J86-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic24:10	González-Aguilera *et al.*, in press
J88-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic25:10	González-Aguilera *et al.*, in press
J247-2	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic26:09	González-Aguilera *et al.*, in press
J249-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic27:09	González-Aguilera *et al.*, in press
J254-1	Viçosa	Nov 2009	*Solanum lycopersicum*	BR:Vic28:09	González-Aguilera *et al.*, in press
J259-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic29:10	González-Aguilera *et al.*, in press
J260-2	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic30:10	González-Aguilera *et al.*, in press
J261-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic31:10	González-Aguilera *et al.*, in press
J262-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic32:10	González-Aguilera *et al.*, in press
J263-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic33:10	González-Aguilera *et al.*, in press

J264-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic34:10		González-Aguilera *et al.*, in press
J266-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic35:10		González-Aguilera *et al.*, in press
J267-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic36:10		González-Aguilera *et al.*, in press
J271-1	Viçosa	Dec 2010	*Solanum lycopersicum*	BR:Vic40:10		González-Aguilera *et al.*, in press
Macroptilium yellow spot virus (MaYSV)						
Oaf1	Olho d'Água das Flores	Jul 2010	*Macroptilium lathyroides*	BR:Oaf1:10	JN419013	Silva *et al.* (2012)
Oaf2	Olho d'Água das Flores	Jul 2010	*Macroptilium lathyroides*	BR:Oaf2:10	JN419014	Silva *et al.* (2012)
Bas1	Barra de Santana	Jan 2010	*Macroptilium lathyroides*	BR:Bas1:09	JN419005	Silva *et al.* (2012)
Bat1	Batalha	Jul 2010	*Macroptilium lathyroides*	BR:Bat1:10	JN419012	Silva *et al.* (2012)
Ced1	Cedro	Dec 2009	*Macroptilium lathyroides*	BR:Ced1:09	JN419007	Silva *et al.* (2012)
Deg1	Delmiro Gouveia	Jul 2010	*Calopogonium mucunoides*	BR:Deg1:10	JN419016	Silva *et al.* (2012)
Inp1	Inhapi	Jul 2010	*Macroptilium lathyroides*	BR:Inp1:10	JN419018	Silva *et al.* (2012)
Inp2	Inhapi	Jul 2010	*Canavalia* sp.	BR:Inp2:10	JN419019	Silva *et al.* (2012)
Mac1	Maceió	May 2010	*Macroptilium lathyroides*	BR:Mac1:10	JN419009	Silva *et al.* (2012)
Mac5	Maceió	Aug 2010	*Macroptilium lathyroides*	BR:Mac5:10	JN419022	Silva *et al.* (2012)
Pdi1	Palmeira dos Índios	Jul 2010	*Macroptilium lathyroides*	BR:Pdi1:10	JN419020	Silva *et al.* (2012)
Pir1	Piranhas	Jul 2010	*Calopogonium mucunoides*	BR:Pir1:10	JN419015	Silva *et al.* (2012)
RC56	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf19:11		this study
RC57	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf20:11		this study
RC59	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf21:11		this study
RC60	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf22:11		this study
RC62	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf23:11		this study
RC62	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb1:11		this study
RC65	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf1:11		this study
RC66	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf2:11		this study
RC67	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf7:11		this study
RC68	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Crb2:11		this study
RC70	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf3:11		this study

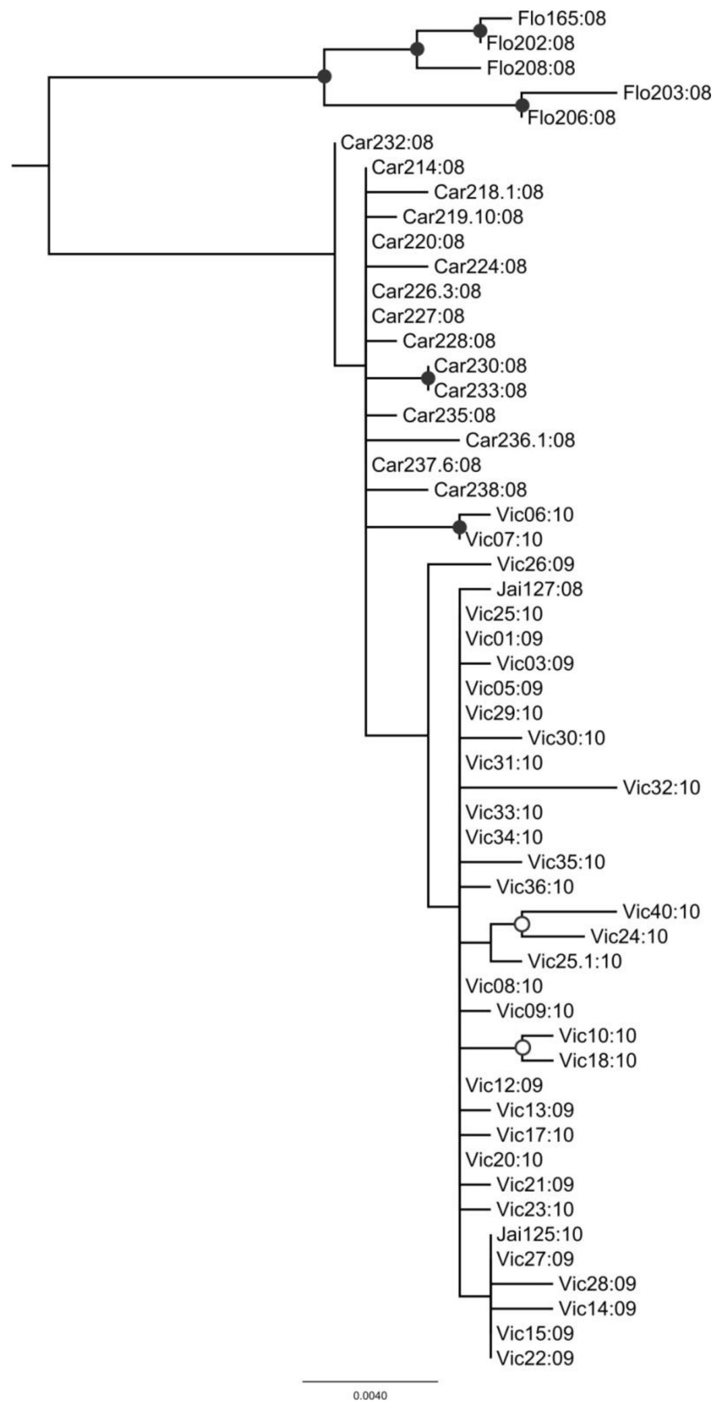
RC71	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf4:11	this study
RC71	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Crb3:11	this study
RC72	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf5:11	this study
RC77	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf8:11	this study
RC78	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf6:11	this study
RC80	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf9:11	this study
RC82	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb4:11	this study
RC83	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf10:11	this study
RC84	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf11:11	this study
RC86	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb5:11	this study
RC88	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf12:11	this study
RC89	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf13:11	this study
RC90	Olho d'Água das Flores	Jul 2011	*Phaseolus vulgaris*	BR:Oaf14:11	this study
RC91	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb6:11	this study
RC94	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf24:11	this study
RC95	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf25:11	this study
RC98	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb7:11	this study
RC102	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb8:11	this study
RC109	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb9:11	this study
RC112	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb10:11	this study
RC118	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb11:11	this study
RC125	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb12:11	this study
RC137	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb13:11	this study
RC141	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb14:11	this study
RC142	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb15:11	this study
RC146	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb16:11	this study
RC153	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb17:11	this study
RC159	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb18:11	this study

RC164	Craibas	Jul 2011	*Phaseolus vulgaris*	BR:Crb19:11	this study
RC172	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf15:11	this study
RC173	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf16:11	this study
RC183	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf17:11	this study
RC229	Olho d'Água das Flores	Jul 2011	*Macroptilium lathyroides*	BR:Oaf18:11	this study

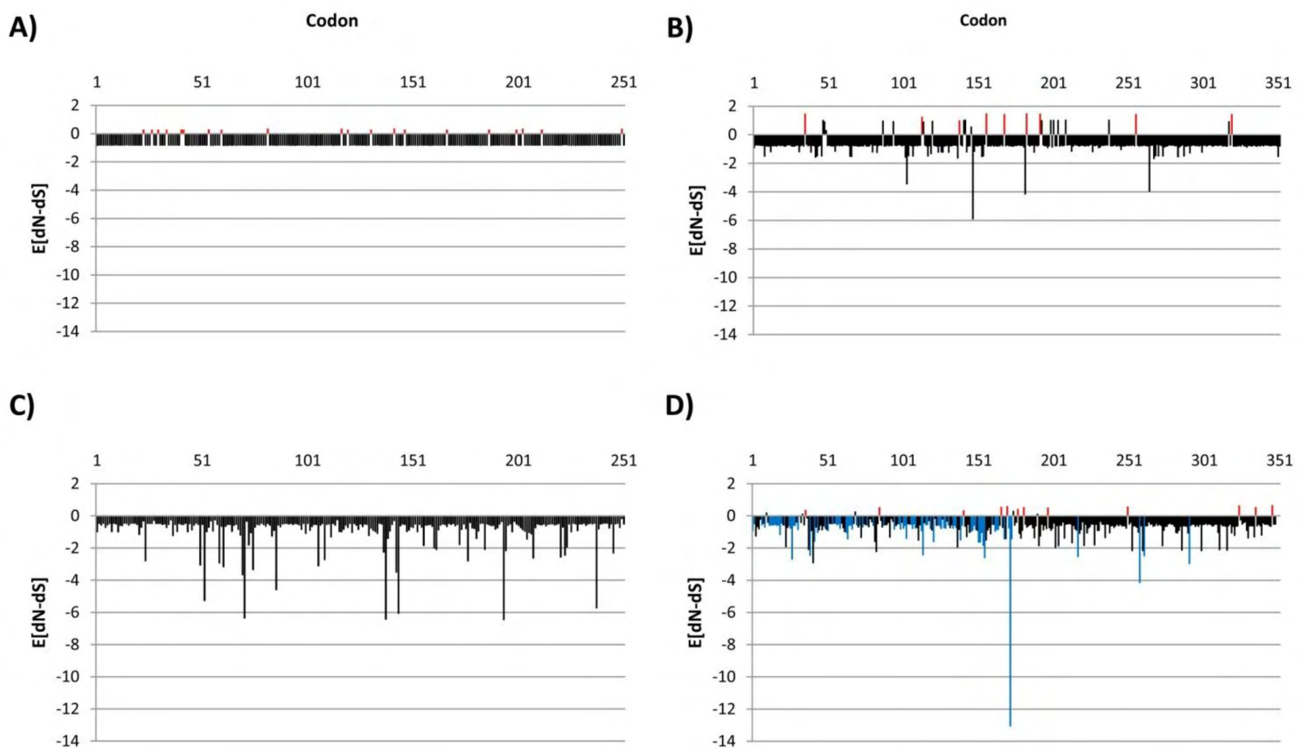
1 **Supplementary Table S2**. Nucleotide substitution models used in this work.

Population	ORF	Phylogenetic reconstruction	Selection analysis (DataMonkey)
ToSRV	CP	TrN + Γ*	TrN
	Rep	TVM + I	HKY
MaYSV	CP	TIM + I + Γ	TrN
	Rep	TVM + I + Γ	HKY

2 *HKY: Hasegawa Kishino-Yano Model, TIM: Transition Model, TrN: Tamura-Nei Model, TVM: Transversion
3 Model, I: Proportion of invariant sites, Γ: Gamma distribution of rates among sites.



Supplementary Figure S1. Midpoint-rooted maximum likelihood trees based on the ToSRV CP nucleotide sequences, excluding the recombinant isolate Vic:20. Nodes with bootstrap values equal or higher than 80% are indicated by filled circles, and those with values lower than 80% and higher 50% by empty circles.



Supplementary Figure S2. dN-dS values calculated across the sites of the CP **(A, C)** and Rep sequences **(B, D)** of isolates from the ToSRV **(A, B)** and MaYSV **(C, D)** populations using the Randon Effects of Likelihood method (REL). Sites with statistical evidence of negative selection are indicated by blue bars, while those under positive selection are indicated by red bars. Neutrally evolving sites are indicated by black bars.

CHAPTER 3

THE RELATIVE CONTRIBUTION OF RECOMBINATION AND MUTATION TO THE QUASISPECIES NATURE OF BEGOMOVIRUS POPULATIONS

Lima, A.T.M., Seah, Y.M., Silva, F.N., Castillo-Urquiza, G.P., Duffy, S. & Zerbini, F.M. The relative contribution of recombination and mutation to the quasispecies nature of begomovirus populations.

The relative contribution of recombination and mutation to the quasispecies nature of begomovirus populations

A.T.M. Lima[a], Y.M. Seah[b], F.N. Silva[a], G.P.C. Urquiza[a], S. Duffy[b#] and F.M. Zerbini[a#]

[a]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-000, Brazil

[b]Dep. of Ecology, Evolution and Natural Resources, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

#Corresponding author: Siobain Duffy

Phone: Fax: 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

1 **Abstract**

2 Begomoviruses (single stranded DNA plant viruses) are responsible for serious agricultural
3 threats. Previous studies have shown that begomovirus populations exhibit a high level of
4 within-host molecular variability and may evolve as quickly as RNA viruses. Although the
5 recombination-prone nature of begomoviruses has been extensively demonstrated, no work
6 has attempted to determine the relative contribution of recombination and mutation to the
7 standing molecular variability of begomovirus populations. By expanding the use of a novel
8 phylogeny-based partitioning method to begomovirus populations, we showed that the
9 diversification of these populations is predominantly driven by mutational dynamics, albeit at
10 different extents. We estimated the molecular variability levels of our populations and
11 observed that they were similar to those of plant RNA viruses, even though begomoviruses
12 replicate using the supposedly proof-reading DNA polymerases from their hosts. A conserved
13 uneven distribution of molecular variability levels across the length of the CP and Rep genes
14 due to recombination was readily evident from our analyses, suggesting a significant
15 contribution of this evolutionary process to the standing molecular variability. By mapping
16 the recombination events onto maximum likelihood trees based on Rep and CP sequences, we
17 assessed recombination as the source of up to 50% of all inferred substitutions. Our results
18 indicate that the evolution of some begomovirus populations might be significantly
19 dependent on the recombination-prone nature of their genomes in addition to their rapid
20 mutational dynamics.

21

Introduction

The theoretical concept of quasispecies was initially developed to mathematically model the evolution of RNA molecules (and other biological macromolecules) (Eigen, 1971). This concept was modified from that originally proposed by Eigen and has been rather liberally applied to describe the evolutionary dynamics of RNA viruses (Holmes, 2010). In virology, it refers to the heterogeneous distribution of genomic variants generated during the viral replication cycle as a result of high mutation rates, rapid replicative kinetics and large population sizes of RNA viruses (Biebricher & Eigen, 2006). These genetic variants are organized around one or more stable consensus sequences known as master sequences (Domingo, 2002; Lauring & Andino, 2010). The first experimental evidence of the quasispecies nature of an RNA virus was provided by studies using the Qβ bacteriophage (Domingo *et al.*, 1978) and an increasing body of evidence points to a similar dynamics for many important viral pathogens (Bukh *et al.*, 1995; Steinhauer *et al.*, 1989; Wain-Hobson, 1992).

The high mutation rates of RNA viruses (Holland *et al.*, 1982) are a consequence of their error prone- RNA-dependent RNA polymerases (RdRp's), and it has been suggested that these viral populations may thoroughly explore their sequence space (all possible combination of mutant sequences) (Eigen *et al.*, 1988) in a wider and faster manner than viruses with genomes based on DNA, which replicate with proof-reading DNA-dependent DNA polymerases (DdDp's). Consequently, it has been proposed that populations of RNA viruses possess a higher adaptive capacity compared to DNA viruses, which would have important implications on strategies to control diseases (Gerrish & Garcia-Lerma, 2003).

Nevertheless, previous studies have shown that ssDNA viruses evolve as quickly as RNA viruses (Drake, 1991; Duffy *et al.*, 2008; Shackelton & Holmes, 2006; Shackelton *et al.*, 2005). Single stranded DNA plant viruses of the families *Geminiviridae* and *Nanoviridae*

47 exhibit high levels of within-host molecular variability (Ge *et al.*, 2007; Grigoras *et al.*, 2010;
48 van der Walt *et al.*, 2008), and substitution rates inferred for begomoviruses (whitefly-
49 transmitted geminiviruses) are similar to those of RNA viruses (Duffy & Holmes, 2008;
50 Duffy & Holmes, 2009). As a consequence, the concept of quasispecies has been expanded to
51 harbor a number of fast evolving viruses, including ssDNA plant viruses (Ge *et al.*, 2007;
52 Isnard *et al.*, 1998).

53 Hypotheses have been formulated in order to explain the high molecular variability in
54 ssDNA virus populations. It has been speculated that low fidelity DNA repair polymerases
55 may replicate these viruses and/or spontaneous biochemical reactions which act preferentially
56 on single-stranded nucleic acids (deamination, oxidation and methylation of bases) might
57 contribute to the variability (Duffy & Holmes, 2008; Duffy & Holmes, 2009; Duffy *et al.*,
58 2008; van der Walt *et al.*, 2008). Although mutational dynamics are a primary factor in the
59 diversification of viral populations (Balol *et al.*, 2010; García-Arenal *et al.*, 2003; Roossinck,
60 1997), they do not explain all the standing molecular variability, since other evolutionary
61 processes including recombination might contribute significantly (Hull, 2009; Martin *et al.*,
62 2011a).

63 Recombination impacts the evolution of several families of viruses (Chare & Holmes,
64 2006) (Bonnet *et al.*, 2005; Fan *et al.*, 2007; Heath *et al.*, 2006; Martin *et al.*, 2011a; Varsani
65 *et al.*, 2006) and has been extensively documented for geminiviruses (Briddon *et al.*, 1996;
66 García-Andrés *et al.*, 2007; Padidam *et al.*, 1999b; Pita *et al.*, 2001). In fact, the devastating
67 mastrevirus *Maize streak virus* (MSV) in Africa, as well as begomoviruses associated with
68 important disease complexes in Europe, Asia and Africa (tomato yellow leaf curl, cotton leaf
69 curl and cassava mosaic diseases, respectively) seem to evolve largely by recombination
70 (Berrie *et al.*, 2001; García-Andrés *et al.*, 2007; García-Andrés *et al.*, 2006; Monci *et al.*,
71 2002; Navas-Castillo *et al.*, 2000; Pita *et al.*, 2001; Sanz *et al.*, 2000; Varsani *et al.*, 2008).

72 Although the mechanics of recombination in ssDNA viruses remains unknown, it is
73 speculated that their high recombination frequency may be a result of a recombination-
74 dependent replication mechanism (RDR) (Jeske *et al.*, 2001).

75 While the contribution of recombination to the molecular variability in geminivirus
76 populations is evident, most studies have focused on detecting recombination without
77 determining the relative contribution of recombination and mutation (Berrie *et al.*, 2001;
78 Davino *et al.*, 2009; García-Andrés *et al.*, 2007; Lozano *et al.*, 2009; Monci *et al.*, 2002;
79 Padidam *et al.*, 1999b; Pita *et al.*, 2001; Saunders *et al.*, 2001; Zhou *et al.*, 1997). We have
80 recently developed a partitioning method of variability based on phylogeny (Lima et al.,
81 *submitted*). By expanding the use of this novel method of sequence analysis to datasets of
82 begomoviruses collected from around the world (available in public databases) we were able
83 to estimate the minimal relative contribution of recombination to begomovirus evolutionary
84 dynamics. Our results confirm that, inasmuch as mutation is the main source of variation for
85 most begomovirus populations, recombination can dramatically contribute to their standing
86 genetic variability.

87

88 **Materials and Methods**

89 *Sequence datasets*

90 Fifteen datasets each of CP and Rep genes from begomoviruses (a total of 1786 sequences
91 from 15 distinct species) and their cognate full-length DNA-A sequences were downloaded
92 from the Genbank database using Taxbrowser (www.ncbi.nlm.nih.gov) between June and
93 July 2012. All genome sequences were organized to begin at the nicking site in the invariant
94 nonanucleotide of the origin of replication (5'-TAATATT//AC-3').

95

96

97 *Multiple sequence alignments and phylogenetic analysis*

98 Multiple sequence alignments were prepared for the full- length DNA-A and for the CP and
99 Rep sequences using Muscle (Edgar, 2004), and visually corrected using MEGA 5.0 (Tamura
100 *et al.*, 2011). Maximum likelihood (ML) trees were inferred using PAUP v. 4.0 (Swofford,
101 2003) with the best fit model of nucleotide substitution determined using Modeltest (Posada
102 & Crandall, 1998) by the Akaike Information Criterion (AIC) (Supplementary Table S1). The
103 heuristic ML search was initiated with a neighbor-joining tree, with optimization using the
104 tree-bisection-reconnection algorithm. The robustness of each individual branch was
105 estimated from 2,000 nearest neighbor interchange bootstrap replicates. Trees were visualized
106 and edited using FigTree (tree.bio.ed.ac.uk/software/figtree).

107

108 *Variability indices*

109 The average pairwise number of nucleotide differences per site (nucleotide diversity, π) was
110 calculated using DnaSP v. 5.10 (Rozas *et al.*, 2003). This same index was also calculated on a
111 sliding window of 100 bases with a step size of 10 bases, in order to estimate the nucleotide
112 diversity across the length of the CP and Rep datasets.

113

114 *Recombination analysis*

115 Recombination analysis was performed for full-length DNA-A sequences using the RDP
116 (Martin & Rybicki, 2000), Geneconv (Padidam *et al.*, 1999b), Bootscan (Martin *et al.*, 2005),
117 Maximum Chi Square (Smith, 1992), Chimaera (Posada & Crandall, 2001), Sister Scan
118 (Gibbs *et al.*, 2000) and 3Seq (Boni *et al.*, 2007) methods as implemented in RDP version
119 3.44 (Martin *et al.*, 2010) (RDP project files available from the authors upon request).
120 Statistical significance was inferred by *p*-values lower than a Bonferroni-corrected cutoff.

121 Only recombination events detected by at least four of the analysis methods available in the
122 program were considered reliable.
123
124 *Detection of positive and negative selection at amino acid sites*
125 To detect sites under positive and negative selection and estimate the contribution of adaptive
126 selection to the standing molecular variability, we analyzed the CP and Rep data sets using
127 two ML-based methods implemented in DataMonkey (www.datamonkey.org): Single
128 Likelihood Ancestor Counting (SLAC) and Partitioning for Robust Inference of Selection
129 (PARRIS) (Kosakovsky-Pond & Frost, 2005; Scheffler *et al.*, 2006). Due to the recombinant-
130 prone nature of the begomovirus genomes, we searched for breakpoints in sequences
131 implicated as recombinant by GARD prior to running these analyses. The SLAC method was
132 also used to estimate the mean dN/dS ratios (ω) for the CP and Rep datasets from
133 begomovirus populations based on the inferred GARD-corrected phylogenetic trees. These
134 methods were applied under the nucleotide substitution models determined in DataMonkey
135 webserver.
136
137 *Relative contribution of recombination and mutation in the genetic variability of*
138 *begomovirus populations*
139 A simple phylogeny based aproach has been developed (Lima *et al.*, *submitted*) to determine
140 the relative contribution of recombination and mutation to the total molecular variability in
141 begomovirus populations. We identified groups of sequences descended from a shared
142 recombination event (based on RDP analysis), and frequently these formed clades on
143 midpoint-rooted CP and Rep ML trees. The sequence of the ancestral node of each of these
144 clades reflected the recombination event (ancestral state reconstruction by marginal ML
145 method in PAUP*), but its parental node did not. Wherever possible, we assigned each

146 recombination event to the branch between these nodes. We did not consider the minor/major
147 parent information from RDP analyses to be definitive, and instead considered the portion of
148 the sequence which differed the most from the parental nodes on the ML tree to be the
149 recombined, introduced block. In the case of recombination events associated with only one
150 sequence, the event was assigned to the branch leading to the corresponding tip. Using
151 PAUP*, we identified all substitutions that occurred over each ML tree, and calculated η, the
152 total number of substitutions over the phylogeny. We then looked at the location within the
153 gene of substitutions that occurred on branches associated with recombination. Substitutions
154 that occurred in the region likely introduced by recombination were added to η recombination
155 (η_r), the total number of substitutions on the phylogeny likely due to recombination. All other
156 substitutions, including those on branches associated with recombination but outside of the
157 region implicated in the recombination event, were added to ηmutation (η_m). Thus, $\eta_{total} = \eta_r$
158 $+ \eta_m$.
159

160 **Results**

161 *Molecular variability of begomovirus populations*

162 The sample sizes in most begomovirus populations were similar (between 34 and 53
163 isolates), allowing for a direct comparison amongst them (Figure 1 and Supplementary Table
164 S1). The exceptions were the AgEV and TYLCCNV datasets (N=21 and 26, respectively)
165 and the EACMV and TYLCV datasets (N=157 and 222, respectively). Nevertheless, the
166 average pairwise number of nucleotide differences (π) for the AgEV dataset (0.04172) was
167 similar to that from other datasets harboring larger sample sizes (*eg*, CLCuGV, N=39, π =
168 0.04290; TYLCV, π = 0.04049). On the other hand, the standing molecular variability
169 estimated for the EACMV and TYLCV datasets (π = 0.05672 and 0.04049, respectively)
170 were lower than those from datasets harboring smaller sample sizes (*eg*, BhYVIV, π =

171 0.06293; BhYMIV, π = 0.06502). The standing molecular variability estimated for the
172 TYLCCNV dataset (the second smallest from our analysis) was markedly high (π = 0.10414).
173 Three groups harboring distinct molecular variability levels were observed: a less
174 variable group (π values of up to 0.03), comprised of two datasets (CLCuBuV and
175 TYLCTV); an intermediate group (π values from 0.03 to 0.06) harboring most datasets
176 (ACMV, AgEV, CLCuGV, CLCuMV, EACMV, MYMIV, ToLCTV and TYLCV) and an
177 highly variable group (π values larger than 0.06) comprised of five datasets (BhYVIV,
178 BhYMIV, SPLCV, ToLCNDV and TYLCCNV).
179 The haplotype diversity estimated for all fifteen DNA-A datasets was similar and
180 close to 1, indicating that most isolates were unique within each population (Suppl. Table
181 S1).
182
183 *Molecular variability across the begomovirus genome/genes is not evenly distributed*
184 We also calculated the nucleotide diversity for the CP and Rep datasets, to assess the
185 individual contribution of these genes to the standing molecular variability in the full-length
186 DNA-A (Figure 1). Four distinct patterns were readily observed: pattern 1 (P1), in which
187 molecular variability levels in both genes from the same population were similar (CLCuBuV
188 and MYMIV); pattern 2 (P2), in which the molecular variability of the Rep gene was slightly
189 higher than the CP gene (AgEV, BhYVIV, CLCuMV, SPLCV, ToLCNDV, ToLCTV,
190 TYLCTV and TYLCCNV); pattern 3 (P3), in which the variability of the Rep gene was
191 markedly higher than that observed for the CP gene (ACMV, BhYMIV, and TYLCV); and
192 pattern 4 (P4), in which the molecular variability level of the CP gene was slightly higher
193 than that of the Rep gene (CLCuGV and EACMV). These results indicate that the CP and
194 Rep genes contribute distinctly to the total standing molecular variability of begomovirus
195 genomes.

196 The uneven distribution of the molecular variability amongst the CP and Rep genes
197 prompted us to further examine the molecular variability levels across the length of both
198 genes (Suppl. Figures S1 and S2). Our analysis revealed increased levels of molecular
199 variability at the central portion and/or 3'-terminal region of the CP (and only rarely at the 5'-
200 terminal region) and at the 5'-terminal region of the Rep gene (and only rarely at the 3'-
201 terminal region) in almost all begomovirus populations. The only exception was the
202 CLCuBuV population whose molecular variability levels in both genes were evenly
203 distributed (Suppl. Figures S1E and S2E). The other populations showed evidence of an
204 uneven distribution of the variability levels in at least one gene (CP or Rep). Our results
205 indicate that these highly variable portions in both genes were also responsible by the uneven
206 molecular variability levels observed amongst the CP and Rep genes in most begomovirus
207 populations.

208 The four patterns of molecular variability distribution between the CP and Rep genes
209 (P1-P4) could be explained based on their highly variable portions. In P1, variability levels
210 across the CLCuBuV CP and Rep length were clearly even, whereas in the MYMIV
211 population there was a slight increase of the molecular variability levels in the 5'-terminal
212 region of both genes, despite yielding similar molecular variability levels (Suppl. Figures S1I
213 and S2I). In P2, both genes showed highly variable portions (with a more striking variability
214 in the Rep 5'-terminal region; Suppl. Figures S1B, C, G, J-N and S2 B, C, G, J-N). P3 was
215 characterized by an even distribution of variability levels across the CP gene and a highly
216 variable Rep 5'-terminal region (Suppl. Figures S1A, D, O and S2A, D, O). In P4, in addition
217 to the highly variable 5'-terminal regions of both CLCuGV genes (Suppl. Figures S1F and
218 S2F), the CP gene showed an even higher variability in its 3'-terminal region. Whereas the
219 EACMV Rep showed a slightly more variable segment in its 3'-terminal region, the CP gene

220 was markedly more variable in both the central and 3'-terminal regions (Suppl. Figures S1H
221 and S2H).

222 Together, these results suggest a distinct interplay of the evolutionary processes in
223 shaping the molecular variability across distinct begomovirus genes.

224

225 *Synonymous site variation in the highly variable portions of the CP and Rep genes*

226 We accessed the putative role of positive and negative selection in shaping the uneven
227 distribution of the molecular variability levels across the CP and Rep genes. The dN/dS ratio
228 (ω) estimated for both genes in all begomovirus populations exhibit values lower than 1,
229 indicating negative selection as the predominant type of selection (Table 1). However, there
230 was wide variation amongst genes/populations (from 0.058772 to 0.684067 for the CLCuGV
231 and CLCuBuV CP datasets, respectively) indicating distinct selective constraints. There was
232 evidence of stronger negative selection acting on the CP gene in all begomovirus populations.
233 The only exception was the CLCuBuV population, in which the ω value for the CP was
234 slightly higher than that of the Rep gene (0.684067 and 0.615212, respectively). In addition,
235 ω values estimated for both CLCuBuV genes were much higher compared to those from
236 other populations (although always lower than 1), indicating a more relaxed negative
237 selection.

238 A single positively selected site was detected by SLAC in the MYMIV Rep and
239 TLCNDV CP datasets (codon positions 181 and 2, respectively). However, the evidence was
240 weak, since no positively selected sites were detected by PARRIS in any dataset.

241 Overall, SLAC detected a larger number of sites with statistical evidence of negative
242 selection from most variable datasets (Figure 2). For example, a considerably larger number
243 of sites under negative selection in the Rep gene were observed for the populations with the
244 most variable Rep datasets (ACMV, BhYVIV, BhYMIV and ToLCNDV, with 65, 48, 69 and

245 123 negatively selected sites, respectively). In addition, our analysis revealed strong evidence
246 of a high synonymous substitution rate in the highly variable regions of both genes (although
247 not always with statistical support for negative selection; Suppl. Figures S3 and S4).

248 Together, these results indicate that the highly variable regions of the Rep and CP
249 genes are the consequence of high local synonymous site variation. In addition, we concluded
250 that adaptive selection does not contribute to the high molecular variability levels in the CP
251 central/3'-terminal and Rep 5'-terminal regions.

252

253 *The mosaic structure of begomovirus genomes*

254 Well supported recombination events were detected in all fifteen datasets, which was
255 expected considering the notorious recombination-prone nature of begomovirus genomes.
256 However, the number of recombination events detected in each dataset was rather variable
257 (Suppl. Table S2), suggesting that this evolutionary process has contributed distinctly to the
258 evolutionary dynamics of different begomovirus populations. A single recombination event
259 ($p=2.27\times10^{-41}$; Suppl. Table S2) was detected in the ACMV dataset, while 26 unique
260 recombination events were detected in the BhYVMV dataset. A high number of
261 recombination events were also detected in the BhYVIV, SPLCV and ToLCTV datasets (16
262 unique events in each dataset). There was no correlation between the number of unique
263 recombination events and the standing molecular variability estimated for the full-length
264 DNA-A. Five unique recombination events were detected in the TYLCCNV (the most
265 variable from our analysis, $\pi = 0.10414$) and AgEV datasets (considerably less variable, $\pi =$
266 0.04172), both harboring similar sample sizes (N=26 and 21, respectively). In addition, five
267 unique events were also detected in the ToLCNDV dataset (the second most variable, $\pi =$
268 0.07661; N=52), similar to the MYMIV and TYLCTV datasets (four events each), both less
269 variable and with smaller sample sizes (N=42 and 39, respectively).

270 A non-random distribution of recombination breakpoints was readily observed, with
271 evidence of two hotspots: in the Rep 5'-terminal region and (less frequently) in the CP
272 central/3'-terminal region. Interestingly, there was a correlation between the location of
273 breakpoints and the uneven distribution of the molecular variability levels across the length
274 of both genes. For example, a well-supported recombination event detected in the EACMV
275 dataset (event 1, $p=1.4\times10^{-39}$) showed both breakpoints in the CP 5'- and 3'-terminal regions
276 (nt positions 201 and 689 in the CP gene sequence, Suppl. Table S2). The segment between
277 the recombination breakpoints showed higher variability compared to the CP 5'-terminal
278 region (Suppl. Figure S1). In addition, in the CLCuGV dataset, the single well supported
279 recombination event (event 4, $p=1.49\times10^{-06}$), whose breakpoint was located in the Rep 5'-
280 terminal region (nt position 2408 in the full-length DNA-A sequence and 299 in the Rep gene
281 sequence), clearly correlated with the high level of variability observed in this region.
282 Together, these results indicate that the highly variable regions of the CP and Rep
283 genes are a consequence of high synonymous site variation due to recombination. However,
284 the even distribution of variability levels across both genes does not indicate the absence of
285 recombination events. In contrast to the highly even distribution of variability levels in both
286 CLCuBuV genes (Suppl. Figures S1E and S2E), six recombination events were detected,
287 showing at least one recombination breakpoint in the Rep 5'-region and CP central/3'-
288 terminal regions (Suppl. Table S2).
289
290 *Long branches in begomovirus phylogeny*
291 Isolates sharing well-supported recombination events frequently formed clades in
292 maximum likelihood (ML) phylogenetic trees for the CP and Rep genes (Suppl. Figure S5
293 and Suppl. Table S2). In addition, these clades were often connected to others by long
294 branches that represented the substitutions acquired by the ancestral recombination event. For

295 example, in the ACMV Rep tree, a clade comprised of four isolates (accession numbers
296 HE616777, HE616779, HE616780 and HE616781) was separated from others with a
297 significant genetic distance. A well-supported recombination event shared by all isolates in
298 this clade was detected spanning the entire Rep gene (event 1, breakpoints 1290 and 38, *p*-
299 value=2.27×10^{-41}; Supplementary Figure S5A/Supplementary Table S2). In the TYLCCNV
300 Rep tree, a long branch leading to isolates JN082237, JN082233, AJ971265 and AJ971524
301 was readily evident. A well-supported recombination event (event 1, *p*-value = 5.20×10^{-84};
302 Supplementary Figure S5N; Supplementary Table S2) spanning the 5'-region of the Rep gene
303 and the 5'-end of the common region was detected in these isolates.

304 We also observed long branches associated with recombination events involving the
305 CP gene sequence. In the EACMV CP tree, two clades separated by a significant genetic
306 distance were clearly associated to a well-supported recombination event (event 1, *p*-
307 value=$1,40 \times 10^{-39}$; Supplementary Figure S5H; Supplementary Table S2) shared by 52
308 isolates and spanning the central and 3'-regions of the CP gene (breakpoint positions 536 and
309 1023). A clade comprised of two isolates (JF502367 and JF502368) and separated by a long
310 branch was also clearly associated with a recombination event spanning the 5' and central
311 regions of the CLCuBuV CP (event 9, *p*-value = 2.68×10^{-06}; Supplementary Figure S5E;
312 Supplementary Table S2).

313 Long branches associated with recombination events detected in a single sequence
314 were also observed. In the CLCuBuV CP ML tree, a long branch leading to the isolate
315 JF510461 was also clearly related to a well-supported recombination event spanning a large
316 region of the CP gene (event 1, breakpoint positions 109 and 1055, *p*-value = 1.75×10^{-06};
317 Supplementary Figure S5E; Supplementary Table S2). Another long branch leading to the
318 isolate DQ866128, representing a recombination event with breakpoints within the Rep gene

319 (event 1, *p*-value = 1.48×10^{-51}; Supplementary Figure S5M; Supplementary Table S2) was
320 observed in the TYLCTV Rep tree.

321 Although we often assigned long branches to recombination events in the
322 phylogenetic trees, some long branches were not associated to recombinant sequences. For
323 example, no statistically supported recombination event was detected in the ACMV isolate
324 represented by accession number JN053426. However, a long branch leading to this isolate
325 was observed in the ACMV CP tree.

326 Together, these results indicate that long branches are often, but not always,
327 associated to recombination events in begomovirus phylogeny.

328

329 *Relative contribution of mutation and recombination to the quasispecies nature of*
330 *begomovirus populations*

331 We mapped all substitutions over the ML trees constructed for CP and Rep genes and
332 counted the number of substitutions on branches which were clearly associated to
333 recombination and mutation, in order to estimate their relative contribution to the standing
334 molecular variability in each begomovirus dataset. Overall, a higher number of substitutions
335 was inferred from the Rep trees compared to the CP trees (Figures 3A and 3B).

336 We observed a wide variation in the relative contribution of each process amongst
337 genes/populations. While the ACMV CP seems to evolve largely due to mutations (only two
338 out of 312 substitutions over the ML tree were due to recombination, a relative contribution
339 of 0.64%; Figure 3A), the CLCuMV CP and, markedly, the CLCuBuV and CLCuGV CP
340 genes depend on both process to generate variability (relative contributions of recombination:
341 34.98, 47.81 and 47.45%, respectively; Figure 3A). We also observed a marked relative
342 contribution of recombination in the SPLCV and TYLCTV CP genes (23.90 and 25.84%,
343 respectively; Figure 3A). In other CP datasets the relative contribution of recombination

344 ranged from 4.42 to 15.84%. These results indicate that the relative contribution of
345 recombination may, in some cases, be equivalent to that of mutation for generating variability
346 in the CP gene.

347 The relative contribution of recombination to molecular variability levels in the Rep
348 gene varied widely amongst datasets. However, we still observed an equivalent relative
349 contribution of recombination in the CLCuBuV Rep gene. Together, these results indicate
350 that the evolution of both CLCuBuV genes is equally dependent on both evolutionary
351 processes. In contrast to the large relative contribution of recombination to the molecular
352 variability in the CLCuGV and CLCuMV CP genes, their Rep genes seem to evolve largely
353 due to mutation.

354 We also observed a high relative contribution of recombination to the molecular
355 variability levels in the ACMV and TYLCCNV Rep genes (33.72 and 39.31%, respectively)
356 and, to a lesser extent, in the SPLCV and TYLCTV Rep genes (25.06 and 32.66%,
357 respectively). Once more, our results indicate a distinct interplay of the evolutionary
358 processes in shaping the molecular variability level amongst genes/populations. Additionally,
359 we conclude that some begomoviruses may largely to depend on the recombination-prone
360 nature of their genomes to evolve.

361

362 **Discussion**

363 Recombination plays an important role in randomizing the variation created by the
364 evolutionary process of mutation (García-Arenal *et al.*, 2003; Worobey & Holmes, 1999). As
365 a consequence, in addition to the rapid mutational dynamics of viral populations (Duffy &
366 Holmes, 2008; Holland *et al.*, 1982; Shackelton & Holmes, 2006; Shackelton *et al.*, 2005),
367 the occurrence of recombination may significantly accelerate their evolution by maximizing
368 the number of combinations amongst the pre-existing alleles. The impact of this process on

369 the evolutionary dynamics of viruses from several families has been extensively documented
370 (Chare & Holmes, 2006; Fan *et al.*, 2007; Heath *et al.*, 2006; Varsani *et al.*, 2006), including
371 for the circular, ssDNA plant geminiviruses (Padidam *et al.*, 1999a; Pita *et al.*, 2001).
372 Recombination events involving strains (Kirthi *et al.*, 2002), species (Monci *et al.*, 2002;
373 Padidam *et al.*, 1999b; Zhou *et al.*, 1997) and genera (Briddon *et al.*, 1996) indicate a
374 significant role of recombination in geminivirus diversification.

375 Biological and epidemiological implications of the occurrence of recombination in
376 geminivirus populations have been documented (Monci *et al.*, 2002; Varsani *et al.*, 2008),
377 suggesting that recombination may be involved in the spread/emergence of new variants of
378 these pathogens (Monjane *et al.*, 2011). It was shown that a natural recombinant between
379 *Tomato yellow leaf curl Sardinia virus* (TYLCSV) and *Tomato yellow leaf curl virus*
380 (TYLCV), exhibiting distinct phenotypic properties from their parents (namely, a wider host
381 range and higher efficiency of insect transmission) became prevalent in common bean in
382 Spain (Monci *et al.*, 2002). More importantly, epidemics caused by geminiviruses around the
383 world frequently involve recombinant viruses (García-Andrés *et al.*, 2007; Idris & Brown,
384 2002; Pita *et al.*, 2001; Sanz *et al.*, 2000; Varsani *et al.*, 2008; Zhou *et al.*, 1997).

385 Although the recombination-prone nature of begomoviruses has been exhaustively
386 demonstrated, no work has attempted to determine the relative contribution of recombination
387 and mutation to the standing molecular variability of begomovirus populations. In this work,
388 by expanding the use of our novel phylogeny-based partitioning method for datasets available
389 in public databases (comprising about 1,800 sequences) we showed that the diversification of
390 begomovirus populations is predominantly driven by mutational dynamics, albeti at different
391 extents. In fact, an equivalent contribution of recombination was demonstrated in some cases.

392 We initiated our analyses by estimating the molecular variability levels in the CP and
393 Rep genes and in full-length DNA-A datasets from begomoviruses sampled around the world

394 and responsible for important epidemics in crops. The pairwise number of nucleotide
395 differences (π) estimated for some datasets was similar to that of plant RNA virus populations
396 which use an error-prone RdRp for replicating their genomes (García-Arenal *et al.*, 2001).
397 For example, the EACMV CP dataset showed molecular variability levels (π = 0.08353)
398 similar to those estimated for the *Rice tungro spherical virus* (RTSV; family *Sequiviridae*,
399 genus *Waikavirus*) CP1 and CP2 genes (π = 0.088 and 0.090, respectively) {Azzam, 2000
400 #10324}. Additionally, molecular variability levels in the EACMV CP were higher than
401 those of the *Citrus tristeza virus* (CTV; family *Closteroviridae*, genus *Closterovirus*) CP (π =
402 0.03792) {Rubio, 2001 #10326}. The nucleotide diversity value obtained for the TYLCCNV
403 CP (π = 0.09712) was close to that estimated for the *Yam mosaic virus* CP (YMV, family
404 *Potyviridae*, genus *Potyvirus*) (Bousalem *et al.*, 2000). Even datasets showing lower
405 molecular variability levels, *eg*. TYLCTV CP (π = 0.02046) and TYLCV CP (π = 0.01756)
406 were similar, for example, to the plant RNA virus *Groundnut rosette virus* (RRV; genus
407 *Umbravirus*) CP gene (π = 0.018) {Deom, 2000 #10327}. Our results provide further
408 evidence that begomovirus populations are highly variable even though these viruses
409 replicate using the supposedly proof-reading DNA polymerases from their hosts (Gutierrez,
410 1999; 2000; Hanley-Bowdoin *et al.*, 1999). In addition, the wide variation of nucleotide
411 diversity values from our analysis suggests an interplay of distinct evolutionary processes in
412 shaping the molecular variability levels in each population/gene.

413 It has been shown that the main type of selection acting on the most viral genes is
414 purifying selection (García-Andrés *et al.*, 2007; García-Arenal *et al.*, 2003). However, the
415 wide variation in molecular variability levels from both begomovirus genes (Rep and CP)
416 prompted us to assess the possible contribution of adaptive selection, since each gene might
417 be under different selective constraints amongst different datasets. Surprisingly, very high ω
418 values were found for the CLCuBuV CP and Rep genes (ω = 0.684067 and 0.615212,

419 respectively), the least variable datasets from our analyses. The ω value estimated for the
420 CLCuBuV CP was much higher even compared to non-vector-borne plant RNA viruses such
421 as *Prunus necrotic ringspot virus* (PNRSV; family *Bromoviridae*, genus *Ilarvirus*; ω = 0.375)
422 and *Pepper mild mottle virus* (Genus *Tobamovirus*, ω = 0.301) (Chare & Holmes, 2004). This
423 value was also more than two times higher than that estimated for a *Tomato severe rugose*
424 *mosaic virus* CP dataset (Lima *et al*., *submitted*). Curiously, the ω value estimated for the
425 CLCuBuV CP was higher than that of its Rep gene, an unusual pattern since stronger
426 negative selection seems to act on the CP gene in most datasets. In fact, evidence of stronger
427 negative selection has been reported for structural genes across virus families. The ω value
428 estimated for a CTV CP dataset (ω = 0.02739) was markedly lower than that estimated for its
429 methyltransferase protein involved in virus replication (ω = 0.12302) {Rubio, 2001 #10326},
430 revealing markedly distinct selective constraints amongst structural and replication-related
431 genes. In addition, the ω value estimated for both genes of the most variable datasets were
432 smaller and/or similar to those of datasets exhibiting lower variability levels.
433 A conserved uneven distribution of molecular variability levels across the length of
434 the CP and Rep genes was readily evident from our analyses, suggesting that in addition to
435 the distinct contribution of the CP and Rep genes for the standing molecular variability of the
436 full-length DNA-A, there was also a distinct interplay of the evolutionary processes in
437 shaping the molecular variability in specific regions of both genes. The central/3'-terminal
438 region of the CP gene and the 5 '-terminal region of the Rep gene were often more variable
439 than other regions of these genes in most datasets analyzed.
440 Although purifying selection seems to act on the CP and Rep genes of all datasets, we
441 also evaluated the possible contribution of adaptive selection to the highly variable regions in
442 the CP and Rep genes, since a small fraction of codons could be under different selective
443 constraints. These analyses did not uncover evidence of adaptive selection. On the contrary, a

444 high number of negatively selected sites co-localized with the highly variable regions on the
445 CP and Rep genes. Additionally, our analyses revealed a high synonymous site variation in
446 these regions, suggesting strong purifying selection.

447 Our results strongly indicate the recombination as the evolutionary process
448 responsible for shaping the uneven molecular variability levels across the CP and Rep genes.
449 We observed a clear correlation between the uneven distribution of the molecular variability
450 levels in both genes and the location of recombination breakpoints.

451 The non-random location of recombination breakpoints has been shown to be a
452 conserved feature amongst ssDNA viruses which use a rolling circle mechanism for
453 replicating their genomes (Lefeuvre *et al.*, 2007a; Lefeuvre *et al.*, 2007b; Martin *et al.*,
454 2011b; Prasanna & Rai, 2007). In fact, most recombination events detected from our analyses
455 had at least one breakpoint in the 5'- or 3'-ends of the origin of replication and in the 5'-end of
456 the Rep gene. This region of the Rep gene is responsible for encoding the Rep catalytic
457 domain, which spans three motifs conserved in rolling circle replication-associated proteins
458 (Ilyina & Koonin, 1992) and the DNA binding specificity determinants involved in high
459 affinity DNA binding (Londono *et al.*, 2010 9477). It has been shown that recombinants
460 whose intra-genomic interactions network is not disrupted by recombination are favored by
461 selection (Martin *et al.*, 2011b). In fact, the exchange of regions involved in highly specific
462 interactions as linked blocks seems to increase the frequency in which they work
463 appropriately in a new genomic context.

464 The 5' and (less frequently) 3'-ends of the common region were also targeted by a
465 number of recombination events. The mechanistic aspects that make this region
466 recombination-prone are unknown. However, its role during the viral replication cycle
467 (through rolling circle replication and/or recombination-dependent replication) could be at
468 least part of the reason (Lefeuvre *et al.*, 2009). Furthermore, it has been observed that

469 recombination tend not to occur within coding sequences, as it could potentially disrupt gene
470 function (Lefeuvre *et al.*, 2007a). In agreement, the segment between the CP and Ren genes
471 has also been implicated as a recombination hotspot (Lefeuvre *et al.*, 2007a; Lefeuvre *et al.*,
472 2009).

473 The viability of recombinant genomes carrying breakpoints at the interface between
474 the CP and Ren genes and the 3'-end of the common region has been demonstrated for
475 begomoviruses causing cassava mosaic disease (CMD) under natural conditions. A well-
476 supported recombination event (p=1.04×10^{-26}) was detected in these isolates, having as major
477 and minor parents the begomoviruses ACMV and CLCuGV, respectively (Tiendrebeogo *et*
478 *al.*, 2012). This event was also detected in our analysis and considerably increased the
479 molecular variability levels of the ACMV Rep gene.

480 Although genomes containing breakpoints in the CP gene have been less frequently
481 detected in our analysis (in comparison to those with breakpoints in the 5'-end of the Rep
482 gene), we also observed a markedly high number of breakpoints falling into the CP gene
483 central and 3'-regions. We observed a significant contribution of these events to the molecular
484 variability levels in the EACMV and CLCuGV CP datasets. EACMV isolates carrying a
485 recombinant event in their CP gene have been implicated in the severe epidemics of CMD in
486 Uganda during the 1990's (Pita *et al.*, 2001).

487 Our results suggest that recombination explains a considerable fraction of the
488 molecular variability of begomovirus populations by significantly increasing the molecular
489 variability levels in the CP and Rep genes. However, there was no direct correlation between
490 the molecular variability levels of the viral populations and the number of recombination
491 events detected, indicating that the relative contribution of the recombination is not a function
492 of the number of unique recombination events.

493 We have recently developed a novel phylogeny-based partitioning method to
494 discriminate the individual contribution of mutation and recombination to the evolutionary
495 dynamics of begomovirus populations. Using this method, we have demonstrated that the
496 relative contribution of recombination may be equivalent to that of mutation in specific
497 regions of the viral genome, although the mutational dynamics has been indicated as the
498 primary source of variation for most viral genes (Lima *et al*. 2012, *submitted*). Once again,
499 we took advantage of the fact that recombinant isolates often form distinct clades on
500 maximum likelihood trees, separated by long branches whose associated substitutions were
501 clearly due to recombination, and we counted the number of substitutions within blocks
502 inferred as recombinants.

503 The relative contribution of mutation and recombination was widely variable among
504 populations/viral genes. Contrary to our expectations, we did not observe a significantly
505 higher relative contribution of recombination for the Rep gene datasets, although it has been
506 observed in absolute terms. Interestingly, our analyses indicated recombination as a source of
507 up to 50% of all inferred substitutions for the CLCuGV CP dataset and up to 40% for the
508 TYLCCNV Rep dataset. Therefore, it seems that the evolution of some begomovirus
509 populations is dependent on the recombination-prone nature of their genomes in addition to
510 their rapid mutational dynamics. On the other hand, we also observed that highly variable
511 populations might evolve largely as a function of their mutational dynamics (*eg*, in the
512 TYLCCNV Rep and CP genes datasets more than 90% of the substitutions are due to
513 mutation). However, it should be noted that our estimations reflect the minimal relative
514 contribution of recombination and therefore do not take into account the fraction of
515 recombination events which are not statistically detectable in the datasets.

516 Contrary to the premise that the rapid evolution of viral populations is essentially a
517 consequence of their rapid mutational dynamics, we have shown that some begomovirus

518 populations depend largely on the recombination-prone nature of their genomes to evolve.
519 Additionally, we show that a considerable fraction (up to 50%) of the substitutions over the
520 begomovirus phylogeny is undoubtedly associated with recombination. Our results have
521 important implications on the reconstruction of the evolutionary history of these pathogens,
522 since the occurrence of recombination violates the assumptions of phylogenetic
523 reconstruction methods and significantly affects their accuracy (Posada & Crandall, 2002
524 10071). It may be time to move away from bifurcating trees for intra- and interspecific
525 studies of begomoviruses, towards network-based models.
526

527 **References**

528

529 **Balol, G. B., Divya, B. L., Basavaraj, S., Sundaresha, S., Mahesh, Y. S., Erayya &**
530 **Huchannanavar, S. D. (2010).** Sources of genetic variation in plant virus
531 populations. *J Pure Appl Microbiol* **4**, 803-808.
532 **Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001).** Complete nucleotide sequence and
533 host range of South African cassava mosaic virus: further evidence for recombination
534 amongst begomoviruses. *J Gen Virol* **82**, 53-58.
535 **Biebricher, C. K. & Eigen, M. (2006).** What is a quasispecies? *Quasispecies: Concept and*
536 *Implications for Virology* **299**, 1-31.
537 **Boni, M. F., Posada, D. & Feldman, M. W. (2007).** An exact nonparametric method for
538 inferring mosaic structure in sequence triplets. *Genetics* **176**, 1035-1047.
539 **Bonnet, J., Fraile, A., Sacristan, S., Malpica, J. M. & Garcia-Arenal, F. (2005).** Role of
540 recombination in the evolution of natural populations of Cucumber mosaic virus, a
541 tripartite RNA plant virus. *Virology* **332**, 359-368.
542 **Bousalem, M., Douzery, E. J. P. & Fargette, M. (2000).** High genetic diversity, distant
543 phylogenetic relationships and intraspecies recombination events among natural
544 populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution.
545 *J Gen Virol* **81**, 243-255.
546 **Briddon, R. W., Bedford, I. D., Tsai, J. H. & Markham, P. G. (1996).** Analysis of the
547 nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly
548 top virus, suggests a recombinant origin. *Virology* **219**, 387-394.
549 **Bukh, J., Miller, R. H. & Purcell, R. H. (1995).** Genetic heterogeneity of hepatitis C virus:
550 quasispecies and genotypes. *Seminars in liver disease* **15**, 41-63.
551 **Chare, E. R. & Holmes, E. C. (2004).** Selection pressures in the capsid genes of plant RNA
552 viruses reflect mode of transmission. *J Gen Virol* **85**, 3149-3157.
553 **Chare, E. R. & Holmes, E. C. (2006).** A phylogenetic survey of recombination frequency in
554 plant RNA viruses. *Arch Virol* **151**, 933-946.
555 **Davino, S., Napoli, C., Dellacroce, C., Miozzi, L., Noris, E., Davino, M. & Accotto, G. P.**
556 **(2009).** Two new natural begomovirus recombinants associated with the tomato

557 yellow leaf curl disease co-exist with parental viruses in tomato epidemics in Italy.
558 *Virus Res* **143**, 15-23.
559 **Domingo, E. (2002).** Quasispecies theory in virology. *J Virol* **76**, 463-465.
560 **Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. (1978).** Nucleotide sequence
561 heterogeneity of an RNA phage population. *Cell* **13**, 735-744.
562 **Drake, J. W. (1991).** A constant rate of spontaneous mutation in DNA-based microbes. *Proc*
563 *Natl Acad Sci USA* **88**, 7160-7164.
564 **Duffy, S. & Holmes, E. C. (2008).** Phylogenetic evidence for rapid rates of molecular
565 evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. *J*
566 *Virol* **82**, 957-965.
567 **Duffy, S. & Holmes, E. C. (2009).** Validation of high rates of nucleotide substitution in
568 geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *J*
569 *Gen Virol* **90**, 1539-1547.
570 **Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008).** Rates of evolutionary change in
571 viruses: Patterns and determinants. *Nat Rev Genet* **9**, 267-276.
572 **Edgar, R. C. (2004).** MUSCLE: A multiple sequence alignment method with reduced time
573 and space complexity. *Bmc Bioinformatics* **5**, 1-19.
574 **Eigen, M. (1971).** Selforganization of Matter and Evolution of Biological Macromolecules.
575 *Naturwissenschaften* **58**, 465-&.
576 **Eigen, M., Winkler-Oswatitsch, R. & Dress, A. (1988).** Statistical geometry in sequence
577 space: a method of quantitative comparative sequence analysis. *Proc Natl Acad Sci U*
578 *S A* **85**, 5913-5917.
579 **Fan, J., Negroni, M. & Robertson, D. L. (2007).** The distribution of HIV-1 recombination
580 breakpoints. *Infect Genet Evol* **7**, 717-723.
581 **García-Andrés, S., Accotto, G. P., Navas-Castillo, J. & Moriones, E. (2007).** Founder
582 effect, plant host, and recombination shape the emergent population of begomoviruses
583 that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology*
584 **359**, 302-312.
585 **García-Andrés, S., Monci, F., Navas-Castillo, J. & Moriones, E. (2006).** Begomovirus
586 genetic diversity in the native plant reservoir *Solanum nigrum*: Evidence for the
587 presence of a new virus species of recombinant nature. *Virology* **350**, 433-442.
588 **García-Arenal, F., Fraile, A. & Malpica, J. M. (2001).** Variability and genetic structure of
589 plant virus populations. *Annual Review of Phytopathology* **39**, 157-186.
590 **García-Arenal, F., Fraile, A. & Malpica, J. M. (2003).** Variation and evolution of plant
591 virus populations. *Int Microbiol* **6**, 225-232.
592 **Ge, L. M., Zhang, J. T., Zhou, X. P. & Li, H. Y. (2007).** Genetic structure and population
593 variability of tomato yellow leaf curl China virus. *J Virol* **81**, 5902-5907.
594 **Gerrish, P. J. & Garcia-Lerma, J. G. (2003).** Mutation rate and the efficacy of
595 antimicrobial drug treatment. *The Lancet infectious diseases* **3**, 28-32.
596 **Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000).** Sister-scanning: a Monte Carlo
597 procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573-
598 582.
599 **Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H. J. & Gronenborn,**
600 **B. (2010).** High variability and rapid evolution of a nanovirus. *J Virol* **84**, 9105-9117.
601 **Gutierrez, C. (1999).** Geminivirus DNA replication. *Cellular and Molecular Life Sciences*
602 **56**, 313-329.
603 **Gutierrez, C. (2000).** Geminiviruses and the plant cell cycle. *Plant Mol Biol* **43**, 763-772.
604 **Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S. & Robertson, D. (1999).**
605 Geminiviruses: Models for plant DNA replication, transcription, and cell cycle
606 regulation. *Critical Reviews in Plant Sciences* **18**, 71-106.

607 **Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. (2006).** Recombination patterns
608 in aphthoviruses mirror those found in other picornaviruses. *J Virol* **80**, 11827-11832.
609 **Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982).**
610 Rapid evolution of RNA genomes. *Science* **215**, 1577-1585.
611 **Holmes, E. C. (2010).** The RNA virus quasispecies: fact or fiction? *J Mol Biol* **400**, 271-273.
612 **Hull, R. (2009).** *Comparative Plant Virology*. London: Elsevier.
613 **Idris, A. M. & Brown, J. K. (2002).** Molecular analysis of Cotton leaf curl virus-Sudan
614 reveals an evolutionary history of recombination. *Virus Genes* **24**, 249-256.
615 **Ilyina, T. V. & Koonin, E. V. (1992).** Conserved sequence motifs in the initiator proteins for
616 rolling circle DNA replication encoded by diverse replicons from eubacteria,
617 eucaryotes and archaebacteria. *Nucleic Acids Res* **20**, 3279-3285.
618 **Isnard, M., Granier, M., Frutos, R., Reynaud, B. & Peterschmitt, M. (1998).**
619 Quasispecies nature of three maize streak virus isolates obtained through different
620 modes of selection from a population used to assess response to infection of maize
621 cultivars. *J Gen Virol* **79**, 3091-3099.
622 **Jeske, H., Lutgemeier, M. & Preiss, W. (2001).** DNA forms indicate rolling circle and
623 recombination-dependent replication of Abutilon mosaic virus. *Embo Journal* **20**,
624 6158-6167.
625 **Kirthi, N., Maiya, S. P., Murthy, M. R. & Savithri, H. S. (2002).** Evidence for
626 recombination among the tomato leaf curl virus strains/species from Bangalore, India.
627 *Arch Virol* **147**, 255-272.
628 **Kosakovsky-Pond, S. L. & Frost, S. D. W. (2005).** Not so different after all: A comparison
629 of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-
630 1222.
631 **Lauring, A. S. & Andino, R. (2010).** Quasispecies Theory and the Behavior of RNA
632 Viruses. *PLoS Pathog* **6**.
633 **Lefeuvre, P., Lett, J. M., Reynaud, B. & Martin, D. P. (2007a).** Avoidance of protein fold
634 disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.
635 **Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009).** Widely conserved
636 recombination patterns among single-stranded DNA viruses. *J Virol* **83**, 2697-2707.
637 **Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A.,**
638 **Becker, N., Reynaud, B. & Lett, J. M. (2007b).** Begomovirus 'melting pot' in the
639 south-west Indian Ocean islands: Molecular diversity and evolution through
640 recombination. *J Gen Virol* **88**, 3458-3468.
641 **Londono, A., Riego-Ruiz, L. & Arguello-Astorga, G. R. (2010).** DNA-binding specificity
642 determinants of replication proteins encoded by eukaryotic ssDNA viruses are
643 adjacent to widely separated RCR conserved motifs. *Arch Virol* **155**, 1033-1046.
644 **Lozano, G., Trenado, H. P., Valverde, R. A. & Navas-Castillo, J. (2009).** Novel
645 begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and
646 *Ipomoea indica*: Taxonomic and phylogenetic implications. *J Gen Virol* **90**, 2550-
647 2562.
648 **Martin, D. & Rybicki, E. P. (2000).** RDP: detection of recombination amongst aligned
649 sequences. *Bioinformatics* **16**, 562-563.
650 **Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P. & Varsani, A.**
651 **(2011a).** Recombination in Eukaryotic Single Stranded DNA Viruses. *Viruses-Basel*
652 **3**, 1699-1738.
653 **Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B.,**
654 **Vincent, C., Reynaud, B. & Lett, J. M. (2011b).** Complex recombination patterns
655 arising during geminivirus coinfections preserve and demarcate biologically important
656 intra-genome interaction networks. *PLoS Pathog* **7**, e1002203.

657 **Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010).**
658 RDP3: A flexible and fast computer program for analyzing recombination.
659 *Bioinformatics* **26**, 2462-2463.
660 **Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005).** A modified bootscan
661 algorithm for automated identification of recombinant sequences and recombination
662 breakpoints. *AIDS Res Hum Retroviruses* **21**, 98-102.
663 **Monci, F., Sanchez-Campos, S., Navas-Castillo, J. & Moriones, E. (2002).** A natural
664 recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and
665 *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming
666 prevalent in Spanish populations. *Virology* **303**, 317-326.
667 **Monjane, A. L., van der Walt, E., Varsani, A., Rybicki, E. P. & Martin, D. P. (2011).**
668 Recombination hotspots and host susceptibility modulate the adaptive value of
669 recombination during maize streak virus evolution. *BMC Evolutionary Biology* **11**.
670 **Navas-Castillo, J., Sanchez-Campos, S., Noris, E., Louro, D., Accotto, G. P. & Moriones,**
671 **E. (2000).** Natural recombination between Tomato yellow leaf curl virus-is and
672 Tomato leaf curl virus. *J Gen Virol* **81**, 2797-2801.
673 **Padidam, M., Beachy, R. N. & Fauquet, C. M. (1999a).** A phage single-stranded DNA
674 (ssDNA) binding protein complements ssDNA accumulation of a geminivirus and
675 interferes with viral movement. *J Virol* **73**, 1609-1616.
676 **Padidam, M., Sawyer, S. & Fauquet, C. M. (1999b).** Possible emergence of new
677 geminiviruses by frequent recombination. *Virology* **265**, 218-224.
678 **Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M.**
679 **(2001).** Recombination, pseudorecombination and synergism of geminiviruses are
680 determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen*
681 *Virol* **82**, 655-665.
682 **Posada, D. & Crandall, K. A. (1998).** MODELTEST: Testing the model of DNA
683 substitution. *Bioinformatics* **14**, 817-818.
684 **Posada, D. & Crandall, K. A. (2001).** Evaluation of methods for detecting recombination
685 from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757-
686 13762.
687 **Posada, D. & Crandall, K. A. (2002).** The effect of recombination on the accuracy of
688 phylogeny estimation. *Journal of Molecular Evolution* **54**, 396-402.
689 **Prasanna, H. C. & Rai, M. (2007).** Detection and frequency of recombination in tomato-
690 infecting begomoviruses of South and Southeast Asia. *Virol J* **4**, 111.
691 **Roossinck, M. J. (1997).** Mechanisms of plant virus evolution. *Annual Review of*
692 *Phytopathology* **35**, 191-209.
693 **Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003).** DnaSP: DNA
694 polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**,
695 2496-2497.
696 **Sanz, A. I., Fraile, A., García-Arenal, F., Zhou, X., Robinson, D. J., Khalid, S., Butt, T.**
697 **& Harrison, B. D. (2000).** Multiple infection, recombination and genome
698 relationships among begomovirus isolates found in cotton and other plants in
699 Pakistan. *J Gen Virol* **81**, 1839-1849.
700 **Saunders, K., Bedford, I. D. & Stanley, J. (2001).** Pathogenicity of a natural recombinant
701 associated with ageratum yellow vein disease: implications for geminivirus evolution
702 and disease aetiology. *Virology* **282**, 38-47.
703 **Scheffler, K., Martin, D. P. & Seoighe, C. (2006).** Robust inference of positive selection
704 from recombining coding sequences. *Bioinformatics* **22**, 2493-2499.
705 **Shackelton, L. A. & Holmes, E. C. (2006).** Phylogenetic evidence for the rapid evolution of
706 human B19 erythrovirus. *J Virol* **80**, 3666-3669.

707 **Shackelton, L. A., Parrish, C. R., Truyen, U. & Holmes, E. C. (2005).** High rate of viral
708 evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci*
709 *U S A* **102**, 379-384.
710 **Smith, J. M. (1992).** Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.
711 **Steinhauer, D. A., de la Torre, J. C., Meier, E. & Holland, J. J. (1989).** Extreme
712 heterogeneity in populations of vesicular stomatitis virus. *J Virol* **63**, 2072-2080.
713 **Swofford, D. L. (2003).** PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
714 Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
715 **Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011).**
716 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood,
717 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-
718 2739.
719 **Tiendrebeogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A.,**
720 **Villemot, J., Traore, V. S., Konate, G., Traore, A. S., Barro, N., Reynaud, B.,**
721 **Traore, O. & Lett, J. M. (2012).** Evolution of African cassava mosaic virus by
722 recombination between bipartite and monopartite begomoviruses. *Virol J* **9**, 67.
723 **van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E. & Rybicki, E. P. (2008).**
724 Experimental observations of rapid Maize streak virus evolution reveal a strand-
725 specific nucleotide substitution bias. *Virol J* **5**, 104.
726 **Varsani, A., Shepherd, D. N., Monjane, A. L., Owor, B. E., Erdmann, J. B., Rybicki, E.**
727 **P., Peterschmitt, M., Briddon, R. W., Markham, P. G., Oluwafemi, S., Windram,**
728 **O. P., Lefeuvre, P., Lett, J. M. & Martin, D. P. (2008).** Recombination, decreased
729 host specificity and increased mobility may have driven the emergence of maize
730 streak virus as an agricultural pathogen. *J Gen Virol* **89**, 2063-2074.
731 **Varsani, A., van der Walt, E., Heath, L., Rybicki, E. P., Williamson, A. L. & Martin, D.**
732 **P. (2006).** Evidence of ancient papillomavirus recombination. *The Journal of general*
733 *virology* **87**, 2527-2531.
734 **Wain-Hobson, S. (1992).** Human immunodeficiency virus type 1 quasispecies in vivo and ex
735 vivo. *Curr Top Microbiol Immunol* **176**, 181-193.
736 **Worobey, M. & Holmes, E. C. (1999).** Evolutionary aspects of recombination in RNA
737 viruses. *J Gen Virol* **80**, 2535-2545.
738 **Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J. &**
739 **Harrison, B. D. (1997).** Evidence that DNA-A of a geminivirus associated with
740 severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J*
741 *Gen Virol* **78**, 2101-2111.
742
743

Table 1. Mean dN/dS for the CP and Rep genes of begomovirus populations retrieved from the Genbank database.

Population	CP	Rep
ACMV	0.204844	0.212786
AgEV	0.068430	0.291057
BhYVIV	0.164179	0.273681
BhYMIV	0.133052	0.228139
CLCuBuV	0.684067	0.615212
CLCuGV	0.058772	0.178010
CLCuMV	0.124233	0.312296
EACMV		
MYMIV	0.140563	0.208555
SPLCV	0.074497	0.185277
ToLCNDV	0.109989	0.181742
ToLCTV	0.143665	0.229201
TYLCTV	0.102353	0.260096
ToYLCCV	0.102831	0.235654
TYLCV		

Supplementary Table S1. Molecular variability in begomovirus populations retrieved from the GenBank database.

Population	Number of sequences	DNA-A H_d	DNA-A π	CP π	Rep π
ACMV	45	1.000 (±0.005)	0.04816 (±0.00665)	0.03013 (±0.00162)	0.06352 (±0.01340)
AgEV	21	0.995 (±0.016)	0.04172 (±0.00330)	0.03057 (±0.00522)	0.03541 (±0.00269)
BhYVIV	46	0.995 (±0.006)	0.06293 (±0.00569)	0.05010 (±0.00727)	0.07249 (±0.01034)
BhYMIV	40	0.996 (±0.007)	0.06502 (±0.00611)	0.03895 (±0.00566)	0.09400 (±0.01012)
CLCuBuV	53	0.998 (±0.005)	0.01698 (±0.00180)	0.01551 (±0.00405)	0.01740 (±0.00217)
CLCuGV	39	0.969 (±0.020)	0.04290 (±0.00308)	0.05102 (±0.00720)	0.04257 (±0.00313)
CLCuMV	34	0.995 (±0.009)	0.05535 (±0.00572)	0.04836 (±0.00798)	0.06210 (±0.00612)
EACMV	157	0.999 (±0.001)	0.05672 (±0.00141)	0.08353 (±0.00271)	0.05038 (±0.00162)
MYMIV	42	1.000 (±0.005)	0.03894 (±0.00135)	0.03774 (±0.00176)	0.03614 (±0.00144)
SPLCV	42	0.997 (±0.007)	0.07602 (±0.00696)	0.06109 (±0.00533)	0.07690 (±0.00951)
ToLCNDV	52	0.999 (±0.004)	0.07661 (±0.00999)	0.05353 (±0.00814)	0.07446 (±0.00785)
ToLCTV	39	1.000 (±0.006)	0.03958 (±0.00408)	0.02510 (±0.00260)	0.04071 (±0.00616)
TYLCTV	35	1.000 (±0.007)	0.02643 (±0.00650)	0.02046 (±0.00585)	0.03065 (±0.00763)
ToYLCCV	26	0.994 (±0.013)	0.10414 (±0.00774)	0.09712 (±0.00576)	0.10569 (±0.01350)
TYLCV	222	0.999 (±0.001)	0.04049 (±0.00306)	0.01756 (±0.00082)	0.05669 (±0.00494)

Figure Legends

Figure 1. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) for the CP and Rep sequences from begomovirus datasets retrieved from the Genbank database.

The standard deviation for each π value is shown.

Figure 2. Number of sites under statistically significant negative selection in the CP and Rep genes, as detected by the Single Likelihood Ancestor Counting (SLAC) method.

Figure 3. Relative contribution of mutation and recombination to the standing molecular variability of the CP (A) and Rep (B) genes of begomovirus datasets retrieved from the Genbank database.

Supplementary Figure S1. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across the CP sequences from begomovirus datasets retrieved from the GenBank database.

Supplementary Figure S2. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across the Rep sequences from begomovirus datasets retrieved from the Genbank database.

Supplementary Figure S3. dN-dS values calculated across the codons of the CP sequences for begomovirus datasets retrieved from the Genbank database using the Single Likelihood Ancestor Counting (SLAC) method.

Supplementary Figure S4. dN-dS values calculated across the codons of the Rep sequences for begomovirus datasets retrieved from the Genbank database using the Single Likelihood Ancestor Counting (SLAC) method.

Supplementary Figure S5. Midpoint-rooted maximum likelihood trees based on the CP **(left)** and Rep **(right)** nucleotide sequences from begomovirus datasets retrieved from the Genbank database. (A*) African cassava mosaic virus*, ACMV; (B) *Ageratum enation virus*, AgEV; (C) *Bhendi yellow vein India virus*, BhYVIV; (D) *Bhendi yellow mosaic India virus*, BhYVMIV; (E) *Cotton leaf curl Burewala virus*, CLCuBuV; (F) *Cotton leaf curl Gezira virus*, CLCuGV; (G) *Cotton leaf curl Multan virus*, CLCuMV; (H) *East African cassava mosaic virus*, EACMV; (I) *Mungbean yellow mosaic India virus*, MYMIV; (J) *Sweet potato leaf curl virus*, SPLCV; (K) *Tomato leaf curl New Delhi virus*, ToLCNDV; (L) *Tomato leaf curl Taiwan virus*, ToLCTV; (M) *Tomato yellow leaf curl Thailand virus*, TYLCTV; (N) *Tomato yellow leaf curl China virus*, TYLCCNV; (O) *Tomato yellow leaf curl virus*, TYLCV.

Figure 1

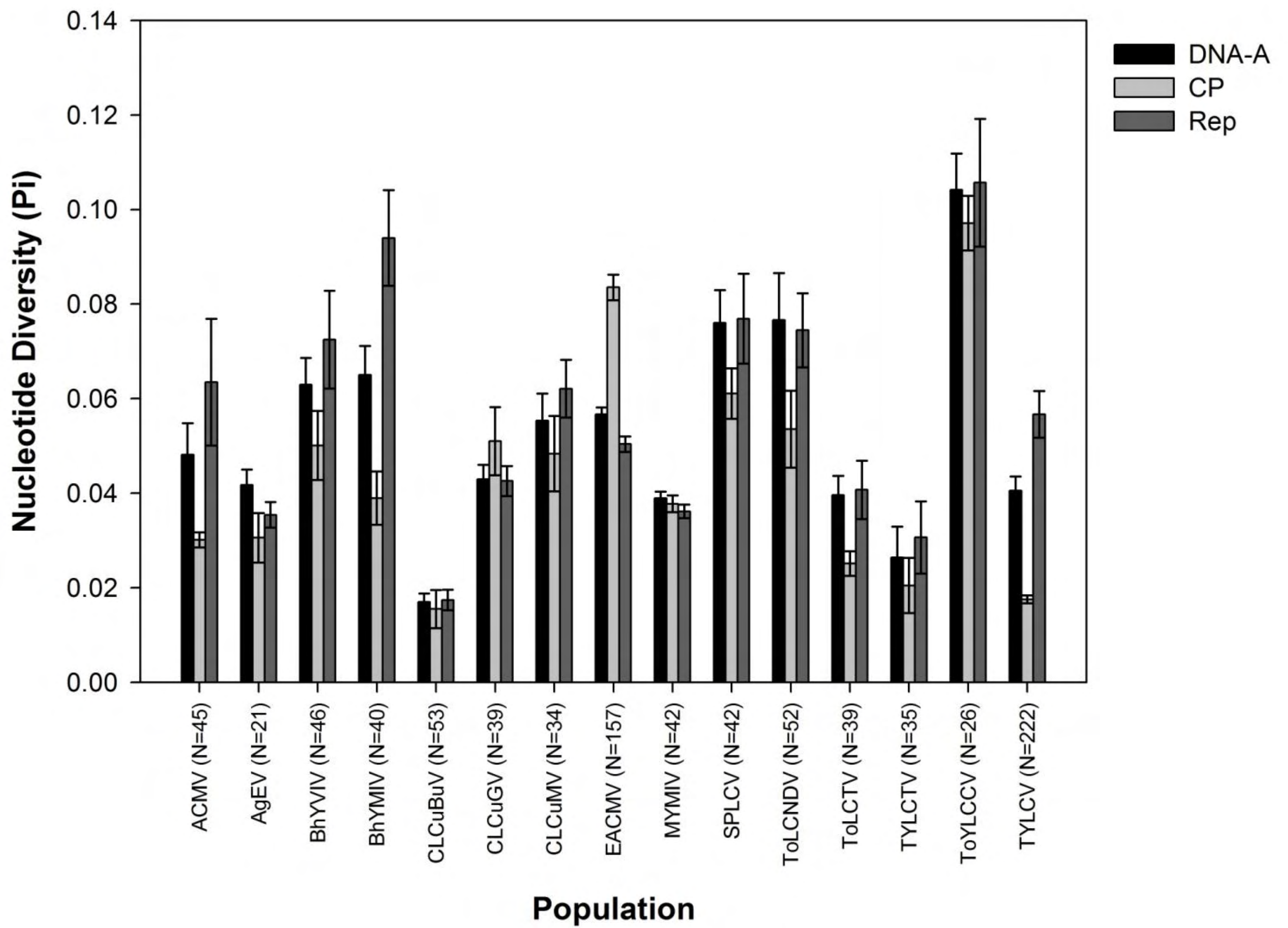


Figure 2

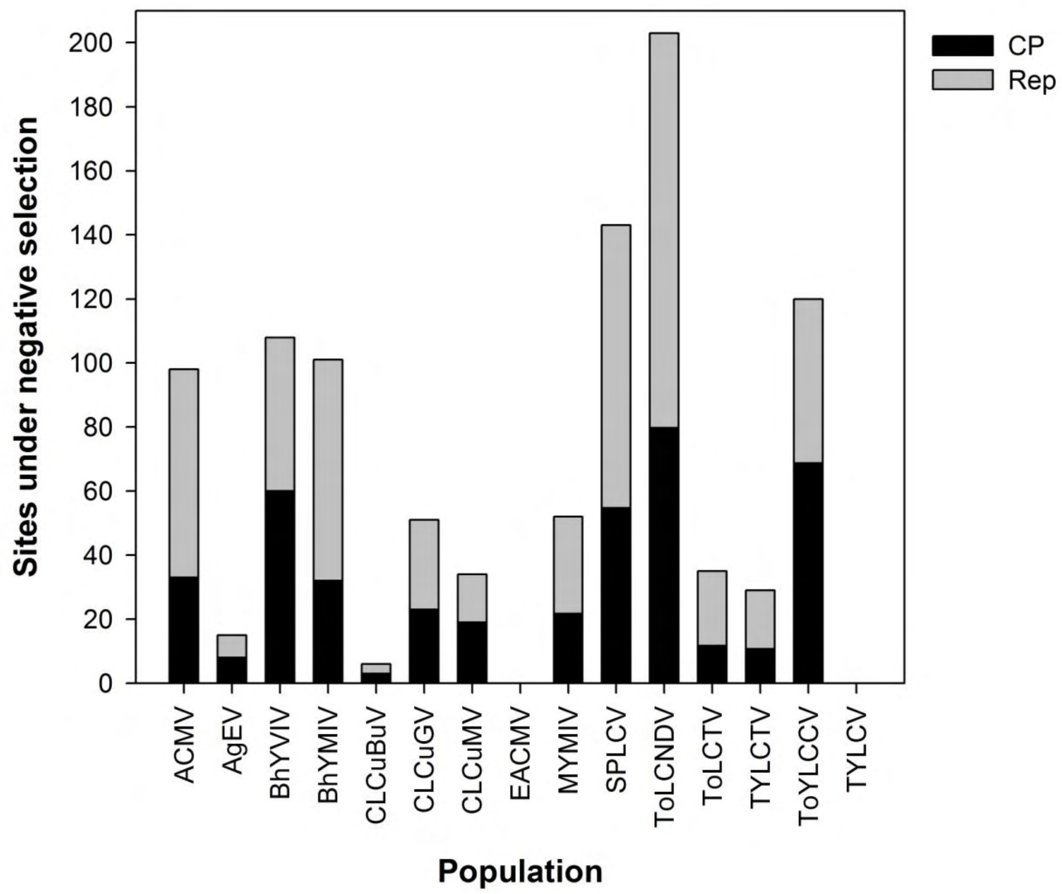
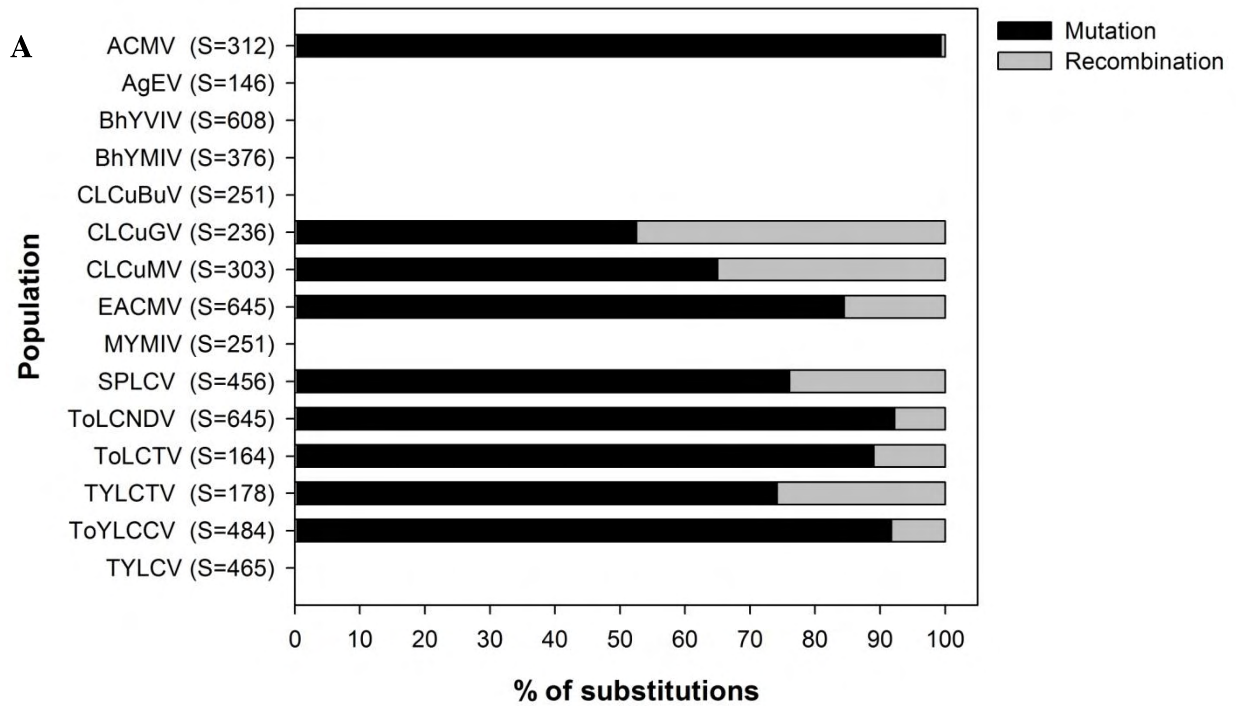
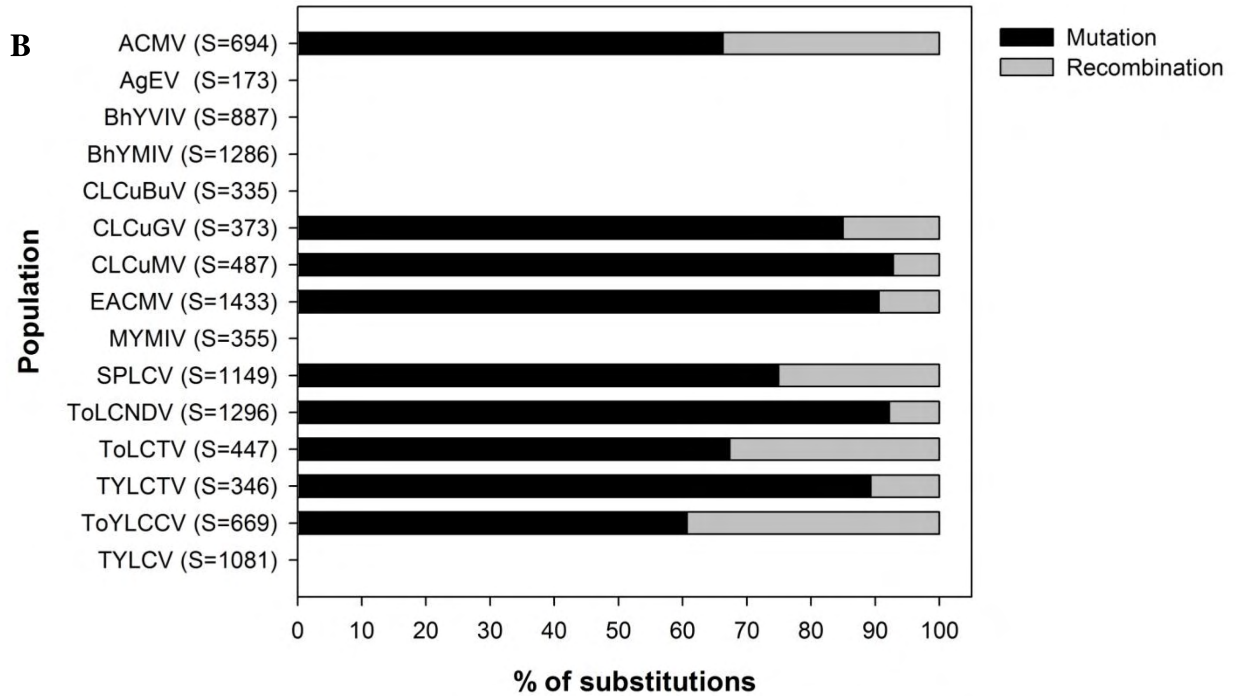
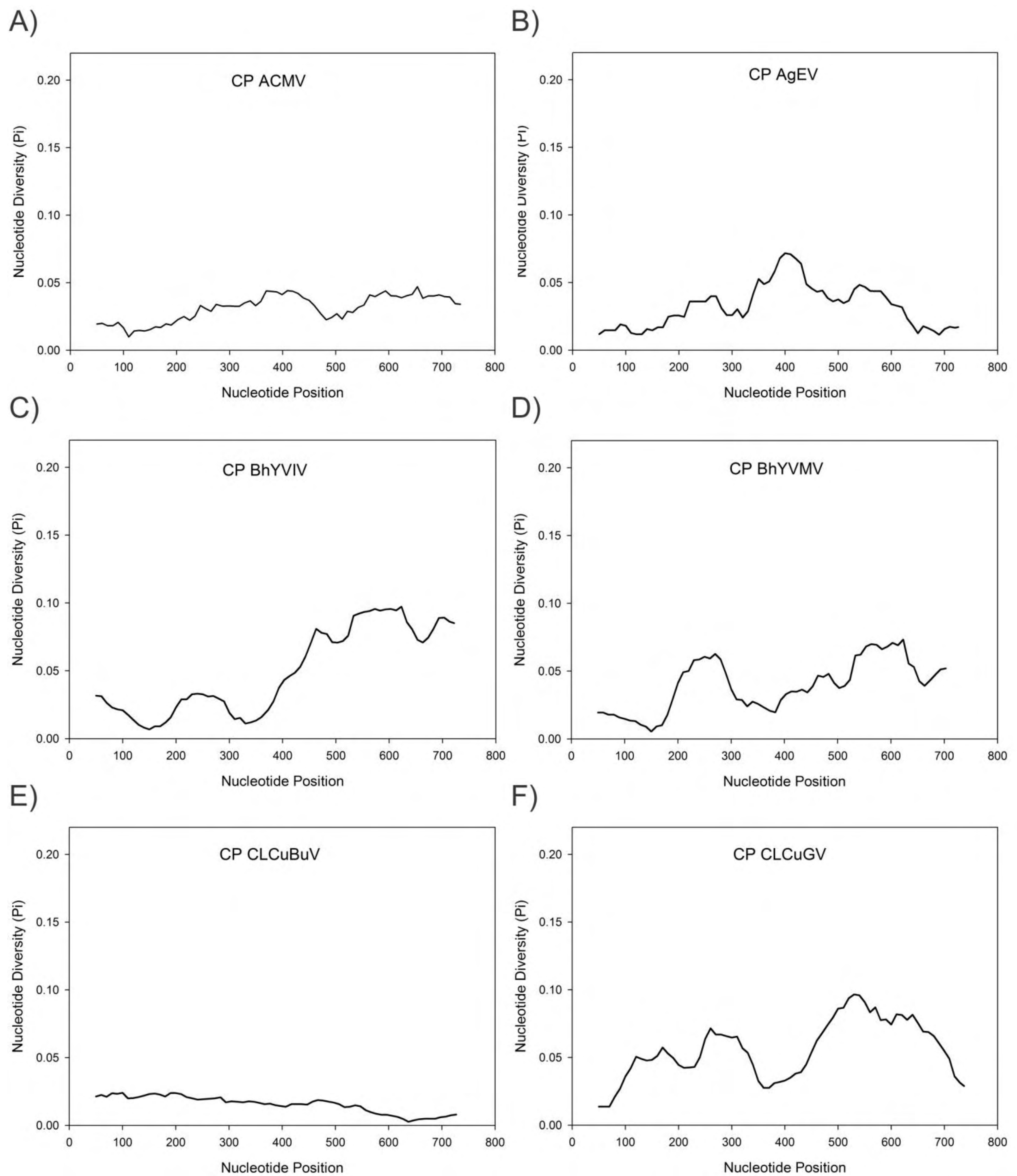


Figure 3

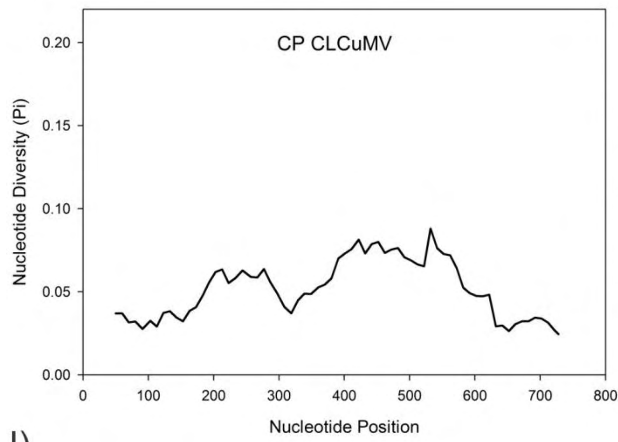




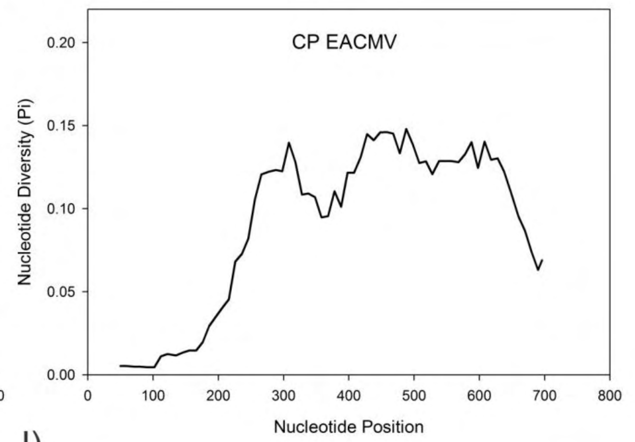
Supplementary Figure S1



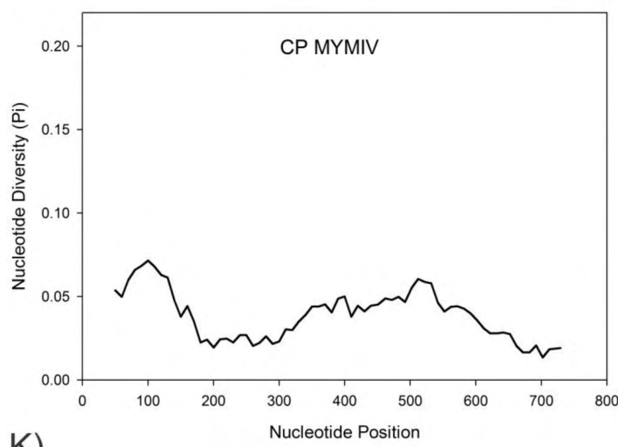
G)



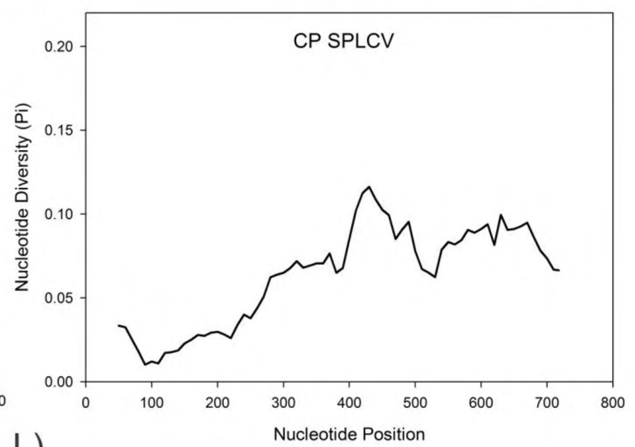
H)



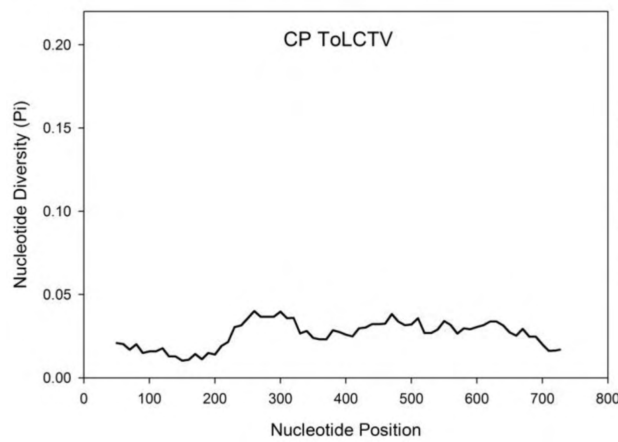
I)



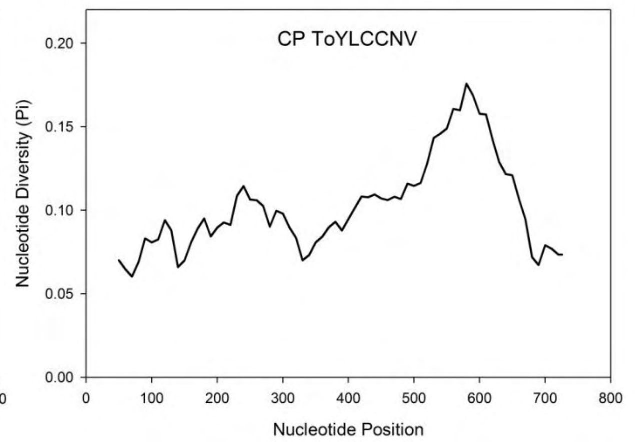
J)



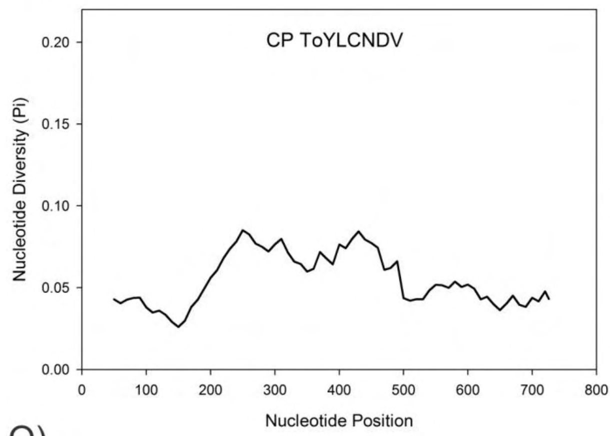
K)



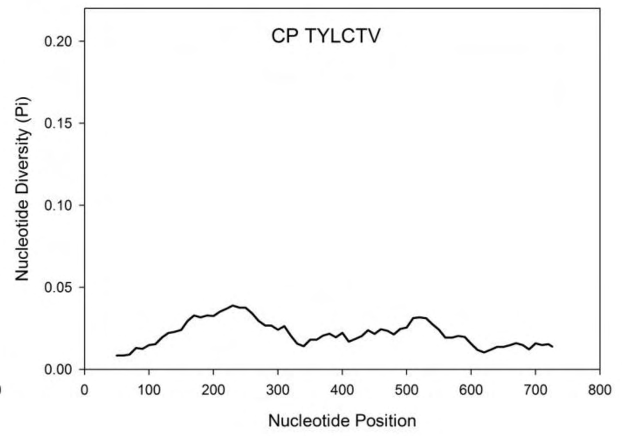
L)



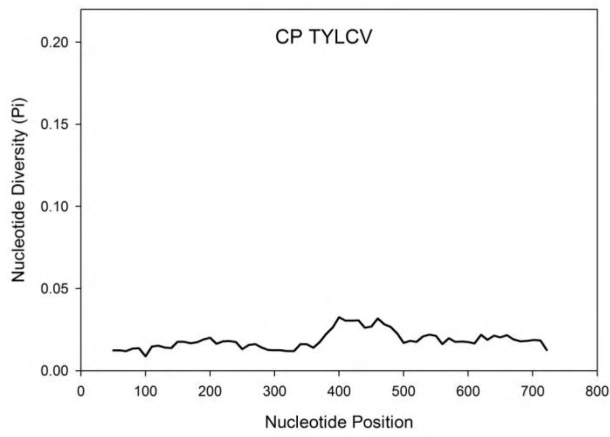
M)



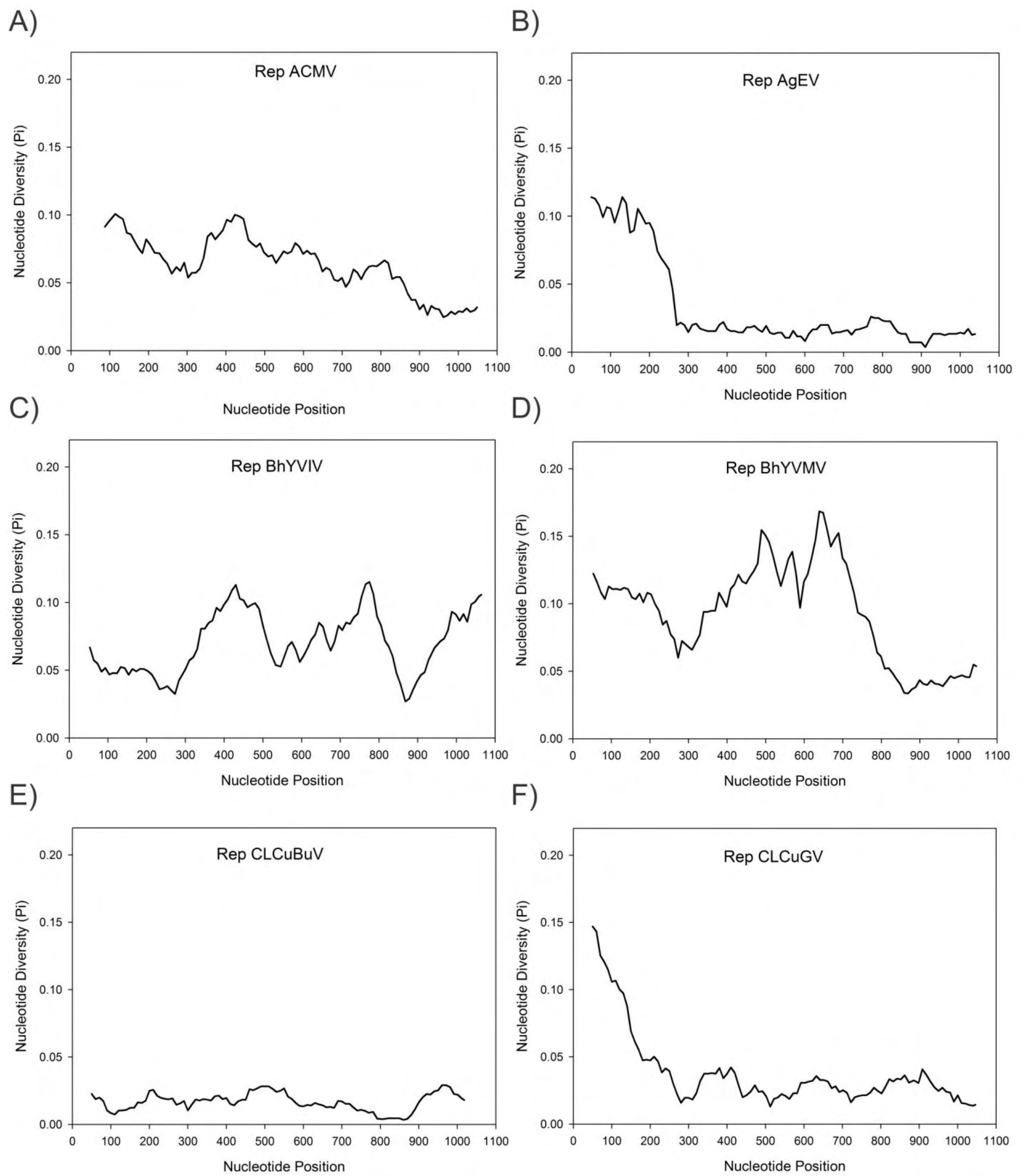
N)



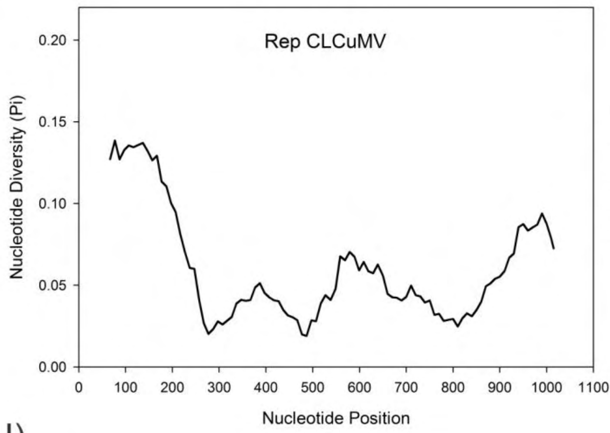
O)



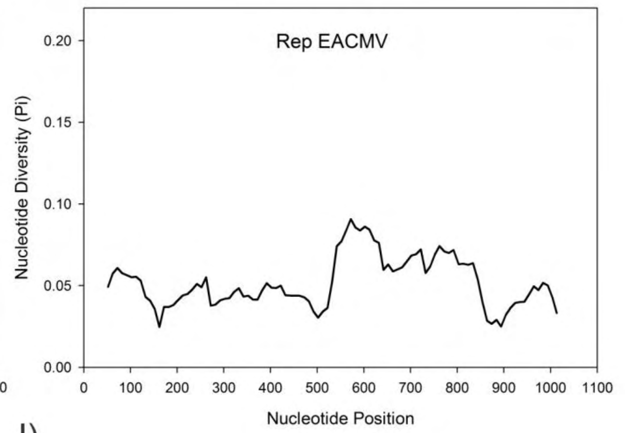
Supplementary Figure S2



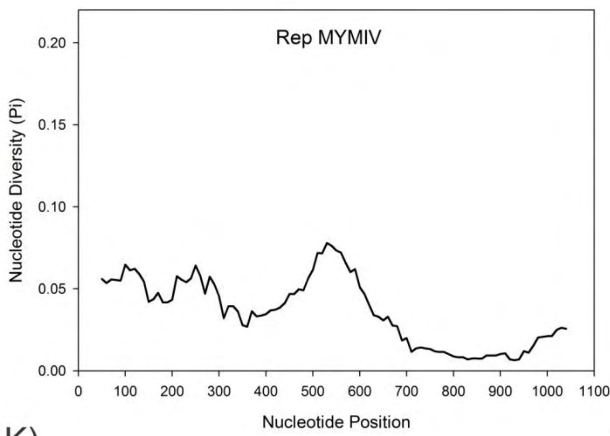
G)



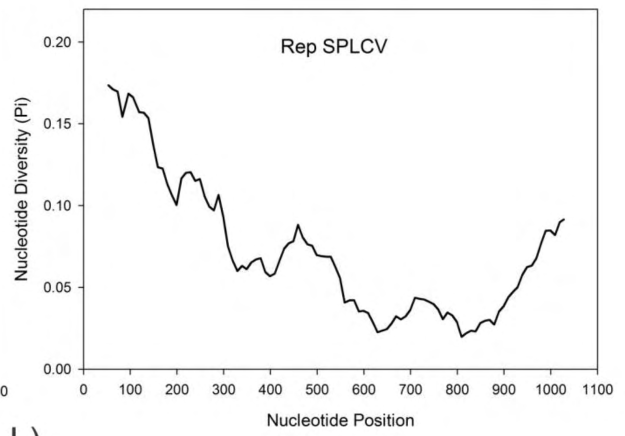
H)



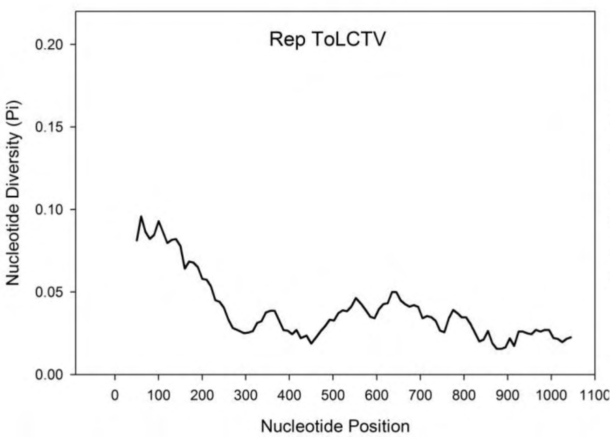
I)



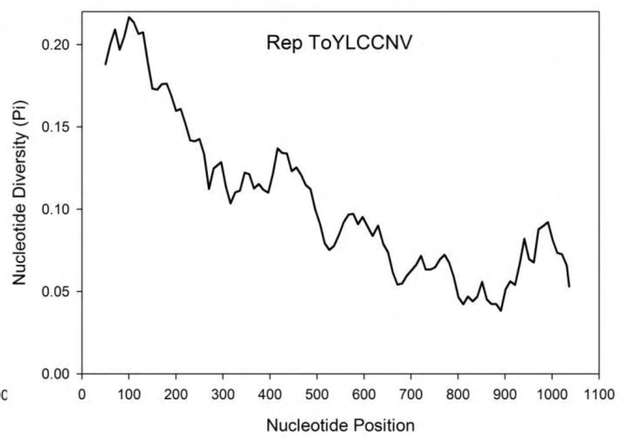
J)



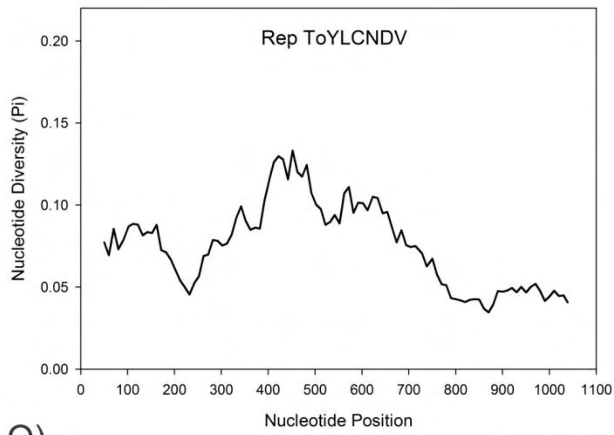
K)



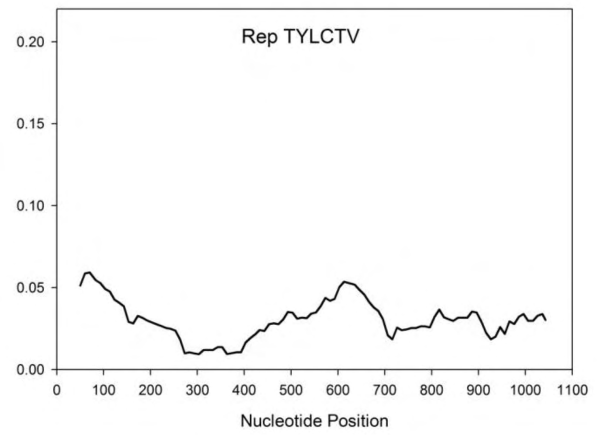
L)



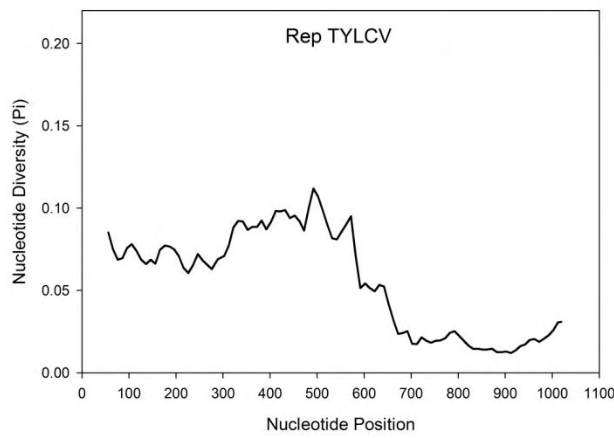
M)



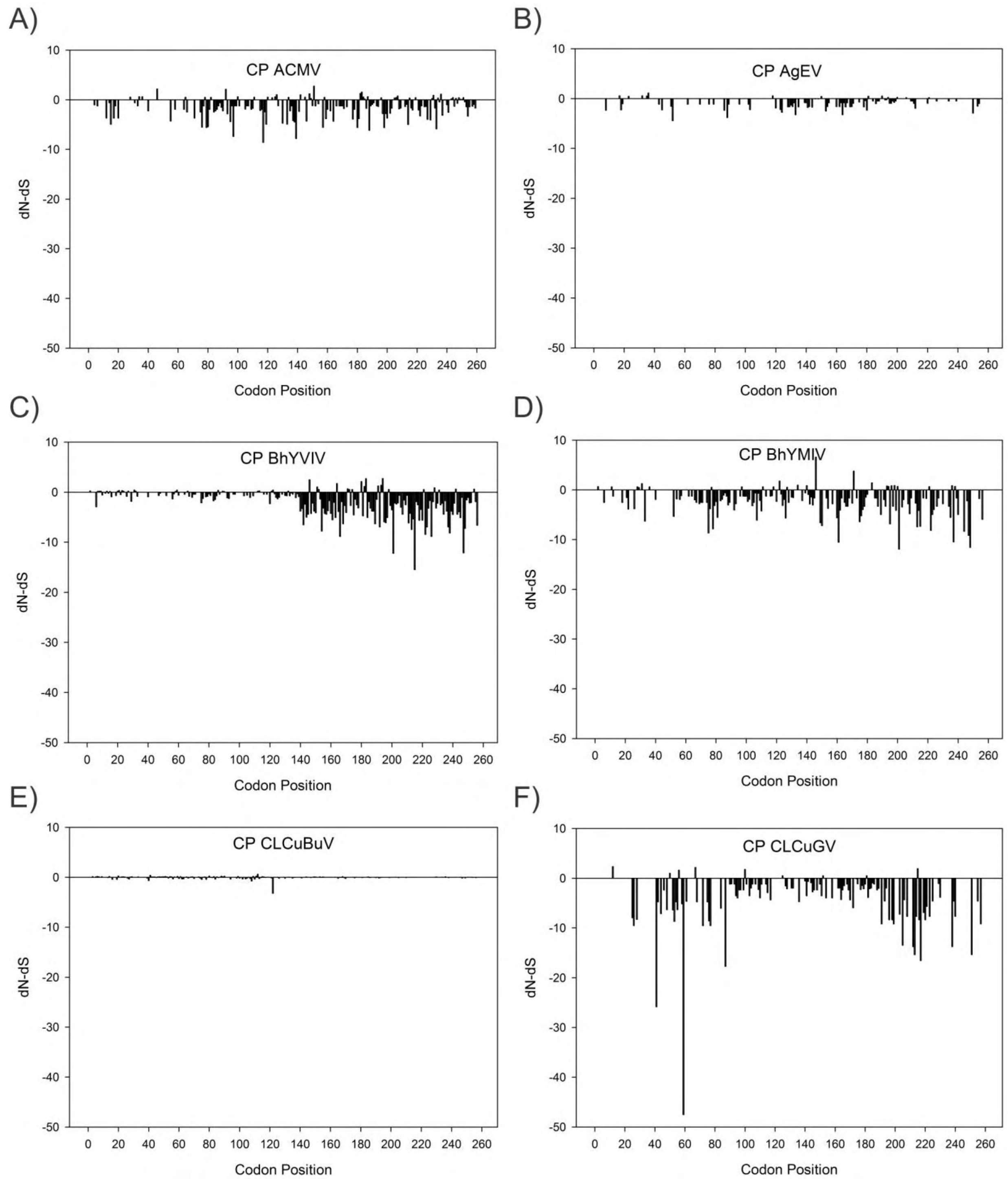
N)



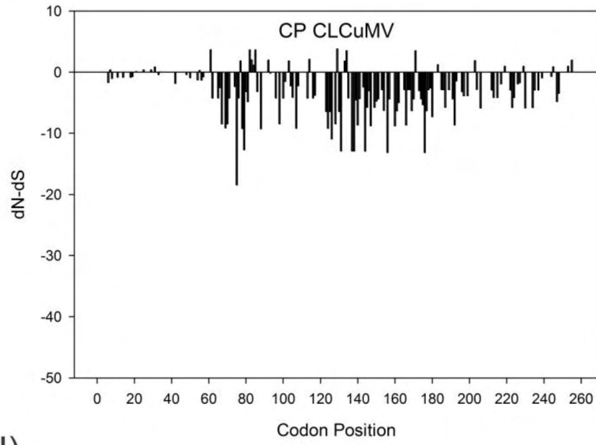
O)



Supplementary Figure S3

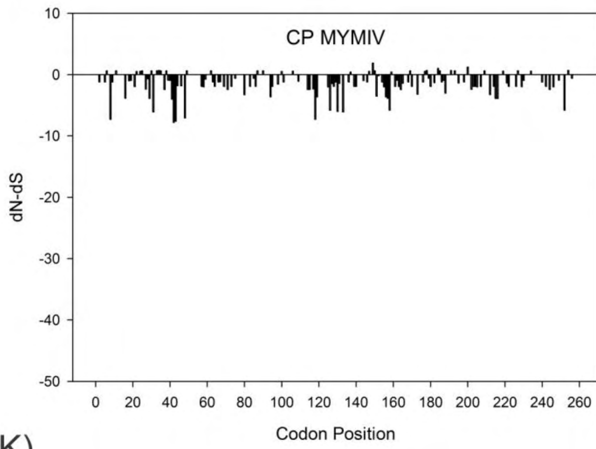


G)

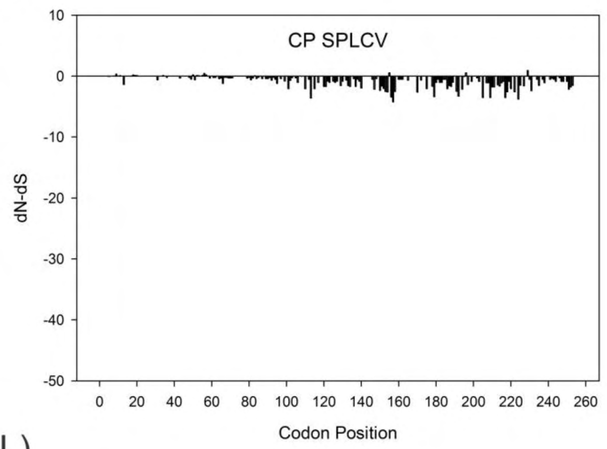


H)

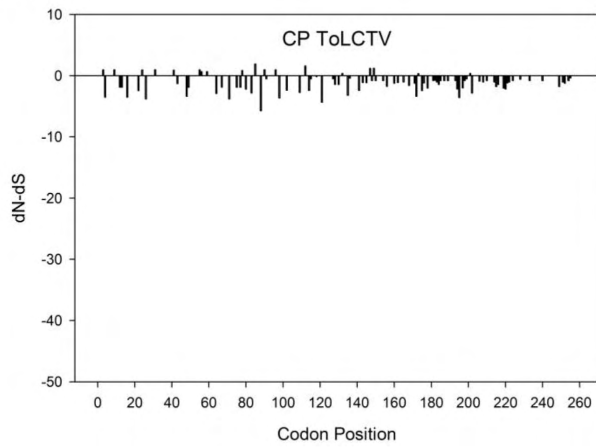
I)



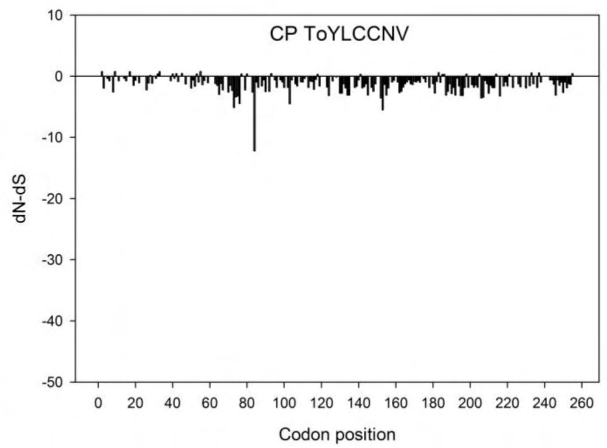
J)



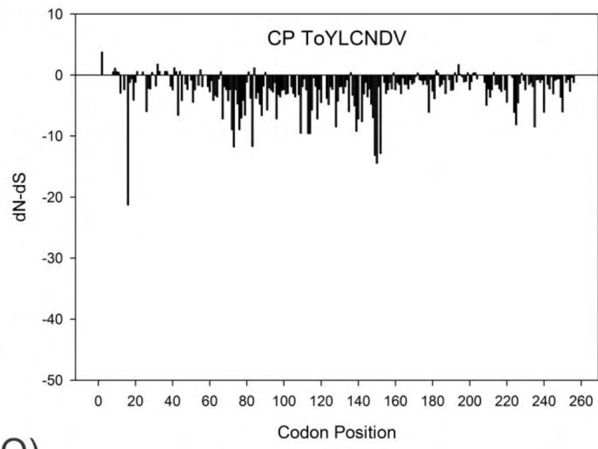
K)



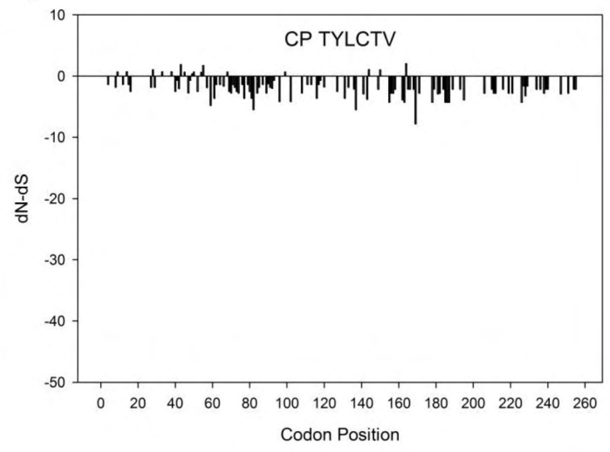
L)



M)

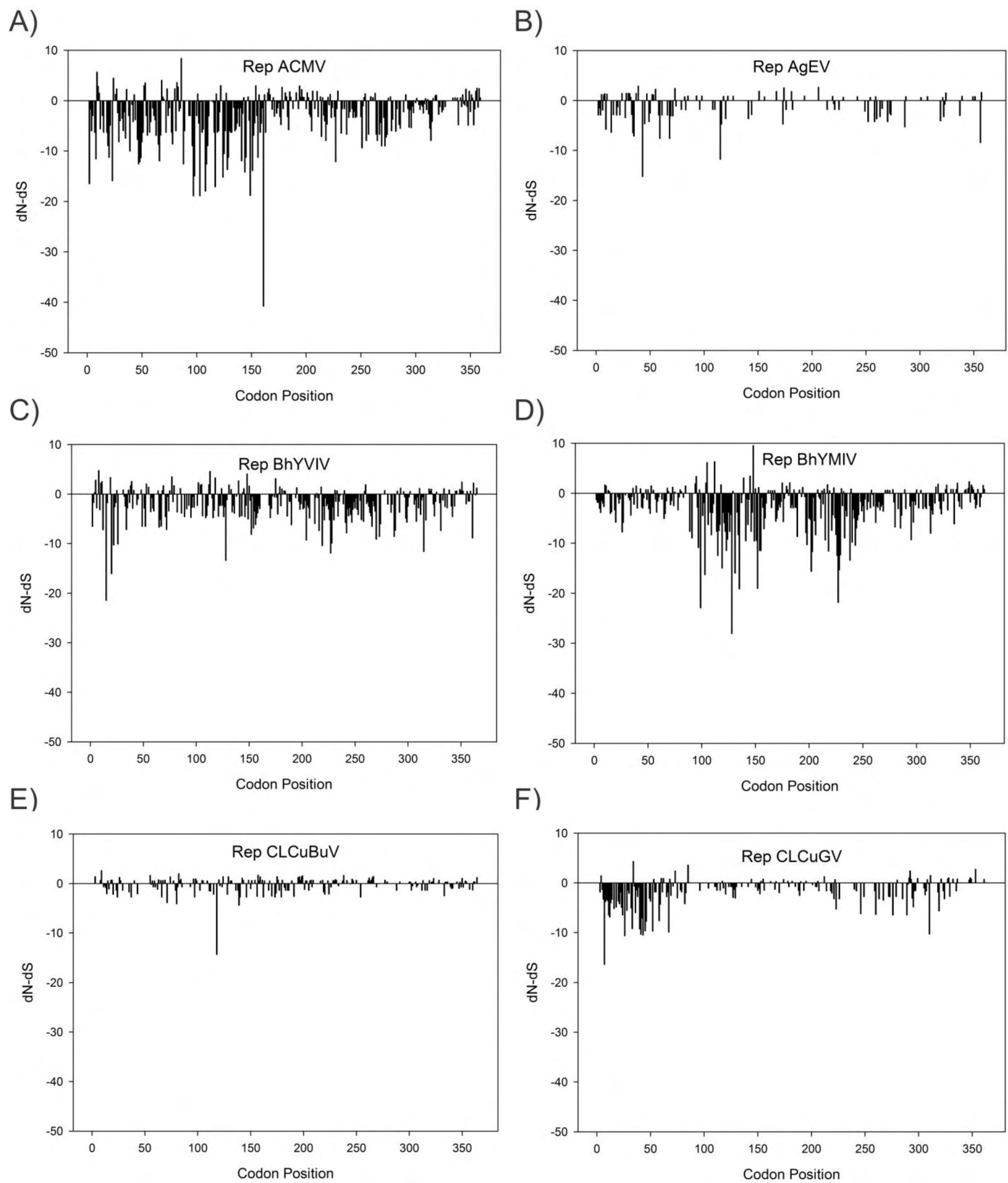


N)

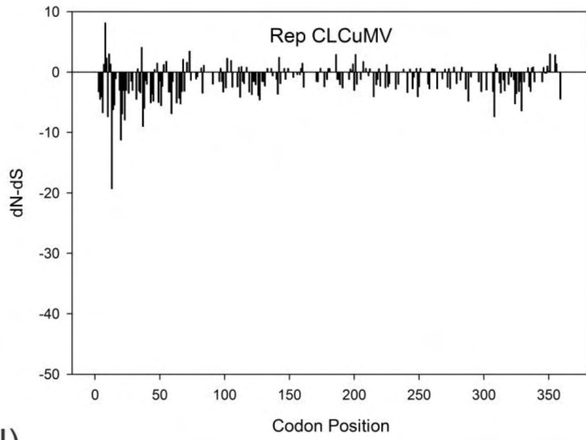


O)

Supplementary Figure S4

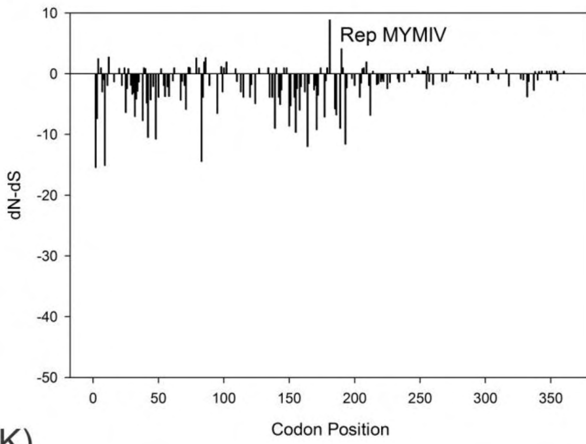


G)

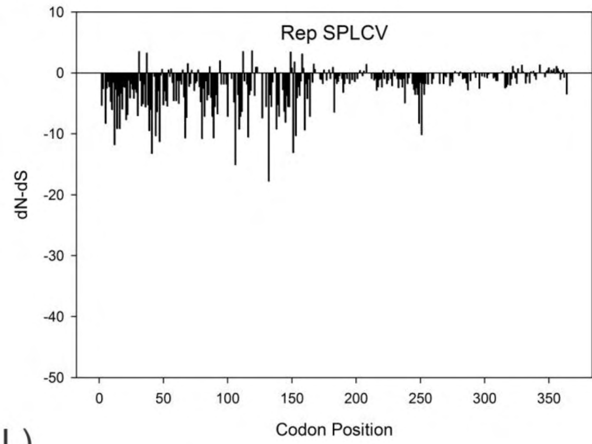


H)

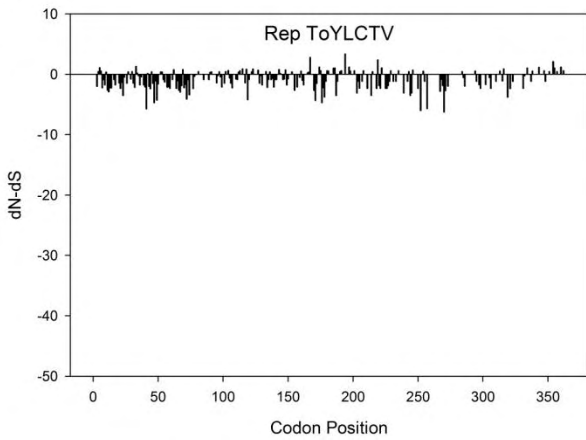
I)



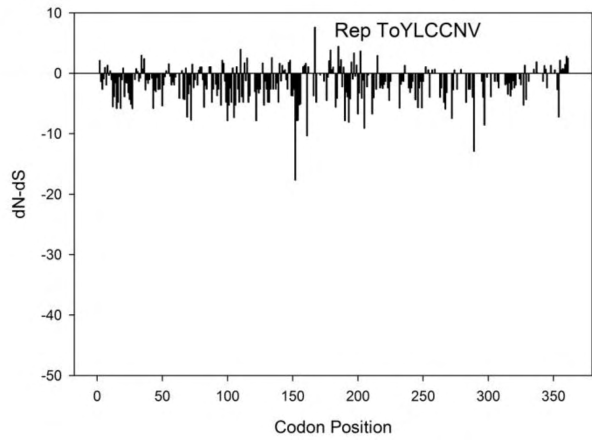
J)



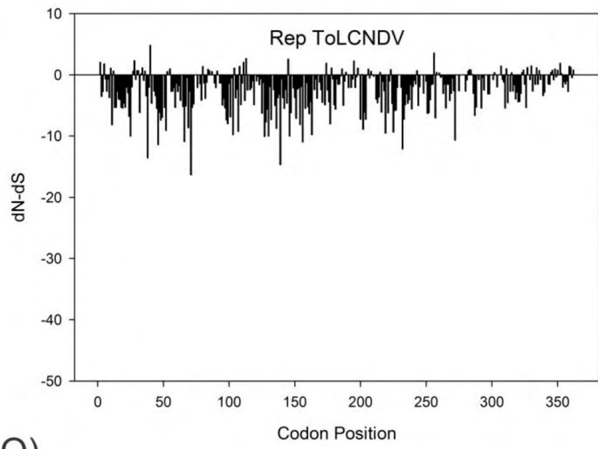
K)



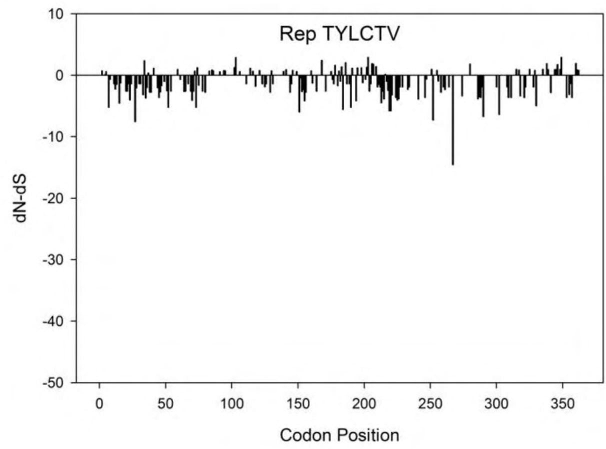
L)



M)



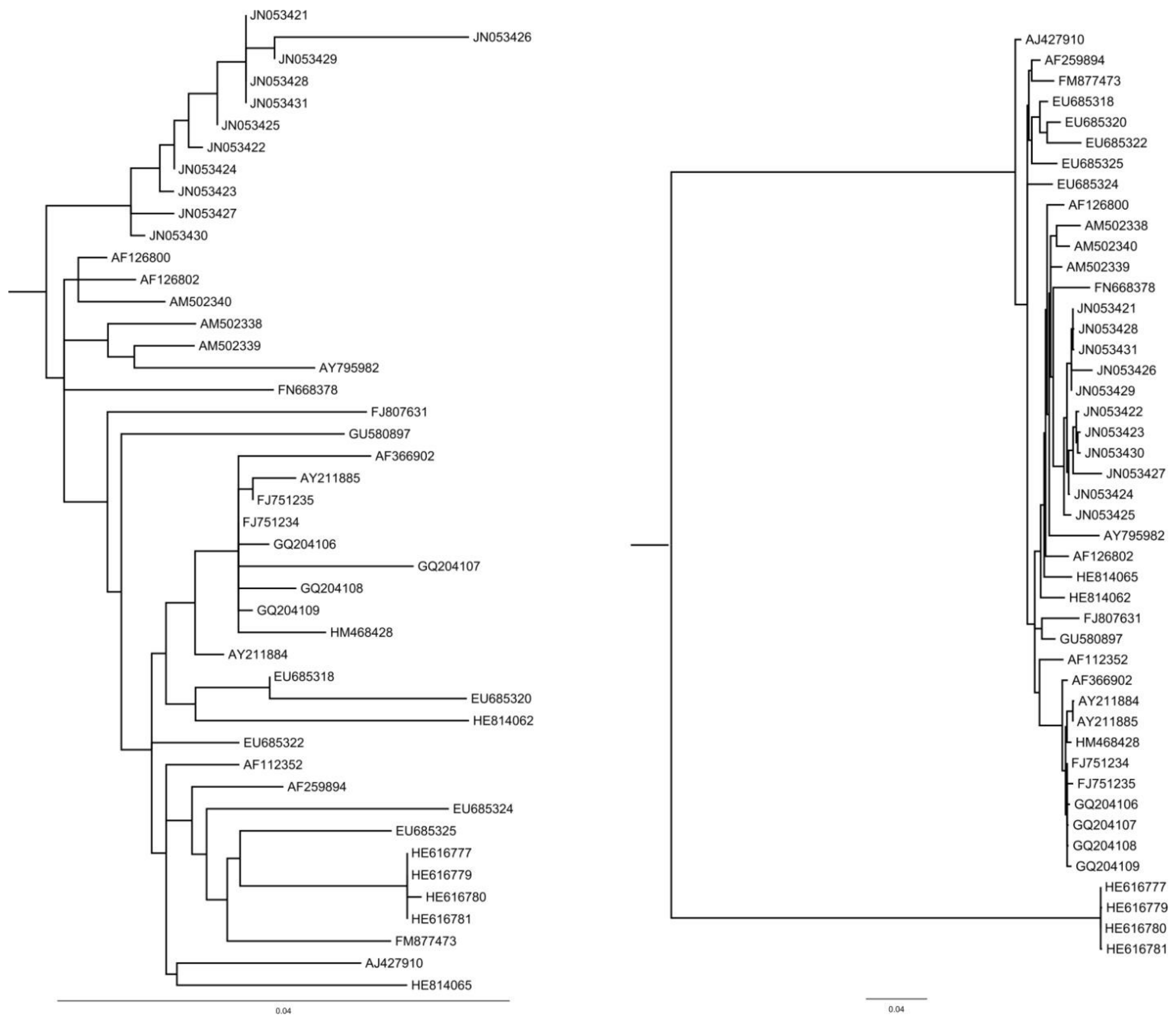
N)



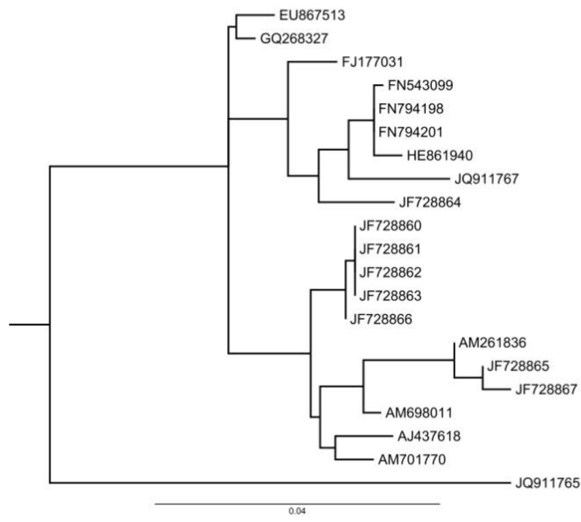
O)

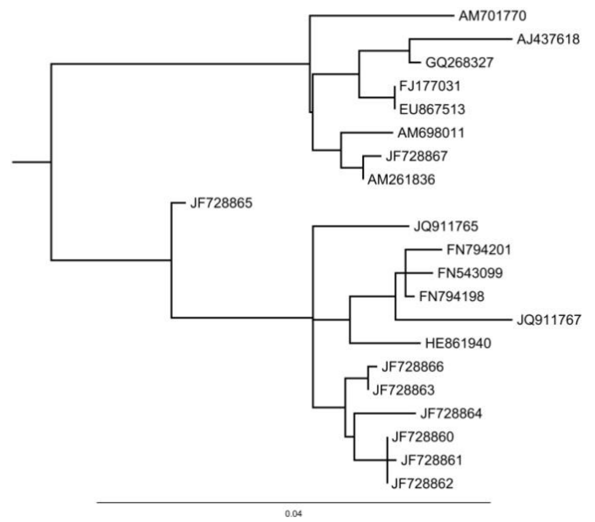
Supplementary Figure S5

A)



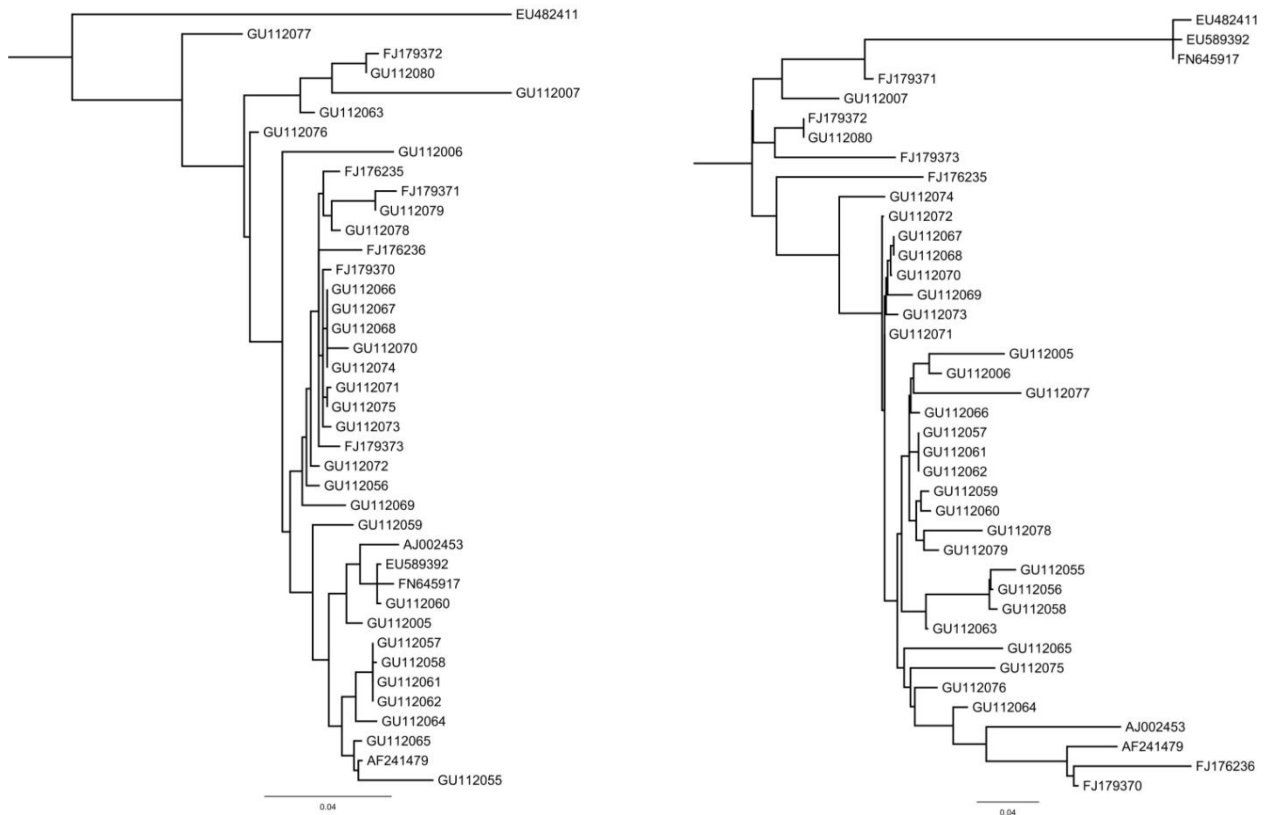
B)



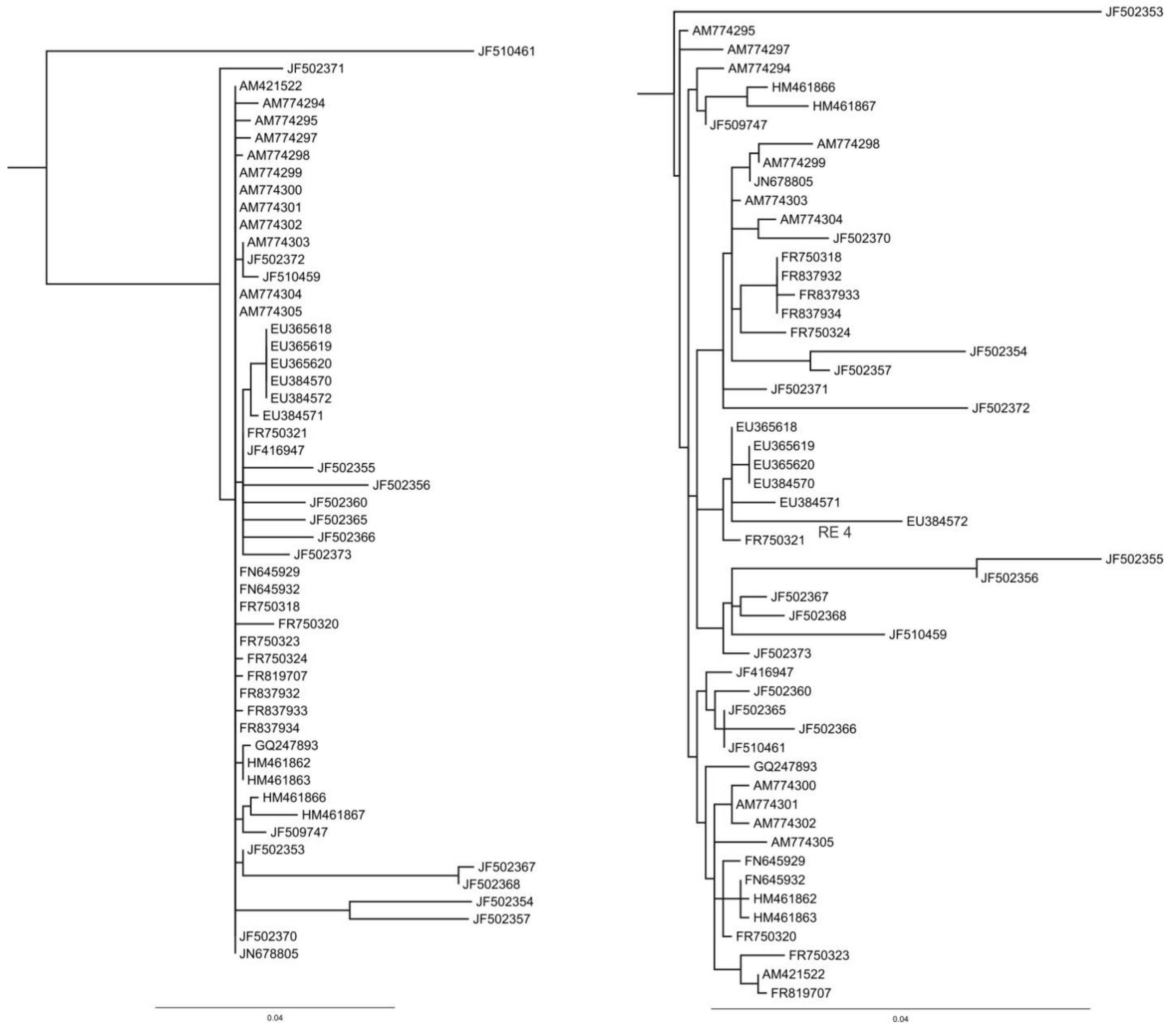


C)

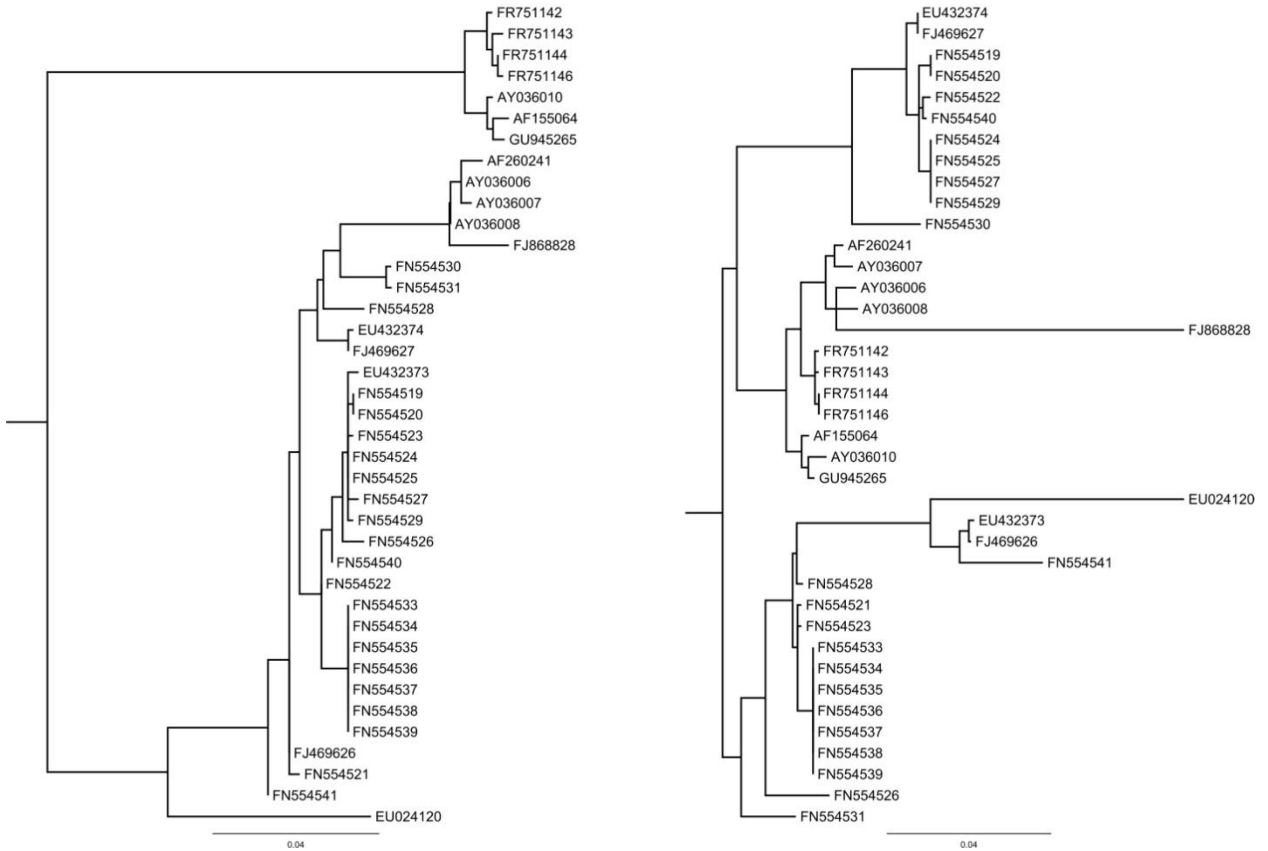
D)



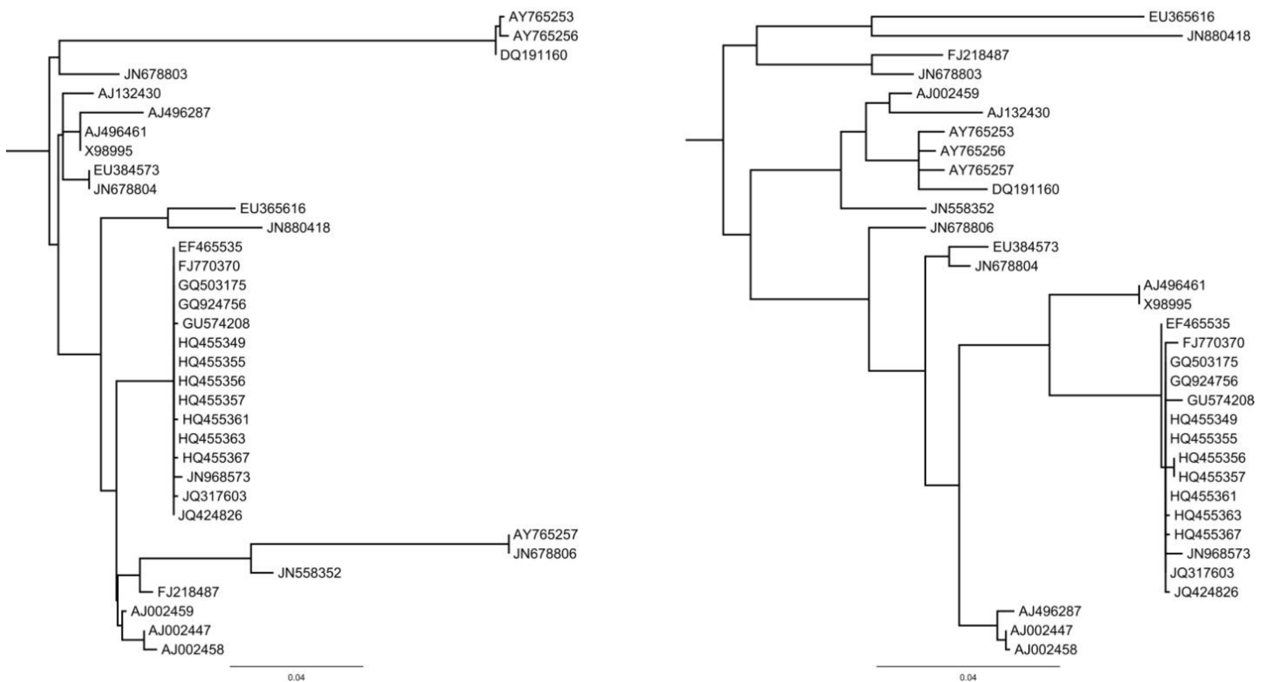
E)



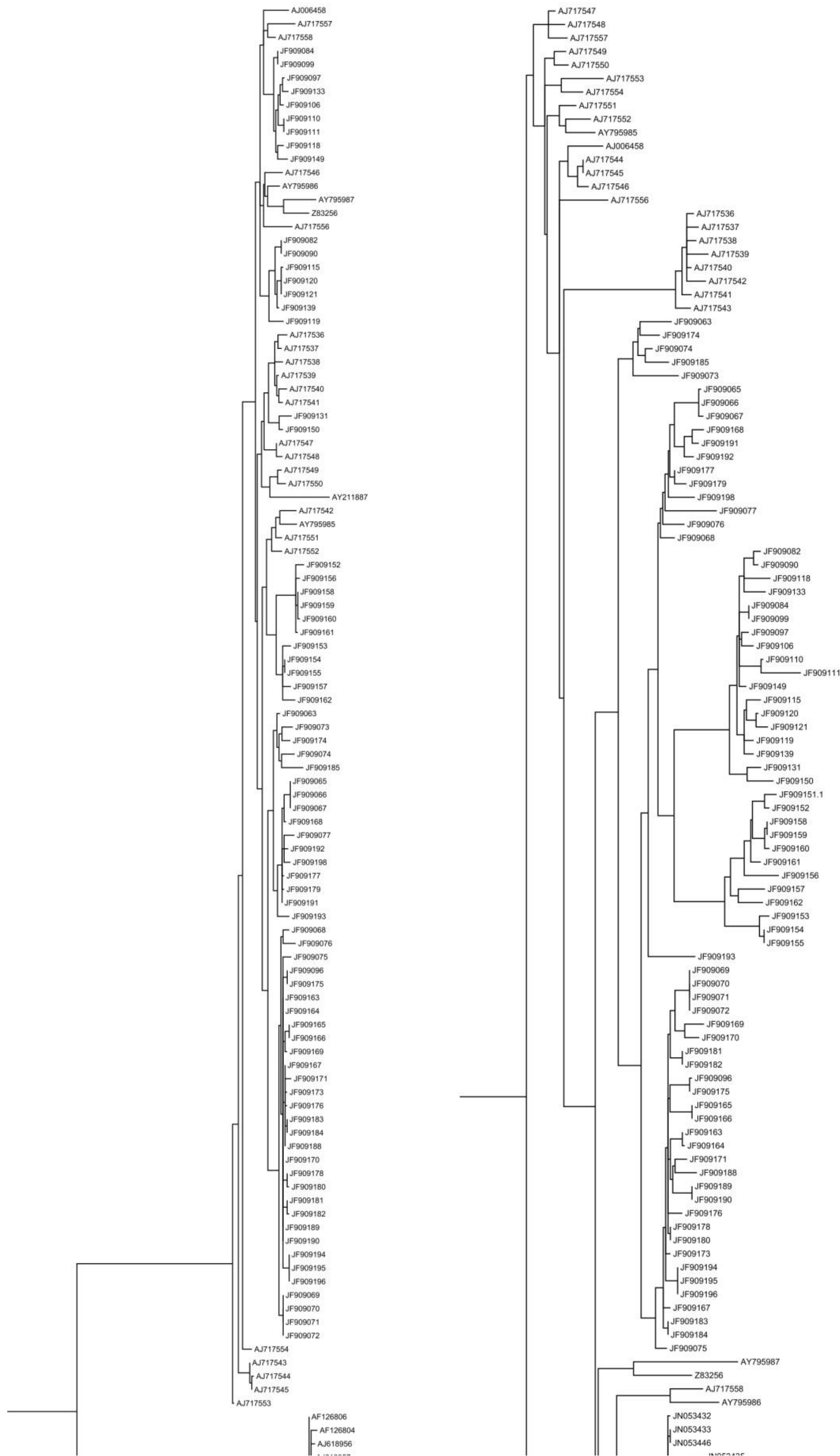
F)



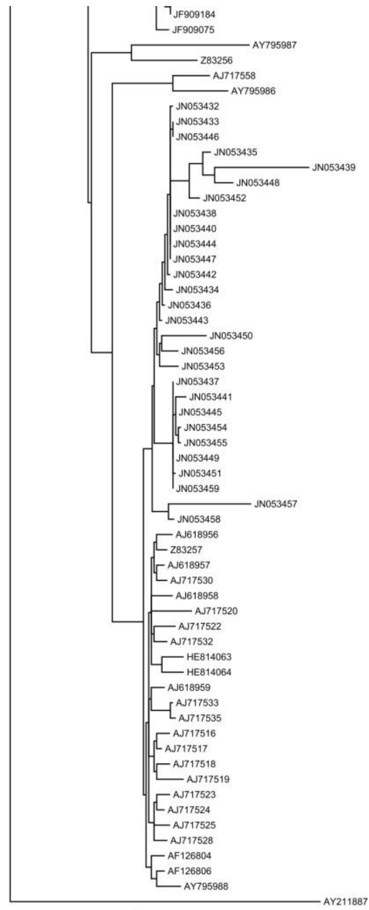
G)



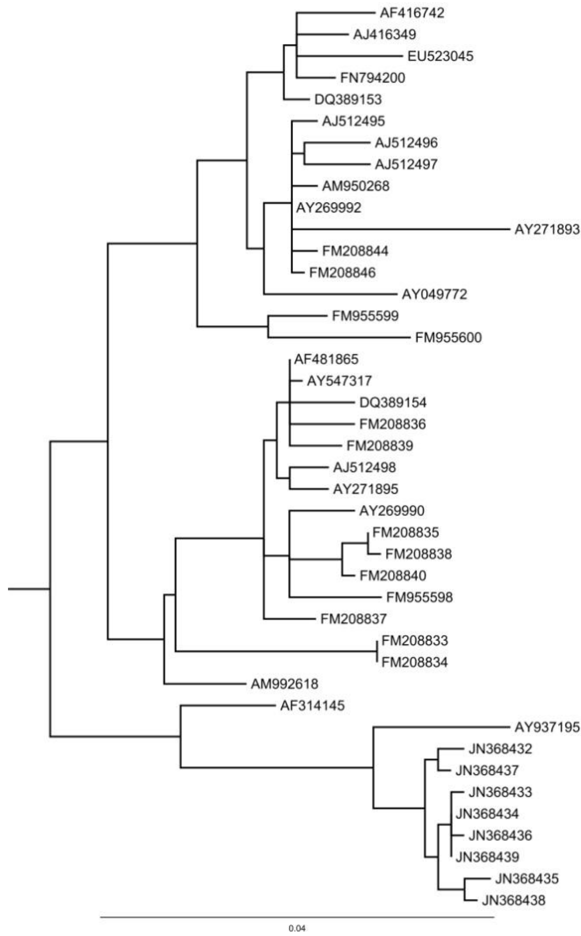
H)

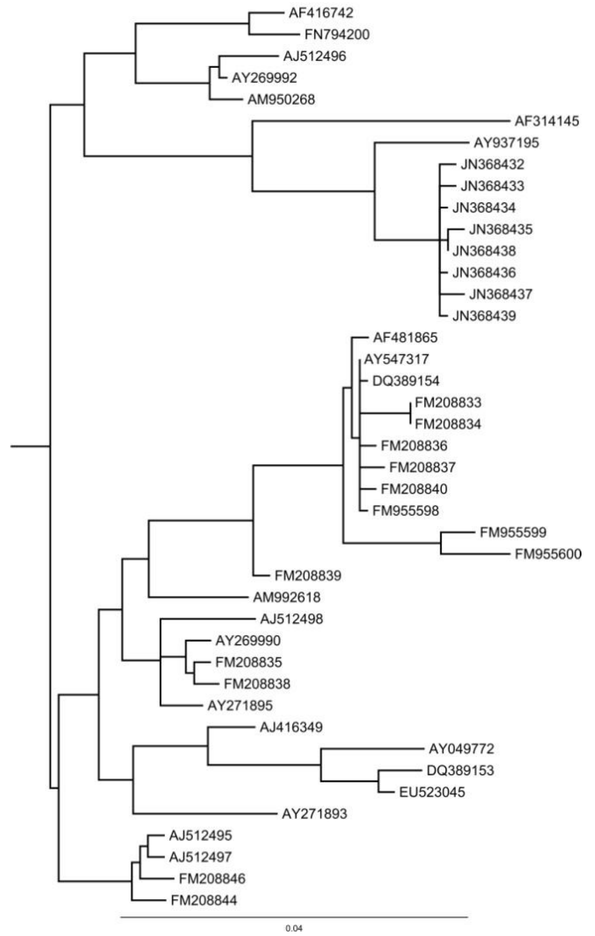


JF909072
AJ717554
AJ717543
AJ717544
AJ717545
AJ717553
AF126806
AF126804
AJ618956
AJ618957
AJ618958
AJ618959
AJ717516
AJ717517
AJ717518
AJ717519
AJ717520
JN053457
JN053458
AJ717522
AJ717523
AJ717524
AJ717525
AJ717526
AJ717528
AJ717530
JN053436
JN053450
AJ717532
AJ717533
AJ717535
AY795988
HE814063
HE814064
JN053432
JN053434
JN053435
JN053440
JN053446
JN053447
JN053452
JN053433
JN053438
JN053442
JN053443
JN053444
JN053456
JN053437
JN053441
JN053448
JN053445
JN053449
JN053451
JN053454
JN053455
JN053459
JN053439
JN053453
Z83257
0.04

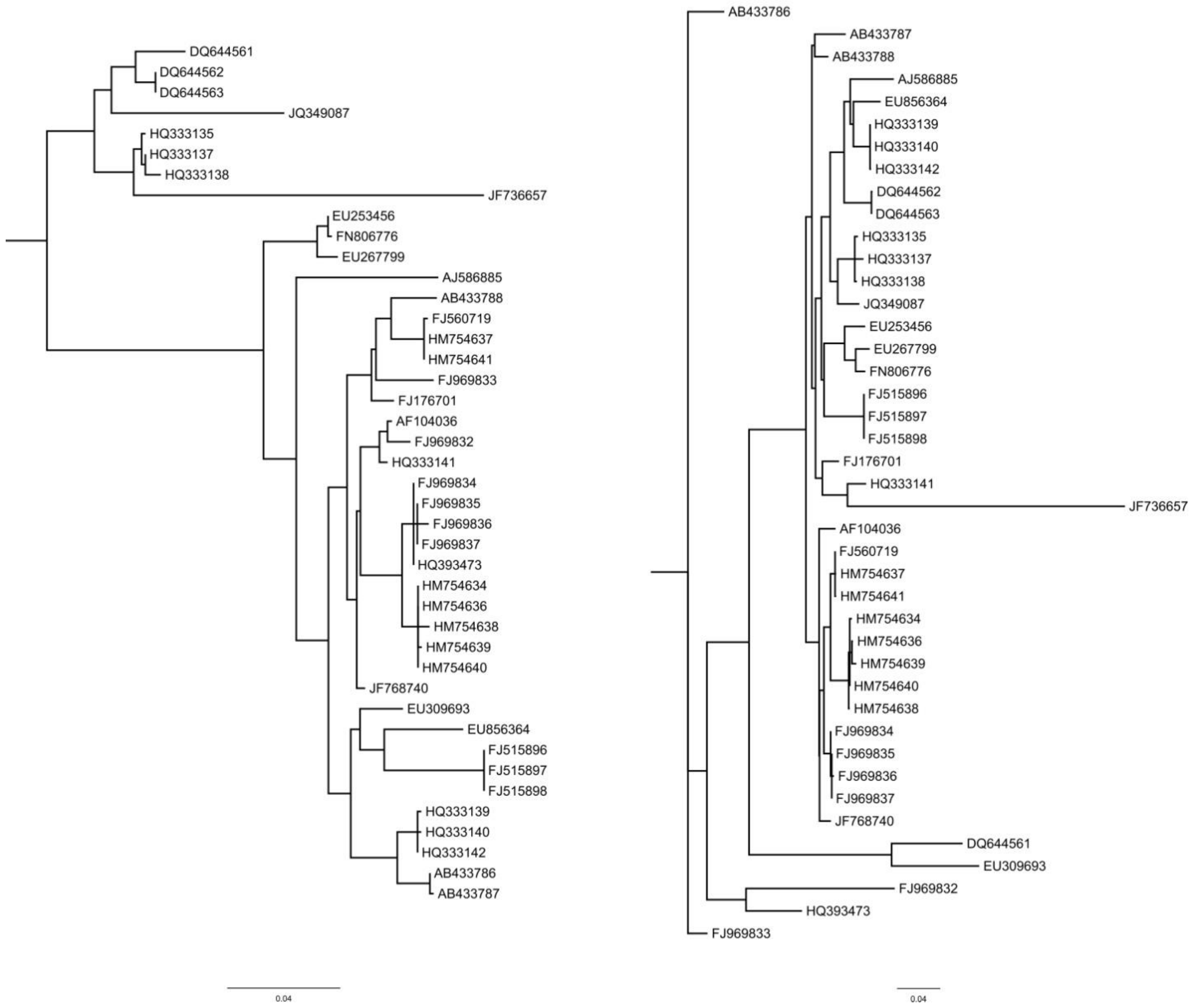

JF909184
JF909075
AY795987
Z83256
AJ717558
AY795986
JN053432
JN053433
JN053446
JN053435
JN053439
JN053448
JN053452
JN053438
JN053440
JN053444
JN053447
JN053442
JN053434
JN053436
JN053443
JN053450
JN053456
JN053453
JN053437
JN053441
JN053445
JN053454
JN053455
JN053449
JN053451
JN053459
JN053457
JN053458
AJ618956
Z83257
AJ618957
AJ717530
AJ618958
AJ717520
AJ717522
AJ717532
HE814063
HE814064
AJ618959
AJ717533
AJ717535
AJ717516
AJ717517
AJ717518
AJ717519
AJ717523
AJ717524
AJ717525
AJ717528
AF126804
AF126806
AY795988
AY211887
0.04

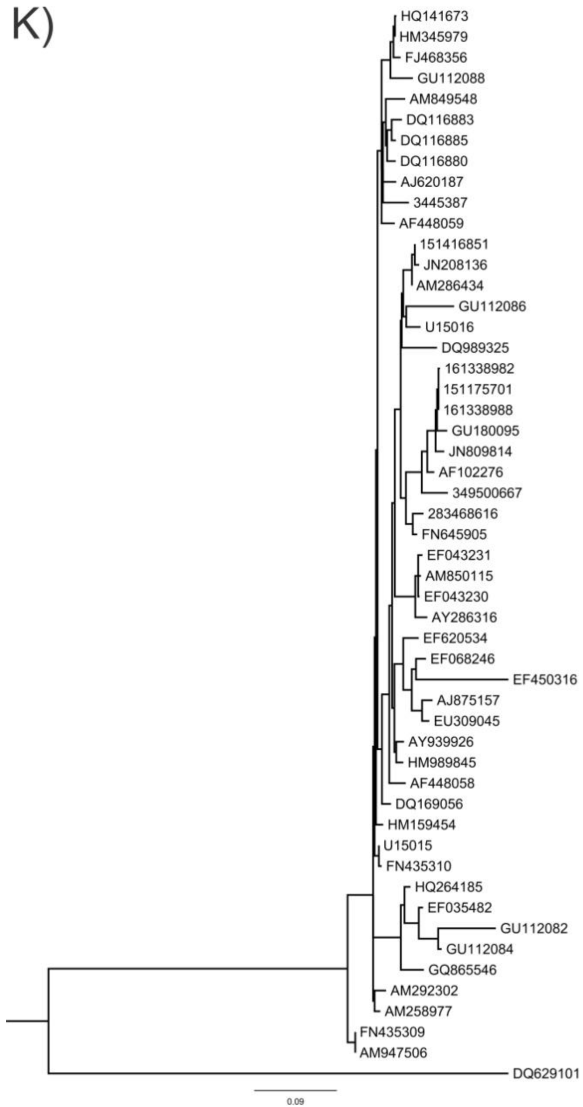
I)

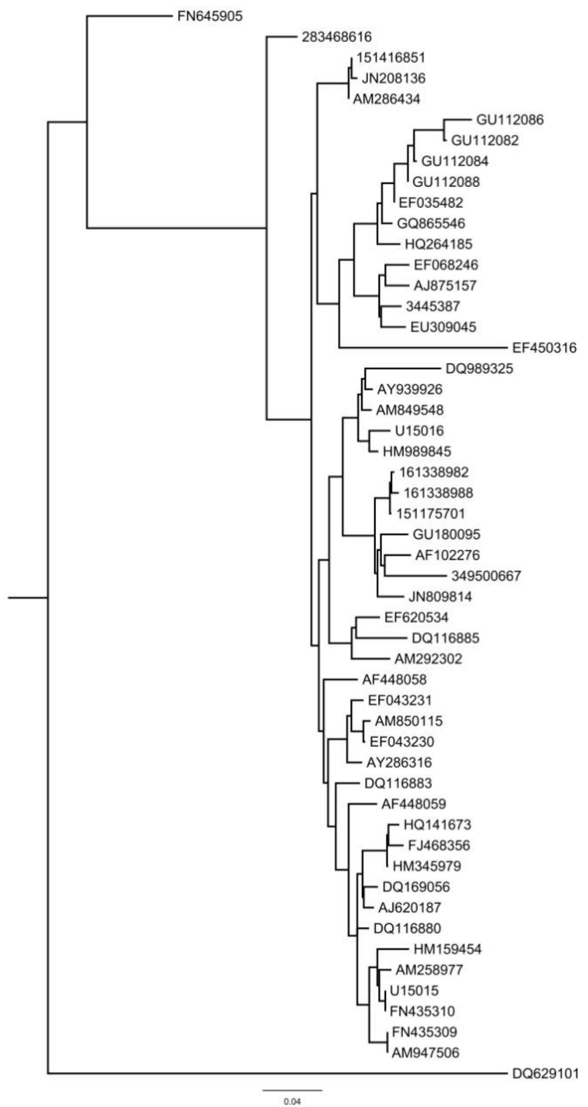




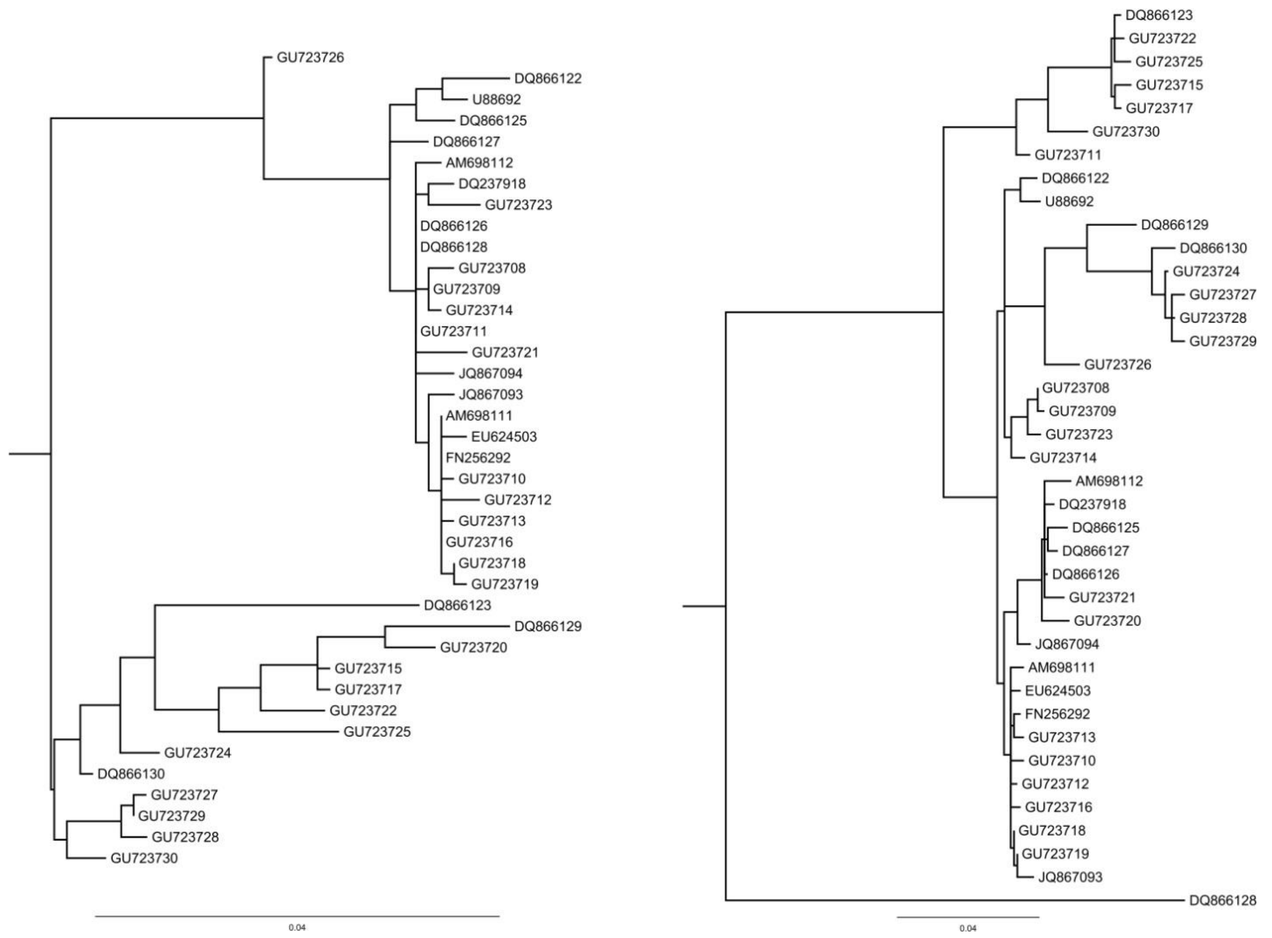
J)



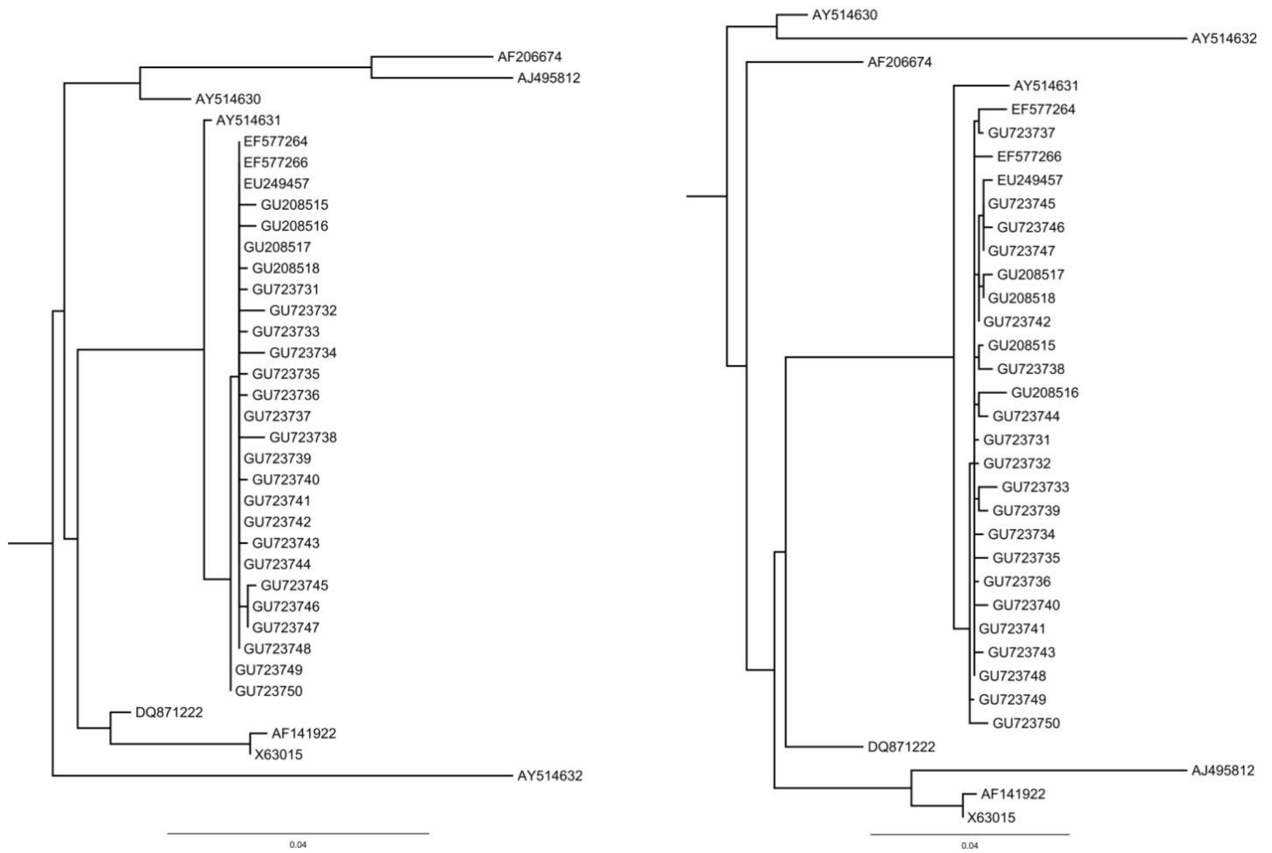

K)
HQ141673
HM345979
FJ468356
GU112088
AM849548
DQ116883
DQ116885
DQ116880
AJ620187
3445387
AF448059
151416851
JN208136
AM286434
GU112086
U15016
DQ989325
161338982
151175701
161338988
GU180095
JN809814
AF102276
349500667
283468616
FN645905
EF043231
AM850115
EF043230
AY286316
EF620534
EF068246
EF450316
AJ875157
EU309045
AY939926
HM989845
AF448058
DQ169056
HM159454
U15015
FN435310
HQ264185
EF035482
GU112082
GU112084
GQ865546
AM292302
AM258977
FN435309
AM947506
DQ629101
0.09


FN645905
283468616
151416851
JN208136
AM286434
GU112086
GU112082
GU112084
GU112088
EF035482
GQ865546
HQ264185
EF068246
AJ875157
3445387
EU309045
EF450316
DQ989325
AY939926
AM849548
U15016
HM989845
161338982
161338988
151175701
GU180095
AF102276
349500667
JN809814
EF620534
DQ116885
AM292302
AF448058
EF043231
AM850115
EF043230
AY286316
DQ116883
AF448059
HQ141673
FJ468356
HM345979
DQ169056
AJ620187
DQ116880
HM159454
AM258977
U15015
FN435310
FN435309
AM947506
DQ629101
0.04

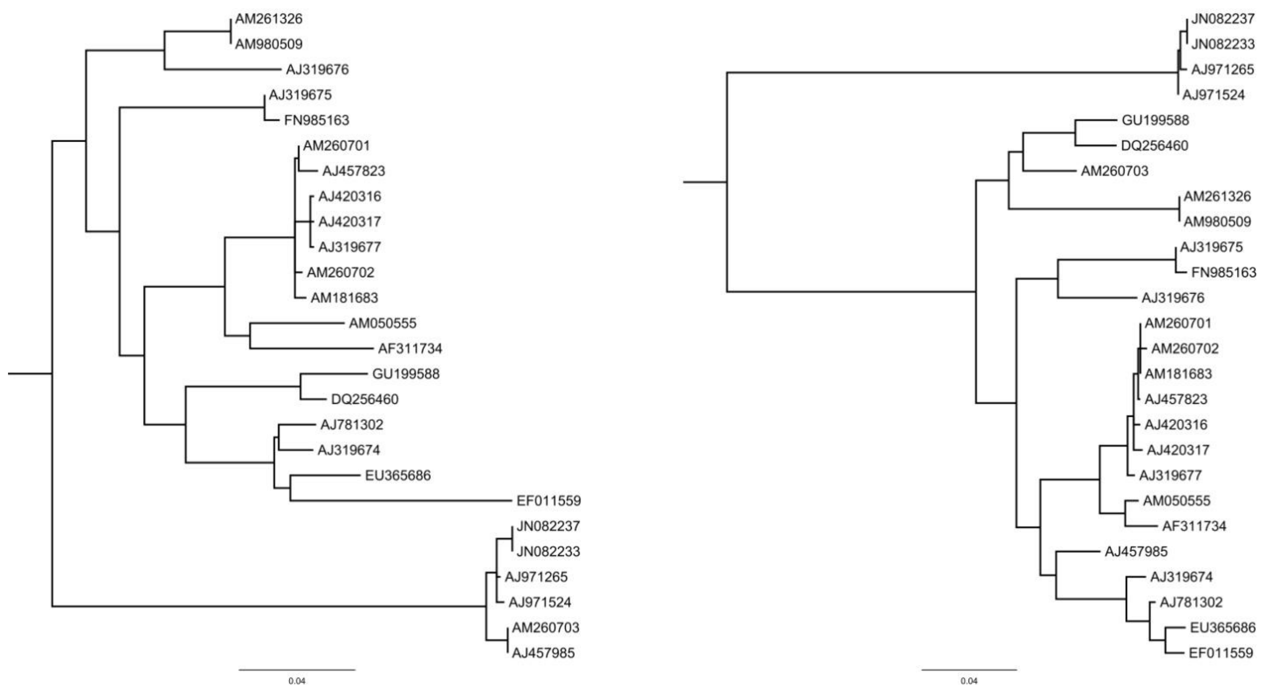
L)



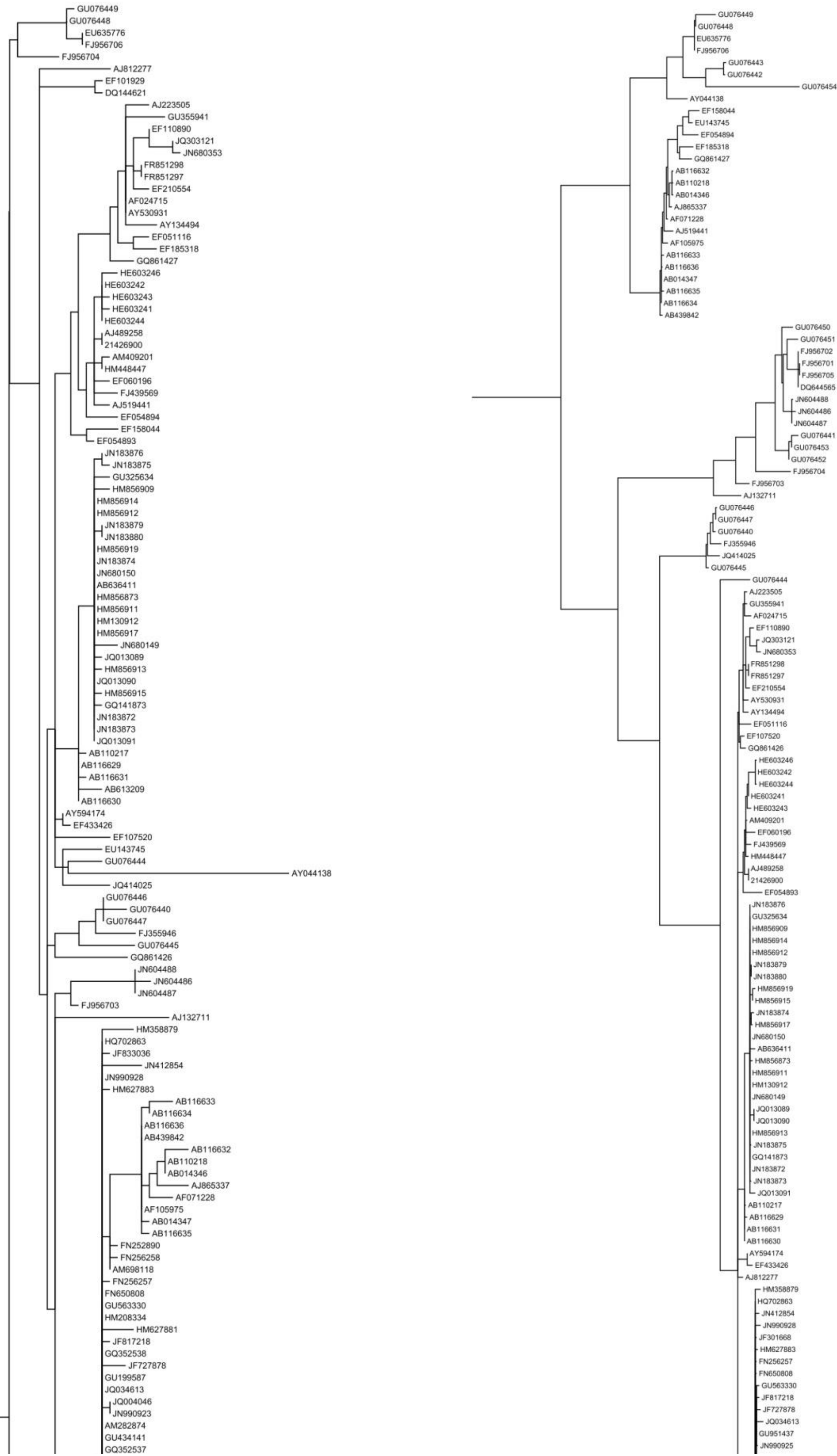
M)

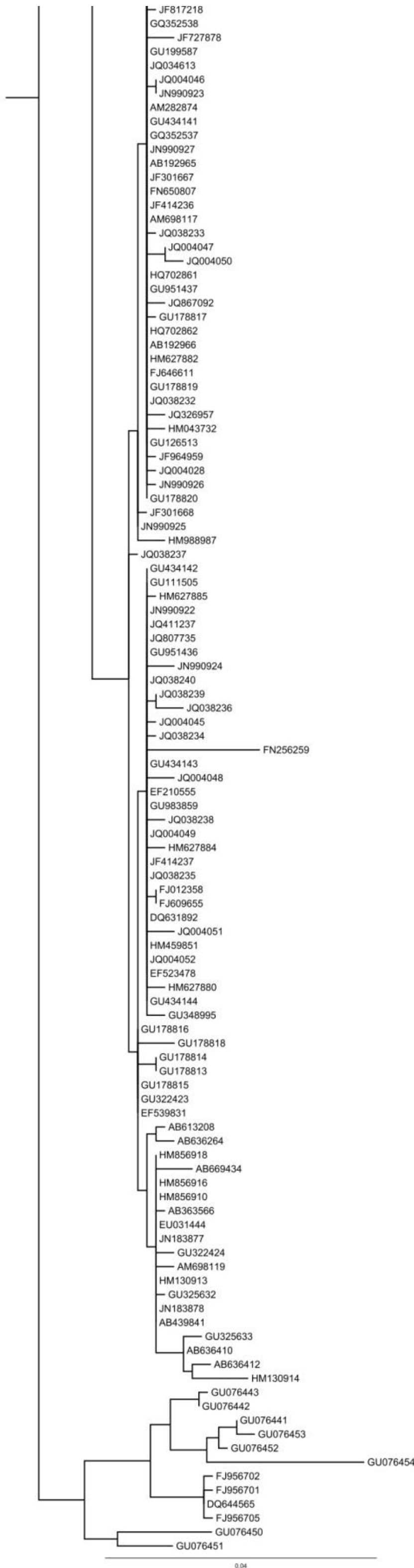


N)

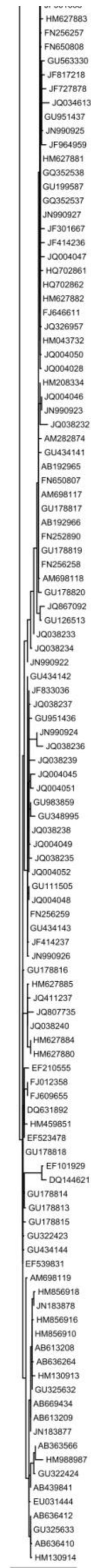


O)





JF817218
GQ352538
JF727878
GU199587
JQ034613
JQ004046
JN990923
AM282874
GU434141
GQ352537
JN990927
AB192965
JF301667
FN650807
JF414236
AM698117
JQ038233
JQ004047
JQ004050
HQ702861
GU951437
JQ867092
GU178817
HQ702862
AB192966
HM627882
FJ646611
GU178819
JQ038232
JQ326957
HM043732
GU126513
JF964959
JQ004028
JN990926
GU178820
JF301668
JN990925
HM988987
JQ038237
GU434142
GU111505
HM627885
JN990922
JQ411237
JQ807735
GU951436
JN990924
JQ038240
JQ038239
JQ038236
JQ004045
JQ038234
FN256259
GU434143
JQ004048
EF210555
GU983859
JQ038238
JQ004049
HM627884
JF414237
JQ038235
FJ012358
FJ609655
DQ631892
JQ004051
HM459851
JQ004052
EF523478
HM627880
GU434144
GU348995
GU178816
GU178818
GU178814
GU178813
GU178815
GU322423
EF539831
AB613208
AB636264
HM856918
AB669434
HM856916
HM856910
AB363566
EU031444
JN183877
GU322424
AM698119
HM130913
GU325632
JN183878
AB439841
GU325633
AB636410
AB636412
HM130914
GU076443
GU076442
GU076441
GU076453
GU076452
GU076454
FJ956702
FJ956701
DQ644565
FJ956705
GU076450
GU076451
0.04



HM627883
FN256257
FN650808
GU563330
JF817218
JF727878
JQ034613
GU951437
JN990925
JF964959
HM627881
GQ352538
GU199587
GQ352537
JN990927
JF301667
JF414236
JQ004047
HQ702861
HQ702862
HM627882
FJ646611
JQ326957
HM043732
JQ004050
JQ004028
HM208334
JQ004046
JN990923
JQ038232
AM282874
GU434141
AB192965
FN650807
AM698117
GU178817
AB192966
FN252890
GU178819
FN256258
AM698118
GU178820
JQ867092
GU126513
JQ038233
JQ038234
JN990922
GU434142
JF833036
JQ038237
GU951436
JN990924
JQ038236
JQ038239
JQ004045
JQ004051
GU983859
GU348995
JQ038238
JQ004049
JQ038235
JQ004052
GU111505
JQ004048
FN256259
GU434143
JF414237
JN990926
GU178816
HM627885
JQ411237
JQ807735
JQ038240
HM627884
HM627880
EF210555
FJ012358
FJ609655
DQ631892
HM459851
EF523478
GU178818
EF101929
DQ144621
GU178814
GU178813
GU178815
GU322423
GU434144
EF539831
AM698119
HM856918
JN183878
HM856916
HM856910
AB613208
AB636264
HM130913
GU325632
AB669434
AB613209
JN183877
AB363566
HM988987
GU322424
AB439841
EU031444
AB636412
GU325633
AB636410
HM130914
0.04

Supplementary Table S2. Recombination detected by RDP in begomovirus populations retrieved from the GenBank database.

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
ACMV							
1	HE616779	1290 (1327)	38 (38)	EU685325	Unknown	GBMC**S**3	2.27×10^{-41}
	HE616777			FJ807631	Unknown		
	HE616780			GU580897	Unknown		
	HE616781			HE814062	Unknown		
AgEV							
1	JQ911765	135 (148)	2407 (2420)	JF728865	Unknown	GBMCS**3**	9.49×10^{-30}
	JQ911767			HE861940	Unknown		
	JF728864			FN794198	Unknown		
	JF728860			FN794198	Unknown		
	JF728862			FN794198	Unknown		
	JF728861			FN794198	Unknown		
	JF728863			FN794198	Unknown		
	JF728866			FN794198	Unknown		
	HE861940			FN794198	Unknown		
	FN794201			FN794198	Unknown		
	FN543099			FN794198	Unknown		
	FN794198			FN794198	Unknown		
2	AM261836	2382 (2393)	2717 (2740)	JF728865	AM701770	GBMCS**3**	6.78×10^{-27}
	AM698011			JF728867	AJ437618		
	JF728867				GQ268327		
3	JF728860	939 (952)	(?)2407 [(?)2420]	AJ437618	JQ911765	GBMCS**3**	2.37×10^{-15}
	JQ911767			AM701770	HE861940		
	JF728864			GQ268327	FN794201		
	JF728862			EU867513	FN543099		
	JF728861			FJ177031	FN543099		
	JF728863			AM698011	FN543099		
	JF728866			AM698011	FN543099		
	HE861940			AM698011	FN543099		
	FN794201			AM698011	FN543099		
	FN543099			AM698011	FN543099		
	FN794198			AM698011	FN543099		
	AM261836			AM698011	FN543099		
	AM698011			AM698011	FN543099		
	JF728865			AM698011	FN543099		
	JF728867			AM698011	FN543099		
4	AM701770	208 (220)	(?)832 [(?)844]	FJ177031	JF728867	MC**S**3	2.67×10^{-10}
	AJ437618			GQ268327	AM261836		
5	FJ177031			JQ911765	Unknown	MC**S**3	7.95×10^{-09}
	AM701770			JQ911765	Unknown		
	AJ437618			JQ911765	Unknown		
	GQ268327			JQ911765	Unknown		
	EU867513			JQ911765	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
BhYVIV							
1	GU112026	2396 (2418)	1051(1056)	GU112018	Unknown	RGBMCS**3**	2.05×10^{-71}
	GU112025			GU112050	Unknown		
	GU112035			GU112031	Unknown		
	GU112024			GU112037	Unknown		
	GU112023			GU112015	Unknown		
	GU112034			GU112013	Unknown		
	GU112033			GU112022	Unknown		
	GU112036			GU112019	Unknown		
	GU112040			GU112027	Unknown		
	GU112041			GU112014	Unknown		
	GU112039			GU112012	Unknown		
	GU112020			GU112016	Unknown		
	GU112021			GU112017	Unknown		
2	GU112009	1691(1701)	2544(2563)	GU112044	Unknown	RGBMCS**3**	3.60×10^{-64}
	GU112008			GU112031	Unknown		
	GU112054			GU112037	Unknown		
	GU112011			GU112048	Unknown		
	GU112051			GU112032	Unknown		
	GU112052			GU112042	Unknown		
	GU112053			GU112043	Unknown		
3	GU112013	710(715)	1246(1251)	GU112019	GU112031	GBMCS**3**	1.39×10^{-35}
4	GU112026	(?)2396 [(?)2418]	64(64)	GU112021	Unknown	GMCS**3**	4.56×10^{-35}
	GU112008			GU112054	Unknown		
	GU112025			GU112011	Unknown		
	GU112035			GU112009	Unknown		
	GU112024			GU112051	Unknown		
	GU112023			GU112052	Unknown		
	GU112034			GU112053	Unknown		
	GU112033			GU112049	Unknown		
5	GU112049	161(166)	1914(1927)	GU112016	Unknown	GBMC**S**3	9.64×10^{-37}
	GU112043			GU112048	Unknown		
	GU112045			GU112032	Unknown		
6	GU112015	1703 (1713)	109 (109)	GU112017	Unknown	GBMC**S**3	3.81×10^{-33}
7	GU112048	1297 (1302)	1911 (1924)	GU112017	GU112037	RGBMCS**3**	2.91×10^{-30}
	GU112008			GU112054	GU112008		
	GU112011			GU112032	GU112011		
	GU112009			GU112015	GU112009		
	GU112051			GU112013	GU112052		
	GU112052			GU112022	GU112053		
	GU112053			GU112018	GU112031		
	GU112038			GU112019	GU112034		
	GU112046			GU112027	GU112033		
	GU112047			GU112014	GU112036		
	GU112045			GU112012	GU112044		
8	GU112042	1063 (1068)	1965 (1978)	GU112019	GU112037	GBMCS**3**	1.63×10^{-23}
	GU112031			GU112054	GU112037		
	GU112032			GU112032	GU112037		
	GU112044			GU112015	GU112037		
	GU112043			GU112038	GU112037		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
BhYVIV							
9	GU112032	33 (33)	481 (486)	GU112009	Unknown	GBMCS**3**	8.56×10^{-23}
	GU112042			GU112008	Unknown		
	GU112043			GU112054	Unknown		
10	GU112024	(?)1113 [(?)1118]	(?)1669 [(?)1679]	GU112037	Unknown	GBMCS**3**	3.68×10^{-17}
	GU112026			GU112031	Unknown		
	GU112025			GU112035	Unknown		
	GU112023			GU112033	Unknown		
	GU112032			GU112048	Unknown		
	GU112013			GU112036	Unknown		
	GU112028			GU112044	Unknown		
	GU112022			GU112042	Unknown		
	GU112020			GU112043	Unknown		
	GU112021			GU112040	Unknown		
	GU112018			GU112041	Unknown		
	GU112019			GU112039	Unknown		
	GU112027			GU112038	Unknown		
	GU112014			GU112046	Unknown		
	GU112012			GU112047	Unknown		
	GU112016			GU112047	Unknown		
	GU112017			GU112047	Unknown		
	GU112030			GU112047	Unknown		
	GU112029			GU112047	Unknown		
11	GU112029	2513 (2535)	586 (591)	GU112047	Unknown	GBMCS**3**	1.08×10^{-13}
	GU112009			GU112052	Unknown		
	GU112037			GU112053	Unknown		
	GU112026			GU112015	Unknown		
	GU112025			GU112013	Unknown		
	GU112035			GU112044	Unknown		
	GU112024			GU112038	Unknown		
	GU112023			GU112046	Unknown		
	GU112034			GU112045	Unknown		
	GU112033			GU112022	Unknown		
	GU112036			GU112018	Unknown		
	GU112040			GU112019	Unknown		
	GU112041			GU112027	Unknown		
	GU112039			GU112014	Unknown		
	GU112028			GU112012	Unknown		
	GU112020			GU112016	Unknown		
	GU112021			GU112030	Unknown		
12	GU112054	2060 (2082)	2429 (2454)	GU112018	GU112008	GMC**3**	1.01×10^{-11}
	GU112052			GU112050	GU112009		
13	GU112035	1626 (1631)	1753 (1766)	GU112044	Unknown	GMCS**3**	6.40×10^{-11}
14	GU112008	716 (722)	(?)1993 [(?)1298]	Unknown	GU112038	GMC**3**	2.92×10^{-09}
	GU112011				GU112054		
15	GU112048	2542 (2564)	599 (604)	Unknown	GU112052	GBM**3**	1.36×10^{-07}
	GU112017				GU112022		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
BhYVIV							
16	GU112050	1476 (1481)	1558 (1563)	Unknown	GU112034	GMCS**3**	3.70×10^{-05}
	GU112049				GU112008		
BYVMV							
1	EU482411			FN645917	Unknown	GMCS**3**	9.28×10^{-70}
2	GU112060	2558 (2609)	1175 (1222)	GU112079	FN645917	GBMCS**3**	3.73×10^{-56}
	GU112078			FJ176235	EU589392		
	GU112005			FJ179371	EU589392		
	GU112059			GU112077	EU589392		
	GU112058			FJ179372	EU589392		
	GU112055			GU112080	EU589392		
	GU112056			GU112078	EU589392		
3	AF241479	1612 (1662)	2474 (2527)	Unknown	FJ179370	GBMCS**3**	4.57×10^{-42}
4	GU112056	(?)1176 [(?)1223]	2312 (2363)	FN645917	GU112060	GBMCS**3**	8.86×10^{-42}
	FJ176236			EU482411	AJ002453		
	GU112063			EU589392	FJ176235		
	GU112058			EU589392	GU112007		
	GU112055			EU589392	FJ179371		
5	FJ179371	1674 (1722)	2508 (2559)	GU112079	Unknown	RGBMCS**3**	1.06×10^{-39}
	FJ179373			GU112078			
6	GU112005	(?)1343 [(?)1390]	1867 (1915)	GU112057	EU589392	GBMCS**3**	1.02×10^{-35}
	GU112006			FJ176235	EU482411		
7	FJ176235	674 (721)	1929 (1977)	Unknown	GU112070	RGBMCS**3**	1.20×10^{-30}
8	GU112065	643 (690)	2073 (2121)	GU112079	Unknown	GBMCS**3**	1.15×10^{-23}
	GU112005			FJ179371	Unknown		
	GU112064			FJ179372	Unknown		
	GU112057			GU112080	Unknown		
	GU112061			GU112080	Unknown		
	GU112062			GU112080	Unknown		
9	FJ179370			Unknown	GU112079	RGBMCS**3**	1.16×10^{-23}
	GU112075			Unknown	GU112079		
	GU112074			Unknown	GU112079		
	GU112069			Unknown	GU112079		
	GU112073			Unknown	GU112079		
	GU112071			Unknown	GU112079		
	GU112070			Unknown	GU112079		
10	GU112078	(?)2558 [(?)2609]	316 (363)	GU112071	Unknown	GMCS**3**	1.22×10^{-22}
11	FJ179371	1911 (1959)	(?)2508 [(?)2559]	EU589392	FJ179373	GMS**3**	6.82×10^{-21}
12	FJ179372	1887 (1935)	2306 (2357)	GU112077	FJ179373	GBMC**3**	1.47×10^{-20}
	GU112080			GU112078	GU112007		
	GU112074			FJ176236			
13	GU112056	25 (45)	(?)2508 [(?)2559]	Unknown	GU112072	GBMCS**3**	2.21×10^{-20}
	AJ002453			Unknown	AJ002453		
	AF241479			Unknown	GU112077		
	GU112058			Unknown	FJ176236		
	GU112055			Unknown	FJ179370		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
BYVMV							
14	GU112007	2553 (2607)	2717 (2813)	Unknown	GU112068	R**G**BMC3	1.67×10^{-18}
	FJ179372			Unknown	GU112077		
	GU112080			Unknown	FJ179372		
15	FN645917	1045 (1092)	(?)2534 [(?)2585]	GU112072	Unknown	GBMCS**3**	4.50×10^{-17}
	EU589392			FJ179373	Unknown		
	FJ176236			FJ179371	Unknown		
16	AJ002453	1192 (1239)	1601 (1649)	GU112072	EU589392	**G**BMC3	8.23×10^{-17}
	GU112007			GU112077	EU482411		
17	GU112077	2545 (2596)	2702 (2796)	Unknown	GU112072	R**G**BMCS3	5.93×10^{-15}
18	FJ176235	(?)1930 [(?)1978]	2141 (2189)	GU112076	EU482411	GBMCS**3**	1.83×10^{-14}
19	FJ179373	(?)328 [(?)375]	1294 (1341)	Unknown	GU112072	GBMS**3**	2.01×10^{-13}
	FJ179371			Unknown	FJ176235		
	GU112079			Unknown	FJ179371		
20	GU112076	958 (1005)	(?)1869 [(?)1917]	GU112068	Unknown	GBMS**3**	1.32×10^{-11}
21	GU112075	(?)1303 [(?)1304]	1905 (1907)	GU112067	Unknown	GBMC**S**3	6.58×10^{-11}
22	GU112063	2451 (2502)	(?)2546 [(?)2597]	GU112079	GU112055	GBMCS**3**	1.85×10^{-08}
23	GU112063	2672 (2742)	(?)130 [(?)177]	GU112076	FN645917	GMC**3**	2.16×10^{-08}
	FJ176236			GU112007	EU589392		
	GU112058			FJ179372	GU112055		
24	GU112063	681 (728)	(?)1175 [(?)1222]	GU112067	GU112007	BMC**S**3	2.93×10^{-09}
	FJ179372			EU589392	GU112007		
	GU112080			FN645917	GU112007		
25	FJ179370	(?)1278 [(?)1325]	1640 (1325)	Unknown	GU112063	GMC**3**	2.41×10^{-07}
26	FJ176235	(?)2740 [(?)2832]	(?)642 [(?)689]	Unknown	GU112065	BCS**3**	1.29×10^{-06}
CLCuBuV							
1	JF510461	109 (109)	1055 (1060)	JF502365	Unknown	GBMCS**3**	1.75×10^{-06}
2	JF502353	954 (959)	80 (80)	JF502367	Unknown	MC**S**3	2.17×10^{-14}
3	JF509747	860 (865)	1521 (1526)	HM461866	Unknown	GBMCS**3**	1.04×10^{-11}
4	EU384572	1834 (1843)	2246 (2255)	EU365618	JF502367	GBMC**3**	2.60×10^{-11}
5	JF502356	2319 (2331)	465 (465)	Unknown	JF502360	GBMCS**3**	1.33×10^{-10}
	JF502355			Unknown	JF510461		
6	JF502372	1452 (1458)	2697 (2709)	JF510459	Unknown	MCS**3**	1.04×10^{-08}
7	JF502356	1320 (1326)	1907 (1919)	Unknown	JF502368	GMC**3**	3.64×10^{-07}
8	JF502354	1011 (1016)	2256 (2267)	Unknown	JF502367	GMC**3**	4.41×10^{-06}
	JF502357			Unknown	JF502368		
9	JF502367	338 (338)	821 (826)	JF502366	Unknown	GB**M**C3	2.57×10^{-05}
	JF502368			JF502355	Unknown		
10	JF502366	(?)1325 [(?)1331]	2315 (2327)	Unknown	JF502365	GMC**S**3	2.68×10^{-06}

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
CLCuGV							
1	AY036008	19 (19)	1702 (1724)	FR751144	FJ868828	RGBMCS**3**	1.48×10^{-36}
	AY036006			FR751142	FJ868828		
	AY036007			FR751143	FJ868828		
	AF260241			FR751146	FJ868828		
	FN554526			AF155064	FJ868828		
	FN554533			AY036010	FJ868828		
	FN554534			GU945265	FJ868828		
	FN554535			FN554526	FJ868828		
	FN554536			FN554533	FJ868828		
	FN554537			FN554534	FJ868828		
	FN554538			FN554535	FJ868828		
	FN554539			FN554536	FJ868828		
	FN554528			FN554537	FJ868828		
	FN554521			FN554538	FJ868828		
	FN554523			FN554539	FJ868828		
	EU432374			FN554528	FJ868828		
	FJ469627			FN554521	FJ868828		
	FN554522			FN554523	FJ868828		
	FN554540			FN554531	FJ868828		
	FN554519			FN554531	FJ868828		
	FN554520			FN554531	FJ868828		
	FN554527			FN554531	FJ868828		
	FN554524			FN554531	FJ868828		
	FN554525			FN554531	FJ868828		
	FN554529			FN554531	FJ868828		
	FN554530			FN554531	FJ868828		
	FN554531			FN554531	FJ868828		
2	AF155064	380 (390)	1301 (1319)	EU432373	Unknown	GBMC**S**3	1.29×10^{-11}
	FR751142			EU024120	Unknown		
	FR751143			FN554541	Unknown		
	FR751144			FJ469626	Unknown		
	FR751146			FN554526	Unknown		
	AY036010			FN554533	Unknown		
	GU945265			FN554534	Unknown		
3	EU024120	615 (626)	911 (922)	FN554541	Unknown	GMC**3**	4.11×10^{-08}
4	FJ868828	(?)1774 [(?)1795]	2408 (2430)	Unknown	AY036007	GBMCS**3**	1.49×10^{-06}
CLCuMV							
1	JN678806	1781 (1814)	16 (16)	AY765257	Unknown	GBMCS**3**	7.77×10^{-44}
	AJ496287			AY765257	Unknown		
	AJ002447			AY765257	Unknown		
	AJ002458			AY765257	Unknown		
	AJ496461			AY765257	Unknown		
	X98995			AY765257	Unknown		
	EU384573			AY765257	Unknown		
	JN678804			AY765257	Unknown		
2	AY765256	2700 (2764)	929(947)	AJ002459	Unknown	RGBMCS**3**	2.11×10^{-31}
	DQ191160			FJ218487	Unknown		
	AY765253			JN678803	Unknown		
	AY765257			JN558352	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
CLCuMV							
3	JN558352	1915 (1949)	720 (751)	Unknown	AJ002459	GBMCS**3**	3.12×10^{-25}
4	JN678803	2206 (2241)	241 (265)	Unknown	FJ218487	GBMCS**3**	5.63×10^{-24}
5	DQ191160	1360 (1378)	2547 (2570)	EU365616	AY765257	GBMC**S**3	2.02×10^{-27}
	AY765253			EU365616	JN558352		
	AY765256			EU365616	AJ002459		
6	AY765253	403 (433)	851 (882)	EU384573	Unknown	GMCS**3**	1.61×10^{-18}
	DQ191160			EU365616	Unknown		
	AY765256			JN968573	Unknown		
7	JN678803	524 (552)	1343 (1376)	JN678806	AJ496461	GBMC**S**3	1.72×10^{-18}
	EU365616			EU365616	JN968573		
	FJ218487			JN558352	FJ770370		
	AJ496287			AY765257	GU574208		
	AJ002447			AY765257	HQ455367		
	AJ002458			AY765257	HQ455356		
	EU384573			AY765257	HQ455357		
	JN678804			AY765257	HQ455355		
	AJ002459			AY765257	JQ317603		
	AJ132430			AY765257	GQ503175		
8	JN678806	2110 (2146)	2437 (2473)	EU384573	AY765257	GMC**3**	4.61×10^{-17}
9	FJ218487	2206 (2241)	224 (250)	AY765257	Unknown	BM**S**3	8.59×10^{-19}
10	AJ002447	2120 (2156)	(?)2670 [(?)2729]	AY765257	JN968573	GC**S**3	3.29×10^{-14}
	AJ496287			JN558352	FJ770370		
	AJ002458			AJ002459	GU574208		
	AJ496461			AJ132430	HQ455367		
	X98995			DQ191160	HQ455356		
	EU384573			AY765253	HQ455357		
10	JN678804	2120 (2156)	(?)2670 [(?)2729]	AY765256	HQ455355	GC**S**3	3.29×10^{-14}
	JN678806			AY765256	JQ317603		
11	AJ496461	(?)17 [(?)17]	(?)394 [(?)424]	AJ002458	JN678806	GBMCS**3**	2.87×10^{-14}
	EU365616			AJ496287	EU365616		
	X98995			AJ002447	EU384573		
	EU384573			AJ002459	JN678804		
	JN678804				AJ132430		
	AJ132430						
12	JN678806	(?)725 [(?)755]	1028 (1058)	JN880418	Unknown	GMC**S**	2.44×10^{-09}
	AY765257			JN880418	Unknown		
13	JN968573	(?)1525 [(?)1559]	1758 (1792)	JN678803	Unknown	MC**S**3	6.39×10^{-08}
	FJ770370			JN678803	Unknown		
	GU574208			JN678803	Unknown		
	HQ455367			JN678803	Unknown		
	HQ455356			JN678803	Unknown		
	HQ455357			JN678803	Unknown		
	HQ455355			JN678803	Unknown		
	JQ317603			JN678803	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
CLCuMV							
13	GQ503175	(?)1525 [(?)1559]	1758 (1792)	JN678803	Unknown	MC**S**3	6.39×10^{-08}
	GQ924756			JN678803	Unknown		
	HQ455349			JN678803	Unknown		
	JQ424826			JN678803	Unknown		
	EF465535			JN678803	Unknown		
	HQ455363			JN678803	Unknown		
	HQ455361			JN678803	Unknown		
	AJ496461			JN678803	Unknown		
	X98995			JN678803	Unknown		
EACMV							
1	AJ717532	1023 (1036)	536 (547)	Unknown	AY795986	GBMCS**3**	1.40×10^{-39}
	JN053439			Unknown	JF909156		
	JN053457			Unknown	JF909152		
	JN053458			Unknown	JF909161		
	AJ717518			Unknown	JF909158		
	AJ618958			Unknown	JF909159		
	AJ717533			Unknown	JF909160		
	AJ717535			Unknown	JF909154		
	JN053450			Unknown	JF909155		
	AJ717520			Unknown	JF909153		
	Z83257			Unknown	JF909157		
	AJ717522			Unknown	JF909162		
	AY795988			Unknown	JF909082		
	AJ618956			Unknown	JF909090		
	JN053448			Unknown	JF909119		
	AJ717519			Unknown	JF909139		
	JN053441			Unknown	JF909115		
	JN053455			Unknown	JF909120		
	JN053451			Unknown	JF909121		
	JN053454			Unknown	JF909118		
	JN053445			Unknown	JF909133		
	JN053449			Unknown	JF909110		
	JN053459			Unknown	JF909111		
	JN053437			Unknown	JF909149		
	JN053453			Unknown	JF909084		
	JN053435			Unknown	JF909099		
	JN053456			Unknown	JF909097		
	JN053452			Unknown	JF909106		
	JN053436			Unknown	JF909131		
	JN053443			Unknown	JF909150		
	JN053434			Unknown	AJ717558		
	JN053442			Unknown	JF909193		
	JN053446			Unknown	JF909185		
	JN053438			Unknown	JF909073		
	JN053444			Unknown	JF909074		
	JN053433			Unknown	JF909063		
	JN053432			Unknown	JF909174		
	JN053440			Unknown	JF909075		
	JN053447			Unknown	JF909096		

Event	Recombinant [1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
EACMV							
1	AJ717530	1023 (1036)	536 (547)	Unknown	JF909175	GBMCS**3**	1.40×10^{-39}
	HE814063			Unknown	JF909165		
	HE814064			Unknown	JF909166		
	AJ717516			Unknown	JF909188		
	AJ717517			Unknown	JF909189		
	AJ618959			Unknown	JF909190		
	AJ618957			Unknown	JF909069		
	AJ717528			Unknown	JF909070		
	AJ717523			Unknown	JF909071		
	AJ717525			Unknown	JF909072		
	AJ717524			Unknown	JF909181		
	AF126804			Unknown	JF909182		
	AF126806			Unknown	JF909169		
2	AY211887	2795 (2813)	1841 (1854)	JF909167	Unknown	GBMC**S**3	7.65×10^{-35}
3	AJ717543	525 (534)	1848 (1861)	AJ717541	AJ717545	GBMCS**3**	6.74×10^{-21}
4	AJ717543	(?)1885 [(?)1898]	2078 (2091)	AJ717516	Unknown	GMC**3**	1.71×10^{-05}
	AJ717542			AY211887	Unknown		
	AJ717541			JN053457	Unknown		
	AJ717537			JN053458	Unknown		
	AJ717536			AJ717518	Unknown		
	AJ717538			AJ618958	Unknown		
	AJ717539			JN053450	Unknown		
	AJ717540			AJ717520	Unknown		
5	AJ717557	1462 (1475)	39 (40)	AY795986	AJ717548	MC**S**3	6.29×10^{-12}
MYMIV							
1	AY547317	1993 (2011)	1 (10)	AY271895	FM208833	GBMC**S**3	7.43×10^{-10}
	FM208839			FM208839	FM955599		
	FM955598			AJ512498	FM955600		
	FM208840			AY269990	FM208834		
	FM208837			FM208835	FM208834		
	DQ389154			FM208838	FM208834		
	AF481865			FM208838	FM208834		
	FM208836			FM208838	FM208834		
2	FN794200	460 (466)	2241 (2249)	AJ512496	AJ416349	MCS**3**	2.87×10^{-07}
	AF416742			FM208844	AJ416349		
3	AJ416349	274 (280)	1745 (1753)	AY049772	Unknown	M**C**S3	5.51×10^{-05}
	AF416742			DQ389153	Unknown		
	FN794200			DQ389153	Unknown		
4	AJ416349	2440 (2448)	(?)17 [(?)17]	Unknown	DQ389153	GBMC**S**3	1.77×10^{-05}
SPLCV							
1	HQ393473	8 (8)	2059 (2085)	Unknown	FJ969834	RGBMCS**3**	5.72×10^{-60}
	FJ969833				EU267799		
2	AB433786	2250 (2276)	2818 (2890)	AB433787	Unknown	GBMCS**3**	2.78×10^{-54}

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
SPLCV							
3	DQ644562	1574 (1601)	2792 (2879)	DQ644561	AJ586885	RGBMCS**3**	2.55×10^{-51}
	HQ333138			EU309693	EU856364		
	HQ333135			AB433788	EU253456		
	HQ333137			AB433788	FN806776		
	FJ515896			AB433788	AB433787		
	FJ515897			AB433788	HQ333142		
	FJ515898			AB433788	HQ333139		
	JQ349087			AB433788	HQ333140		
	DQ644563			AB433788	AB433788		
	EU856364			AB433788	HM754639		
	EU267799			AB433788	HM754638		
	EU253456			AB433788	HM754636		
	FN806776			AB433788	HM754634		
	AB433787			AB433788	HM754640		
	HQ333142			AB433788	FJ969836		
	HQ333139			AB433788	FJ969835		
	HQ333140			AB433788	FJ969837		
	HQ333141			AB433788	FJ560719		
	AB433788			AB433788	HM754637		
	HM754639			AB433788	AF104036		
	HM754638			AB433788	AF104036		
	HM754636			AB433788	AF104036		
	HM754634			AB433788	AF104036		
	HM754640			AB433788	AF104036		
	FJ969836			AB433788	AF104036		
	FJ969834			AB433788	AF104036		
	FJ969835			AB433788	AF104036		
	FJ969837			AB433788	AF104036		
	HM754641			AB433788	AF104036		
	FJ560719			AB433788	AF104036		
	HM754637			AB433788	AF104036		
	FJ176701			AB433788	AF104036		
	AF104036			AB433788	AF104036		
	JF768740			AB433788	AF104036		
4	FJ969832	2175 (2204)	2755 (2824)	HQ333139	Unknown	GCS**3**	3.87×10^{-28}
5	JF736657	2125 (2188)	2726 (2841)	Unknown	EU253456	GMCS**3**	2.45×10^{-25}
					AJ586885		
6	AB433787	1296 (1303)	(?)1531 [(?)1554]	HM754641	AJ586885	RGBMCS**3**	1.02×10^{-20}
	AB433786			EU309693	AJ586885		
7	DQ644562	117 (120)	1086 (1094)	HM754638	HQ333137	GBMCS**3**	1.58×10^{-19}
	DQ644561			EU309693	HQ333138		
	JQ349087			FJ969832	HQ333135		
	DQ644563			FJ969833			
8	EU253456	2428 (2483)	(?)2751 [(?)2866]	FN806776	Unknown	GMCS**3**	3.06×10^{-11}
9	HQ333141	1875 (1901)	2489 (2515)	AB433788	Unknown	GMCS**3**	2.27×10^{-09}
10	AJ586885	1121 (1129)	1193 (1201)	Unknown	EU856364	**G**BMC3	6.09×10^{-09}

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
SPLCV							
11	JQ349087	459 (466)	867 (874)	AJ586885	Unknown	BMCS3	7.80×10^{-09}
	DQ644561			AJ586885	Unknown		
	HQ333138			AJ586885	Unknown		
	HQ333135			AJ586885	Unknown		
	HQ333137			AJ586885	Unknown		
	DQ644562			AJ586885	Unknown		
	DQ644563			AJ586885	Unknown		
12	HM754641	(?)2827 [(?)2916]	389 (396)	Unknown	HQ333141	GBMCS3	3.66×10^{-09}
	HM754639			Unknown	HQ393473		
	HM754638			Unknown	HM754639		
	HM754636			Unknown	FJ969837		
	HM754634			Unknown	FJ969837		
	HM754640			Unknown	FJ969837		
	FJ560719			Unknown	FJ969837		
	HM754637			Unknown	FJ969837		
13	DQ644563	1718 (1745)	(?)2792 [(?)2879]	HM754634	Unknown	GMCS3	1.51×10^{-08}
	FJ969832			HQ333138	Unknown		
	AJ586885			HQ333135	Unknown		
	HQ333138			HM754639	Unknown		
	HQ333135			HM754638	Unknown		
	DQ644562			HM754636	Unknown		
	EU856364			HM754640	Unknown		
	HQ333142			HM754641	Unknown		
	HQ333139			FJ560719	Unknown		
	HQ333140			HM754637	Unknown		
14	AB433787	1770 (1796)	(?)2625 [(?)2654]	FJ969836	AB433788	GMS3	1.36×10^{-06}
	AB433786			FJ969834	FJ176701		
15	EU309693	(?)2813 [(?)2893]	702 (708)	Unknown	FJ176701	GBMCS3	1.24×10^{-05}
				Unknown	AF104036		
				Unknown	JF768740		
16	AF104036	(?)2825 [(?)2914]	740 (747)	FJ969834	HQ333139	MCS3	2.56×10^{-05}
	FJ969832			HQ393473	HQ333140		
ToLCNDV							
1	FN645905	163 (195)	2063 (2113)	Unknown	283468616	RGBMCS3	1.32×10^{-36}
2	DQ989325	2074 (2118)	835 (873)	Unknown	AY939926	GBMCS3	1.07×10^{-22}
3	GU112086	1681 (1722)	638 (674)	Unknown	GU112082	RGBMCS3	1.16×10^{-22}
4	GU112088	1286 (1322)	75 (102)	FJ468356	GU112082	GBMCS3	7.13×10^{-22}
	GQ865546			AF102276	GQ865546		
	EF035482			AY286316	EF035482		
	GU112084			EF043231	GU112084		
5	DQ989325	(?)986 [(?)1024]	1366 (1404)	AM286434	DQ629101	BMCS3	7.87×10^{-12}
6	DQ116883	632 (670)	1461 (1500)	AM947506	Unknown	GBMCS3	5.15×10^{-11}
	DQ116885			AY286316	Unknown		
7	GU112082	543 (581)	1461 (1500)	HQ264185	Unknown	GBMCS3	6.80×10^{-11}

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
ToLCNDV							
8	AM258977	1871 (1920)	2680 (2736)	AM849548	Unknown	BMC**S**3	1.13×10^{-14}
	DQ116883			AM849548	Unknown		
	AF448058			AM849548	Unknown		
	AM286434			AM849548	Unknown		
	151416851			AM849548	Unknown		
	JN208136			AM849548	Unknown		
	AF448059			AM849548	Unknown		
	AM947506			AM849548	Unknown		
	FN435309			AM849548	Unknown		
	FJ468356			AM849548	Unknown		
	HQ141673			AM849548	Unknown		
	HM345979			AM849548	Unknown		
	AJ620187			AM849548	Unknown		
	HM159454			AM849548	Unknown		
	FN435310			AM849548	Unknown		
	U15015			AM849548	Unknown		
	DQ116880			AM849548	Unknown		
	DQ169056			AM849548	Unknown		
9	DQ116885	2311 (2361)	188 (220)	AM292302	Unknown	GMC**3**	2.74×10^{-07}
ToLCTV							
1	DQ866128	2697 (2709)	2016 (2017)	Unknown	DQ237918	GBMCS**3**	1.48×10^{-51}
2	GU723711	2098 (2098)	2736 (2749)	EU624503	GU723722	GBMCS**3**	1.16×10^{-26}
3	GU723717	1979 (1980)	111 (112)	DQ866129	Unknown	GBMCS**3**	3.16×10^{-21}
	GU723730			DQ866129	Unknown		
	DQ866123			DQ866129	Unknown		
	GU723715			DQ866129	Unknown		
	GU723725			DQ866129	Unknown		
	GU723722			DQ866129	Unknown		
	GU723720			DQ866129	Unknown		
4	GU723726	610 (611)	2186 (2190)	GU723724	GU723708	GBMC**S**3	4.33×10^{-20}
5	GU723725	1021 (1023)	(?)1959 [(?)1961]	GU723724	DQ866129	BM**S**3	4.75×10^{-11}
	DQ866123			DQ866130	GU723715		
	GU723717			DQ866130	GU723715		
	GU723715			DQ866130	GU723715		
	GU723722			DQ866130	GU723715		
ToYLCCNV							
1	AJ971524	2066 (2073)	2734 (2756)	AJ457985	Unknown	GBMCS**3**	5.20×10^{-84}
	JN082233			AM260703	Unknown		
	JN082237			AJ457823	Unknown		
	AJ971265			AM260702	Unknown		
2	EF011559	2693 (2711)	543 (548)	AJ781302	Unknown	RGBMCS **3**	5.65×10^{-35}
3	AJ457985	2367 (2377)	2672 (2690)	AM260703	AJ781302	GBMCS**3**	4.37×10^{-19}
4	GU199588	1041 (1048)	2672 (2690)	AJ781302	AM260703	RMCS**3**	2.31×10^{-12}
	DQ256460			AJ457823			
5	AM261326	1211 (1219)	2654 (2669)	AJ319676	AM260703	RB**M**CS	3.42×10^{-09}
	AM980509			AJ319675			

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCTV							
1	AY514632	578 (579)	1902 (1909)	Unknown	DQ871222	GBMCS3	9.56×10^{-16}
	AY514630				AF141922		
2	AY514632	2498 (2505)	295 (296)	Unknown	AY514630	GBMCS**3**	2.22×10^{-10}
3	X63015	2228 (2239)	2718 (2733)	DQ871222	AJ495812	GBMCS**3**	2.49×10^{-09}
	AF141922			AF206674			
4	AJ495812	1976 (1983)	2704 (2714)	AF206674	Unknown	GBMCS**3**	3.80×10^{-08}
				AF206674			
TYLCV							
1	GU076442	2123 (2154)	2683 (2739)	GU076452	AY044138	RGBMCS**3**	9.33×10^{-40}
	GU076454			FJ956703	GQ861427		
	GU076443			FJ956704	EF054894		
	GU076449			GU076453	EF185318		
	FJ956706			GU076441	EU143745		
	EU635776			FJ956705	EF158044		
	GU076448			FJ956701	AF105975		
2	JN604488	2701 (2745)	2119 (2137)	AJ132711	Unknown	GBMCS**3**	3.14×10^{-35}
	FJ956703			AJ132711	Unknown		
	FJ956704			AJ132711	Unknown		
	JN604487			AJ132711	Unknown		
	JN604486			AJ132711	Unknown		
	GU076452			AJ132711	Unknown		
	GU076453			AJ132711	Unknown		
	GU076441			AJ132711	Unknown		
	FJ956705			AJ132711	Unknown		
	FJ956701			AJ132711	Unknown		
	FJ956702			AJ132711	Unknown		
	DQ644565			AJ132711	Unknown		
	GU076450			AJ132711	Unknown		
	GU076451			AJ132711	Unknown		
3	AJ519441	2720 (2756)	2014 (2026)	Unknown	21426900	GBMCS**3**	4.49×10^{-30}
	GQ861427			Unknown	DQ144621		
	EF054894			Unknown	EF101929		
	EF185318			Unknown	GQ861426		
	EU143745			Unknown	AJ812277		
	EF158044			Unknown	EF051116		
	AF105975			Unknown	EF107520		
	AB439842			Unknown	AY134494		
	AB116633			Unknown	GU355941		
	AB116635			Unknown	AY530931		
	AB116636			Unknown	AF024715		
	AB116634			Unknown	AJ223505		
	AB014347			Unknown	EF210554		
	AJ865337			Unknown	EF110890		
	AF071228			Unknown	AY594174		
	AB116632			Unknown	EF433426		
	AB014346			Unknown	AB116629		
	AB110218			Unknown	AB110217		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
4	GU076448	1137 (1168)	1707 (1738)	JN604486	Unknown	RGMCS**3**	3.97×10^{-18}
	AJ132711			FJ956703	Unknown		
	JQ414025			FJ956704	Unknown		
	GU076445			JN604487	Unknown		
	FJ355946			JN604488	Unknown		
	GU076446			GU076450	Unknown		
	GU076447			GU076451	Unknown		
	GU076440			GU076451	Unknown		
	GU076444			GU076451	Unknown		
	DQ144621			GU076451	Unknown		
	EF101929			GU076451	Unknown		
	GQ861426			GU076451	Unknown		
	AJ812277			GU076451	Unknown		
	AB613209			GU076451	Unknown		
	HM988987			GU076451	Unknown		
	AM698119			GU076451	Unknown		
	HM130914			GU076451	Unknown		
	GU325633			GU076451	Unknown		
	AB636412			GU076451	Unknown		
	AB636410			GU076451	Unknown		
	AB669434			GU076451	Unknown		
	HM130913			GU076451	Unknown		
	JN183877			GU076451	Unknown		
	GU325632			GU076451	Unknown		
	HM856910			GU076451	Unknown		
	HM856916			GU076451	Unknown		
	JN183878			GU076451	Unknown		
	HM856918			GU076451	Unknown		
	AB636264			GU076451	Unknown		
	AB613208			GU076451	Unknown		
	AB439841			GU076451	Unknown		
	EU031444			GU076451	Unknown		
	AB363566			GU076451	Unknown		
	GU322424			GU076451	Unknown		
	JN990926			GU076451	Unknown		
	JQ038232			GU076451	Unknown		
	JQ867092			GU076451	Unknown		
	JN412854			GU076451	Unknown		
	GU126513			GU076451	Unknown		
	GU178820			GU076451	Unknown		
	GU178819			GU076451	Unknown		
	GU178817			GU076451	Unknown		
	AM698118			GU076451	Unknown		
	FN256258			GU076451	Unknown		
	FN252890			GU076451	Unknown		
	AB192965			GU076451	Unknown		
	AB192966			GU076451	Unknown		
	AM282874			GU076451	Unknown		
	GU434141			GU076451	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
4	JQ004047	1137 (1168)	1707 (1738)	GU076451	Unknown	RGMCS**3**	3.97×10^{-18}
	JQ034613			GU076451	Unknown		
	HM358879			GU076451	Unknown		
	JF964959			GU076451	Unknown		
	GU563330			GU076451	Unknown		
	JQ038233			GU076451	Unknown		
	JQ004050			GU076451	Unknown		
	JQ326957			GU076451	Unknown		
	JF727878			GU076451	Unknown		
	JF414236			GU076451	Unknown		
	JN990925			GU076451	Unknown		
	FJ646611			GU076451	Unknown		
	HM043732			GU076451	Unknown		
	JN990928			GU076451	Unknown		
	JF301668			GU076451	Unknown		
	HM627881			GU076451	Unknown		
	JF301667			GU076451	Unknown		
	JF817218			GU076451	Unknown		
	JQ004046			GU076451	Unknown		
	JN990923			GU076451	Unknown		
	FN650807			GU076451	Unknown		
	AM698117			GU076451	Unknown		
	HM208334			GU076451	Unknown		
	JN990927			GU076451	Unknown		
	HM627883			GU076451	Unknown		
	HM627882			GU076451	Unknown		
	FN650808			GU076451	Unknown		
	JQ004028			GU076451	Unknown		
	FN256257			GU076451	Unknown		
	GU951437			GU076451	Unknown		
	GQ352538			GU076451	Unknown		
	HQ702861			GU076451	Unknown		
	GQ352537			GU076451	Unknown		
	HQ702862			GU076451	Unknown		
	GU199587			GU076451	Unknown		
	FN256259			GU076451	Unknown		
	JQ807735			GU076451	Unknown		
	JQ038238			GU076451	Unknown		
	JQ038236			GU076451	Unknown		
	JQ038239			GU076451	Unknown		
	HM627885			GU076451	Unknown		
	JQ038240			GU076451	Unknown		
	HM627884			GU076451	Unknown		
	HM627880			GU076451	Unknown		
	JQ411237			GU076451	Unknown		
	GU434144			GU076451	Unknown		
	HM459851			GU076451	Unknown		
	EF210555			GU076451	Unknown		
	DQ631892			GU076451	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
4	FJ609655	1137 (1168)	1707 (1738)	GU076451	Unknown	RGMCS**3**	3.97×10^{-18}
	FJ012358			GU076451	Unknown		
	EF523478			GU076451	Unknown		
	GU178814			GU076451	Unknown		
	GU178813			GU076451	Unknown		
	GU178818			GU076451	Unknown		
	GU178816			GU076451	Unknown		
	GU178815			GU076451	Unknown		
	EF539831			GU076451	Unknown		
	GU322423			GU076451	Unknown		
	JN990924			GU076451	Unknown		
	JQ038235			GU076451	Unknown		
	JF414237			GU076451	Unknown		
	JQ004048			GU076451	Unknown		
	GU434143			GU076451	Unknown		
	GU434142			GU076451	Unknown		
	GU111505			GU076451	Unknown		
	JQ004051			GU076451	Unknown		
	JQ038234			GU076451	Unknown		
	JN990922			GU076451	Unknown		
	JF833036			GU076451	Unknown		
	JQ004045			GU076451	Unknown		
	GU348995			GU076451	Unknown		
	JQ038237			GU076451	Unknown		
	GU951436			GU076451	Unknown		
	GU983859			GU076451	Unknown		
	JQ004049			GU076451	Unknown		
	JQ004052			GU076451	Unknown		
	EF051116			GU076451	Unknown		
	EF107520			GU076451	Unknown		
	FR851298			GU076451	Unknown		
	FR851297			GU076451	Unknown		
	AY134494			GU076451	Unknown		
	GU355941			GU076451	Unknown		
	JN680353			GU076451	Unknown		
	JQ303121			GU076451	Unknown		
	AY530931			GU076451	Unknown		
	AF024715			GU076451	Unknown		
	AJ223505			GU076451	Unknown		
	EF210554			GU076451	Unknown		
	EF110890			GU076451	Unknown		
	AY594174			GU076451	Unknown		
	EF433426			GU076451	Unknown		
	AB116629			GU076451	Unknown		
	AB110217			GU076451	Unknown		
	AB116631			GU076451	Unknown		
	AB116630			GU076451	Unknown		
	AB636411			GU076451	Unknown		
	HM856919			GU076451	Unknown		
	JQ013091			GU076451	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
4	JQ013089	1137 (1168)	1707 (1738)	GU076451	Unknown	RGMCS**3**	3.97×10^{-18}
	JN680149			GU076451	Unknown		
	JN183880			GU076451	Unknown		
	JN183879			GU076451	Unknown		
	GU325634			GU076451	Unknown		
	JN183874			GU076451	Unknown		
	JN183875			GU076451	Unknown		
	GQ141873			GU076451	Unknown		
	HM856917			GU076451	Unknown		
	HM856915			GU076451	Unknown		
	HM856909			GU076451	Unknown		
	JQ013090			GU076451	Unknown		
	HM856873			GU076451	Unknown		
	HM856913			GU076451	Unknown		
	JN183876			GU076451	Unknown		
	JN183873			GU076451	Unknown		
	HM856914			GU076451	Unknown		
	JN183872			GU076451	Unknown		
	HM130912			GU076451	Unknown		
	HM856912			GU076451	Unknown		
	JN680150			GU076451	Unknown		
	HM856911			GU076451	Unknown		
	EF054893			GU076451	Unknown		
	EF060196			GU076451	Unknown		
	FJ439569			GU076451	Unknown		
	HE603246			GU076451	Unknown		
	HE603242			GU076451	Unknown		
	HE603244			GU076451	Unknown		
	HE603243			GU076451	Unknown		
	HE603241			GU076451	Unknown		
	AJ489258			GU076451	Unknown		
	21426900			GU076451	Unknown		
	HM448447			GU076451	Unknown		
	AM409201			GU076451	Unknown		
	GQ861427			GU076451	Unknown		
	EF054894			GU076451	Unknown		
	EF185318			GU076451	Unknown		
	EU143745			GU076451	Unknown		
	EF158044			GU076451	Unknown		
	AJ519441			GU076451	Unknown		
	AF105975			GU076451	Unknown		
	AB439842			GU076451	Unknown		
	AB116633			GU076451	Unknown		
	AB116635			GU076451	Unknown		
	AB116636			GU076451	Unknown		
	AB116634			GU076451	Unknown		
	AB014347			GU076451	Unknown		
	AJ865337			GU076451	Unknown		
	AF071228			GU076451	Unknown		
	AB116632			GU076451	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
4	AB014346	1137 (1168)	1707 (1738)	GU076451	Unknown	RGMCS**3**	3.97×10^{-18}
	AB110218			GU076451	Unknown		
	GU076449			GU076451	Unknown		
	FJ956706			GU076451	Unknown		
	EU635776			GU076451	Unknown		
5	GU076444	2432 (2444)	2548 (2560)	GU076440	GQ861427	GBMC3	4.88×10^{-17}
	DQ144621			JQ414025	GU076454		
	EF101929			GU076445	AY044138		
	GQ861426			FJ355946	EF054894		
	AJ812277			GU076446	EF185318		
	AB613209			GU076447	EU143745		
	HM988987			GU076447	EF158044		
	AM698119			GU076447	AJ519441		
	HM130914			GU076447	AF105975		
	GU325633			GU076447	AB439842		
	AB636412			GU076447	AB116633		
	AB636410			GU076447	AB116635		
	AB669434			GU076447	AB116636		
	HM130913			GU076447	AB116634		
	JN183877			GU076447	AB014347		
	GU325632			GU076447	AJ865337		
	HM856910			GU076447	AF071228		
	HM856916			GU076447	AB116632		
	JN183878			GU076447	AB014346		
	HM856918			GU076447	AB110218		
	AB636264			GU076447	GU076442		
	AB613208			GU076447	GU076443		
	AB439841			GU076447	GU076449		
	EU031444			GU076447	FJ956706		
	AB363566			GU076447	EU635776		
	GU322424			GU076447	GU076448		
	JN990926			GU076447	GU076448		
	JQ038232			GU076447	GU076448		
	JQ867092			GU076447	GU076448		
	JN412854			GU076447	GU076448		
	GU126513			GU076447	GU076448		
	GU178820			GU076447	GU076448		
	GU178819			GU076447	GU076448		
	GU178817			GU076447	GU076448		
	AM698118			GU076447	GU076448		
	FN256258			GU076447	GU076448		
	FN252890			GU076447	GU076448		
	AB192965			GU076447	GU076448		
	AB192966			GU076447	GU076448		
	AM282874			GU076447	GU076448		
	GU434141			GU076447	GU076448		
	JQ004047			GU076447	GU076448		
	JQ034613			GU076447	GU076448		
	HM358879			GU076447	GU076448		
	JF964959			GU076447	GU076448		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
5	GU563330	2432 (2444)	2548 (2560)	GU076447	GU076448	GBMC3	4.88×10^{-17}
	JQ038233			GU076447	GU076448		
	JQ004050			GU076447	GU076448		
	JQ326957			GU076447	GU076448		
	JF727878			GU076447	GU076448		
	JF414236			GU076447	GU076448		
	HQ702863			GU076447	GU076448		
	JN990925			GU076447	GU076448		
	FJ646611			GU076447	GU076448		
	HM043732			GU076447	GU076448		
	JN990928			GU076447	GU076448		
	JF301668			GU076447	GU076448		
	HM627881			GU076447	GU076448		
	JF301667			GU076447	GU076448		
	JF817218			GU076447	GU076448		
	JQ004046			GU076447	GU076448		
	JN990923			GU076447	GU076448		
	FN650807			GU076447	GU076448		
	AM698117			GU076447	GU076448		
	HM208334			GU076447	GU076448		
	JN990927			GU076447	GU076448		
	HM627883			GU076447	GU076448		
	HM627882			GU076447	GU076448		
	FN650808			GU076447	GU076448		
	JQ004028			GU076447	GU076448		
	FN256257			GU076447	GU076448		
	GU951437			GU076447	GU076448		
	GQ352538			GU076447	GU076448		
	HQ702861			GU076447	GU076448		
	GQ352537			GU076447	GU076448		
	HQ702862			GU076447	GU076448		
	GU199587			GU076447	GU076448		
	FN256259			GU076447	GU076448		
	JQ807735			GU076447	GU076448		
	JQ038238			GU076447	GU076448		
	JQ038236			GU076447	GU076448		
	JQ038239			GU076447	GU076448		
	HM627885			GU076447	GU076448		
	JQ038240			GU076447	GU076448		
	HM627884			GU076447	GU076448		
	HM627880			GU076447	GU076448		
	JQ411237			GU076447	GU076448		
	GU434144			GU076447	GU076448		
	HM459851			GU076447	GU076448		
	EF210555			GU076447	GU076448		
	DQ631892			GU076447	GU076448		
	FJ609655			GU076447	GU076448		
	FJ012358			GU076447	GU076448		
	EF523478			GU076447	GU076448		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
5	GU178814	2432 (2444)	2548 (2560)	GU076447	GU076448	GBMC3	4.88×10^{-17}
	GU178813			GU076447	GU076448		
	GU178818			GU076447	GU076448		
	GU178816			GU076447	GU076448		
	GU178815			GU076447	GU076448		
	EF539831			GU076447	GU076448		
	GU322423			GU076447	GU076448		
	JN990924			GU076447	GU076448		
	JQ038235			GU076447	GU076448		
	JF414237			GU076447	GU076448		
	JQ004048			GU076447	GU076448		
	GU434143			GU076447	GU076448		
	GU434142			GU076447	GU076448		
	GU111505			GU076447	GU076448		
	JQ004051			GU076447	GU076448		
	JQ038234			GU076447	GU076448		
	JN990922			GU076447	GU076448		
	JF833036			GU076447	GU076448		
	JQ004045			GU076447	GU076448		
	GU348995			GU076447	GU076448		
	JQ038237			GU076447	GU076448		
	GU951436			GU076447	GU076448		
	GU983859			GU076447	GU076448		
	JQ004049			GU076447	GU076448		
	JQ004052			GU076447	GU076448		
	EF051116			GU076447	GU076448		
	EF107520			GU076447	GU076448		
	FR851298			GU076447	GU076448		
	FR851297			GU076447	GU076448		
	AY134494			GU076447	GU076448		
	GU355941			GU076447	GU076448		
	JN680353			GU076447	GU076448		
	JQ303121			GU076447	GU076448		
	AY530931			GU076447	GU076448		
	AF024715			GU076447	GU076448		
	AJ223505			GU076447	GU076448		
	EF210554			GU076447	GU076448		
	EF110890			GU076447	GU076448		
	AY594174			GU076447	GU076448		
	EF433426			GU076447	GU076448		
	AB116629			GU076447	GU076448		
	AB110217			GU076447	GU076448		
	AB116631			GU076447	GU076448		
	AB116630			GU076447	GU076448		
	AB636411			GU076447	GU076448		
	HM856919			GU076447	GU076448		
	JQ013091			GU076447	GU076448		
	JQ013089			GU076447	GU076448		
	JN680149			GU076447	GU076448		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
5	JN183880	2432 (2444)	2548 (2560)	GU076447	GU076448	GBMC3	4.88×10^{-17}
	JN183879			GU076447	GU076448		
	GU325634			GU076447	GU076448		
	JN183874			GU076447	GU076448		
	JN183875			GU076447	GU076448		
	GQ141873			GU076447	GU076448		
	HM856917			GU076447	GU076448		
	HM856915			GU076447	GU076448		
	HM856909			GU076447	GU076448		
	JQ013090			GU076447	GU076448		
	HM856873			GU076447	GU076448		
	HM856913			GU076447	GU076448		
	JN183876			GU076447	GU076448		
	JN183873			GU076447	GU076448		
	HM856914			GU076447	GU076448		
	JN183872			GU076447	GU076448		
	HM130912			GU076447	GU076448		
	HM856912			GU076447	GU076448		
	JN680150			GU076447	GU076448		
	HM856911			GU076447	GU076448		
	EF054893			GU076447	GU076448		
	EF060196			GU076447	GU076448		
	FJ439569			GU076447	GU076448		
	HE603246			GU076447	GU076448		
	HE603242			GU076447	GU076448		
	HE603244			GU076447	GU076448		
	HE603243			GU076447	GU076448		
	HE603241			GU076447	GU076448		
	AJ489258			GU076447	GU076448		
	21426900			GU076447	GU076448		
	HM448447			GU076447	GU076448		
	AM409201			GU076447	GU076448		
6	EF110890	(?)2015 [(?)2027]	2116 (2128)	FJ355946	Unknown	GBM**3**	5.64×10^{-13}
	GU076444			AJ132711	Unknown		
	DQ144621			JQ414025	Unknown		
	EF101929			GU076445	Unknown		
	GQ861426			GU076446	Unknown		
	AJ812277			GU076447	Unknown		
	AB613209			GU076440	Unknown		
	HM988987			GU076440	Unknown		
	AM698119			GU076440	Unknown		
	HM130914			GU076440	Unknown		
	GU325633			GU076440	Unknown		
	AB636412			GU076440	Unknown		
	AB636410			GU076440	Unknown		
	AB669434			GU076440	Unknown		
	HM130913			GU076440	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
6	JN183877	(?)2015 [(?)2027]	2116 (2128)	GU076440	Unknown	GBM**3**	5.64×10^{-13}
	GU325632			GU076440	Unknown		
	HM856910			GU076440	Unknown		
	HM856916			GU076440	Unknown		
	JN183878			GU076440	Unknown		
	HM856918			GU076440	Unknown		
	AB636264			GU076440	Unknown		
	AB613208			GU076440	Unknown		
	AB439841			GU076440	Unknown		
	EU031444			GU076440	Unknown		
	AB363566			GU076440	Unknown		
	GU322424			GU076440	Unknown		
	JN990926			GU076440	Unknown		
	JQ038232			GU076440	Unknown		
	JQ867092			GU076440	Unknown		
	JN412854			GU076440	Unknown		
	GU126513			GU076440	Unknown		
	GU178820			GU076440	Unknown		
	GU178819			GU076440	Unknown		
	GU178817			GU076440	Unknown		
	AM698118			GU076440	Unknown		
	FN256258			GU076440	Unknown		
	FN252890			GU076440	Unknown		
	AB192965			GU076440	Unknown		
	AB192966			GU076440	Unknown		
	AM282874			GU076440	Unknown		
	GU434141			GU076440	Unknown		
	JQ004047			GU076440	Unknown		
	JQ034613			GU076440	Unknown		
	HM358879			GU076440	Unknown		
	JF964959			GU076440	Unknown		
	GU563330			GU076440	Unknown		
	JQ038233			GU076440	Unknown		
	JQ004050			GU076440	Unknown		
	JQ326957			GU076440	Unknown		
	JF727878			GU076440	Unknown		
	JF414236			GU076440	Unknown		
	HQ702863			GU076440	Unknown		
	JN990925			GU076440	Unknown		
	FJ646611			GU076440	Unknown		
	HM043732			GU076440	Unknown		
	JN990928			GU076440	Unknown		
	JF301668			GU076440	Unknown		
	HM627881			GU076440	Unknown		
	JF301667			GU076440	Unknown		
	JF817218			GU076440	Unknown		
	JQ004046			GU076440	Unknown		
	JN990923			GU076440	Unknown		
	FN650807			GU076440	Unknown		
	AM698117			GU076440	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
6	HM208334	(?)2015 [(?)2027]	2116 (2128)	GU076440	Unknown	GBM**3**	5.64×10^{-13}
	JN990927			GU076440	Unknown		
	HM627883			GU076440	Unknown		
	HM627882			GU076440	Unknown		
	FN650808			GU076440	Unknown		
	JQ004028			GU076440	Unknown		
	FN256257			GU076440	Unknown		
	GU951437			GU076440	Unknown		
	GQ352538			GU076440	Unknown		
	HQ702861			GU076440	Unknown		
	GQ352537			GU076440	Unknown		
	HQ702862			GU076440	Unknown		
	GU199587			GU076440	Unknown		
	FN256259			GU076440	Unknown		
	JQ807735			GU076440	Unknown		
	JQ038238			GU076440	Unknown		
	JQ038236			GU076440	Unknown		
	JQ038239			GU076440	Unknown		
	HM627885			GU076440	Unknown		
	JQ038240			GU076440	Unknown		
	HM627884			GU076440	Unknown		
	HM627880			GU076440	Unknown		
	JQ411237			GU076440	Unknown		
	GU434144			GU076440	Unknown		
	HM459851			GU076440	Unknown		
	EF210555			GU076440	Unknown		
	DQ631892			GU076440	Unknown		
	FJ609655			GU076440	Unknown		
	FJ012358			GU076440	Unknown		
	EF523478			GU076440	Unknown		
	GU178814			GU076440	Unknown		
	GU178813			GU076440	Unknown		
	GU178818			GU076440	Unknown		
	GU178816			GU076440	Unknown		
	GU178815			GU076440	Unknown		
	EF539831			GU076440	Unknown		
	GU322423			GU076440	Unknown		
	JN990924			GU076440	Unknown		
	JQ038235			GU076440	Unknown		
	JF414237			GU076440	Unknown		
	JQ004048			GU076440	Unknown		
	GU434143			GU076440	Unknown		
	GU434142			GU076440	Unknown		
	GU111505			GU076440	Unknown		
	JQ004051			GU076440	Unknown		
	JQ038234			GU076440	Unknown		
	JN990922			GU076440	Unknown		
	JF833036			GU076440	Unknown		
	JQ004045			GU076440	Unknown		
	GU348995			GU076440	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
6	JQ038237	(?)2015 [(?)2027]	2116 (2128)	GU076440	Unknown	GBM**3**	5.64×10^{-13}
	GU951436			GU076440	Unknown		
	GU983859			GU076440	Unknown		
	JQ004049			GU076440	Unknown		
	JQ004052			GU076440	Unknown		
	EF051116			GU076440	Unknown		
	EF107520			GU076440	Unknown		
	FR851298			GU076440	Unknown		
	FR851297			GU076440	Unknown		
	AY134494			GU076440	Unknown		
	GU355941			GU076440	Unknown		
	JN680353			GU076440	Unknown		
	JQ303121			GU076440	Unknown		
	AY530931			GU076440	Unknown		
	AF024715			GU076440	Unknown		
	AJ223505			GU076440	Unknown		
	EF210554			GU076440	Unknown		
	AY594174			GU076440	Unknown		
	EF433426			GU076440	Unknown		
	AB116629			GU076440	Unknown		
	AB110217			GU076440	Unknown		
	AB116631			GU076440	Unknown		
	AB116630			GU076440	Unknown		
	AB636411			GU076440	Unknown		
	HM856919			GU076440	Unknown		
	JQ013091			GU076440	Unknown		
	JQ013089			GU076440	Unknown		
	JN680149			GU076440	Unknown		
	JN183880			GU076440	Unknown		
	JN183879			GU076440	Unknown		
	GU325634			GU076440	Unknown		
	JN183874			GU076440	Unknown		
	JN183875			GU076440	Unknown		
	GQ141873			GU076440	Unknown		
	HM856917			GU076440	Unknown		
	HM856915			GU076440	Unknown		
	HM856909			GU076440	Unknown		
	JQ013090			GU076440	Unknown		
	HM856873			GU076440	Unknown		
	HM856913			GU076440	Unknown		
	JN183876			GU076440	Unknown		
	JN183873			GU076440	Unknown		
	HM856914			GU076440	Unknown		
	JN183872			GU076440	Unknown		
	HM130912			GU076440	Unknown		
	HM856912			GU076440	Unknown		
	JN680150			GU076440	Unknown		
	HM856911			GU076440	Unknown		
	EF054893			GU076440	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
6	EF060196	(?)2015 [(?)2027]	2116 (2128)	GU076440	Unknown	GBM**3**	5.64×10^{-13}
	FJ439569			GU076440	Unknown		
	HE603246			GU076440	Unknown		
	HE603242			GU076440	Unknown		
	HE603244			GU076440	Unknown		
	HE603243			GU076440	Unknown		
	HE603241			GU076440	Unknown		
	AJ489258			GU076440	Unknown		
	21426900			GU076440	Unknown		
	HM448447			GU076440	Unknown		
	AM409201			GU076440	Unknown		
7	JN604486	18 (19)	987 (1004)	GU076450	Unknown	GMC**S**3	1.28×10^{-13}
	JN604487			GU076452	Unknown		
	JN604488			GU076453	Unknown		
8	AB116632	2607 (2620)	(?)2650 [(?)2027]	AF105975	Unknown	GMCS**3**	1.40×10^{-09}
	GQ861427			AB439842	Unknown		
	EF054894			AB116633	Unknown		
	EF185318			AB116635	Unknown		
	EU143745			AB116636	Unknown		
	EF158044			AB116634	Unknown		
	AJ519441			AB014347	Unknown		
	AJ865337			AB014347	Unknown		
	AF071228			AB014347	Unknown		
	AB014346			AB014347	Unknown		
	AB110218			AB014347	Unknown		
9	AB116636	16 (19)	(?)969 [(?)983]	GU076452	Unknown	MC**S**3	6.22×10^{-12}
	AJ132711			GU076453	Unknown		
	JQ414025			GU076441	Unknown		
	GU076445			GU076441	Unknown		
	FJ355946			GU076441	Unknown		
	GU076446			GU076441	Unknown		
	GU076447			GU076441	Unknown		
	GU076440			GU076441	Unknown		
	GU076444			GU076441	Unknown		
	DQ144621			GU076441	Unknown		
	EF101929			GU076441	Unknown		
	GQ861426			GU076441	Unknown		
	AJ812277			GU076441	Unknown		
	AB613209			GU076441	Unknown		
	HM988987			GU076441	Unknown		
	AM698119			GU076441	Unknown		
	HM130914			GU076441	Unknown		
	GU325633			GU076441	Unknown		
	AB636412			GU076441	Unknown		
	AB636410			GU076441	Unknown		
	AB669434			GU076441	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
9	HM130913	16 (19)	(?)969 [(?)983]	GU076441	Unknown	MC**S**3	6.22×10^{-12}
	JN183877			GU076441	Unknown		
	GU325632			GU076441	Unknown		
	HM856910			GU076441	Unknown		
	HM856916			GU076441	Unknown		
	JN183878			GU076441	Unknown		
	HM856918			GU076441	Unknown		
	AB636264			GU076441	Unknown		
	AB613208			GU076441	Unknown		
	AB439841			GU076441	Unknown		
	EU031444			GU076441	Unknown		
	AB363566			GU076441	Unknown		
	GU322424			GU076441	Unknown		
	JN990926			GU076441	Unknown		
	JQ038232			GU076441	Unknown		
	JQ867092			GU076441	Unknown		
	JN412854			GU076441	Unknown		
	GU126513			GU076441	Unknown		
	GU178820			GU076441	Unknown		
	GU178819			GU076441	Unknown		
	GU178817			GU076441	Unknown		
	AM698118			GU076441	Unknown		
	FN256258			GU076441	Unknown		
	FN252890			GU076441	Unknown		
	AB192965			GU076441	Unknown		
	AB192966			GU076441	Unknown		
	AM282874			GU076441	Unknown		
	GU434141			GU076441	Unknown		
	JQ004047			GU076441	Unknown		
	JQ034613			GU076441	Unknown		
	HM358879			GU076441	Unknown		
	JF964959			GU076441	Unknown		
	GU563330			GU076441	Unknown		
	JQ038233			GU076441	Unknown		
	JQ004050			GU076441	Unknown		
	JQ326957			GU076441	Unknown		
	JF727878			GU076441	Unknown		
	JF414236			GU076441	Unknown		
	HQ702863			GU076441	Unknown		
	JN990925			GU076441	Unknown		
	FJ646611			GU076441	Unknown		
	HM043732			GU076441	Unknown		
	JN990928			GU076441	Unknown		
	JF301668			GU076441	Unknown		
	HM627881			GU076441	Unknown		
	JF301667			GU076441	Unknown		
	JF817218			GU076441	Unknown		
	JQ004046			GU076441	Unknown		
	JN990923			GU076441	Unknown		
	FN650807			GU076441	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
9	AM698117	16 (19)	(?)969 [(?)983]	GU076441	Unknown	MC**S**3	6.22×10^{-12}
	HM208334			GU076441	Unknown		
	JN990927			GU076441	Unknown		
	HM627883			GU076441	Unknown		
	HM627882			GU076441	Unknown		
	FN650808			GU076441	Unknown		
	JQ004028			GU076441	Unknown		
	FN256257			GU076441	Unknown		
	GU951437			GU076441	Unknown		
	GQ352538			GU076441	Unknown		
	HQ702861			GU076441	Unknown		
	GQ352537			GU076441	Unknown		
	HQ702862			GU076441	Unknown		
	GU199587			GU076441	Unknown		
	FN256259			GU076441	Unknown		
	JQ807735			GU076441	Unknown		
	JQ038238			GU076441	Unknown		
	JQ038236			GU076441	Unknown		
	JQ038239			GU076441	Unknown		
	HM627885			GU076441	Unknown		
	JQ038240			GU076441	Unknown		
	HM627884			GU076441	Unknown		
	HM627880			GU076441	Unknown		
	JQ411237			GU076441	Unknown		
	GU434144			GU076441	Unknown		
	HM459851			GU076441	Unknown		
	EF210555			GU076441	Unknown		
	DQ631892			GU076441	Unknown		
	FJ609655			GU076441	Unknown		
	FJ012358			GU076441	Unknown		
	EF523478			GU076441	Unknown		
	GU178814			GU076441	Unknown		
	GU178813			GU076441	Unknown		
	GU178818			GU076441	Unknown		
	GU178816			GU076441	Unknown		
	GU178815			GU076441	Unknown		
	EF539831			GU076441	Unknown		
	GU322423			GU076441	Unknown		
	JN990924			GU076441	Unknown		
	JQ038235			GU076441	Unknown		
	JF414237			GU076441	Unknown		
	JQ004048			GU076441	Unknown		
	GU434143			GU076441	Unknown		
	GU434142			GU076441	Unknown		
	GU111505			GU076441	Unknown		
	JQ004051			GU076441	Unknown		
	JQ038234			GU076441	Unknown		
	JN990922			GU076441	Unknown		
	JF833036			GU076441	Unknown		
	JQ004045			GU076441	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
9	GU348995	16 (19)	(?)969 [(?)983]	GU076441	Unknown	MC**S**3	6.22×10^{-12}
	JQ038237			GU076441	Unknown		
	GU951436			GU076441	Unknown		
	GU983859			GU076441	Unknown		
	JQ004049			GU076441	Unknown		
	JQ004052			GU076441	Unknown		
	EF051116			GU076441	Unknown		
	EF107520			GU076441	Unknown		
	FR851298			GU076441	Unknown		
	FR851297			GU076441	Unknown		
	AY134494			GU076441	Unknown		
	GU355941			GU076441	Unknown		
	JN680353			GU076441	Unknown		
	JQ303121			GU076441	Unknown		
	AY530931			GU076441	Unknown		
	AF024715			GU076441	Unknown		
	AJ223505			GU076441	Unknown		
	EF210554			GU076441	Unknown		
	EF110890			GU076441	Unknown		
	AY594174			GU076441	Unknown		
	EF433426			GU076441	Unknown		
	AB116629			GU076441	Unknown		
	AB110217			GU076441	Unknown		
	AB116631			GU076441	Unknown		
	AB116630			GU076441	Unknown		
	AB636411			GU076441	Unknown		
	HM856919			GU076441	Unknown		
	JQ013091			GU076441	Unknown		
	JQ013089			GU076441	Unknown		
	JN680149			GU076441	Unknown		
	JN183880			GU076441	Unknown		
	JN183879			GU076441	Unknown		
	GU325634			GU076441	Unknown		
	JN183874			GU076441	Unknown		
	JN183875			GU076441	Unknown		
	GQ141873			GU076441	Unknown		
	HM856917			GU076441	Unknown		
	HM856915			GU076441	Unknown		
	HM856909			GU076441	Unknown		
	JQ013090			GU076441	Unknown		
	HM856873			GU076441	Unknown		
	HM856913			GU076441	Unknown		
	JN183876			GU076441	Unknown		
	JN183873			GU076441	Unknown		
	HM856914			GU076441	Unknown		
	JN183872			GU076441	Unknown		
	HM130912			GU076441	Unknown		
	HM856912			GU076441	Unknown		
	JN680150			GU076441	Unknown		
	HM856911			GU076441	Unknown		

Event	Recombinant[1]	Recombination breakpoints		Parents		Methods[2]	*P*-value[3]
		Initial	Final	Major	Minor		
TYLCV							
9	EF054893	16 (19)	(?)969 [(?)983]	GU076441	Unknown	MC**<u>S</u>**3	6.22×10^{-12}
	EF060196			GU076441	Unknown		
	FJ439569			GU076441	Unknown		
	HE603246			GU076441	Unknown		
	HE603242			GU076441	Unknown		
	HE603244			GU076441	Unknown		
	HE603243			GU076441	Unknown		
	HE603241			GU076441	Unknown		
	AJ489258			GU076441	Unknown		
	21426900			GU076441	Unknown		
	HM448447			GU076441	Unknown		
	AM409201			GU076441	Unknown		
	AY044138			GU076441	Unknown		
	GQ861427			GU076441	Unknown		
	EF054894			GU076441	Unknown		
	EF185318			GU076441	Unknown		
	AJ519441			GU076441	Unknown		
	AF105975			GU076441	Unknown		
	AB439842			GU076441	Unknown		
	AB116633			GU076441	Unknown		
	AB116635			GU076441	Unknown		
	AB116634			GU076441	Unknown		
	AB014347			GU076441	Unknown		
	AJ865337			GU076441	Unknown		
	AF071228			GU076441	Unknown		
	AB116632			GU076441	Unknown		
	AB014346			GU076441	Unknown		
	AB110218			GU076441	Unknown		

[1] Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?) Indicates that the breakpoint could not be precisely pinpointed.
[2] R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, CHIMAERA; S, SisScan; 3, 3SEQ.
[3]The reported *P*-value is for the program in bold, underlined type and is the lowest *P*-value calculated for the event in question.