LARISSA GOULART ZANARDO

NEW INSIGHTS OF *Cowpea mild mottle virus* (CPMMV) INFECTION IN SOYBEAN: THE INVOLVEMENT OF VIRAL REPLICASE IN SYMPTOMS INDUCTION AND ITS INTERACTION WITH HOST FACTORS

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

VIÇOSA
MINAS GERAIS-BRASIL
2017

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

T
Z27n
2017

Zanardo, Larissa Goulart, 1987-
New insights of *Cowpea mild mottle virus* (CPMMV) infection in soybean : the involvement of viral replicase in symptoms induction and its interaction with host factors / Larissa Goulart Zanardo. – Viçosa, MG, 2017.
x, 219f. : il. (algumas color.) ; 29 cm.

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

1. Soja - Doenças e pragas. 2. *Carlavirus*. 3. Vírus - Genética. 4. Recombinação (Genética). I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22 ed.633.349

LARISSA GOULART ZANARDO

NEW INSIGHTS OF *Cowpea mild mottle virus* (CPMMV) INFECTION IN SOYBEAN: THE INVOLVEMENT OF VIRAL REPLICASE IN SYMPTOMS INDUCTION AND ITS INTERACTION WITH HOST FACTORS

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

Pedro Augusto Braga dos Reis

Sérgio Hermínio Brommonschenkel

Danielle Ribeiro de Barros

Claudine Márcia Carvalho
(Coorientadora)

Francisco Murilo Zerbini Júnior
(Orientador)

AGRADECIMENTOS

Primeiramente à Deus pelo dom da vida, pelas oportunidades e possibilidade de recomeço a cada dificuldade.

Aos meus pais por serem um porto seguro, por me ouvirem e me apoiarem a cada dia, e principalmente, por me fazerem acreditar que era possível!

Aos meus irmãos pelo apoio, incentivo, conselhos e brincadeiras durante todos os anos de estudo.

Ao Bruno pelo amor e cuidado, pelo incentivo e paciência, pelos conselhos e por sonhar esse sonho comigo.

À Universidade Federal de Viçosa (UFV) pela formação e oportunidade.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pela concessão das bolsas de estudos.

À Prof. Claudine Márcia de Carvalho por compartilhar seus conhecimentos e experiências, por acreditar e incentivar os trabalhos, pela dedicação, amizade e orientação valiosa.

Ao Prof. Francisco Murilo Zerbini, pela oportunidade e confiança, pelos ensinamentos, pela disponibilidade e suporte durante todo o curso.

Ao Prof. Eduardo S. G. Mizubuti pela paciência, atenção e disponibilidade, pelos ensinamentos e por compartilhar um pouco do seu conhecimento.

Aos amigos Talita, Fernanda, Diogo, César e Sarah, em especial, pelos incentivos constantes, pelos conselhos, conversas e brincadeiras diárias, pela ajuda nos trabalhos e por compartilharem suas experiências. Muito Obrigada!

Ao Dr. Murilo Siqueira Alves por todo apoio e amizade, pelas repetições experimentais e biológicas e por acreditar e incentivar todos os trabalhos.

Aos amigos do Laboratório de Virologia Vegetal Adriana, Alison, Ana Carolina, André, Angélica, Camila, Fábio, Glória, Hermano, Igor, João Paulo, Josiane, Márcio, Osvaldo, Patrícia, Renan, Roberta, Silvia e Tarsiane pela amizade, companheirismo, brincadeiras e por tornarem o ambiente de trabalho mais agradável.

A todos que contribuíram e torceram por mim durante toda essa etapa, minha sincera gratidão.

SUMÁRIO

ABSTRACT

ZANARDO, Larissa Goulart, D.Sc., Universidade Federal de Viçosa, February, 2017. **New insights of *Cowpea mild mottle virus* (CPMMV) infection in soybean: The involvement of viral replicase in symptoms induction and its interaction with host factors.** Advisor: Francisco Murilo Zerbini Júnior. Co-advisers: Claudine Márcia Carvalho and Murilo Siqueira Alves.

Cowpea mild mottle virus (CPMMV, family *Betaflexiviridae* and genus *Carlavirus*) is an emerging virus in Brazil. The virus is transmitted by the whitefly *Bemisia tabaci* and it infects hosts preferentially of *Fabaceae* family, being found in different crops in almost all the continents in the world. In this work, several aspects about biology and genetics of CPMMV were dicussed and the CPMMV-host relationship was explored. CPMMV causes varied symptoms in plants in the fields and in greenhouse. The nature of this symptoms variation was determinate in this work using successive inoculations of a CPMMV Brazilian isolate causing necrosis in the CD206 soybean cultivar. Successive inoculations led to the isolate, which previously had caused necrosis to induce mild symptoms (mosaic and vein clearing). This was verified after six successive inoculations of a soybean sample from the field and also from a local lesion of a viral isolate, and thus submitted to a genetic bottleneck, in *Nicotiana benthamiana* leaves. We have shown that altering of symptoms pattern increased the fitness of the virus and vector, suggesting that the induction of different symptoms by viral variants is an adaptive advantage. The viral region involved in the induction of CPMMV symptoms was viral replicase (encoded by ORF1), different sites distributed throughout the recombinant blocks of ORF1 were associated with phenotype changes. In this work, also shought to the understanding about the replication of CPMMV in soybean identifying factors necessary for replication. To date, all that has been known about betaflexiviruses's replication is that it occurs in the cytoplasm. None host factor was known. Here, two host proteins able to interact with the RNA-dependent RNA polymerase domain (RdRp) from a yeast two-hybrid vector-constructed soybean cDNA library were indentifyed. For the library construction, soybean plants cv. CD206 with CPMMV infection after 0, 3, 7 and 14 days of inoculation were used. The proteins identified were: the CSC1 type ERD4 (GmERD4) and an inositol methyltransferase (GmIMT). The interaction was confirmed by yeast two-hybrid assays and by Bimolecular Fluorescence Complementation (BiFC) assays. Prediction and subcellular localization of both proteins and the RdRp domain were also performed. GmERD4 was located in the endoplasmic

reticulum (ER), GmIMT presented cytoplasmic location while the RdRp domain induced punctual structures in the plasma membrane and on the ER network. Analysis of the expression of both genes in soybean plants infected with CPMMV demonstrated that the they were induced after 3 and 7 days of inoculation. Overexpression of *GmERD4* and *GmIMT* was performed on soybean protoplasts infected with CPMMV and it was found that overexpression of *GmERD4* increased viral accumulation after 7 days of infection. Viral replicase is involved in different viral infection processes, acting not only on genome replication, but also on the induction of symptoms. This reinforces the multifunctional nature of viral proteins.

RESUMO

ZANARDO, Larissa Goulart, D.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Novas descobertas sobre a infecção do *Cowpea mild mottle virus* (CPMMV) em soja: O envolvimento da replicase viral na indução de sintomas e sua interação com fatores do hospedeiro.** Orientador: Francisco Murilo Zerbini Júnior. Coorientadores: Claudine Márcia de Carvalho e Murilo Siqueira Alves.

O *Cowpea mild mottle virus* (CPMMV, família *Betaflexiviridade* e gênero *Carlavirus*) é um vírus emergente no Brasil. O vírus é transmitido pela mosca-branca *Bemisia tabaci* e infecta hospedeiros preferencialmente da família *Fabaceae*, sendo encontrado em diferentes culturas em quase todos os continentes do mundo. Nesse trabalho, vários aspectos da biologia e genética do CPMMV foram discutidos e a relação CPMMV-hospedeiro foi explorada. O CPMMV causa sintomas muito variados em plantas campo e em casa de vegetação. A natureza dessa variação de sintomas foi determinada nesse trabalho utilizando-se inoculações sucessivas de um isolado de CPMMV brasileiro causador de necrose na cultivar de soja CD206. As inoculações sucessivas levaram o isolado que antes causava necrose a induzir sintomas brandos (mosaicos e clareamento de nervuras). Isso foi verificado após seis inoculações sucessivas de uma amostra vegetal de soja proveniente do campo e também a partir do isolado viral oriundo de uma lesão local, e submetido, portanto a um gargalo genético, em folhas de *Nicotiana benthamiana*. A alteração do padrão de sintomas aumentou adaptabilidade do vírus e vetor, sugerindo que a indução de diferentes sintomas pelos variantes virais é uma vantagem adaptativa. A região viral envolvida na indução de sintomas pelo CPMMV foi a replicase viral (codificada pela ORF1) e diferentes sítios distribuídos ao longo de blocos recombinantes da ORF1 foram associados às alterações do fenótipo. Nesse trabalho, também se buscou compreender um aspectos da replicação do CPMMV em soja, identificando-se fatores necessários à replicação. Até o momento tudo que se sabia sobre a replicação dos betaflexivírus é que ela ocorria no citoplasma. Nenhum fator do hospedeiro era conhecido. Aqui foram identificadas duas proteínas do hospedeiro capazes de interagir com o domínio RNA polimerase dependente de RNA (RdRp) a partir de uma biblioteca de cDNA de soja construída em vetor de duplo-híbrido de leveduras. Para a construção da biblioteca foram utilizadas plantas de soja cv. CD206 com infecção pelo CPMMV após 0, 3, 7 e 14 dias de inoculação. As proteínas identificadas foram: a ERD4 tipo CSC1 (GmERD4) e uma inositol metiltransferase (GmIMT). A interação foi confirmada por ensaios de duplo-

híbrido de leveduras e por ensaios de complementação de fluorescência bimolecular (BiFC). A predição e a localização subcelular de ambas as proteínas e do domínio RdRp também foram realizados. GmERD4 localizou-se no retículo endoplasmático (RE) e GmIMT apresentou localização citoplasmática, enquanto o domínio RdRp induziu estruturas pontuais na membrana plasmática e sobre a rede do RE. A análise da expressão de ambos os genes em plantas de soja infectadas com o CPMMV, demonstrou que eles foram induzidos após 3 e 7 dias da inoculação. A superexpressão de *GmERD4* e *GmIMT* foi realizada em protoplastos de soja infectados com o CPMMV e foi verificado que a superexpressão de *GmERD4* aumentou o acúmulo viral após 7 dias de infecção. A replicase viral está envolvida em diferentes processos da infecção viral, atuando não apenas na replicação dos genomas, mas também na indução de sintomas. Isso reforça a natureza multifuncional das proteínas virais.

INTRODUÇÃO

As infecções causadas por vírus, de modo geral, causam grandes perdas à produção agrícola e são de difícil controle. A necrose da haste da soja causada pelo *Cowpea mild mottle virus* (CPMMV, família *Betaflexiviridade*, gênero *Carlavirus*) se encaixa bem nesse cenário. O CPMMV foi identificado há cerca de 40 anos e mesmo assim pouco se sabe sobre aspectos da biologia e genética desse vírus. O vírus passou a ser um problema na cultura da soja brasileira a partir dos anos 2000, causando sintomas tão variados em campo, que a necrose da haste chegou a ser considerada como causada por diferentes patógenos. Entre os sintomas mais severos estão a necrose da haste e folhas, queima do broto e nanismo. Sintomas como mosaico, clareamento de nervuras e bolhosidades são também comumente observados.

Grandes perdas econômicas na produção de soja foram associadas à necrose da haste em diferentes estados brasileiros produtores de soja, especialmente Goiás e Mato Grosso, sendo, portanto a virose uma grande ameaça à produção de soja. Além da severidade dos sintomas, o vírus é um grande problema por ser transmitido pela mosca branca *Bemisia tabaci* de maneira não persistente. Isso significa que a transmissão pode ocorrer durante a picada de prova, o que aumentam as chances de disseminação do vírus entre diferentes hospedeiros, por um vetor de difícil controle. Um estudo inicial com diferentes isolados de CPMMV coletados em campos de soja em vários estados brasileiros foi realizado e evidenciou-se a existência de duas estirpes de CPMMV (BR1 e BR2). Variações no genoma foram observadas e análises filogenéticas mostraram a ocorrência de agrupamentos de acordo com o padrão de sintomas induzidos.

A natureza da variação dos sintomas até então era desconhecida. Determinar a causa dessa variação foi uns dos objetivos desse trabalho. A indução de diferentes sintomas por vírus já foi demonstrada estar associada a poucas alterações na sequencia nucleotídeos, geralmente uma ou duas mutações. Dessa forma, os diferentes sintomas causados pelo CPMMV poderiam estar também associados a diferentes sítios dos variantes virais (mais adaptados ao hospedeiro), os quais se mantiveram no campo e foram disseminados.

Para que a infecção viral tenha sucesso o vírus precisa superar as respostas de defesa, recrutar a maquinaria do hospedeiro para a realização da tradução, replicação e movimento viral. Como isso é feito e os fatores do hospedeiro necessários para o sucesso da infecção viral ainda são desconhecidos para os betaflexivírus. A interação dos carlavírus com seus hospedeiros é desconhecida, tudo que sabe para sobre a replicação é que ela ocorre no citoplasma. Nenhum fator do hospedeiro necessário à replicação é conhecido e nem o local específico de ocorrência da replicação no citoplasma.

Assim, esse trabalho buscou responder parte dessas perguntas, esclarecendo aspectos da interação vírus-planta (CPMMV-Soja). Visto que o entendimento do processo de infecção viral pelo CPMMV poderá contribuir para o desenvolvimento de medidas do controle da doença mais eficientes

CHAPTER 1

Cowpea mild mottle virus – an emerging carlavirus

Zanardo, L. G., Carvalho, C. M. *Cowpea mild mottle virus* – an emerging carlavirus. Tropical Plant Pathology. Submitted.

Title: *Cowpea mild mottle virus* – an emerging carlavirus

Zanardo, L. G., Carvalho, C. M.

Departamento de Fitopatologia /BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, Brazil, 36570-000

* Corresponding author: claudine.carvalho@ufv.br

Phone: (+55-31) 3899-1087; Fax: (+55-31) 3899-2240;

Keywords: CPMMV, *Carlavirus*, emerging pathogens, *Bemisia tabaci*

Abstract

In this review, we describe and discuss the biology and aspects of infection of an emerging pathogen, the RNA virus *Cowpea mild mottle virus* (CPMMV). The review takes a global perspective but has an emphasis on Brazilian soybean in which CPMMV causes stem necrosis. Emergence or re-emergence of pathogens are normally associated with changes in ecology, transmission routes, host range, climate and pathogen variation. Since its first description in Ghana in 1973, CPMMV has spread across the world and, although it principally infects Fabaceous plants, is also able to infect hosts from Solanaceae and Lamiaceae. Besides the problem in soybean, CPMMV is re-emerging in genetically modified common bean. To limit the impact of CPMMV and combat future outbreaks, it is necessary to understand the ecological and evolutionary factors responsible for its emergence. We identify surveillance as a key defense against CPMMV, as CPMMV is transmitted non-persistently by the whitefly *Bemisia tabaci*, there is the possibility of seed transmission and the virus is able to cause asymptomatic infections. We discuss the potential for development of resistant crop lines and identify key areas for future research.

INTRODUCTION

The increasing incidence of infectious diseases in plants, due to emergent pathogens, has hindered agricultural production in recent years (Henning *et al.*, 2014; Woolhouse, 2002). Emerging infectious diseases are caused by pathogens that have increased their geographical distribution and host range, have changed their virulence, have newly evolved or have recently been discovered (Anderson *et al.*, 2004; Stukenbrock & Mcdonald, 2008; Imam *et al.*, 2016; Woolhouse, 2002). Viruses are the major taxonomic group of emergent pathogens of humans, wildlife and plants, causing almost half of all reported emergences in all three sets of hosts (Anderson *et al.*, 2004; Dobson & Foufopoulos, 2001; Woolhouse, 2002). Viruses are a major problem in many crops, vegetable and ornamental plants, decreasing product yields and qualities. Every year throughout the world, they cause economic losses estimated at up to US$ 60 billion, with losses to food crops alone up to $20 billion cited by Zhao *et al.*, 2017.

A number of studies have been carried out with different viruses to understand how they emerge (Bedhomme *et al.*, 2014; Dobson & Foufopoulos, 2001; Woolhouse, 2002; Yates *et al.*, 2006). It has been shown that the emerging and reemerging viruses can come from within the host population itself; from the external environment and from populations of other host species (Woolhouse, 2002). The early steps of virus emergence can be influenced by evolutionary and ecological factors such as host diversity, climate, vectors, pathogen controls methods and most importantly, genetic variation in viral populations (Anderson *et al.*, 2004; Bedhomme *et al.*, 2014).

RNA viruses in particular evolve quickly and display high variation, due to their quick replication and short generation times, large effecttives population sizes and lack of proofreading (Lalic *et al.*, 2011). Here, we discuss *Cowpea mild mottle virus* (CPMMV), a single-stranded positive-sense RNA virus that is able to infect peanuts (*Arachis hypogaea*) soybean (*Glycine max*), beans (*Phaseolus vulgaris*) and other crops across the world (Menzel *et al.*, 2010; Zanardo *et al.*, 2014a). Of these crops, soybean is the most important on a global scale, being the fourth most produced and consumed grain in the world and the principal oilseed crop (Hirakuri & Lazzarotto, 2014). Brazil is the world's second largest producer of soybeans (Embrapa, 2014) and it is estimated that about 103.5 million tons will be produced in the 2016/2017 season (Reuters, 2017).

CPMMV is the etiological agent of soybean stem necrosis disease. This virus is considered as an emergent pathogen in Brazil for several reasons: (1) a host change occurred; it was first described in Brazil in 1983 infecting common bean cv. Jalo (*Phaseolus vulgaris*) (Costa *et al.*, 1983), but the disease only became a major economic problem to Brazilian soybean production in the 2000's (Almeida *et al.*, 2003; Zanardo *et al.*, 2014a); (2) the virus has been found in soybean across Brazil (in states from all five regions, North, Northeast, Midwest, Southeast and South) from 2000 to 2010, implying that it had quickly spread in the field (Almeida, 2008; Almeida *et al.*, 2003; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b); (3) CPMMV is highly virulent, with symptoms varying from mild (mosaic, crinkled and blistered leaves and vein clearing) to severe (necrosis, dwarfism and bud blight) in soybean plants (Almeida *et al.*, 2003; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b). As an emergent pathogen, CPMMV deserves special attention in order to prevent

new outbreaks in different crops and countries. This is therefore the focus of the present review.

No single cost-effective or efficient control measure can be recommended to prevent CPMMV infection in all cases and in the different regions across the globe. The non-adoption of effective control practices to restrict the pathogen has allowed the virus to spread since its initial description in cowpea (*Vigna unguiculata*) in Ghana (Brunt & Kenten, 1973). The need for a CPMMV-resistant soybean cultivar for commercial release is urgent but its development will depend upon improved understanding of the virus and its interaction with its hosts. In this paper, therefore, we provide a comprehensive review of current knowledge of CPMMV. We discuss key aspects of its biology, describe its spread across the world with a particular focus on Brazil, document its recent emergence in common bean, describe the phylogenetic relationships among CPMMV isolates, highlight discuss possible control measures and key research areas for the future.

BIOLOGY OF THE PATHOGEN *Cowpea mild mottle virus* (CPMMV)

Genomic organization

Cowpea mild mottle virus (CPMMV) is a single-stranded positive-sense RNA virus of the family *Betaflexiviridae* (see below for further taxonomic information) with a genome of 8,200 nucleotides with a cap structure [m7GpppG] linked to the 5´ terminus and a poly-A tail at the 3′end (Menzel *et al.*, 2010; Adams *et al.*, 2012; Zanardo *et al.*, 2014a). CPMMV belongs to the genus *Carlavirus* and has a genomic organization typical of the genus, with six open reading frames (ORFs) (Figure 1A): ORF1 encodes a putative replicase protein (predicted at 211.6 kDa), with four conserved motifs: methyltransferase, C23 peptidase,

RNA helicase and RNA-dependent RNA polymerase (RdRp) (Zanardo *et al.*, 2014a). ORFs 2, 3 and 4 encode the triple gene block proteins (TGB1-3p, with 25, 11.6 and 7.5 kDa), that are essential for virus movement. ORF 5 encodes the coat protein (CP, 32 kDa). Finally, ORF 6 encodes a cysteine-rich protein (CRP; 15,2 kDa) with nucleic acid-binding activity (NABP). CPMMV is encapsidated by CP subunits in flexuous filamentous particles (10-15 x 610-700 nm) and can be found in the cytoplasm of palisade, mesophyll, parenchyma and epidermal cells in soybean and *Nicotiana clevelandii* (Almeida *et al.*, 2003; Brunt *et al.*, 1983). These particles form aggregates in the form of sheets or bundles (Figure 1B) and often brush-like inclusions (Figure 1 C) that are typical of carlavirus infections (Almeida *et al.*, 2003; Almeida *et al.*, 2005; Brunt *et al.*, 1983).

Viral proteins are often multifunctional, potentially acting at more than one moment of viral infection. Each function is essential for viral survival and can be dependent on viral species, host and vector. Carlaviruses encode just six proteins, but two of these have more than one known function. TGBp1, for example, has been shown to suppress systemic silencing in the case of the carlavirus *Potato virus M* (PVM) (Senshu *et al.*, 2011). This is in addition to its role in viral movement and it is possible that CPMMV TGBp1 acts similarly. Many functions have been attributed to the CRP protein encoded by ORF6. CRPs of PVM and *Potato virus H* (PVH) act as systemic and local suppressors of RNA silencing (Li *et al.*, 2013; Senshu *et al.*, 2011), although the CRP of *Chrysanthemum virus* B (CVB) does not (Lukhovitskaya *et al.*, 2009; Lukhovitskaya *et al.*, 2013). The latter exhibited nuclear localization and acted as a transcription factor regulating cell proliferation and mesophyll tissue growth in tobacco plants (Lukhovitskaya *et al.*, 2013). The CRP of CPMMV contains a motif for a putative C-4-type zinc finger (ZF) and an adjacent putative nuclear

localization signal (NLS) (Menzel *et al.*, 2010; Zanardo *et al.*, 2014a), suggesting that it might act in a similar fashion to CVB CRP, without excluding the possibility that it acts also as a suppressor of RNA silencing. Besides these functions, CRP has also been suggested to be a symptom determinant for several carlaviruses. In an experiment in which *Potato virus X* (PVX) expressed the CRPs of six different carlaviruses [PVM, *Daphne virus S* (DVS), CVB, *Lily symptomless virus* (LSV), *Helleborus net necrosis virus* (HeNNV), and *Narcissus common latent virus* (NLV)], different symptoms were induced *Nicotiana occidentalles* with increased accumulation of PVX. The increase of viral accumulation (suggesting silencing suppression activity) and expression of symptoms have not been correlated, but it was shown that symptom induction is dependent on the ZF and NLS domains (Fujita *et al.*, 2016).

Much remains unknown about the functions of CPMMV proteins. PVM, CVB and a few other viruses have received some attention, but studies with CPMMV are at a very early stage. This is equally true for genome expression, which we discuss next.

Viral genome expression

Little is known of carlavirus genome expression, but in this section we use the information that is available to propose a model for CPMMV. For carlaviruses, the only ORF translated from the full length genomic RNA is ORF1, that encodes the replicase protein (Adams *et al.*, 2012; Adams & Kreuze, 2016) (Figure 1A). For *Blueberry scorch virus* (BBScV) (Lawrence *et al.*, 1995), and probably other carlaviruses, the product of ORF1 (replicase) is proteolytically processed between the helicase and RNA-dependent RNA polymerase domains with a fragment of about 30-40 kDa being removed from the replicase (Adams *et*

al., 2012). Processing of the carlavirus replicase has also been suggested for other carlaviruses: *Helenium virus S* (HelSV) (Foster & Mills, 1990a), *Carnation latent virus* (CLV) (Meehan & Mills, 1991), *Potato virus S* (PVS) (Foster & Mills, 1992) and PVM (Tavantzis, 1991). In the case of CPMMV, the replicase autoproteolysis is mediated by the C23 peptidase domain (Figure 1A) (Zanardo *et al.*, 2014a). The CPMMV replicase would be translated directly from the viral ssRNA+, during or after the decapsidation of the viral particle, giving origin to the polyprotein, which would be proteolytically processed by the C23 peptidase domain. Two fragments are formed: the first includes the methyltransferase, helicase and C23 peptidase domains, while the second includes only the RdRp domain (Figure 1A). Processing of the polyprotein probably favors the individual activity of the viral replicase domains: while the methyltransferase and helicase domains act in RNA capping and in the opening of the secondary structures of viral RNA, the RdRp domain can polymerize the viral genomes independently. The polymerization activity of RdRp can also depend on more than one subunit, so the replicase cleavage may allow the dimerization or tetramerization of the RdRp domains. Moreover, the proteolytic processing can facilitate interaction with host proteins and other viral proteins required for replication.

The expression of carlavirus 3' ORFs is via subgenomic (sg) RNAs (Figure 1A). This was first suggested by Mackenzie *et al.*, (1989) and Rupasov *et al.* (1989) in studies with PVS and PVM, respectively, and was confirmed by Foster & Mills (1990b), who verified that the expression of PVM CP was by sgRNA of 1.3 kb. The presence of sgRNAs was also verified in plants systemically infected by CLV and in purified particles of HelSV using northern hybridization. For CLV and HelSV, two dsRNA species equivalent to 2.1 and 1.6 kb (Meehan & Mills, 1991) and 3.3 and 1.5 kb (Foster & Mills, 1990a), respectively, were

visualized in northern hybridization. The larger sgRNA (sgRNA1) would be responsible for the translation of the TGBs while the smaller sgRNA (sgRNA2) would be responsible for translating CP and CRP (Figure 1A) (Almeida *et al.*, 2003; Foster & Mills, 1990a; Foster & Mills, 1990b; Adams *et al.*, 2012).

Regarding the regulation of carlavirus expression, the only available information is that the 5'-untranslated leader sequences of the genomic RNA and the sgRNA2 of PVS have been shown to act as efficient enhancers of replicase and CP translation, respectivelly (Turner & Foster, 1997; Turner *et al.*, 1999). This probably applies also to regulation of CPMMV expression, but at present this has not been investigated.

As shown here, the process of gene expression and regulation of carlaviruses remains obscure. Few studies have been performed and none of these involve CPMMV. Given the current importance of this emerging virus, much still needs to be done to clarify the entire process of CPMMV infection. Which viral and hosts proteins are involved in the different processes, how they act to favor viral infection and to avoid host defense mechanisms are essential questions that are without answers at the moment.

Taxonomy

The family *Betaflexiviridae*, to which carlaviruses such as CPMMV belong, is a member of the order *Tymovirales*. Until recently *Betaflexiviridae* comprised seven genera (*Carlavirus, Capillovirus, Citrivirus, Foveavirus, Trichovirus, Vitivirus* and *Tepovirus*) distinguished mainly by the phylogeny of the replicase and the number and arrangement of other genes (coding movement proteins, capsidial protein and additional proteins) (Adams *et al.*, 2012).

The genus *Carlavirus* has the highest number of species described in the *Betaflexiviridae*, at 47 compared with at most seven in any other genera (Adams & Kreuze, 2016). The large number of host species infected by members of the *Carlavirus* genus includes important crops such as bean, blueberry, chrysanthemum, garlic, grape, melon, soybean and many others (Lawrence *et al.*, 1995; Meng & Li, 2010; Nagata *et al.*, 2005; Pramesh & Baranwal, 2013; Rodríguez-Pardina *et al.*, 2004; Singh *et al.*, 2012; Zanardo *et al.*, 2014a).

New phylogenetic analyses of the replicase gene have revealed two major branches within the *Betaflexiviridae* that correlate with the type of movement protein ('30K-like' or TGB – Triple Gene Block). This led to a proposal to subdivide *Betaflexiviridae* into two subfamilies, *Quinvirinae* and *Trivirinae*, along with the creation of four new genera: *Chordovirus*, *Divavirus*, *Robigovirus*, and *Prunevirus* (Adams & Kreuze, 2016). The proposed subfamily *Quinvirinae* includes *Carlavirus*, *Foveavirus* and *Robivirus* (all of which have TGB-type as the movement proteins), while the *Trivirinae* subfamily would include *Citrivirus, Capillovirus*, *Chordovirus*, *Divavirus*, *Prunevirus, Tepovirus, Trichovirus* and *Vitivirus* (all of which have 30K-like movement proteins) (Adams & Kreuze, 2016).

The CP and replicase genes have been used to classify the species in the *Betaflexiviridae*. Viruses belongs to the same species in *Betaflexiviridae* are expected to have more than 72% nucleotide identity (80% amino acid identity of encoded proteins) for CP or replicase (Adams *et al.*, 2012; Adams & Kreuze, 2016). Although this criterion has been widely used, there is growing evidence that recombination events can be very frequent in the ORF1 of carlaviruses (replicase) (Singh *et al.*, 2008; Singh *et al.*, 2012; Zanardo *et al.*,

2014a). As this genomic region is used in taxonomic classification, there is a risk of errors in classification occurring in cases where only partial sequences have been determined. For example, when six Brazilian CPMMV isolates were sequenced and compared with a Ghanaian isolate (Zanardo *et al.*, 2014a), the identity values for replicase (ORF1) among the isolates were 60-61% for nt and 58-61% for aa, all values below the taxonomic classification criterion. However, the values for CP were 79% for nt and 95-96% for aa, showing clearly that the Brazilian and Ghanaian isolates belonged to the species CPMMV. If only the viral replicase had been sequenced, the impression would be that the six Brazilian isolates would belong to a new species of *Carlavirus*. The low similarity observed between Brazilian and Ghanaian isolates is attributed to recombination events. So, in our opinion only one of these two genes (replicase or CP) should be considered and CP seems to be the most appropriate region for *Carlavirus* species classification. Alternatively, perhaps the time has come to use whole genome sequences for species demarcation in the *Carlavirus* genus (Zanardo *et al.*, 2014a). Even so, greater attention needs to be paid to recombination events, since much of the variation generated can be attributed to these, not only in the viral replicase, but in different regions of the viral genome.

Besides the problems with the classification criteria for carlavirus species, a further factor of importance is transmission by the vector. For other plant viruses, the nature of vector transmission is sometimes used as one of the criteria used to classify species within a genus; in the case of carlaviruses, there is variation in vector transmission within the genus (Adams *et al.*, 2012).

Transmission mode

In general, carlaviruses are both sap-transmitted and transmitted by aphids (Adams *et al.*, 2012). Several aphid species (*Aphis craccivora, A. glycines, A. gossypii, A.spiraecola*, *Myzus persicae*) were therefore tested as possible vectors of CPMMV yet none of these were able to transmit the virus (El-Hassan *et al.*, 1997; Iwaki *et al.*, 1982; Mansour *et al.*, 1998; Thouvenel *et al.*, 1982). This led to the discovery that CPMMV is an exception among carlaviruses, being transmitted by the whitefly *Bemisia tabaci* (Hemptera: Aleyrodidae) (Iwaki *et al.*, 1982; Jeyanandarajah & Brunt, 1993; Marubayashi *et al.*, 2010; Munyappa & Reddy, 1983; Naidu *et al.*, 1998). This insect is a vector of more than 200 plant viruses recognized as species in the genera *Begomovirus* (*Geminiviridae*), *Crinivirus* (*Closteroviridae*), *Carlavirus* or *Ipomovirus* (*Potyviridae*)(Polston *et al.*, 2013).

The mode of CPMMV transmission by *B. tabaci* is considered to be nonpersistent. The first to suggest the nonpersistent mode was Thouvenel *et al.* (1982). After his study, subsequents studies were directed at determining the transmission mode. In these experiments, that were tested ranged from 10 min to 48 hours to acquisition access periods (AAP) and inoculation access periods (IAP) tested ranged from 5 minutes to 24 hours. Neither ranges led to variations in transmission (El-Hassan *et al.*, 1997; Iizuka *et al.*, 1984; Iwaki *et al.*, 1982; Mansour *et al.*, 1998; Marubayashi *et al.*, 2010; Prado, 2014). Thus, the virus could be transmitted with only 10 min of AAP and 5 min IAP, consistent with nonpersistent transmission (Iwaki *et al.*, 1982; Marubayashi *et al.*, 2010). The experiments also showed that there is no latent or retention period (Mansour *et al.*, 1998).

Regarding the vector *B. tabaci*, the current consensus is that it is a species complex and that there is no morphological character usable to correctly classify them, so the best approach is the sequencing of the mitochondrial cytochrome oxidase 1 (mtCOI) gene (Brown, 2000; De Barro *et al.*, 2011). Dinsdale *et al.* (2010) proposed a threshold divergence of 3.5% in the sequences of the mtCOI gene to separate the 24 morphologically indistinguishable *B. tabaci* species. Based on this report, *B. tabaci* biotype B (as it was previously known) is now part of the species/group Middle East-Asia Minor 1 species (MEAM1). Almeida *et al.* (2005) and Marubayashi *et al.* (2010) showed that the "biotype B" (now the species MEAM 1) was able to transmit CPMMV in beans and soybeans in Brazil. The same was observed by Prado (2014) in soybean plants. The main species of whitefly involved in virus transmission in Brazil is MEAM 1, but two other species have been reported in Brazil: New World (NW, the old A biotype) (Lima *et al.*, 2002) and Mediterranean (MED, the old Q biotype) (Da Fonseca Barbosa *et al.*, 2015). The potential of these two species of whitefly to transmit CPMMV has not yet been evaluated. This information is important as the whitefly species present in the field can define which viral isolates will be transmitted and which hosts are infected.

Quite how or why whiteflies can transmit CPMMV is unknown; among the many carlaviruses described, it is one of the few transmitted by this vector, which includes *Melon yellowing-associated virus* (MYaV) (Nagata *et al*., 2005) and *Cucumber vein-clearing virus* (Menzel *et al.*, 2011). We suspect that the CPMMV capsid protein was obtained by recombination from a different (whitefly-transmitted) virus in mixed infections. Such a recombination event could have allowed CPMMV to be transmitted by whiteflies, to infect

new hosts, normally not infected by carlavirus and to be spread in hosts preferentially colonized by whiteflies.

The ability of CPMMV to be transmitted by seeds was tested by a number of authors but this issue is still unclear. The isolate collected by Brunt & Kenten (1973) in Ghana was transmitted via soybean seeds, cowpea seeds and at a low frequency in common bean seeds. Seed transmission was also detected in a Venezuelan isolate in yardlong bean seeds. However, Brazilian isolates did not show any transmission in seeds (Almeida *et al.*, 2003; Almeida *et al.*, 2005). Seed transmission for CPMMV may be dependent on the viral isolate (or strain) and the host cultivar, which would make it more difficult to control the disease, especially with asymptomatic infections.

As CPMMV can has the potential to spread through seed and by the viruliferous whitefly, the introduction of the virus in new regions is possible, as has been observed in the last years, especially in Brazil. This can occur especially through the accidental importation of infected seed from already infected locales, by migration of whiteflies or by long-distance transport of whiteflies by vehicles along transport routes. CPMMV is therefore a virus that represents a great risk to legume crop species and measures to control the propagation of the virus by vector and by seeds are desirable.

THE EMERGENCE OF CPMMV AS A THREAT TO CROPS

Host range and symptomatology

The natural hosts of CPMMV usually include only species of Fabaceae (Laguna *et al.*, 2006), but other hosts can be infected experimentally. Symptoms are highly variable and depend on the host, season and the viral strain (Menzel *et al.*, 2010; Zanardo *et al.*, 2014a). The diversity of symptoms caused by CPMMV in different hosts is so great that different diseases have been described and attributed to different viruses. Examples of this are groundnut crinkle virus (Durbern & Dollet, 1981), psophocarpus necrotic mosaic (Fortuner *et al.*, 1979) and voandzeia mosaic virus (Fauquet & Thouvenel, 1987) in the Ivory Coast, tomato pale chlorosis disease in Israel (Cohen & Antignus, 1982) and tomato fuzzy vein disease in Nigeria (Brunt & Phillips, 1981). It is now known that all of these viruses were caused by CPMMV. The CPMMV has already been described infecting two solanaceous hosts: tomato (*Solanum lycopersicum*) causing chlorosis and fuzzy vein (Brunt & Phillips, 1981; Cohen & Antignus, 1982) and eggplants (*Solanum melongena*) causing mosaics in Jordan (Mansour et al., 1998).

CPMMV was originally described in cowpea (*V. unguiculata*) causing systemic mottling, chlorotic blotches and leaf malformations, and was therefore denominated as the mild mottling virus of cowpea (Brunt & Kenten, 1973; Thouvenel *et al.*, 1982). In peanut or groundnut (*A. hypogaea*) it was able to cause necrotic lesions, chlorotic rings or line patterns followed by systemic leaf chlorosis, rolling and vein necrosis (Brunt & Kenten, 1973; Mukoye *et al.*, 2015). In pea (*Pisum sativum*) only mosaic and chlorosis were observed (Tavasoli *et al.*, 2008; Thouvenel *et al.*, 1982). Besides those hosts, CPMMV can infect different species of beans, causing the disease denominated bean angular mosaic in common bean (*P. vulgaris*) (Costa *et al.*, 1983) (Figure 2). The occurrence of angular

mosaic was first described in Brazil, due to the pattern of symptoms observed in common bean cv. Jalo. Depending on common bean cultivar, infection may be asymptomatic (Brito *et al.*, 2012; Mink & Keswani, 1987; Mukoye *et al.*, 2015; Tavasoli *et al.*, 2008; Zanardo *et al.*, 2014a). In mung bean (*Vigna mung*) and in green grams (*Vigna radiata*) mild symptoms such as vein clearing and mottle were described (Mink & Keswani, 1987; Mukoye *et al.*, 2015; Tavasoli *et al.*, 2008; Thouvenel *et al.*, 1982), while in jack bean (*Canavalia ensiformis*) more severe symptoms such as vein mosaic, chlorosis, apical necrosis and malformation can occur. In *Macrotyloma uniflorum*, one of the lesser known beans, CPMMV may induce necrosis (Naidu *et al.*, 1998).

CPMMV has received particular attention in Brazil. Here, the most affected crops are soybeans and common bean. Symptoms in soybean are highly variable (Figure 2) when compared to other hosts. Mild symptoms that have been observed includes: mosaic, chlorosis, vein clearing and mottling (Almeida *et al.*, 2003; Brito *et al.*, 2012; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b), while severe symptoms includes: bud blight, dwarfing, leaf and stem necrosis, with these diverse symptoms sometimes occurring in the same soybean cultivar (Almeida *et al.*, 2005; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b) (Figure 2). The disease associated with CPMMV in Brazilian soybean was named stem necrosis of soybean (Almeida *et al.*, 2003).

The occurrence of symptomless CPMMV infections in soybean has also been described experimentally in some cultivars (Almeida *et al.*, 2003; Zanardo *et al.*, 2014a). The occurrence of asymptomatic infections, allied to nonpersistent whitefly transmission and the possibility of transmission by seeds probably contributed to the spread of the virus in

Brazil and in the world. Initially silent CPMMV infections in soybean would not have been controlled, while crop rotation with beans in the off-season periods would also have favored the maintenance of the virus in the field and facilitated its spread.

Recently, CPMMV was described in two new host, in natural infections: chia (*Salvia hispanica*, Lamiaceae) (Celli *et al.*, 2016) and dixie ticktrefoil (*Desmodium tortuosum*, Fabaceae) (Henning *et al.*, 2014). It was able to induce yellowing, mottling and blistering in leaves, and leaf and stem deformation in chia (Celli *et al.*, 2016) and mottle in dixie ticktrefoil. The presence of CPMMV in chia is a particular surprise as it is not a fabaceous plant, and this suggests that either the virus is expanding its host range, as a consequence perhaps of its increased prevalence in the field, or that it has always been present in other host families, but has only now been detected inducing symptoms. The latter is particularly likely if only asymptomatic infections had happened in these hosts.

Host range studies have shown that CPMMV is able to infect diverse hosts under experimental conditions. Thus, some African isolates infected tomato (*Solanum lycopersicum*), causing mottling and inconspicuous banding of minor veins (Brunt *et al.*, 1983; Brunt & Kenten, 1973) or asymptomatic infection (Tavasoli *et al.*, 2008), and *Nicotiana clevelandii*, causing chlorosis. Infection in these hosts was observed only for isolates from Ghana and the Ivory Coast (Brunt & Kenten, 1973; Thouvenel *et al.*, 1982). In *Nicotiana benthamiana, N. debney, Chenopodium amaranticolor* and *C. quinoa*, only chlorotic local lesions have been observed (Figure 2) (Almeida *et al.*, 2005; Brito *et al.*, 2012; Naidu *et al.*, 1998; Tavasoli *et al.*, 2008; Thouvenel *et al.*, 1982; Zanardo *et al.*, 2014a) and no reports of systemic infection were made in any of these hosts for any viral

isolates. The induction of local lesions in these hosts was observed for all CPMMV isolates tested, suggesting that these hosts can be used as indicator plants for diagnosis.

The capacities to induce different symptoms in the same host and infect hosts from different families are very interesting characteristics of CPMMV. The nature of these differential infections and the genomic regions involved in the induction of symptoms have not been determined, requiring further study. This promises to be challenging given the diversity of symptoms caused.

The spread of CPMMV globally and in Brazil

Following its original description in Ghana in 1973 (Brunt & Kenten, 1973) in cowpea, CPMMV was subsequently described infecting soybeans in Thailand and the Ivory Coast (Iwaki *et al.*, 1982; Thouvenel *et al.*, 1982) (Figure 3). Although the incidence was high in these regions, infections were commonly asymptomatic or inconspicuous, with the virus being considered of little importance in Africa (Brunt & Kenten, 1973). As a result, no control measures were undertaken, and probably the virus persisted in different hosts causing silent infections of minor interest.

A decade later, new outbreaks were noticed in east Africa, across Asia and in Brazil (Figure 3). In Kenya and Tanzania, CPMMV was found infecting groundnuts, soybean and mung bean, with losses of 64-80% of groundnut production reported from Kenya (Iwaki *et al.*, 1982; Mink & Keswani, 1987). The virus was subsequently described in Sudan infecting irrigated groundnuts during the 1992-1994 seasons (Figure 3). Over a period of three years,

the incidence in groundnut fields usually ranged between 10 and 50%, but in some fields amounted to 100% (El-Hassan *et al.*, 1997), with losses of 10-100%.

CPMMV was reported in Asia infecting soybean in Taiwan and Thailand (Iwaki, 1986; Iwaki *et al.*, 1982; Tavasoli *et al.*, 2008), groundnut in Malaysia, Indonesia and India (Iizuka *et al.*, 1984; Iwaki, 1986) and eggplants, soybean and cowpea in Jordan, Iran and Taiwan, respectively (Chang *et al.*, 2013; Ghorbani *et al.*, 2008; Tavasoli *et al.*, 2008). All infections showed typical symptoms of CPMMV infection.

The first description of CPMMV in the Americas was in 1983 in Brazil (Figure 3) (Costa *et al.*, 1983). The simplest scenario for its introduction to the Americas is in infected bean or soybean seeds from Africa. No trace has yet been found to provide information to map the introduction of the virus in Brazil; as not all isolates are available for sequencing, this task becomes a major challenge. The first description in Brazil is old, but, as in Ghana, the virus was considered of minor importance for almost 20 years. This only changed in the 2000-01 season when CPMMV became a primary factor limiting yield in soybean in Goiás state (Almeida *et al.*, 2003). Three years later, it was found in soybean across Brazil (Almeida, 2008; Almeida *et al.*, 2003; Almeida *et al.*, 2005). By 2010, CPMMV was present in a number of regions of the states of Paraná, Minas Gerais, Goiás, Mato Grosso, Bahia, Tocantins and Maranhão (Almeida, 2008; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b). The only Brazilian state that produces soybean but has not reported the virus is Rio Grande do Sul (RS) (Almeida, 2008). In January of 2014, we collected 200 soybean samples with symptoms of systemic necrosis in fields of the municipalities of Vacaria, Capão Bonito do Sul and Lagoa Vermelha (all in RS). All samples were used for indirect ELISA tests (Clark

426 *et al.*, 1986) using a polyclonal antiserum (Carvalho *et al.*, 2013) to confirm CPMMV
427 infections and none of the samples were positive (Zanardo upublished data). It is possible
428 then that RS represents a limit to the range of the virus. As Brazilian isolates are apparently
429 not transmitted by seeds and the winter in RS is relatively harsh, it is possible that the virus
430 vector is not able to survive in large numbers, so reducing the incidence of the virus.
431
432 CPMMV has also been described in other regions of the Americas. In Argentina (Salta
433 province) the virus was identified during the summers of 2001-02, 2005 and 2014, infecting
434 common bean, soybean and chia, respectively (Laguna *et al.*, 2006; Rodríguez- Pardina *et*
435 *al.*, 2004; Celli *et al.*, 2016). It was also found infecting yardlong bean (*V. unguiculata*
436 subsp. *sesquipedalis*) in Venezuela in 2012 (Brito *et al.*, 2012) and it has recently been
437 described in Puerto Rico infecting soybean in experimental fields (Brown & Rodrigues,
438 2014). In some soybean lines in Puerto Rico the virus was observed in 100% of plants,
439 suggesting the extreme potential of this disease to undermine the present and future
440 cultivation of soybean in Puerto Rico either in winter nurseries or during other times of the
441 year (Brown & Rodrigues, 2014).
442
443 The virus is also present in bean and whiteflies (*Bemisia tabaci*) in Florida-USA, but it has
444 not been found yet in others states of the US mainland (Rosario *et al.*, 2014). It is possible
445 that the virus was introduced in Florida by seed (beans and/or soybeans) or ornamental
446 plants from Puerto Rico or South America, because the viruses from Florida, Puerto Rico
447 and Brazil are phylogenetically close (Figure 4A and B). Although the transmission of
448 CPMMV by soybean seed in Brazil is still controversial, the transmission by seeds of
449 common bean has not yet been ruled out. So, there is the possibility of transmission to

Florida from Brazilian seeds. The presence of CPMMV in Florida could represent a risk for the production of soybean in the US, and the US is the largest producer in the world.

Phylogenetic relationship and viral strains

Given the global spread and distribution of CPMMV, insights may arise from a phylogenetic analysis of the available isolates. With this in mind, a phylogenetic tree was constructed of all of the complete genome sequences of carlaviruses available in Genbank, using Bayesian inference (Figure 4A). This showed clear clustering of CPMMV isolates in a quite distinct branch from other carlaviruses. This could have implications for the taxonomy of the group, a subject for future consideration (especially as CPMMV is transmitted by whiteflies while the other carlaviruses are transmitted by aphids). At the level of complete genome sequences, there is an indication that the original isolate, from Ghanaian cowpea, is distinct from the isolates from the Americas (from soybean, common bean or from whiteflies). There is an indication of two CPMMV strains, as described by Zanardo et al., (2014a) and these may be associated with possible recombination events within ORF1 (viral replicase). Unfortunately, no other full sequences are available at present.

The clustering among CPMMV isolates was clearly independent of hosts; there are isolates from different hosts (soybean, bean and whiteflies) clustering in clade 2 (Figure 4A). There are more sequences of CP available than of the full genome and a phylogenetic analysis of these is presented in Figure 4B. Here, it can be observed that there are clusters among CPMMV isolates independent of the host plants (soybean, beans, peanuts and also whitefly vectors). The phylogenetic analysis of CP shows two Indian isolates identified in peanuts

which were the most divergent, clearly clustering in a different clade. The phylogeny indicates that these are genetically quite distant from other CPMMV isolates (Figure 4B). Even the isolate from Ghana, considered the oldest, but recently sequenced, clustered with isolates sequenced in Brazil and India (Figure 4B).

Within clade 2 of the CP tree, there are two groups (figure 4B): the first group includes isolates from South America (Brazil and Argentina), India, Puerto Rico, Taiwan and Florida in U.S., while the second group included isolates from India, Ghana and Brazil. There is no obvious geographical pattern to this tree and likewise, host plants seem not to be a factor. This is a strong indication that a single isolate or variants of this isolate has been propagated over time to different host and regions, with genetic divergence arising subsequent to its spread

The two groups within clade 2 of the CP phylogeny suggest that there are different viral strains (1 and 2) (Figure 4B). The existence of two CPMMV strains (CPMMV-BR1 and CPMMV-BR2) in Brazil has already been proposed by Zanardo et al., 2014b, based on partial sequences (ORF2-6). A further result of these studies was that the topologies of the phylogenetic trees were associated with symptoms induced in soybean plants: the isolates causing severe symptoms in soybean tended to cluster together while those causing mild symptoms also clustered together (Zanardo *et al.*, 2014b). This suggests that symptom variation can be explained by genetic variation.

The above information indicates that the phylogenetic clustering among CPMMV isolates cannot be explained by host or geography. It is possible that recombination events are

generating the observed diversity and any phylogenetic patterns are best related to induction of different symptoms patterns within a given host. Above this level, there do appear to be at least two phylogenetically distinct CPMMV strains, and above the species level, it seems that the whitefly-vectored CPMMV is distinct from the other, aphid-vectored, carlaviruses.

The recent emergence of CPMMV infecting transgenic beans in Brazil

While the focus of this review is how CPMMV has emerged as a problem in Brazilian soybean, we have sought to emphasize the risk of its emergence in other situations, such as other crop plants or in other geographical regions. Indeed, we are able to justify this by describing its recent emergence (or re-emergence) in common bean (*P. vulgaris*) in Brazil. Specifically, CPMMV has emerged in a transgenic line of common bean that was released for commercial production in 2011 (Faria *et al.*, 2016). This line is resistant to another virus, *Bean golden mosaic virus* (BGMV; *Geminiviridae*: *Begomovirus*) that is a major constraint to bean production in Brazil. It is the first line released after some four decades of research seeking resistance to BGMV, so the emergence of CPMMV infecting this line only two years after its release was met with dismay.

Common beans are a staple food and represent a very important source of protein in the daily diet for Brazilians and the states of Paraná, Minas Gerais, Bahia, São Paulo and Goiás are the most important producer, described by Faria *et al.*, 2016. BGMV has been a key pathogen of common beans since the 1970's (Fiallos, 2010). This virus, like CPMMV, is transmitted by whiteflies and the main symptoms of the disease are a yellow-golden mosaic of the leaves, variable levels of curling, generally stunted growth, and distorted pods,

symptoms that vary among cultivated genotypes and with time of infection (Faria *et al.*, 2016).

When fields are infected, losses due to BGMV range from 40 to 100% (Morales, 2006). Because of this, breeding for resistance to golden mosaic started soon after its recognition as a problem in the 1970's. However, even after 25 years of breeding it was not possible to release a highly resistant commercial common bean cultivar (Faria *et al.*, 2016). Due to the lack of results from classic breeding efforts, allied to the tremendous difficulty of controlling the whitefly vector, work began on transgenic beans in the early 1990's (Aragão & Rech, 2001; Aragão *et al.*, 1998; Faria *et al.*, 2006). This research program also took considerable time, but in 2007, eighteen transgenic lines were obtained using an intron-hairpin construction of the BGMV *AC1* gene (rep), based on the RNAi technique. Plants with up to 93% resistance were obtained even when exposed to high levels of viruliferous whiteflies (Bonfim *et al.*, 2007). In the field, homozygous plants showed 100% immunity to BGMV, while neighboring non-transgenic plants showed severe symptoms characteristic of golden mosaic disease (Aragao & Faria, 2009). The transgenic plants had normal phenotypes and no off-target or epigenetic effects were observed; a line denominated 5.1 of BGMV-resistant GM common bean developed by Embrapa (Empresa Brasileira de Pesquisa Agropecuaria), was selected for biosafety analyses and proved to be safe for cultivation and human consumption (Aragão & Faria, 2010; Faria *et al.*, 2016). In 2011, the Brazilian Biosafety Committee (CTNBio) approved the commercial release of line (http://ctnbio.mcti.gov.br/liberacaocomercial#/liberacao-comercial/consultar-processo) (Faria *et al.*, 2016).

The expectation was that problems with BGMV infection in common bean had finally been resolved and that this virus would no longer be a problem for commercial production. In 2013, however, CPMMV was detected in common bean, for the first time since its description in beans in Brazil in 1983 (Costa *et al.*, 1983). Common bean fields in the Midwest, Northeast and Southern regions of Brazil were reported as being infected by CPMMV and the virus was particularly noticeable in the BGMV-resistant GM lines (Faria *et al.*, 2016). The symptoms of CPMMV in these GM plants were crinkled leaves, with vein enations visible as necrotic veins on the abaxial surface of the leaf (Faria *et al.*, 2016). In inoculated plants, CPMMV infection was confirmed using electron microscopy (feather-like aggregates of CPMMV virions were observed), ELISA tests and next generation sequencing (NGS).

To date, two isolates of CPMMV from BGMV-resistant GM common bean have been sequenced, one from the Northeast of Brazil (Faria *et al.*, 2016) and the other from the state of Paraná in the South (Milanesi *et al.*, 2015). Both exhibited high sequence similarity (93-99%) with Brazilian isolates from soybean. This suggests that CPMMV may have been present for a long time in common beans in mixed infections with BGMV, remaining undetected while BGMV remained at a high prevalence. Another (non-exclusive) possibility is that the virus now found in common bean shifted hosts from soybean recently.

Whatever the reasons for the emergence of CPMMV in common bean, it represents a tremendous challenge. This virus has been reducing soybean yields for about 15 years and to date there are no resistant soybean cultivars available (Almeida *et al.*, 2003; Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b). The BGMV-resistant common bean could serve as a

starting point to develop resistance also to CPMMV. As there are common bean cultivars that are not infected by CPMMV (Zanardo *et al.*, 2014a), classical breeding may still be an interesting option. Recently, it was shown that six Brazilian CPMMV isolates infecting soybean, which showed high similarity with CPMMV isolates infecting beans, are unable to infect *P. vulgaris* cv. Ouro Negro and *P. vulgaris* cv. Manteigão (Zanardo *et al.*, 2014a). Both of these hosts could be good candidates to be used in breeding programs against CPMMV infection. As a final consideration, the relationship *in planta* between BGMV and CPMMV infections has not been studied, and such work could help to explain how CPMMV infections may have remained undetected in common beans (if that is what really happened).

Control mechanisms and resistance

CPMMV presents distinct challenges for control or management. The first challenge is control of its spread and emergence in new crops. It seems to have been introduced accidentally in many different countries; possible routes are via asymptomatic infected plants (including possibly ornamentals), infected seeds, or the whitefly vector. Movement of plant material across international borders is already controlled to varying degrees while specific control of seeds aimed at CPMMV would require greater evidence of its potential for seedborne transmission. Effective control of the spread of the whitefly vector also seems to be difficult. For the time being, it is probably most important to monitor new infections due to CPMMV and act as these arise. CPMMV has the potential for rapid spread, vector control is difficult and, while there are at least two CPMMV strains, there is little indication of population structuring in the virus according to geography or host plants that might suggest it could be restricted to regions or crop species. It is transmitted in a

non-persistent mode by a highly polyphagous and generally abundant vector. CPMMV is an RNA virus, which are known for their high rates of evolution, and recombination seems to be important. For all of these reasons, CPMMV represents a threat to fabaceous and other crops across the world, and current knowledge offers little prospect of control of its geographical spread or of its emergence in new crops.

While the above discussion is not very optimistic, it is important that the spread and emergence of CPMMV be understood so that action can be taken when it does threaten agricultural production. Although it is far from a quick or easy solution, the development of resistant cultivars probably represents the best strategy. The screening of resistance sources in soybean, common bean and peanut is crucial to prepare for impending new outbreaks and to control the virus where it has been found. Resistance must be durable and effective for both viral strains. Suryanto *et al.* (2014) tried to identify the number and roles of CPMMV resistance genes in Indonesian soybean cultivars. They showed that the resistance properties of soybean were controlled by two duplicate recessive or dominant (depending on the cultivar combination) epistatic genes and its inheritance was additive. The fact that inheritance of resistance is additive means that this resistance has a high genetic value and can be inherited, which are two important characteristic for breeding programs of soybean crops. Despite this initial work, no resistant cultivar is available at the moment for virus control.

In Brazil, work on resistance or tolerance to CPMMV is still at an early stage. Mituti & Almeida (2006) suggested that the use of tolerant cultivars could be a good option to control the virus. It was shown that the soybean cultivars BRSMT Pintado and BRS 133,

which are tolerant to CPMMV, accumulated more viruses in initial infection than the susceptible cultivars CD 206 and BRS 136, especially at low temperatures. However, 16 days after inoculation, the tolerant cultivars were able to recover from infection, reducing the viral titer and the chances of viral propagation. This indicates that the use of tolerant cultivars might help to limit the spread of CPMMV in the field in Brazil; given there is no resistant cultivar. However, the fact that these two cultivars have a high viral titer at the beginning of the infection may favor, even for a short period, the transmission of the virus. As the virus is transmitted in a nonpersistent manner, this is a particular concern and so tolerant cultivars should be used with caution.

So, while tolerant cultivars are an option, resistance is more desirable. In an effort to identify sources of CPMMV resistance, Almeida (2008) analyzed the reaction to CPMMV infection in 170 soybean cultivars available in Brazil using mechanical inoculation and analyses based on symptomatology. He found eight cultivars resistant to CPMMV, 23 with moderate resistance and 139 that were susceptible. In another recent study in Brazil, 50 soybean cultivars were analyzed and 92% showed tolerance to CPMMV while only 8% were susceptible (Brizola *et al.*, 2015). These results show that the virus is able to infect a variety of Brazilian soybean cultivars, but even so there are resistant or tolerant cultivars available that could be used in a breeding program to prevent or restrict CPMMV infection in Brazilian fields. This would have the added benefit of reducing the prevalence of the virus, and so its potential to emerge in new crops.

In general more studies are needed to understand the epidemiology of CPMMV and its interaction with its host plants, to find resistant cultivars and ensure that any resistance is

durable. It is also necessary to alert growers to the problem and ensure they can recognize the symptoms of CPMMV disease in its numerous hosts (not only legumes), as well as paying attention to CPMMV transmission through seeds and its spread by the whitefly vector.

PERSPECTIVES

Although CPMMV was first described a considerable time ago, it remains a challenge for virologists. Its recent emergence, especially in Brazilian soybean and now common bean, means that it has begun to attract more attention. The fact that it is now found infecting hosts of new families and in different geographical regions further demonstrates the emerging nature of this pathogen and potential future risks. Meanwhile, the fact that it can at least in some instances be transmitted by seeds, and is also transmitted by a notorious vector in *B. tabaci*, mean that its control promises to be extremely difficult.

Key areas for future research are the basic biology of CPMMV, in particular how its genome ultimately contributes to the infection process and to symptom expression, how it evolves and its interaction with host plants and vectors.

As described above, very little is known about how CPMMV infection occurs, particularly which viral proteins are necessary in each moment of the infectious process. Unraveling this process is a necessary task for greater understanding of the pathogen and one possible route is the discovery of host proteins used by the virus. This strategy promises to be informative also to efforts to breed resistance to the virus and to ensure any resistance is durable.

The high degree of variation in symptoms caused within the same soybean cultivar is a very interesting characteristic of CPMMV that provides clear evidence of significant molecular variability within CPMMV populations. Quantifying and characterizing this variation is an essential task to make future resistant cultivars durable in the field as it is probable that the variation of symptoms observed in soybean plants is directly associated with viral genomic variation. The viral proteins are subject to different evolutionary forces and induce different phenotypes. The nature of the viral proteins and the exact locations of the symptom determinants are completely unknown for CPMMV.

The ability of the whitefly to transmit CPMMV has been very little investigated. Among the many carlaviruses described, CPMMV is one of the few transmitted by this vector and it would be fascinating to discover how this became possible. As there may be vector proteins involved in this process, this understanding might offer avenues to restrict vector borne transmission in the future. A particular mystery is how highly virulent isolates of CPMMV can be efficiently transmitted in the field as necrosis of the plant is presumably lethal to developing vectors. There may well be variations in transmission of isolates that are related to the symptoms they cause and this may help to explain why symptoms can be so variable in the field.

We hope soon to unravel some of these mysteries for CPMMV and help to understand much more about CPMMV infections, especially in Brazil.

690 **ACKNOWLEDGEMENTS**

691 We are grateful to Prof. Elliot Watanabe Kitajima (ESALQ) for assistance with the
692 transmission electron micrographs and to the Nucleus of Support for Research in Electron
693 Microscopy of ESALQ. We also thank Simon Elliot for comments on the manuscript.

694

695 **REFERENCES**

696 Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN,
697 Ryu KH, Saldarelli P, Yoshikawa N Family Betaflexiviridae. pp.920-941 In: KING AMQ,
698 ADAMS MJ, CARSTENS EB & LEFKOWITZ EJ (Eds.) Virus taxonomy. Ninth Report of
699 the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic
700 Press. 2012.
701
702 Adams M, Kreuze J (2016) Revision of family Betaflexiviridae, order Tymovirales. ICTV
703 online. Avaliable at: www.ictvonline.org/. Acessed on December 10, 2016.
704
705 Almeida AMR, Piuga FF, Kitajima EW, Gaspar JO, Valentin N, Benato LC, Marin SRR,
706 Bineck E, Belintani P, Nunes Junior J, Hoffmann L, Meyer MC (2003) Necrose da haste da
707 soja. Série Documentos - Embrapa soja 221:1-48.
708
709 Almeida AMR, Piuga FF, Marin SRR, Kitajima EW, Gaspar JO, Oliveira TG, Moraes TG
710 (2005) Detection and partial characterization of a carlavirus causing stem necrosis of
711 soybean in Brazil. Fitopatologia Brasileira 30:191-194.
712
713 Almeida AMR (2008) Viroses da Soja no Brasil: sintomas, etiologia e controle. Série
714 Documentos - Embrapa soja 306:1-62.
715
716 Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004)
717 Emerging infectious diseases of plants: pathogen pollution, climate change and
718 agrotechnology drivers. Trends in Ecology & Evolution 19:535-544.
719
720 Aragao FJ, Faria JC (2009) First transgenic geminivirus-resistant plant in the field. Nature
721 Biotechnology 27:1086-1088.
722
723 Aragão FJL, Faria JC (2010) Proposta de liberação comercial de feijoeiro geneticamente
724 modificado resistente ao mosaico dourado. Evento Embrapa 5.1 (EMB-PV051-1) CTNBio,
725 Avaliable on: ctnbio.mcti.gov.br/liberacaocomercial#/liberacao-comercial/consultar-
726 processo.
727
728 Aragão FJL, Rech EL (2001) Transgenic common bean (*Phaseolus vulgaris*). In:
729 Biotechnology in Agriculture and Forestry-Transgenic Crops II. Y. P 269-283.
730

Aragão FJL, Ribeiro SG, Barros LMG, Brasileiro ACM, Maxwell D, P. R, E. L.,, Faria JC (1998) Transgenic beans (*Phaseolus vulgaris L.*) engineered to express viral antisense RNAs showed delayed and attenuated symptoms of bean golden mosaic geminivirus. Molecular Breeding 4:491-499.

Bedhomme S, Hillung J, Elena SF (2014) Emerging viruses: why they are not jacks of all trades? Current Opinion of Virology 10:1-6.

Bonfim K, Faria JC, Nogueira EO, Mendes EA, Aragao FJ (2007) RNAi-mediated resistance to Bean golden mosaic virus in genetically engineered common bean (Phaseolus vulgaris). Molecular Plant-Microbe Interactions Journal 20:717-726.

Brito M, Fernandez-Rodriguez T, Garrido MJ, Mejias A, Romano M, Marys E (2012) First report of *Cowpea mild mottle Carlavirus* on yardlong bean (*Vigna unguiculata* subsp. *sesquipedalis*) in Venezuela. Viruses 4:3804-3811.

Brizola DC, Dias LA, Silva JA, Policam MPRM, Almeida AMR (2015) Avaliação da reação de linhagens de soja ao vírus do mosaico comum da soja (*Soybean mosaic virus*) e ao vírus causador da necrose da haste (*Cowpea mild mottle virus*). Série Documentos 359 - X Jornada Acadêmica da Embrapa Soja-Resumos expandidos 80-87.

Brown JK (2000) Molecular markers for the identification and global tracking of whitefly vector-Begomovirus complexes. Virus Research 71:233-260.

Brown JK, Rodrigues JCV (2014) Recovery Plan *Cowpea mild mottle virus* Carlavirus: *Betaflexiviridae*; order *Tymovirales*. 1-24. Avaliable at: www.ars.usda.gov/. Acessed on december 04, 2016.

Brunt AA, Atkey PT, Woods RD (1983) Intracellular occurrence of *Cowpea mild mottle virus* in two unrelated plant species. Intervirology 20:137-142.

Brunt AA, Kenten RH (1973) *Cowpea mild mottle*, a newly recognized virus infecting cowpeas (*Vigna unguiculata*) in Ghana. Annals of Applied Biology 74:67-74.

Brunt AA, Phillips S (1981) Fuzzy vein, a disease of tomato (*Lycopersicon esculenlum*) in western Nigeria induced by *Cowpea mild mottle virus*. Tropical Agriculture 58:117-180.

Carvalho SL, Silva FN, Zanardo LG, Almeida AMR, Zerbini FM, Carvalho CM (2013) Production of polyclonal antiserum against *Cowpea mild mottle virus* coat protein and its application in virus detection. Tropical Plant Pathology 38:49-54.

Celli MG, Perotto MC, Merino MC, Nome CFD, Flores CR, Conci VC (2016) First report of *Cowpea mild mottle virus* in chia (*Salvia hispanica*). Crop Protection 89:1-5.

Chang CA, Chien LY, Tsai CF, Lin YY (2013) First Report of Cowpea mild mottle virus in Cowpea and French Bean in Taiwan. Plant Disease 97:1001.

Clark MF, Lister RM, Bar-Joseph M, Arthur Weissbach HW ELISA techniques. pp.742-766 In: Methods in Enzymology. Academic Press. 1986.

Cohen S, Antignus Y (1982) A noncirculative whitefly-borne virus affecting tomatoes in Israel. Phytoparasitica 10:101-109.

Costa AS, Gaspar JO, Vega J (1983) Mosaico angular do feijão jalo causado por um carlavírus transmitido pela mosca branca *Bemisia tabaci*. Fitopatologia Brasileira 8:325-327.

Da Fonseca BL, Yuki VA, Marubayashi JM, De Marchi BR, Perini FL, Pavan MA, de Barros DR, Ghanim M, Moriones E, Navas-Castillo J, Krause-Sakate R (2015) First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil. Pest Management Science 71:501-504.

De Barro PJ, Liu SS, Boykin LM, Dinsdale AB (2011) Bemisia tabaci: a statement of species status. Annul Reviews of Entomology 56:1-19.

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

Dobson A, Foufopoulos J (2001) Emerging infectious pathogens of wildlife. Philosophical Transactions of the Royal Society of London B Biological Sciences 356:1001-1012.

Durbern J, Dollet M (1981) Groundnut crinkle virus, a new member of *Carlavirus* group. Phytopathology 101:337-347.

El-Hassan S, M., Naidu R, A., Ahmed AH, Murant AF (1997) A Serious Disease of Groundnut Caused by *Cowpea Mild Mottle Virus* in the Sudan. Journal of Phytopathology 145:301-304.

EMBRAPA (2014) Soja em números (safra 2013/2014). Available at: www.embrapa.br/soja/cultivos/soja1/dados-economicos. Acessed on 2015.

Faria J, Aragão F, Souza T, Quintela E, Kitajima E, Ribeiro S (2016) Golden Mosaic of Common Beans in Brazil: Management with a Transgenic Approach. APS Features.

Faria JC, Albino MMC, Dias BBA, Cançado LJ, Cunha NB, Silva LM, Viana GR, Aragão FJL (2006) Partial resistance to Bean golden mosaic virus in a transgenic common bean (*Phaseolus vulgaris L.*) line expressing a mutated rep gene. Plant Science 171:565-571.

Fauquet C, Thouvenel J Plant Viral Diseases in the Ivory Coast. Paris: ORSTOM. 1987.

Fiallos FRG (2010) Doenças causadas por virus na cultura de feijoeiro (*Phaseolus vulgaris L.*). Ciencia y Tecnología 3:1-6.

826 Fortuner R, Fauquet C, Lourd M (1979) Diseases of the winged bean in Ivory Coast. Plant
827 Disease Reporter 63:194-199.
828
829 Foster GD, Mills PR (1990a) Evidence for the role of subgenomic RNA species in the
830 production of *Helenium virus S* coat protein during in vitro translation. Virus Research
831 17:61-70.
832
833 Foster GD, Mills PR (1990b) Evidence for the role of subgenomic RNAs in the production
834 of potato virus S coat protein during in vitro translation. Journal of General Virology
835 71:1247-1249.
836
837 Foster GD, Mills PR (1992) Translation of *Potato virus S* RNA in vitro: Evidence of
838 protein processing. Virus Genes 6:47-52.
839
840 Fujita N, Komatsu K, Ayukawa Y, Matsuo Y, Hashimoto M, Netsu O, Teraoka T, Yamaji
841 Y, Namba S, Arie T (2016) N-terminal region of cysteine-rich protein (CRP) in
842 carlaviruses is involved in the determination of symptom types. Molecular Plant Pathology.
843
844 Ghorbani SGM, Shahraein N, Elahinia SA (2008) Serodiagnosis of cowpea (*Vigna*
845 *unguilata)* viruses in Guilan province, Iran. Iranian Journal of Virology 1:28-31.
846
847 Henning AA, Almeida AMR, Godoy CV, Seixas CDS, Yorinori JT, Costamilan LM,
848 Ferreira LP, Meyer MC, Soares RM, Dias WP (2014) Manual de identificação de doenças
849 da soja. Série Documentos - Embrapa soja 256.
850
851 Hirakuri MH, Lazzarotto PP (2014) O agronegócio da soja nos contextos mundial e
852 brasileiro. Série Documentos - Embrapa soja 349.
853
854 Iizuka N, Rajeshwari R, Reddy DVR, Goto T, Munivappa V, N. Bharathan N, Ghanekar
855 AM (1984) Natural Occurrence of a Strain of *Cowpea mild mottle virus* on Groundnut
856 (*Arachis hypogaea*) in India. Phyphatology 109:245-253.
857
858 Imam J, Singh PK, Shukla P (2016) Plant Microbe Interactions in Post Genomic Era:
859 Perspectives and Applications. Frontiers in Microbiology 7:1488.
860
861 Iwaki M (1986) Soybean crinkle leaf and cowpea mild mottle viruses. International
862 Symposium on Virus Diseases of Rice and Leguminous Crops. Tropical Agricultural
863 Research 19.
864
865 Iwaki M, Thongmeearkon P, Prommin M, Honda Y, Hibi J (1982) Whitefly transmission
866 and some properties of *Cowpea mild mottle virus* on soybean in Thailand. Plant Disease
867 66:265-268.
868
869 Jeyanandarajah P, Brunt AA (1993) The natural occurrence, transmission, properties and
870 possible affinities of *Cowpea mild mottle virus*. Journal of Phytopathology 137:148-156.
871
872 Laguna IG, Arneodo JD, Rodríguez-Pardina P, Fiorona M (2006) *Cowpea mild mottle virus*
873 infecting soybean crops in northwestern Argentina. Fitopatologia Brasileira 31:317-317.

Lalic J, Cuevas JM, Elena SF (2011) Effect of host species on the distribution of mutational fitness effects for an RNA virus. PLoS Genetics 7:e1002378.

Lawrence DM, Rozanov MN, Hillman BI (1995) Autocatalytic processing of the 223-kDa protein of blueberry scorch carlavirus by a papain-like proteinase. Virology 207:127-135.

Li YY, Zhang RN, Xiang HY, Abouelnasr H, Li DW, Yu JL, McBeath JH, Han CG (2013) Discovery and Characterization of a Novel Carlavirus Infecting Potatoes in China. PLoS One 8:e69255.

Lima LHC, Campos L, Moretzsohn MC, Návia D, Oliveira MRV (2002) Genetic diversity of *Bemisia tabaci* (Genn.) populations in Brazil revealed by RAPD markers. Genetics and Molecular Biology 25:217-223.

Lukhovitskaya NI, Ignatovich IV, Savenkov EI, Schiemann J, Morozov SY, Solovyev AG (2009) Role of the zinc-finger and basic motifs of chrysanthemum virus B p12 protein in nucleic acid binding, protein localization and induction of a hypersensitive response upon expression from a viral vector. Journal of General Virology 90:723-733.

Lukhovitskaya NI, Solovieva AD, Boddeti SK, Thaduri S, Solovyev AG, Savenkov EI (2013) An RNA virus-encoded zinc-finger protein acts as a plant transcription factor and induces a regulator of cell size and proliferation in two tobacco species. Plant Cell 25:960-973.

Mackenzie RJ, Tremaine JH, State-Smith R (1989) Organization and interviral homologies of the 3'-terminal portion of potato virus S RNA. Journal of General Virology 70:1053-1063.

Mansour A, Al-Musa A, Vetten HJ, Lesemann DE (1998) Properties of a *Cowpea Mild Mottle Virus* (CPMMV) Isolate from Eggplant in Jordan and Evidence for Biological and Serological Differences Between CPMMV Isolates from Leguminons and Solanaceous Hosts. Journal of Phytopathology 146:539-547.

Marubayashi JM, Yuki VA, Wutke EB (2010) Transmissão do *Cowpea mild mottle virus* pela mosca branca *Bemisia tabaci* biótipo B para plantas de feijão e soja. Summa Phytopathologica 36:158-160.

Meehan BM, Mills PR (1991) Cell-free translation of *Carnation latent virus* RNA and analysis of virus specific dsRNA. Virus Genes 5:175-181.

Meng B, Li C (2010) The capsid protein of Grapevine rupestris stem pitting-associated virus contains a typical nuclear localization signal and targets to the nucleus. Virus Research 153:212-217.

Menzel W, Winter S, Vetten H (2010) Complete nucleotide sequence of the type isolate of *Cowpea mild mottle virus* from Ghana. Archives of Virology 155:2069-2073.

Menzel W, Abang MM, Winter S (2011) Characterization of Cucumber vein-clearing virus, a whitefly (Bemisia tabaci G.)-transmitted carlavirus. Archives of Virology 156:2309-2311.

Milanesi DF, Zanardo LG, Faria JC, Carvalho CM (2015) Detection and whole genome sequencing of CPMMV in common bean resistant to BGMV from Paraná state. Virus Reviews & Research, Brasília, DF, v. 20, p. 212, Oct. 2015. Supplement 1, ref. PIV408. Edição dos Resumos do XXVI Brazilian Congress of Virology, X Mercosur Meeting of Virology, Florianóplis, SC.

Mink GI, Keswani CL (1987) First report of *Cowpea mild mottle virus* on bean and mung bean in Tanzania. Plant Disease 71:557.

Mituti T, Almeida AM (2006) Tolerância ao *Cowpea mild mottle virus* e sua concentração relativa em soja. Série Documentos - Embrapa soja 276:168-170.

Morales FJ (2006) History and current distribution of begomoviruses in Latin America. Advances in Virus Research 67:127-162.

Morales FJ, Jones PG (2004) The ecology and epidemiology of whitefly-transmitted viruses in Latin America. Virus Research 100:57-65.

Mukoye B, Mangeni BC, Leitich RK, Wosula DW, Omayio DO, Nyamwamu PA, Arinaitwe W, Winter S, Abang MM, Were HK (2015) First report and biological characterization of *Cowpea mild mottle virus* (CPMMV) infecting groundnuts in western Kenya. Journal of Agri-Food and Applied Sciences 3:1-5.

Munyappa V, Reddy DVR (1983) Transmission of *Cowpea mild mottle virus* by *Bemisia tabaci* in a nonpersistent manner. Plant Disease 67:391-393.

Nagata T, Alves DM, Inoue-Nagata AK, Tian TY, Kitajima EW, Cardoso JE, de Avila AC (2005) A novel melon flexivirus transmitted by whitefly. Archives of Virology 150:379-387.

Naidu RA, Gowda S, Satyanarayana T, Boyko V, Reddy AS, Dawson WO, Reddy DV (1998) Evidence that whitefly-transmitted *Cowpea mild mottle virus* belongs to the genus *Carlavirus*. Archives of Virology 143:769-780.

Polston JE, De Barro P, Boykin LM (2013) Transmission specificities of plant viruses with the newly identified species of the Bemisia tabaci species complex. Pest Management Science 70:1547-1552.

Prado FG (2014) Effects of *Cowpea mild mottle virus* on *Bemisia tabaci*. – M.Sc Thesis, Universidade Federal de Viçosa. Viçosa, MG, BR.

Pramesh D, Baranwal VK (2013) Molecular characterization of coat protein gene of Garlic common latent virus isolates from India: an evidence for distinct phylogeny and recombination. Virus Genes 47:189-193.

970 Reuters (2017) Colheita de soja do Brasil começa com previsão de recorde de 103,5 mi t.
971 Available at: www.portaldoagronegocio.com.br Acessed on January 10, 2017.
972
973 Rodríguez- Pardina PE, Arneodo JD, Truol GA, Herrera PS, Laguna IG (2004) First record
974 of *Cowpea mild mottle virus* in bean crops in Argentina. Australasian Plant Pathology
975 33:129-130.
976
977 Rosario K, Capobianco H, Ng TF, Breitbart M, Polston JE (2014) RNA viral metagenome
978 of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus
979 in North America. PLoS One 9:e86748.
980
981 Rupasov VV, Morozov SY, Kanyuka KV, Zavriev SK (1989) Partial nucleotide sequence
982 of potato virus M RNA shows similarities to potexviruses in gene arrangement and be
983 encoded amino acid sequences. . Journal of General Virology 70:1861-1869.
984
985 Senshu H, Yamaji Y, Minato N, Shiraishi T, Maejima K, Hashimoto M, Miura C, Neriya
986 Y, Namba S (2011) A dual strategy for the suppression of host antiviral silencing: two
987 distinct suppressors for viral replication and viral movement encoded by *Potato virus M*.
988 Journal of Virology 85:10269-10278.
989
990 Singh A, Mahinghara B, Hallan V, Ram R, Zaidi A (2008) Recombination and
991 phylogeographical analysis of *Lily symptomless virus*. Virus Genes 36:421-427.
992
993 Singh L, Hallan V, Martin D, Ram R, Zaidi A (2012) Genomic sequence analysis of four
994 new *Chrysanthemum virus B* isolates: evidence of RNA recombination. Archives of
995 Virology 157:531-537.
996
997 Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems.
998 Annual Review Phytopathology 46:75-100.
999
1000 Suryanto A, Kuswanto;, Sitompul SM, Kasno A (2014) Estimation Of Number And Genes
1001 Actions Of CPMMV (*Cowpea Mild Mottle Virus*) Disease Resistance Genes On Soybean
1002 Crop. IOSR Journal of Agriculture and Veterinary Science 7:51-57.
1003
1004 Tavantzis SM (1991) Coat protein and protease activity as in vitro translation products of
1005 *Potato carlavirus M*. Archives of Virology 120:241-252.
1006
1007 Tavasoli M, Shahraeen N, Ghorbani S (2008) Detection and some properties of *Cowpea*
1008 *mild mottle virus* isolated from soybean in Iran. Pakistan Journal of Biological Sciences
1009 11:2624-2628.
1010
1011 Thouvenel JC, Monsarrat A, Fauquet C (1982) Isolation of *Cowpea mild mottle virus* from
1012 diseased soybean in the Ivory Coast. Plant Disease 66:336-337.
1013
1014 Turner R, Foster GD (1997) Deletion analysis of a translational enhancer upstream from the
1015 coat protein open reading frame of potato virus S. Archives of Virology 142:167-175.
1016

Turner RL, Glynn M, Taylor SC, Cheung MK, Spurr C, Twell D, Foster GD (1999) Analysis of a translational enhancer present within the 5'-terminal sequence of the genomic RNA of potato virus S. Archives of Virology 7:1451-1461.

Woolhouse ME (2002) Population biology of emerging and re-emerging pathogens. Trends in Microbiology 10:S3-7.

Yates A, Antia R, Regoes RR (2006) How do pathogen evolution and host heterogeneity interact in disease emergence? Proceeding of Royal Society B 273:3075–3083.

Zanardo LG, Silva FN, Bicalho AAC, Urquiza GPC, Lima ATM, Almeida AMR, Zerbini FM, Carvalho CM (2014a) Molecular and biological characterization of *Cowpea mild mottle virus* isolates infecting soybean in Brazil and evidence of recombination. Plant Pathology 63:456-465.

Zanardo LG, Silva FN, Lima AT, Milanesi DF, Castilho-Urquiza GP, Almeida AM, Zerbini FM, Carvalho CM (2014b) Molecular variability of *Cowpea mild mottle virus* infecting soybean in Brazil. Archives of Virology 159:727-737.

Zhao L, Feng C, Wu K, Chen W, Chen Y, Hao X, Wu Y (2017) Advances and prospects in biogenic substances against plant virus: A review. Pesticide Biochemistry and Physiology 135:15-26.

Figure Legends

Figure 1: Genomic organization of *Cowpea mild mottle virus* (CPMMV) and cytoplasmatic inclusions induced in soybean CD206 by CPMMV. **A.** Genomic organization of CPMMV in six ORFs and the putative protein expression strategy. The possible auto proteolysis of viral replicase (coding by ORF1) is in accordance with Lawrence *et al*., 1995. In the viral replicase, four domains are described (Methyltrasferase, Helicase, C23-peptidase and RNA-dependent RNA polymerase). In this model, the 3'ORFs are expressed by subgenomic RNAs (sgRNAs): the sgRNA1 (the largest) is responsible for translation of the triple gene block proteins (TGBs), while the second sgRNA2 (the smaller) is responsible for the capsidial protein (CP) and cysteine rich protein (CRP). **B and C.** Transmission electron micrograph of CPMMV-infected soybean cv. CD206 leaf cell 14 days after mechanical inoculation. **B.** Sheets or bundles-like aggregate of CPMMV virions

in citoplasm. **C.** Brush-like inclusion of CPMMV virions. In both images the virions were immunogold labeled after being exposed to anti-CP-(CPMMV) antibody. The black points indicate the gold particles specifically attached to the filamentous particles forming the inclusions.

Figure 2: Symptoms induced by Brazilian *Cowpea mild mottle virus* (CPMMV) isolates in different hosts. **A.** Mosaic in common bean cv. Jalo; **B.** Mosaic in cowpea cv. B7 Gurguéia; **C.** Local lesions in *N. benthamiana*; **D.** Chlorotic local lesions in *N. debneyi*. **E and F.** Mosaic and vein clearing in soybean CD206; **G.** Systemic necrosis in soybean CD206; **H.** Bud blight, leaf and stem necrosis in soybean CD206.

Figure 3: Global spread of *Cowpea mild mottle virus* (CPMMV) since the first report. The numbers represents chronologically the descriptions of CPMMV in the different countries in the world: 1- Ghana; 2a- Ivory Cost; 2b-Thailand; 3-Kenya; 4-Tanzania; 5a-Malaysia; 5b-Indonesia; 6-Brazil; 7-India; 8-Sudan; 9-Jordan; 10-Argentina; 11-Iran; 12-Taiwan; 13-Venezuela; 14-Puerto Rico and 15- Florida (USA). The letters a and b refer to locations that were described in the same year.

Figure 4: Phylogenetic relationships among *Cowpea mild mottle virus* (CPMMV) isolates and other carlaviruses, based on complete genome sequences or capsidial protein (CP) sequences using Bayesian inference (implemented in MrBayes in CIPRES with 10 million generations) **A.** Complete genome phylogeny with selection of model GTR+I+G. *Indian citrus ringspot virus* (ICRSV, genus *Mandarivirus*, family *Alphaflexiviridae*) was used as the outgroup. The 5' UTR, 3' UTR and intergenic regions were removed from the

alignment and the overlapping coding regions were maintained for complete genome phylogeny. **B.** CP phylogeny with selection of model GTR+G. Support for the nodes is presented as filled circles (posterior probabilities from 0.98 to 1.0) or open circles (posterior probabilities from 0.85 to 0.90). The accession numbers of the sequences are shown next to their acronyms. The black and gray bars indicate the two clades observed in the phylogenies.

Figure 1:

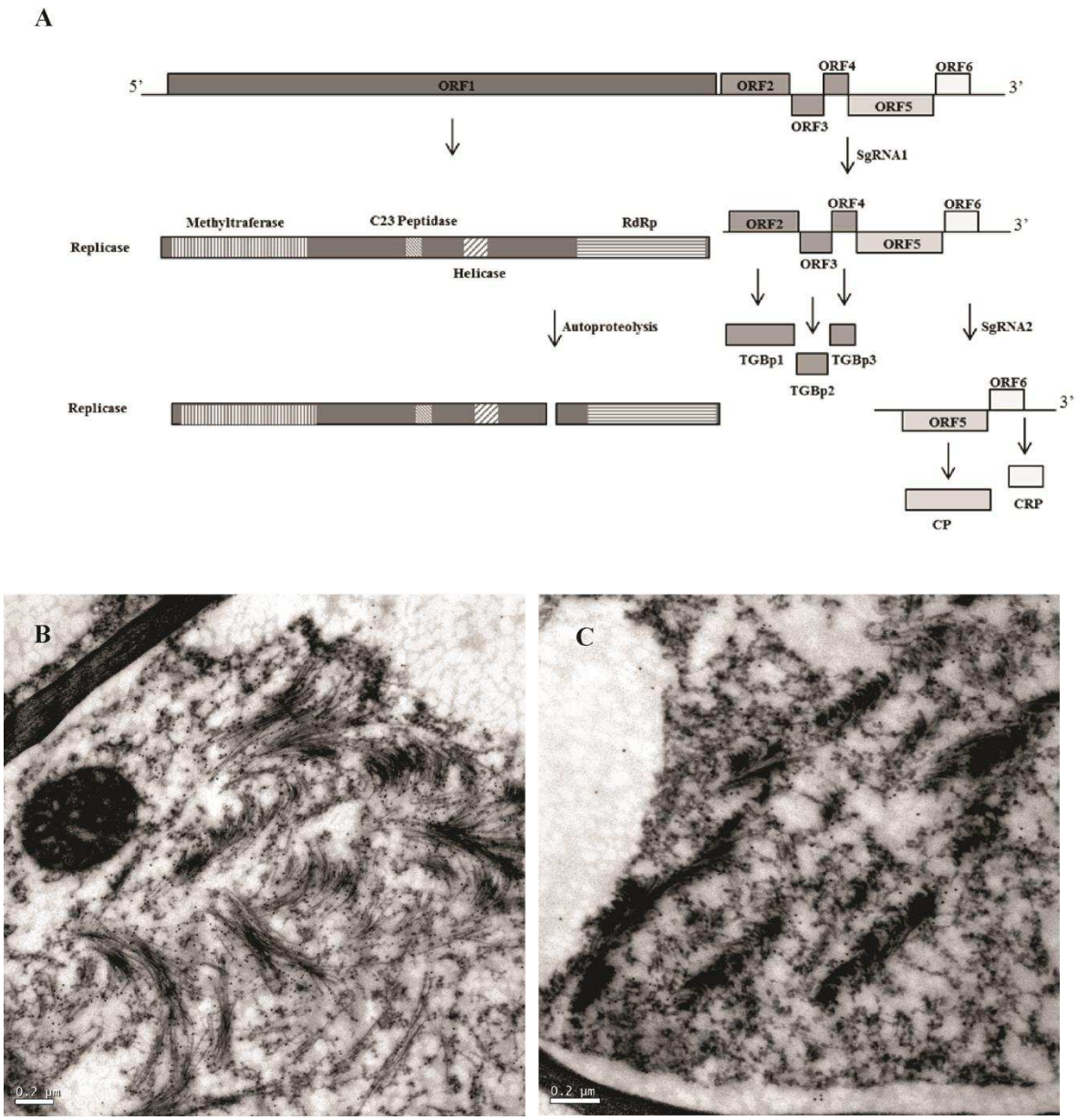

A
5'
ORF1
ORF2
ORF4
ORF6
3'
ORF5
ORF3
SgRNA1
Methyltraferase
C23 Peptidase
RdRp
Replicase
Helicase
ORF2
ORF4
ORF6
3'
ORF5
ORF3
Autoproteolysis
SgRNA2
TGBp1
TGBp3
TGBp2
Replicase
ORF6
3'
ORF5
CRP
CP
B
0.2 µm
C
0.2 µm

Figure 2:

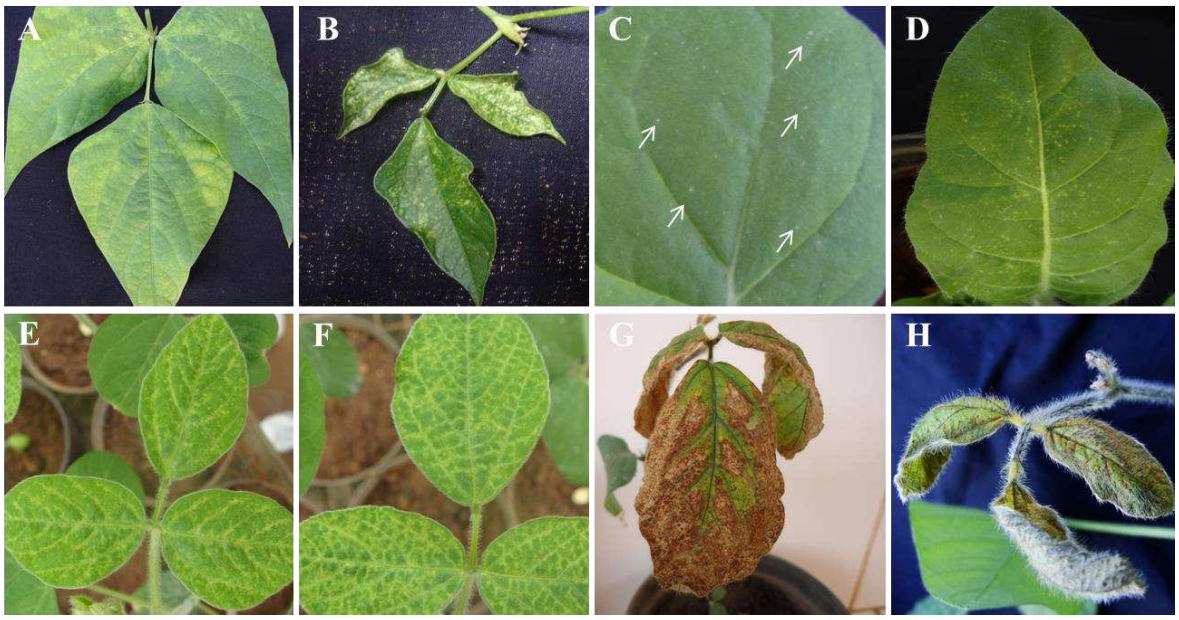


Figure 3:

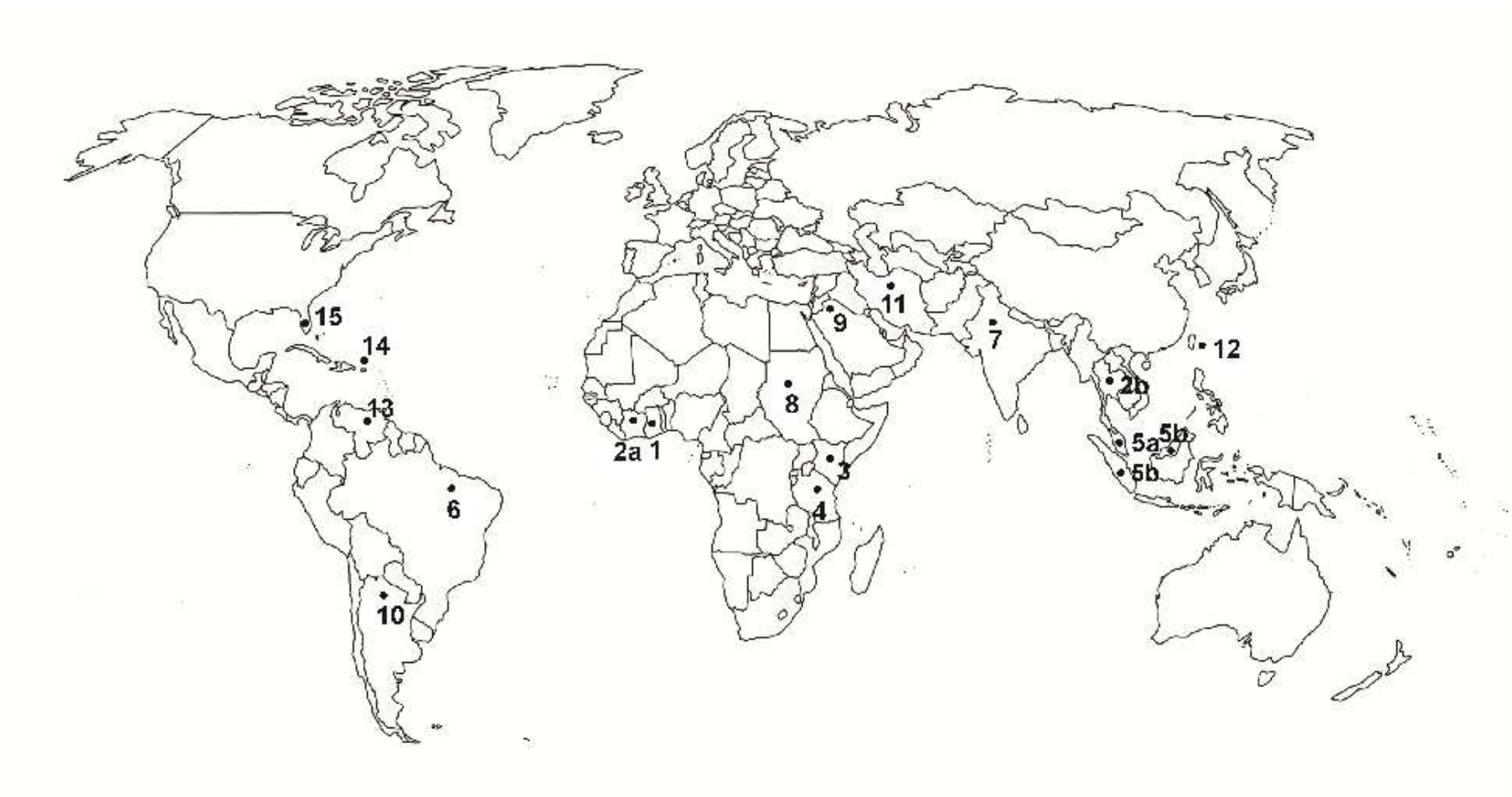
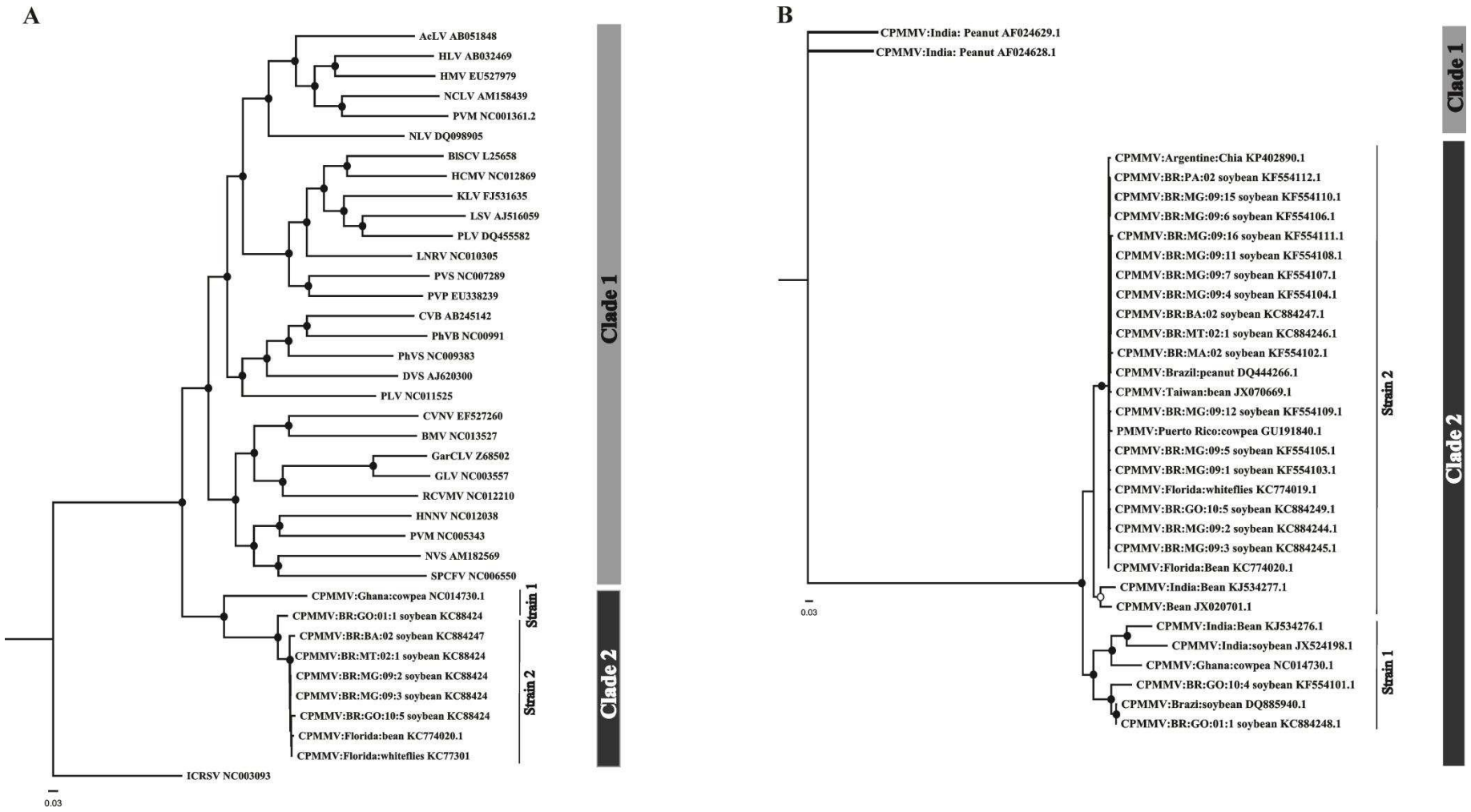


Figure 4:


A
AcLV AB051848
HLV AB032469
HMV EU527979
NCLV AM158439
PVM NC001361.2
NLV DQ098905
BlSCV L25658
HCMV NC012869
KLV FJ531635
LSV AJ516059
PLV DQ455582
LNRV NC010305
PVS NC007289
PVP EU338239
CVB AB245142
PhVB NC00991
PhVS NC009383
DVS AJ620300
PLV NC011525
CVNV EF527260
BMV NC013527
GarCLV Z68502
GLV NC003557
RCVMV NC012210
HNNV NC012038
PVM NC005343
NVS AM182569
SPCFV NC006550
CPMMV:Ghana:cowpea NC014730.1
CPMMV:BR:GO:01:1 soybean KC88424
CPMMV:BR:BA:02 soybean KC884247
CPMMV:BR:MT:02:1 soybean KC88424
CPMMV:BR:MG:09:2 soybean KC88424
CPMMV:BR:MG:09:3 soybean KC88424
CPMMV:BR:GO:10:5 soybean KC88424
CPMMV:Florida:bean KC774020.1
CPMMV:Florida:whiteflies KC77301
ICRSV NC003093
0.03
Clade 1
Clade 2
Strain 1
Strain 2
B
CPMMV:India: Peanut AF024629.1
CPMMV:India: Peanut AF024628.1
CPMMV:Argentine:Chia KP402890.1
CPMMV:BR:PA:02 soybean KF554112.1
CPMMV:BR:MG:09:15 soybean KF554110.1
CPMMV:BR:MG:09:6 soybean KF554106.1
CPMMV:BR:MG:09:16 soybean KF554111.1
CPMMV:BR:MG:09:11 soybean KF554108.1
CPMMV:BR:MG:09:7 soybean KF554107.1
CPMMV:BR:MG:09:4 soybean KF554104.1
CPMMV:BR:BA:02 soybean KC884247.1
CPMMV:BR:MT:02:1 soybean KC884246.1
CPMMV:BR:MA:02 soybean KF554102.1
CPMMV:Brazil:peanut DQ444266.1
CPMMV:Taiwan:bean JX070669.1
CPMMV:BR:MG:09:12 soybean KF554109.1
PMMV:Puerto Rico:cowpea GU191840.1
CPMMV:BR:MG:09:5 soybean KF554105.1
CPMMV:BR:MG:09:1 soybean KF554103.1
CPMMV:Florida:whiteflies KC774019.1
CPMMV:BR:GO:10:5 soybean KC884249.1
CPMMV:BR:MG:09:2 soybean KC884244.1
CPMMV:BR:MG:09:3 soybean KC884245.1
CPMMV:Florida:Bean KC774020.1
CPMMV:India:Bean KJ534277.1
CPMMV:Bean JX020701.1
CPMMV:India:Bean KJ534276.1
CPMMV:India:soybean JX524198.1
CPMMV:Ghana:cowpea NC014730.1
CPMMV:BR:GO:10:4 soybean KF554101.1
CPMMV:Brazi:soybean DQ885940.1
CPMMV:BR:GO:01:1 soybean KC884248.1
0.03
Clade 1
Clade 2
Strain 2
Strain 1

CHAPTER 2

Experimental Evolution of a Plant Virus Reveals Recombination-Driven Reduction in Virulence Accompanied by Increases in Diversity and Viral Fitness.

Zanardo, L. G., Trindade, T. A., Mar, T. B., Alves, M. S., Barbosa, T. M. C., Queiroz, S. S., Lima, R. R. P. N., Milanesi, D. F., Elliot, S. L., Mizubuti, E. S.G., Zerbini, F. M.,Carvalho, C. M. Experimental Evolution of a Plant Virus Reveals Recombination-Driven Reduction in Virulence Accompanied by Increases in Diversity and Viral Fitness.

Experimental Evolution of a Plant Virus Reveals Recombination-Driven Reduction in Virulence Accompanied by Increases in Diversity and Viral Fitness

Larissa G. Zanardo[1], Tiago A. Trindade[2], Talita B. Mar[1], Tarsiane M.C. Barbosa[1], Sarah .S. Queiroz[1], Murilo S. Alves[1], Roberta R.P.N. Lima[1], Diogo F. Milanesi[1], Simon L. Elliot[2], Eduardo S.G. Mizubuti[1], F. Murilo Zerbini[1], Claudine M. Carvalho[1*]

[1]Departamento de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, Brazil.

[2] Departamento de Entomologia, Universidade Federal de Viçosa, Viçosa, MG, Brazil.

Corresponding author: Claudine M. Carvalho, E-mail: claudine.carvalho@ufv.br

Short title: Recombination-Driven Reduction in Virulence in a Plant Virus.

Key words: RNA virus; virus evolution; recombination; evolution of virulence; systemic necrosis; *Carlavirus*; *Cowpea mild mottle virus*; parasite-vector interactions; *Bemisia tabaci*.

ABSTRACT

Some major themes in pathogen evolution are emergence, evolution of virulence, host adaptation and the processes that underlie them. RNA viruses are of particular interest due to their rapid evolution. The *in vivo* molecular evolution of a plant RNA virus was demonstrated here using a single necrotic isolate of *Cowpea mild mottle virus* (CPMMV) and a single soybean genotype submitted to serial inoculations. In the first experiment, the virus lost the capacity to cause necrosis after six passages. In the second, a severe bottleneck was imposed and virulence loss occurred in the second passage. The change to milder symptoms had fitness benefits to the virus (greater accumulation) and to its vector, the whitefly *Bemisia tabaci*. The systemic necrosis symptom showed features of the hypersensitive response, with its appearance dependent on co-chaperones: *GmRAR1*, *GmSGT1* and *GmMAPKKKα* in the MAPK cascade. Genetic polymorphism was highest in ORF1 (viral replicase) and was independent of symptoms. Recombination was a major contributor to this diversity – even with the strong genetic bottleneck, recombination events and hot spots were detected within ORF1. Virulence reduction was associated with different sites in ORF1 obtained through recombination, in both experiments. Overall, the results demonstrated that reduction in virulence was a consequence of the emergence of new variants, driven by recombination. Besides providing details of the biochemical and evolutionary mechanisms behind a reduction in virulence, we propose that this recombination-driven switch in virulence allows the pathogen to adapt to a new host very rapidly and, potentially, switch back.

AUTHOR SUMMARY

A common feature of emerging diseases is a highly virulent pathogen that, with time, evolves towards reduced virulence. This is in accordance with current theory that predicts a tradeoff between virulence and transmission. We examined a plant RNA virus, *Cowpea mild mottle virus*, that has emerged over the past two decades as a major threat to the world's principal oilseed crop, soybean. Some isolates of this virus cause systemic necrosis and consequent death of the host plant, which certainly appears to be maladaptive. In successive inoculations of host plants, we found that the virus switched to lower virulence, with increases in viral fitness within the plant, and in vector fitness on the plant, as a result. A comprehensive population genetical analysis revealed that recombination was the driving force behind the switch to lower virulence. In the field, the pathogen is faced with a variety of host genotypes, so we propose that recombination allows it to adapt extremely rapidly to changing local conditions. This would imply that selection towards lower virulence following emergence is not a continuous or gradual process, and even that highly virulent genotypes may be maintained in the population if they are adaptive under certain conditions.

INTRODUCTION

Last century, Frank Fenner collected decades of data showing evolutionary changes in the virulence of the myxoma virus that had been imported to Australia to control the European rabbit. In that instance, virulence was initially high (this is why the virus was selected), reduced year-on-year and then finally began to increase again, apparently towards some optimal level for the virus [1]. Such new associations between host and parasite (the myxoma virus was isolated in South America) may also arise where the parasite is invasive,

and this can be a major factor contributing to the problem of emergent diseases. Where virulence is initially high, it may be maladaptive, hence the parasite's evolution towards lower virulence (the same applies to a maladaptive low virulence, but these cases usually attract less attention) and in fact the pattern of selection on virulence is expected to change with time [2]. The evolution of virulence has become a major theme in evolutionary ecology and much theory has been developed, based principally upon a supposed trade-off between virulence and transmission [3, 4]. It has also been the subject of a number of studies in experimental evolution, these usually supporting the idea of some optimal level of virulence [3, 5]. Quite often, however, the underlying mechanisms, that could be considered to be genetic, biochemical or physiological, are under-investigated. There are notable exceptions, however, such as a recent study in which a bacterial pathogen of the nematode *Caenorhabditis elegans* evolved towards avirulence in selection experiments, and the authors were able to identify the genetic changes underlying this transition [6].

A number of experimental evolution studies have also been conducted with RNA viruses [7], because variability in such viral populations tends to be high. RNA viruses have rapid replication, short generation times and large effective population sizes that, associated with the lack of proofreading and high mutation and recombination rates, can generate highly polymorphic populations [8]. These factors can lead to rapid within-host evolution [9], adaptation to new niches and hosts, breakdown of host resistance [10] and changes in virulence [11, 12]. High mutation rates, in association with recombination events, offer a range of possibilities that allow viral populations to explore a diverse "genotypic landscape" quickly and to find the most beneficial genotype [13]. Recombination contributes by creating potentially advantageous genotypes and removing deleterious

mutations, boosting the process of emergence. Rates of recombination depend on the prevalence of multiple infections and the ability of the viral replicases to undergo template switching. In summary, the maintenance of genetic variation is highly dependent on mutation and recombination rates, on the distribution of mutational effects on viral fitness and on the strength of selection and genetic drift within populations [14].

The carlavirus *Cowpea mild mottle virus* (CPMMV, family *Betaflexiviridae*, genus *Carlavirus*) is a single-stranded positive-sense RNA virus (~ 8.2 Kb) [15, 16] that has posed a major threat to soybean (*Glycine max*) production in Brazil as it causes the disease soybean stem necrosis. As the name implies, this disease can be highly virulent – the necrosis is systemic and can lead to plant death in 14 days [16]. Curiously, though, there is considerably variability in disease symptoms in a given field, ranging from systemic necrosis to mild mosaic or even asymptomatic plant infection [16-18]. It has been shown that there are two strains of CPMMV in Brazil: CPMMV-BR1 and BR2, but both strains include isolates that cause mild (mosaic) or severe (systemic necrosis) symptoms in a given soybean genotype [18]. While there is clearly a genetic basis for the variations in virulence observed in this system, there has been little indication in the studies conducted to date as to the underlying reasons for this variation.

CPMMV is unusual among carlaviruses in that it is vectored (in a non-persistent manner) by the whitefly *Bemisia tabaci* MEAM 1 [19-22]. This insect is itself an important invasive organism [23] and is notorious for its polyphagy. This polyphagy is understood to be a major factor contributing to the high diversity of plant begomoviruses [24] and may well be important also to the evolution of the CPMMV. While CPMMV was originally described

from cowpea (*Vigna unguiculata*) in West Africa, it has a broad host range encompassing many Fabaceae and the Brazilian strains are quite distinct from the type strain [16, 18]. It seems likely that CPMMV has had to adapt to a variety of different hosts. Akin to Burdon and Thrall's geographical mosaic of coevolution [25], a mosaic of hosts would be expected to contribute considerably to pathogen diversity. In this context, a suitable hypothesis is that the extremely high virulence of some CPMMV isolates to some host genotypes is essentially maladaptive. This could be understood as maladaptively high virulence due to CPMMV being an invasive pathogen in South America, or even a new association having arisen with soybean (both of which represent a novel pattern of selection on CPMMV as an invasive pathogen).

The present study considers a highly virulent isolate of the virus and a susceptible genotype of the host. In selection experiments, repeated passages (mechanical inoculations) of the virus were run through the host with or without the imposition of a genetic bottleneck. After several passages, the necrotic symptoms caused by the virus switched to milder, mosaic symptoms. Viral fitness (accumulation) was examined within the host plant before and after the change in symptoms, and similarly how this change affected performance of the insect vector on the plant. Sequencing of multiple viral genomes before and after the phenotypic change allowed a detailed analysis of the relative contributions of mutation and recombination to the observed changes, as well as the pattern of selective pressure on different ORFs.

RESULTS AND DISCUSSION

Serial Passages Lead to Reduced Virulence

We conducted serial inoculations (or passages) of an isolate of CPMMV in soybean, in three replicated lineages. Infection was by mechanical inoculation and we used only symptomatic plants to inoculate each subsequent generation (the experimental design is explained in detail in Material and Methods and in Fig 1A). Although the number of plants that were successfully infected varied during the first passages (P3, P4 and P5, Fig 1A), all of the infected plants suffered from necrosis – interestingly, these necrotic symptoms were much less extensive in P5 than in the previous passages (Fig 1B). In the subsequent passage, P6, the phenotype changed in all three lineages simultaneously, with infected plants now showing mild symptoms characterized by typical CPMMV mosaic with vein clearing, while only one plant of the 23 infected plants was necrotic (Fig 1A, B).

We chose to repeat the above experiment for the sake of validation, but also submitting the viral population to a genetic bottleneck. For this, we first inoculated plants of *N. benthamiana* to obtain local lesions [16]. As one or a few viral particles are likely responsible for such local lesions, this represents a very strong genetic bottleneck. These local lesions served as inoculum to infect three soybean plants. At 40 dpi, one of these showed necrotic symptoms (Fig 1C, D), although not as extensive as in most of the passages in the previous experiment, while the other two showed no signs of infection. This single plant was used to infect three new soybean plants, as in the first experiment (Fig 1C). Strikingly, these three soybean plants showed mosaic symptoms after 14 dpi (Fig 1C, D). Thus, the switch from severe, necrotic, symptoms to milder mosaic symptoms was repeated in a second experiment, but much more swiftly, presumably due to the imposition of a

genetic bottleneck. We now sought to investigate the fitness consequences of this change, the biochemical and genetic aspects of the systemic necrosis, and the molecular evolutionary processes underlying the phenotypic change.

Viral Accumulation and Vector Performance Increase once CPMMV Becomes Less Virulent

Systemic necrosis is very damaging to the host plant, usually resulting in rapid death. As CPMMV and its whitefly vectors depend upon their host for reproduction and subsequent transmission (or dispersal in the case of the insect) to new hosts, the high virulence that systemic necrosis represents may well be maladaptive for the virus. We thus wished to examine both viral accumulation and vector performance in plants suffering necrosis in comparison with plants exhibiting mosaic symptoms. The first experiment allowed this and in particular allowed a comparison between populations of the same viral isolate before and after it lost its capacity to cause necrosis.

We used real time PCR to determine viral accumulation and thereby estimate within-host fitness [26, 27] in P1, P3 and P6. Viral accumulation in P6 plants (with mosaic) was higher than in P1 and P3 plants (with necrosis) (Fig 2A), while there was no difference in accumulation between P1 and P3, in both of which passages the virus still caused necrosis. This is probably a direct consequence of a high fitness cost to the pathogen resulting from high virulence [28]. Our results with CPMMV are similar to a study of *Soybean mosaic virus* (SMV) in soybean, in which an increase in virulence (the occurrence of necrosis symptoms) resulted in reduced fitness [29]. It is also in line with ecological evolutionary theory, whereby high virulence may ultimately reduce transmission (i.e. fitness) due to negative effects on the host [3, 30].

Considering these results, the maintenance of the virulent isolates in nature would require that they be transmitted during the early stages of the infection. Although nymphal stages of the whitefly are almost entirely sessile, adults are highly motile and take flight at the slightest disturbance. As the virus is transmitted in a nonpersistent fashion, this may facilitate transmission even if the virus is lethal to the plant. Meanwhile, less virulent isolates could be transmitted at early or late stages and these could even represent distinct strategies for transmission. It has been shown that a highly virulent isolate of *Cucumber mosaic virus* (CMV) in Spain, that caused necrosis in tomato, required high transmission rates ensured by a high density of its aphid vector [31].

The effects of viral infection on a plant can have knock-on effects on the performance of vectors and ultimately viral transmission [32-37]. These effects can be positive or negative, and can be observed in important components of the insect's fitness, such as survival, developmental time, fecundity or oviposition [38, 39, 40, 41]. To examine whitefly performance, we first infected soybean plants with inoculum taken from P2 and P5, which resulted in P3 (necrotic) and P6 (mosaic). A third treatment consisted of whiteflies on a completely different CPMMV isolate that causes mosaic symptoms naturally. Finally, there was a blank-inoculated control. There were significant differences in survival of whitefly nymphs among treatments (LME, Likelihood ratio test = 24.2, d.f. = 3, $P < 0.0001$; Fig 2B). Survival on plants infected with the inoculum from P6 and the mosaic strain did not differ significantly from one another or from the control, but survival on P3-infected plants was considerably lower (over 80% of insects had died by day 5, *versus* less than 5% for the other treatments) (Fig 2B). Meanwhile, no nymphs were able to develop on the P3-infected

plants, while this variable was otherwise unaffected by the others treatments (LME, Likelihood ratio test = 5.62, d.f = 3, P = 0.06; Fig 2C).

These results show a clear ecological and evolutionary cost to the virus of inducing systemic necrosis, since the virus is dependent upon whitefly for its transmission to new host plants. It may also be that necrotic symptoms themselves are not adaptive, but represent excessive (maladaptive) virulence on host genotypes to which a given strain is not adapted. Yet the conclusion remains that systemic necrosis does not represent an adaptive trait. Given a patchwork of different soybean varieties in the field, in which CPMMV strains may cause different symptoms, one might expect a dynamic and constantly evolving situation in which a strain that has adapted to one plant genotype infects a new genotype in which it causes necrosis; here, it is selected for reduced virulence, that might then affect its capacity to infect the first genotype successfully. While this is speculative at present, it might warrant further investigation.

*Systemic Necrosis Has Features of the Hypersensitive Response (*HR)

We next turned our attention to the biochemical pathways underlying the systemic necrosis symptoms. We examined biochemical features and gene expression in plants from P3 (with necrosis) and P6 (without necrosis). A characteristic brown colour, indicating H_2O_2 accumulation, was detected in leaves from P3 (necrosis), but not from P6 (mosaic) (Fig 3A).

The genes *GmPR1, GmPR4* and *GmNAC6* are typical defense-related genes that are expressed during the HR [42, 43]. Expression analyses of these three genes showed that

229 they are induced in soybean plants from P1 and P3, but not in P6 (Fig 3B). Similarly, RAR1
230 and SGT1 are co-chaperones of Hsp90, and these proteins have an important role in
231 regulation of the plant immune response [44]: the genes *GmSGT1* and *GmRAR1* were
232 induced in plants inoculated with all viral populations (P1, P3 and P6), irrespective of the
233 symptoms caused (Fig 3C). It can thus be seen that CPMMV induces the defence response
234 pathway activated by RAR1 and SGT1, independent of HR. This means that the plant
235 defense response against the pathogen has been activated, but the pathways, that lead to
236 plant cell death (PCD) or mosaic symptoms are different.
237
238 Next we examined *GmMAPKKKα* and *GmMEKK2*, genes from the MAPK cascade. This
239 cascade plays an important role in many R-gene-mediated defence responses to plant
240 pathogens and has been proposed to act downstream of the elicitor recognition step that
241 involves SGT1 and RAR1 [45]. In plants with systemic necrosis (P1 and P3),
242 *GmMAPKKKα* and *GmMEK2* were induced (Fig 3D). In contrast, only *GmMEK2* was
243 induced in plants showing mosaic (P6). These results suggest that *GmMAPKKKα*
244 expression is necessary for induction of systemic necrosis in CPMMV infection, but
245 *GmMEK2* is not.
246
247 Overall, the systemic necrosis caused by CPMMV infection had several features of HR.
248 Among those common features, we can include ROS and PCD, the induction of the
249 pathogenesis-related genes *GmPR1* and *GmPR*4, and *GmNAC6*, and induction of a part of
250 the MAPK cascade. In future studies, it will be interesting to link this information with the
251 fitness effects observed above.
252

ORF 1 Has High Genetic Polymorphism and is The Region Involved in Population Subdivision

Since the soybean genotype used for the serial passages was the same, it seems reasonable to suspect the emergence of viral variants with higher fitness in both experiments. For the first experiment, we amplified 10 full genomes from each of three passages (P1, P3 and P6) and from the three distinct selection lineages. For the second (bottleneck) experiment, 40 full genomes were amplified, 10 from each passage (BP1 and BP2) and repetition (a, b and c). For all genomes amplified, the six typical carlavirus ORFs were detected: ORF 1 encoding the putative viral replicase; ORFs 2, 3 and 4, encoding the triple gene block proteins (TGBp1, 2 and 3), which are essential for virus movement; ORF 5 encoding the coat protein (CP, 34 kDa); and ORF 6 encoding a cystein-rich protein (CRP) with a zinc finger motif, acting as a silencing supressor [16, 46]. The degree of nucleotide identity for all coding regions ranges from 93 to 100% for nt in the first experiment and from 98 to 100% for nt in the second (S1 and S2 Fig).

For both experiments the genetic variability in all coding regions was evaluated (Table 1 and S1 Table). The sequences from each passage (P1, P3, P6, BP1 and BP2) were assumed to correspond to populations and each experiment was analyzed separately. ORF1 was the most variable region of the CPMMV genome for all populations in both experiments (Table 1 and S1 Table). There was an increase in genetic polymorphism with each passage in the first experiment (Table 1), that was observed only for ORF1 in the second (S1 Table).

Analysis of the first experiment showed that the mean pairwise number of nucleotide differences (π) for ORF1 increased threefold from P1 to P6 (0.017 to 0.051) (Table 1). In the bottleneck experiment, ORF1 showed very small π values that were very similar for BP1 (0.00435) and BP2 (0.00463) (S1 Table). The population-scaled mutation rates (θ-w) ranged from 10^{-4} - 10^{-3} for ORFs 2-6, but for ORF1 this value was 10^{-2} for all passages in the first experiment (Table 1). For the to bottleneck experiment the values of θ-w were close to 10^{-3} for all ORFs except for ORF3 for which was 10^{-2} (S1 Table).

The π value in slinding windown analysis for each individual ORF was also determined (Fig 4, S3 and S4 Fig.) and ORF1 was found to have the greatest variation in both experiments (Fig 4). For the first experiment especially, four regions were observed (A, B, C and D) to be extremely variable (Fig 4A), while for the bottleneck experiment, only regions C and D were detected (Fig 4B). Regions A, C and D are within the methyltransferase, helicase and RNA-dependent-RNA-polymerase (RdRp) domains of the viral replicase, respectively (Fig 4). For the other coding regions a low amount of variation was observed in both experiments, with almost none in the bottleneck experiment (S3 and S4 Fig). Thus, the genetic bottleneck to which the initial inoculum was submitted in experiment 2 clearly affected the genetic diversity of the CPMMV population. For many plant RNA viruses, low levels of variation have been observed, indicating that accumulation of significant differences in virus populations is a slow process, especially when the population is submitted to a genetic bottleneck [47].

High levels of variation are commonly found in flexivirus ORF1 in inter-host analyses [48, 49], but this is the first time that it has been observed in an intra-host analysis. The variation analyses clearly showed that alterations in the CPMMV genome occurred during the serial inoculations, even in the population submitted to a bottleneck. To determine which genomic region could differentiate populations phylogenetic analyses using Bayesian Inference (BI) were performed for each coding region of the CPMMV genome, from both experiments.

In first experiment, the topology of the phylogenetic tree constructed for ORF1 had two clusters (Fig 4C). The first cluster included all isolates from P1 and P3 and 11 isolates from the P6 population. The second cluster only included isolates from P6 (Fig 5C). The phylogenetic trees constructed for the other coding regions showed no such separation according to passage (S5 Fig), so ORF1 was the genomic region that differentiated phylogenetically the CPMMV P1, P3 and P6 populations. Similar results were obtained in the second experiment (Fig 5D and S6 Fig). We therefore used discriminant analysis of principal components (DAPC) to identify the sites that most contributed to differentiation among the populations in both experiments (Fig 5E and F, S7 Fig.). A total of 73 sites were identified and described in four regions (A-D) of the viral replicase (Fig 5E and S2 Table) in the first experiment. The sites which most contribute to population differentiation were in interdomain regions (methyltransferase-peptidase and peptidase-helicase) and within the RdRp domain (Fig 5E and S2 Table). For the bottleneck experiment, a total of 34 sites most contributed to differentiation among the two populations and were in the interdomain region (metiltransferase-peptidase) and peptidase domain (Fig 5F and S2 Table).

ORF1 is a Recombination Hotspot

Recombination events were only detected in ORF1. For the first experiment 38 recombination events were detected (S3 Table) which occurred within at least one of the four ORF1 variable regions (A, B, C and D) (Figs. 4A and 6A). For the bottleneck experiment, four recombination events were detected (S4 Table). Of these, two were of interest: event 1 corresponded to a large recombinant block (30-3896) and included the methyltransferase, helicase and peptidase domains (S4 Table) while event 2 included a small recombinant block (2942-4054) including the peptidase-helicase interdomain and helicase domain.

We analyzed each recombination event in detail. In the first experiment eight of the 38 recombination events detected (events 13, 15, 16, 18, 25, 27, 29 and 37) were shared by more than ten isolates (S3 Table). Curiously, events 13, 15, 18, 25 and 37 were shared by isolates from the three passages (P1, P3 and P6). Of these eight events, three were of most interest: (a) event 15 was detected among the isolates from P1, P3 and the 11 isolates from P6, (b) event 16 occurred in all isolates from P1 and P3 (necrotic isolates), involving the peptidase and helicase domains, and (c) event 29 was detected only in some isolates of P6 (S3 Table). Together, events 15 and 29 can explain the clustering in the phylogenetic tree of 11 isolates from P6 in Fig 4C cluster 1. In the bottleneck experiment, event 1 only included isolates from BP1 (necrotic isolates) and event 2 only included isolates from BP2 (mild isolates) (S4 Table).

Independent phylogenetic trees were constructed, using an alignment in which all recombinant blocks were removed, so that recombination effects under genetic variability

345 could be perceived easily (S8 Fig). In both experiments the topology of the phylogenetic
346 trees changed, new clustering occurred among the isolates (S8 Fig).
347
348 Recombination rates (ρ) were also determined for the passages: P1, P3, P6, BP1 and BP2
349 (Table 2). The highest ρ value in the first experiment was in P6, and in the bottleneck
350 experiment in BP2, the two passages that were associated with changes in symptoms
351 Additionally, the recombination rate by site (ρ/site) was determined along the CPMMV
352 nucleotide sequence (Table 2). The rate ρ/site in P3 was the lowest ($1.75x10^{-2}$) while the
353 highest value was in P6 (0.6324) (Table 2) in the first experiment. In the bottleneck
354 experiment, ρ/site was highest in BP2 (0.318) (Table 2).
355
356 The relation ρ/θ was also determined to identify the contribution of recombination and
357 mutation for variability of CPMMV populations. In P1 and P6, recombination contributes
358 more to variability than mutation, while in P3 the opposite was observed. The relation ρ/θ
359 in the bottleneck experiment also showed that recombination contributes more to variability
360 than mutation (Table 2). Overall, recombination contributes to the high polymorphism in
361 ORF1. Significant higher densities of breakpoints were found in ORF1 when detectable
362 breakpoint positions were plotted on a density map, suggesting that ORF1 is a
363 recombination hotspot (Fig 5). Recombination seems to be responsible for much of the
364 variation found in ORF1 in both experiments.
365
366 *Some Codons Are Under Positive Selection in the Replicase Gene, But Adaptive Selection is*
367 *Not Responsible for Maintaining the Diversity in ORF1*

Selection analyses were performed on each ORF from the CPMMV genome using neutrality tests (Table 3 and S5 Table) for both experiments. In the first experiment, all values were negative, except for ORF1 in P6, and statistically significant deviations from neutrality were observed for different populations and genomic regions (Table 3). Curiously, ORF5 was the only region where all values were negative and statistically significant (Table 3). In ORF1 of P6, the tests were non-significant, so the high polymorphism was not being maintained by selection (Table 3). The same was observed with the bottleneck experiment for BP1 and BP2. The selective constraints in all ORFs from BP1 showed negative values, which were significant only in ORF3 (S5 Table). In BP2, all ORFs had significant negative values except for ORF4 (S5 Table). Thus, for BP2 there is evidence of strong purifying selection or recent population expansion – because the diversity in BP2 was, in general, smaller than in BP1 and the original population was derived from the genetic bottleneck.

Site-by-site selection analyses were also performed for each genomic region in P1, P3 and P6 in the first experiment and in BP1 and BP2 in the bottleneck experiment, using five algorithms: SLAC, FEL, IFEL, REL and PARRIS (Table 3 and S5 Table). In the first experiment, evidence of negative (purifying) selection was observed for most codons under selection in ORF2-6 by all algorithms (Table 3 and S6 Table). For ORF1 there were codons under positive selection, but many were also under negative selection, especially in the P6 population (Table 3 and S6 Table). Strong evidence of positive selection in individual codons in ORF1 from P6 was detected for SLAC and PARRIS (Table 3 and S6 Table). For the bottleneck experiment the site-by-site selection analyses showed that the majority of the codons under selection were subject to negative selection in all ORFs from both

populations (BP1 and BP2) (S7 Table). Sites under positive selection were found in ORF1 (BP1 and 2) and in ORF6 in BP2 (S7 Table). However, the evidence of positive selection was weak; the codons were detected only by IFEL and/or REL (S7 Table).

The codons corresponding to the sites responsible for population subdivision were compared with the codons under selection in the populations to identify coincidences in each experiment. Many sites involved in population subdivision in the first experiment were under selection (Fig 4E, S2 and S6 Tables), but in contrast with the first experiment, no coincidence was identified in the bottleneck experiment (Fig 4E, S2 and S7 Tables).

Together, these results indicate that adaptive selection is not acting to maintain the genetic variability of CPMMV populations.

The Sites Associated with Phenotypic Change Differ Between Experiments 1 and 2, but are Propagated Across Extant Population by Recombination

To try to associate the phenotypic change over successive inoculations with genotypic changes in ORF 1, association and localization tests were performed using the program ALTree [50]. The 70 sequences generated in the first experiment were used in analyses and 67 haplotypes were detected, 39 belonging to the control class (H01-H39, necrotic isolates) and 28 to the case class (H40-H67, mild isolates). In the association test, the phylogenetic tree showed 13 levels of which four (10-13 levels) were significant (data not shown). The tree global *P-value* was calculated and showed a significance value of 1.9998×10^{-4}, showing an association between phenotype and genotype changes (data not shown).

The localization analysis was based on character S that represented the phenotype status (necrotic or mild). Each haplotype was classified as a phenotype state using the character S. An evolution index (Vi) was calculated between each change of each site in ORF1 and the changes of the character S in haplotypes. The sites whose changes were the most correlated to character S were considered as involved in phenotype change. A total of 27 sites were found involved in the phenotype change from necrosis to mosaic (Table 4) in the first experiment. The 27 sites were compared with the sites responsible for population subdivision and 17 were putatively under positive selection (Table 4, S2 and S6 Tables). Interestingly, most sites involved in altering phenotype were located in recombinant blocks involving a number of isolates (events 13, 15, 16, 18, 20 and 25) and after the recombination hotspot 1 (Fig 5 and 6, Table 4 and S3 Table). This suggested that the recombination events, in association with mutation and selection action are involved with the phenotype changes.

The association and localization tests were repeated to sequences from BP1 and BP2 to check if the same sites associated with the phenotypic change in the first experiment would be identified again in the bottleneck experiment. For the association test, we detected 40 haplotypes (from 40 sequences generated), 10 belonging to the control class (H01-H10, necrotic isolates) and 30 to the case class (H30-H40, mild isolates). The phylogenetic tree of association test showed eight levels of which four (5-8 levels) were significant (data not shown). The global *P-value* of the tree was significant at 9.9999×10^{-5}, so there was an association between the phenotype and genotype changes.

In localization analyses, a total of 36 sites were associated with the phenotype change from necrosis to mosaic (Table 4) in the bottleneck experiment. All the sites identified coincided with the sites responsible for population subdivision and none of them are under selection (Table 4 and S6 Table). All these sites were located in recombinant block from event 1 in ORF1 (Fig 6, Table 4 and S7 Table).

None of the sites identified as responsible for the alteration in phenotype by the association test in the first experiment were identified in the second experiment (Table 4). However, all sites identified were located on recombining regions, supporting the likelihood that mutations are being propagated across extant populations by recombination (Fig 6). The results suggest that recombination events within CPMMV ORF1 were responsible for the phenotypic changes observed after successive inoculations, for distributing mutations among individuals in the population. Experiments involving infectious clones and targeted mutation sites may be used to evaluate the induction of symptoms and viral fitness.

It was expected that the change in symptoms caused by CPMMV would be associated with mutations in one or a few sites that altered one or a few amino acids. Our expectations were based on studies showing that the mutation of a single nucleotide causing an amino acid change may be enough to cause drastic changes in the phenotype [51, 52]. We can see in this study that the phenotypic changes in the viral population are governed by a much more complex virus-plant interaction network than might have been expected. In principle, changes are not determined by one or a few mutations, but by a set of them and the easiest and fastest way to acquire these beneficial mutations are by recombination [14].

In the present study, a set of mutations were acquired by different variants of CPMMV through the infections and were propagated via recombination in the serial plant inoculations. Having higher fitness, the frequencies of these variants increased in the population, and could be maintained in the population by not being recognized by the plants' defense mechanisms. Transposing this situation to the field, these new mild variants, accumulating to higher levels within the plant and being more beneficial for the insect vectors, would predominate. This local, or host, adaptation, might however render the variant less well adapted to another host genotype. It would be interesting to determine if the now-mild variant has lower fitness than its ancestor in a different host. If so, we may observe a reversal of the evolutionary change shown here, back towards the more virulent genotype. Given these considerations, we propose that the recombination events that we observe in CPMMV have actually been selected for, as they would permit rapid adaptation to different hosts.

CONCLUSIONS

We have shown that the high virulence of an isolate of CPMMV (its capacity to cause systemic necrosis) evolves to lower virulence (causing milder symptoms such as mosaic) in a consistent pattern across separate lineages and in two experiments. Our study shows that this could be due to local and swift evolutionary processes, even within a single isolate in the case of CPMMV, and potentially for other plant viruses. Clearly, several sites localized within recombinant blocks were associated with the change in symptoms. These alterations in the viral genome generated variants with higher fitness and also allowed the vector to increase its fitness.

MATERIAL AND METHODS

Selection Experiments – Experimental Design

Two selection experiments were conducted, using the same isolate of CPMMV, CPMMV:BR:GO:10:5, access number KC884249.1. This was collected in January 2010 from soybean in Cristalina, Goiás, central-western Brazil (one of the main soybean-producing regions in Brazil) and has previously been described both molecularly and in terms of its host range [16].

To initiate the first experiment, sample leaves were used to mechanically inoculate three plants of soybean cv. CD206 at VC stage (corresponding to 14 days from sowing). Inoculum was obtained by grinding the material in 0.1M phosphate buffer, pH 7.2, with 0.1% sodium sulfite. All inoculated plants in this study were maintained in gauze cages within a greenhouse (means daily temperature of 26 ± 2°C) and were sprayed weekly with insecticides to avoid any possible contamination. All of these plants showed systemic necrosis (of stem and petiole) plus dwarfism after 14 dpi (Fig 1). One of the plants (first passage-P1) was used to inoculate three more plants (P2), and these showed the same symptoms after 14 dpi (Fig 1B). Each of these three plants was then used to inoculate twelve more plants (P3) and these were then considered three separate repetitions. The plants of each repetition with visible signs of infection (necrosis in all cases) after 14 dpi were collected, pooled and used for inoculation of 12 new plants in each repetition. This was done until P6, with phenotypes (i.e. disease symptoms) being recorded and material being stored as described above throughout.

For the second experiment, *Nicotiana benthamiana* plants were mechanically inoculated as above. The local lesions observed in *N. benthamiana* plants were excised and used to inoculate three soybean plants (as above), thus imposing a genetic bottleneck on the virus population. At 40 dpi, only one of these three plants showed necrotic symptoms. Leaf tissue from this this plant was used to inoculate three new soybean plants. Again, these plants were observed for symptoms and material was stored as described above.

Real Time PCR Quantification of Viral Accumulation (Selection Experiment 1)

RNA from plants from P1, P3 and P6 was used for cDNA synthesis (three replicates in each of the three selection lineages bar P1 – just the three replicates) in addition to three independent healthy soybean plants as control. We used 500 ng of total RNA, extracted using the RNeasy Plant Mini Kit (QIAGEN), ORF6R primer [16] and Superscript III reverse transcriptase (Invitrogen), according to the manufacturer's protocol. All procedures for real-time PCR, including tests, validation and experiments, were conducted according to the recommendations of Applied Biosystems, using specific primers (Table S8). All quantifications were performed on StepOnePlus instrument (Applied Biosystems) using SYBR Green PCR Master Mix (Applied Biosystems). This was done using individual tubes and specific primers for ORF5 (CP coding region; Table S8). Cycling conditions were 10 min for 94°C, 40 cycles of 15s at 94°C and 1 min at 60°C. All samples were submitted to thermal denaturation to determine their dissociation curve (melting) and to verify the specificity of amplifications. The standard curve was obtained using increasing amounts of PCR amplified for the CPMMV CP gene. Means were compared using Student's *t* test in Excel.

Vector Performance (Selection Experiment 1)

Bemisia tabaci were obtained from a colony established in 2008 at the Federal University of Viçosa, Minas Gerais, Brazil. The insects were reared on cabbage (*Brassica oleracea* var. capitata) in gauze cages under greenhouse conditions. The whiteflies were identified as *B. tabaci* MEAM1 by partial sequencing of the mitochondrial cytochrome oxidase 1 gene (mtCOI) and comparing with three reference sequences of this species/group in GenBank (access number AF340215, AJ510071 and AJ510079).

For the assays, soybean plants were inoculated (as above) with virus from P2, resulting in P3 (causing necrotic symptoms) and P5, resulting in P6 (causing mosaic symptoms). As a form of positive control, a third set of plants was inoculated with a separate isolate that causes mosaic symptoms in soybean cv. CD206 (CPMMV:BR:GO:01:1; [16]) and as a negative control, further plants were treated only with carborundum and buffer. Plants showing symptoms were used 8-11 days after inoculation and CPMMV infection was subsequently confirmed by RT-PCR (Zanardo et al., 2014a).

Newly emerged nymphs (<24h) were used to evaluate survival and developmental times. Nymphs were obtained by allowing adults to oviposit on healthy soybean plants for 24 h and were then collected with a fine paint brush and placed on the abaxial surface of the first and second trifoliate leaves of the test plants. Ten nymphs were placed per plant, five on each trifoliate leaf. Nymphs that died after 24 h were replaced by others of the same age. Each treatment consisted of 30-40 nymphs placed on 3-4 plants. The experiment was carried out in randomized blocks in a room at 26±2°C and 12:12 photoperiod. Survival and development were recorded daily under stereomicroscope until emergence of adults or

death of nymphs. Dead nymphs were identified by their dark-color and dried appearance. In order to eliminate handling effects on mortality, only nymphs that survived more than 3 days after placing them on the plants were used for survival analyses.

Survival and development data were analyzed using a linear mixed effects model (nlme package of R) [53]. Replicates were used as random factor and the treatments as fixed factor. We started the analyses by including interactions of factors and checking residual distributions. Non-significant interactions and factors were removed from the full model using the ANOVA function. Contrasts among treatments were assessed with the glht function (lsmeans package of R).

Staining Assay 3,3'-diaminobenzidine (DAB) (Selection Experiment 1)

To detect H_2O_2 in necrotic leaf tissues, a staining assay was performed using DAB, according to [54, 55] with modifications. Nine leaves, three from each of three different plants from the third and sixth passage were collected and submitted to staining with DAB solution, at first for 5 minutes (1mg/ml DAB with 10mM Na_2HPO_4 and 0.5µl/ml of Tween20, PH 7.5) using vacuum infiltration and then in the dark for a further 8 hours in the dark, then submerged in a bleaching solution (ethanol:acetic acid:glycerol - 3:1:1) and finally immersed in a boiling water bath (~90-95°C) for 30 minutes. The sample leaves were maintained in 10mM Na_2HPO_4 buffer for subsequent analyses. H_2O_2 was visualized as a reddish brown color.

Quantification of Gene Expression (Selection Experiment 1)

Quantification of gene expression was used with the same total RNA samples as described above and the same protocol for cDNA synthesis and qPCR. In place of the ORF6R primer, oligo (dT) was used. Relative quantification was performed in individual tubes for *GmSGT1*, *GmRAR1*, *GmMAPKKKα*, *GmMEK2*, *GmPR1*, *GmPR4* and *GmNAC6*. All samples were submitted to thermal denaturation to determine their dissociation curve (melting) and to verify the specificity of amplifications. The expression of each gene was tested in three replicates in each of the three selection lineages. Initial tests were performed on serial cDNA dilutions of $10^{0,}$ 10^{-1}, 10^{-2} and 10^{-3} and the dilution with the best efficiency, according to the $E=10^{-1}$/slope equation, was selected. The slope angle was obtained from regression of the amplification reaction. For quantification of gene expression, *GmHelicase* was used as a reference gene to normalize the real-time PCR data. The relative expression and statistical analyses were done using the REST2009 software (https://www.qiagen.com/), according to the manufacturer's instructions.

RT-PCR and CPMMV Cloning

Amplification and cloning of viral genomes was performed from plants of all replicate lineages from P1, P3 and P6 in the first experiment and BP1 and BP2 in the second experiment. Total RNA was extracted from 100 mg of leaf tissue of individual or pooled soybean plants cv. CD206 (14 dpi), depending on the number of plants infected in each repetition and experiment, using the RNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. Reverse transcription (RT) was performed using Superscript III reverse transcriptase (Invitrogen), according to the manufacturer's protocol, using 500 ng of total RNA. PCR amplifications were carried out using Platinum high fidelity DNA polymerase (Invitrogen). The primers and RT-PCR amplifications conditions were as

605 described by Zanardo *et al.* (2014a), except for the 5'-end, for which the primer ORF1F
606 (GAAAAACCCG) and ORF1R (TCTCGTTAGCTGAGGGTT) were used. Amplified PCR
607 products were gel-purified using the Illustra GFX PCR DNA and Gel Band Purification Kit
608 (GE Healthcare), ligated into pGEM-T Easy Vector (Promega) and transformed into
609 *Escherichia coli* DH5α cells. Ten different clones of each amplified region were selected.
610 Plasmid DNA was purified using the Illustra^{TM} plasmid Prep Mini Spin Kit (GE healthcare)
611 and sequenced in both directions with universal primers (M13F/M13R) by Macrogen
612 (Seoul, South Korea).
613
614 *Sequence Assembly, Identification and Alignment*
615 Sequences were assembled using DNA BASER SEQUENCE ASSEMBLER v. 3.5 (Heracle
616 Biosoft) and the ORFs were located using ORF FINDER
617 (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The nucleotide sequences were initially
618 submitted to a BLAST search to check the sequence identification. Nucleotide sequences of
619 each experiment were analyzed independently. Each ORF was aligned using the Muscle
620 module [56] in MEGA v.5, based on amino acid sequences.
621
622 *Description of Molecular Variability and Divergence Among Sequences*
623 Descriptors of molecular variability were estimated using the DnaSP software v. 5.10 [57],
624 giving (i) the total number of segregating sites (s); (ii) G+C content; (iii) mean nucleotide
625 differences between sequences (k); (iv) nucleotide diversity (π); (v) number of mutations;
626 (vi) number of haplotypes (h); (vii) haplotype diversity (Hd) and (viii) Watterson's
627 estimator for the population-scaled mutation rate based on the total number of segregating
628 sites (θ-W).

629

630 The π statistic was also calculated using a sliding window of 100 bases, with a step size of
631 50 bases to estimate the nucleotide diversity throughout the section of the genome running
632 from ORFs 1 to 6. Confidence intervals (CI) of 95 % were estimated for π values based on
633 coalescent simulations (10,000 replicates) using DnaSP v. 5.10.

634

635 *Bayesian Phylogenetic Analyses*

636 Phylogenetic trees were constructed for each of the six ORFs individually (ORF1-6).
637 Phylogenetic relationships were inferred using Bayesian inference (BI) in MrBayes v.3.0
638 [58], with the evolution models selected by MrModeltest v.2.2 [59] using the Akaike
639 information criterion (AIC). Two runs with four Markov Chain Monte Carlo (MCMC)
640 simulations were conducted simultaneously using 20 million generations, except for ORF1
641 of the first experiment (50 million generations), starting from random initial trees. Burn-in
642 was set at 25% from the resulting trees. Trees were visualized with FigTree version 1.3.1
643 (http://tree.bio.ed.ac.uk/software/figtree/) and midpoint rooted.

644

645 *Discriminant Analysis of Principal Components (DAPC)*

646 Multivariate statistical analysis using Discriminant Analysis of Principal Components
647 (DAPC) was performed in the package adegenet implemented in R software [60] using
648 nucleotide sequences from ORF1 of both experiments. The DAPC method is intended to
649 describe the diversity between pre-defined groups of observations. We considered that we
650 had three different populations: population 1 (P1), population 2 (P3) and population 3 (P6)
651 for the first experiment and two different populations: population 1 (BP1) and population 2
652 (BP2) for the second experiment. The algorithm runs on a transformation of the raw data

using Principal Component Analysis (PCA). All of the principal components were retained in order to conserve all of the variation in the original data. The key nucleotides involved in the cluster differentiation were determined and plotted using a threshold of 0.02 in the first experiment and 0.002 in the bottleneck experiment in each discriminant function of DAPC.

Recombination Analyses

Detection of potential recombinant sequences were performed using Recombination Detection Program (RDP) v. 4.7 [61] incorporating the recombination-detecting methods RDP (R), Geneconv (G), Maximum chi square (M), Bootscan (B), Sister scan (S), Chimaera (C) and 3 SEQ (3S). Default settings and a multiple comparison-corrected P-value cutoff of 0.05 were used throughout. Only those recombination events detected by more than four methods were considered credible. The LDhat package in the RDP program was also used to estimate the recombination rate (ρ), determine the recombination rates by site (ρ/site) and estimate the relation of recombination rate by mutation rate (ρ/θ). A breakpoint density plot was constructed and the statistical significance of potential breakpoint hot- and cold-spots was tested using a permutation test with 10,000 permutations, windows size of 200 and step size of 100, with Confidence Intervals (CI) of 95 and 99% in DnaSP v. 5.10. This test allows determination of whether the breakpoint distribution was significantly non-random.

Selection Analyses

To test the occurrence of selection in populations three types of neutrality tests were used: Tajima's D, Fu and Li's D* and F*, using DnaSP v. 5.10 [57]. Gene and site-specific selection were also measured for each viral protein and population in both experiments.

Sites under negative and positive selection in each gene were determined using five different maximum likelihood-based algorithms: Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Internal Fixed Effects Likelihoods (IFEL), Random Effects Likelihood (REL) and Partitioning for Robust Inference of Selection (PARRIS) within the HyPhy software package (http://www.hyphy.org/) implemented in the Datamonkey server (http://www.hyphy.org/) with default conditions. The nucleotide substitution model applied was specific for each ORF and population, and was run before all analyses. Phylogenetic trees corrected for recombination were inferred by GARD (available at the Datamonkey server) and used as input for the selection analysis.

Association Test by Phylogenetic Approach and Localization Analysis

To determine the association between phenotypes (necrotic and mosaic) and genotype changes observed in ORF1 after the passages we conducted an association test and a localization analysis with ALTree program [50]. The program, standing for association detection and localization of susceptibility sites using haplotype phylogenetic trees, is used to perform nested homogeneity tests to compare distributions of affected and unaffected individuals in different clades of a given tree [50]. The objective is to detect if some clades of a phylogenetic tree are more or less enriched in affected or unaffected individuals compared with the rest of the tree. For the association test we determined the haplotypes in DnaSP v.5.10, the necrotic phenotype was chosen as control and the mosaic phenotype as case. Parsimony phylogenetic trees were inferred using PAUP v. 4.0. All tests in Altree were done at phylogenetic tree level. The *P*-value at each tree level was performed by 10,000 permutation procedures to test the difference between the observed distribution of affected or unaffected individuals and the random distribution.

ALTree was also used to localize the susceptibility sites. It includes an algorithm to identify which sites are most likely to be involved in detected association. It was used to localize the sites involved in phenotype changes from necrotic to mosaic symptoms. For this, an *S* character (supplementary character) was added for each affected (case-mosaic phenotype) and unaffected (control-necrotic phenotype) haplotype. After that, haplotypes including character *S* were reconstructed at ancestral nodes using PAUP v. 4.0. Localization of susceptibility sites was done by computing a correlated evolution index (Vi) calculated between each change of site and the changes of the character *S* in two possible directions of change. The sites whose evolution is the most correlated with the *S* character are those most likely involved in the phenotype change. All procedures were performed according to the program manual.

ACKNOWLEDGEMENTS AND FUNDING INFORMATION

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Brazil); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil); Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Brazil) and Vale S.A. S.L.E., E.S.G.M, F.M.Z and C.M.C. were in receipt of CNPq productivity grants.

REFERENCES

1. Fenner F, Fantini B. Biological Control of Vertebrate Pests: The History of Myxomatosis – An Experiment in Evolution. CAB International, Wallingford, UK. 1999.
2. Berngruber TW, Froissart R, Choisy M, Gandon S. Evolution of virulence in emerging epidemics. PLoS pathogens. 2013;9(3):e1003209. PubMed PMID: 23516359.

3. Alizon S, Hurford A, Mideo N, Van Baalen M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J Evol Biol. 2009;22(2):245-59. PubMed PMID: 19196383.
4. Anderson RM, May RM. Coevolution of hosts and parasites. Parasitology. 1982;85:411-26.
5. Froissart R, Doumayrou J, Vuillaume F, Alizon S, Michalakis Y. The virulence-transmission trade-off in vector-borne plant viruses: a review of (non-)existing studies. Philos Trans R Soc Lond B Biol Sci. 2010;365(1548):1907-18. PubMed PMID: 20478886.
6. Jansen G, Crummenerl LL, Gilbert F, Mohr T, Pfefferkorn R, Thanert R, et al. Evolutionary Transition from Pathogenicity to Commensalism: Global Regulator Mutations Mediate Fitness Gains through Virulence Attenuation. Mol Biol Evol. 2015;32(11):2883-96. PubMed PMID: 26199376.
7. Bedhomme S, Lafforgue G, Elena SF. Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. Mol Biol Evol. 2012;29(5):1481-92. PubMed PMID: 22319146.
8. Lalic J, Cuevas JM, Elena SF. Effect of host species on the distribution of mutational fitness effects for an RNA virus. PLoS genetics. 2011;7(11):e1002378. PubMed PMID: 22125497.
9. Cuevas JM, Willemsen A, Hillung J, Zwart MP, Elena SF. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. Mol Biol Evol. 2016;32(5):1132-47. PubMed PMID: 25660377.
10. Hajimorad MR, Eggenberger AL, Hill JH. Loss and gain of elicitor function of *Soybean mosaic virus* G7 provoking Rsv1-mediated lethal systemic hypersensitive response maps to P3. J Virol. 2005;79(2):1215-22. PubMed PMID: 15613348.
11. Hasiów-Jaroszewska B, Borodynko N, Jackowiak P, Figlerowicz M, Pospieszny H. Single mutation converts mild pathotype of the *Pepino mosaic virus* into necrotic one. Virus Res. 2011;159(1):57-61.
12. Tribodet M, Glais L, Kerlan C, Jacquot E. Characterization of Potato virus Y (PVY) molecular determinants involved in the vein necrosis symptom induced by PVYN isolates in infected Nicotiana tabacum cv. Xanthi. J Gen Virol. 2005;86(Pt 7):2101-5. PubMed PMID: 15958688.
13. Burch CL, Chao L. Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature. 2000;406(6796):625-8. PubMed PMID: 10949302.
14. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforgue G, et al. The evolutionary genetics of emerging plant RNA viruses. Mol Plant-Microbe Interact. 2011;24(3):287-93. PubMed PMID: 21294624.
15. Menzel W, Winter S, Vetten H. Complete nucleotide sequence of the type isolate of *Cowpea mild mottle virus* from Ghana. Arch Virol. 2010;155(12):2069-73.
16. Zanardo LG, Silva FN, Bicalho AAC, Urquiza GPC, Lima ATM, Almeida AMR, et al. Molecular and biological characterization of *Cowpea mild mottle virus* isolates infecting soybean in Brazil and evidence of recombination. Plant Pathology 2014;63(2):456-65.
17. Almeida AMR, Piuga FF, Kitajima EW, Gaspar JO, Valentin N, Benato LC, et al. Necrose da haste da soja. Série Documentos 2003;221:1-48.
18. Zanardo LG, Silva FN, Lima AT, Milanesi DF, Castilho-Urquiza GP, Almeida AM, et al. Molecular variability of *Cowpea mild mottle virus* infecting soybean in Brazil. Arch Virol. 2014;159(4):727-37. PubMed PMID: 24142270.

19. Iwaki M, Thongmeearkon P, Prommin M, Honda Y, Hibi J. Whitefly transmission and some properties of *Cowpea mild mottle virus* on soybean in Thailand. Plant Dis. 1982;66:265-8.
20. Naidu RA, Gowda S, Satyanarayana T, Boyko V, Reddy AS, Dawson WO, et al. Evidence that whitefly-transmitted *Cowpea mild mottle virus* belongs to the genus *Carlavirus*. Arch Virol. 1998;143(4):769-80. PubMed PMID: 9638146.
21. Marubayashi JM, Yuki VA, Wutke EB. Transmissão do *Cowpea mild mottle virus* pela mosca branca *Bemisia tabaci* biótipo B para plantas de feijão e soja. Summa Phytopathol. 2010;36:158-60.
22. Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. Ann Entomol Soc Am. 2010;103:196-208.
23. De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. Bemisia tabaci: a statement of species status. Ann Rev Entomol. 2011;56:1-19. PubMed PMID: 20690829.
24. da Silva SJ, Castillo-Urquiza GP, Hora Junior BT, Assuncao IP, Lima GS, Pio-Ribeiro G, et al. High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed Cleome affinis in northeastern Brazil. Arch Virol. 2011;156(12):2205-13. PubMed PMID: 22006043.
25. Burdon JJ, Thrall PH. Spatial and temporal patterns in coevolving plant and pathogen associations. The American Naturalist. 1999;153:S15-33.
26. Carrasco P, Daros JA, Agudelo-Romero P, Elena SF. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. J Virol Meth. 2007;139(2):181-8. PubMed PMID: 17092574.
27. Péréfarres F, Thébaud G, Lefeuvre P, Chiroleu F, Rimbaud L, Hoareau M, et al. Frequency-dependent assistance as a way out of competitive exclusion between two strains of an emerging virus. Proceedings of the Royal Society B: Biological Sciences. 2014;281.
28. Leach JE, Vera Cruz CM, Bai J, Leung H. Pathogen fitness penalty as a predictor of durability of disease resistance genes. Annu Rev Phytopathol. 2001;39:187-224. PubMed PMID: 11701864.
29. Khatabi B, Wen RH, Hajimorad MR. Fitness penalty in susceptible host is associated with virulence of Soybean mosaic virus on Rsv1-genotype soybean: a consequence of perturbation of HC-Pro and not P3. Mol Plant Pathol. 2013;14(9):885-97. PubMed PMID: 23782556.
30. Anderson RM, May RM. Coevolution of hosts and parasites. Parasitology. 1982;85:411-26.
31. Escriu F, Fraile A, Garcia-Arenal F. The evolution of virulence in a plant virus. Evolution. 2003;57(4):755-65. PubMed PMID: 12778545.
32. Mauck KE, De Moraes CM, Mescher MC. Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. Proc Natl Acad Sci U S A. 2010;107(8):3600-5. PubMed PMID: 20133719.
33. Mauck K, Bosque-Pérez NA, Eigenbrode SD, De Moraes CM, Mescher MC. Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses. Funct Ecol. 2012;26:1162-75.
34. Zhang T, Luan JB, Qi JF, Huang CJ, Li M, Zhou XP, et al. Begomovirus-whitefly mutualism is achieved through repression of plant defences by a virus pathogenicity factor. Mol Ecol. 2012;21(5):1294-304. PubMed PMID: 22269032.

35. Luan JB, Yao DM, Zhang T, Walling LL, Yang M, Wang YJ, et al. Suppression of terpenoid synthesis in plants by a virus promotes its mutualism with vectors. Ecol Lett. 2013;16(3):390-8. PubMed PMID: 23279824.
36. Wan G, Jiang S, Wang W, Li G, Tao X, Pan W, et al. *Rice stripe virus* counters reduced fecundity in its insect vector by modifying insect physiology, primary endosymbionts and feeding behavior. Scientific reports. 2015;5:12527. PubMed PMID: 26211618.
37. Mauck KE. Variation in virus effects on host plant phenotypes and insect vector behavior: what can it teach us about virus evolution? Curr Opin Virol. 2016;21:114-23. PubMed PMID: 27644035.
38. McKenzie CL. Effect of *Tomato mottle virus* (ToMoV) on Bemisia tabaci biotype B (Homoptera: *Aleyrodidae*) oviposition and adult survivorship on healthy tomato. Fla Entomol. 2002;85:367-8.
39. Belliure B, Janssen A, Maris PC, Peters D, Sabelis MW. Herbivore arthropods benefit from vectoring plant viruses. Ecol Lett. 2005;8:70-9.
40. Jiu M, Zhou XP, Tong L, Xu J, Yang X, Wan FH, et al. Vector-virus mutualism accelerates population increase of an invasive whitefly. PloS one. 2007;2(1):e182. PubMed PMID: 17264884.
41. Sidhu JS, Mann RS, Butter NS. Deleterious effects of Cotton leaf curl virus on longevity and fecundity of whitefly, Bemisia tabaci (Gennadius). J Entom. 2009;6:62-9.
42. Alves MS, Soares ZG, Vidigal PM, Barros EG, Poddanosqui AM, Aoyagi LN, et al. Differential expression of four soybean bZIP genes during Phakopsora pachyrhizi infection. Funct Integr Genomics. 2014;15(6):685-96. PubMed PMID: 26013145.
43. Faria JA, Reis PA, Reis MT, Rosado GL, Pinheiro GL, Mendes GC, et al. The NAC domain-containing protein, GmNAC6, is a downstream component of the ER stress- and osmotic stress-induced NRP-mediated cell-death signaling pathway. BMC Plant Biol. 2011;11:129. PubMed PMID: 21943253.
44. Takahashi A, Casais C, Ichimura K, Shirasu K. HSP90 interacts with RAR1 and SGT1 and is essential for RPS2-mediated disease resistance in Arabidopsis. Proc Natl Acad Sci U S A. 2003;100(20):11777-82. PubMed PMID: 14504384.
45. Nakagami H, Pitzschke A, Hirt H. Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci. 2005;10(7):339-46. PubMed PMID: 15953753.
46. Martelli GP, Adams MJ, Kreuze JF, Dolja VV. Family *Flexiviridae*: a case study in virion and genome plasticity. Annu Rev Phytopathol. 2007;45:73-100. PubMed PMID: 17362202.
47. Garcia-Arenal F, Fraile A, Malpica JM. Variation and evolution of plant virus populations. Int Microbiol. 2003;6(4):225-32. PubMed PMID: 13680390.
48. Moles M, Delatte H, Farreyrol K, Grisoni M. Evidence that *Cymbidium mosaic virus* (CymMV) isolates divide into two subgroups based on nucleotide diversity of coat protein and replicase genes. Arch Virol. 2007;152:705–15.
49. Alabi OJ, Martin RR, Naidu RA. Sequence diversity, population genetics and potential recombination events in grapevine rupestris stem pitting-associated virus in Pacific North-West vineyards. J Gen Virol. 2010;91(Pt 1):265-76. PubMed PMID: 19759241.
50. Bardel C, Danjean V, Genin E. ALTree: association detection and localization of susceptibility sites using haplotype phylogenetic trees. Bioinformatics. 2006;22(11):1402-3. PubMed PMID: 16595555.

51. Hirata H, Lu X, Yamaji Y, Kagiwada S, Ugaki M, Namba S. A single silent substitution in the genome of Apple stem grooving virus causes symptom attenuation. J Gen Virol. 2003;84(Pt 9):2579-83. PubMed PMID: 12917479.
52. Diveki Z, Salanki K, Balazs E. The necrotic pathotype of the cucumber mosaic virus (CMV) ns strain is solely determined by amino acid 461 of the 1a protein. Mol Plant-Microbe Interact. 2004;17(8):837-45. PubMed PMID: 15305604.
53. Pinheiro J, Bates D, DebRoy S, D S, 3.1-128 RCTnLaNMEMRpv. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-128. http://CRANR-projectorg/package=nlme 2016.
54. Thordal-Christenzen H, Zang Z, Wei Y, Collinge DB. Subcellular localization of H2O2 in plants. H2O2 accumulation in papillae and hypersensitive response during the barley—powdery mildew interaction. The plant J. 1997;11(6):1187–94
55. Daudi A, O'Brien JA. Detection of Hydrogen Peroxide by DAB Staining in *Arabidopsis* Leaves. Bio-protocol. 2012;2:1-4.
56. Edgar R. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BioMed Cent Bioinforma. 2004;5(1):113. PubMed PMID: doi:10.1186/1471-2105-5-113.
57. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19(18):2496-7. PubMed PMID: 14668244.
58. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572-4. doi: 10.1093/bioinformatics/btg180.
59. Nylander JAA. MrModeltest v2. Program distributed by the author. 2004.
60. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403-5. PubMed PMID: 18397895.
61. 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(19):2462-3.

FIGURE CAPTIONS

Fig 1. Experimental design and phenotypic results of the two selection experiments (serial inoculations) using isolate CPMMV:BR:GO:10:5 of *Cowpea mild mottle virus* (CPMMV) in soybean cv. CD206 at the VC stage (14 days of development) and the symptoms presented (14 days post-inoculation) in each passage. **A.** Experimental design for the first experiment, showing the number of plants used at each step (passages P1-6) and the creation of three separate selection lineages. **B.** Representative images of the symptoms observed. From P1 to P5, severe symptoms were observed: systemic necrosis, bud burning and dwarfism, while in P6 symptoms were mild: mosaics and vein clearing. **C.** Schematic representation of serial inoculation of the bottleneck experiment with the CPMMV:BR:GO:10:5 isolate excised from local lesions in *N. benthamiana* and re-inoculated in soybean cv. CD206 at the VC stage in two passages (BP1-BP2). **D.** Symptoms observed over the serial inoculations on each passage after 14 days of inoculation. In BP1, necrotic symptoms were observed in the leaves, while in BP2 symptoms were very mild: mosaics, leaf deformation and vein clearing. The letters a, b and c represent the repetitions. Black vases represent plants with symptoms of systemic necrosis, white vases are plants that showed no viral infection and grey vases the plants that showed mild symptoms such as mosaic.

Fig 2. Viral accumulation and vector fitness in plants following serial passages. A single isolate of *Cowpea mild mottle virus* (CPMMV) was passaged six times through soybean plants. As such, P1 and P3 represent initial passages in which the virus caused systemic necrosis in the plants, while P6 represents a passage in which viral symptoms became milder (mosaic). **A.** Viral accumulation of CPMMV in the P1, P3 and P6 passages. Letters

represent statistically significant difference in means (P< 0.05) analyzed with 95% Confidence Intervals (CI). Error bars are standard deviations. **B.** Survival of whitefly (*Bemisia tabaci*) nymphs on control plants, plants infected with a distinct mosaic strain (C+), plants infected with the necrotic strain (P3) and plants infected with the mild variant (P6). **C.** Mean developmental times of newly emerged nymphs to adulthood on control plants (healthy), plants infected with the distinct mosaic strain (C+) , plants infected with the necrotic isolate (P3) and plants infected with the mild variant (P6). ∞ means no data.

Fig 3. Systemic necrosis induced by *Cowpea mild mottle virus* (CPMMV) has characteristics of the hypersensitive response (HR). **A.** Detection of H_2O_2 (ROS) in necrotic leaf tissues by staining assay using DAB in leaves from healthy plants and plants inoculated with a CPMMV isolate that had been passaged three times (P3, in which the virus still caused necrotic symptoms) or six times (P6, in which the virus had become milder, causing only mosaic symptoms). Data are 14 days post inoculation (dpi). The upper panel shows the leaves before staining with DAB and the lower panel after staining. A characteristic brown color indicates H_2O_2 production. **B-D.** Gene expression in soybean plants from P1, P3 and P6 passages in relation to healthy plants 14 dpi. **B.***GmPR1, GmPR4* and *GmNAC6.* **C.** *GmSGT1 and Gm RAR1* and **D.** *GmMAPKKKα* and *GmMEK2.* The bars represent the average expression level in three plants from each treatment calculated and statistically analysed with the REST 2009 program. * represents treatments statistically different from healthy plants.

Fig 4. Sequences and phylogenetic analyses of ORF1 (viral replicase) of *Cowpea mild mottle virus* (CPMMV) from two serial passage experiments in soybean. **A.** Average

pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across ORF1.Sequences from a data set with 70 CPMMV sequences from the first passage experiment (dark grey line: ten sequences from P1 – first passage, light gray line: 30 sequences from P3- third passage and dotted line: 30 sequences from P6 – sixth passage. **B.** Sequences from data set with 40 CPMMV sequences from the second experiment in which a bottleneck was imposed (dark grey line: ten sequences from BP1- first passage, light gray line: 30 sequences from BP2 – second passage.**C-D.** Phylogenetic relationships of isolate of *Cowpea mild mottle virus* (CPMMV) subjected to passage experiments. Phylogenies are based on the sequences of ORF1 from a first experiment and a subsequent experiment in which a bottleneck was imposed, using Bayesian inference (implemented in MrBayes 3.1). Support for the nodes is presented as filled circles (posterior probabilities from 0.95 to 1.0). The grays bars show the grouping of isolates causing severe necrosis symptoms and the black bars show isolates causing mild mosaic symptoms. **C.** ORF1 tree using 70 sequences from the first experiment with model GTR+I+G and 50 million generations; **D.** ORF1 tree using 40 sequences from the second experiment with model GTR+I and 20 million generations; **E-F**. Contribution of ORF1 sites of an isolate of *Cowpea mild mottle virus* (CPMMV) to population subdivision following two serial passage experiments. Contributions were determined by Discriminant Analysis of Principal Components (DAPC). Heights of each bar a proportional to the contribution of the corresponding site; sites above the threshold (0.02 or 0.002) correspond the most contributing sites. The sites that were associated with the change of phenotypes in association and localization tests are listed in S2 Table. The letters A, B, C and D corresponded to the regions with higher polymorphism previously shown.

Fig 5. Distribution of recombination breakpoints within ORF1 of an isolate of *Cowpea mild mottle virus* (CPMMV) that had been subjected to serial passages in two separate experiments. **A.** ORF1 recombination breakpoints in the first experiment involving 70 sequences, which included P1, P3 and P6 (first, third and sixth passages, respectively). **B.** ORF1 recombination breakpoints in the second experiment involving 40 sequences which includes BP1 and BP2 (first and second passages respectively). All detectable breakpoint positions are indicated by small vertical lines at the top of graph. A 200 nucleotide window was moved along the alignment one nucleotide at a time and the number of breakpoints detected within the window region was counted and plotted (solid line). The upper and lower broken lines indicate 99 and 95% confidence thresholds, respectively, for globally significant breakpoints clusters. Dark and light grey areas, indicate, respectively, local 95 and 99% breakpoint clustering thresholds. Vertical black arrows indicate recombinants hot-spots. The domains in the viral replicase are showed at the top of each graph.

Fig 6. Schematic representation of recombination blocks and the sites in ORF1 associated with phenotypic changes in an isolate of *Cowpea mild mottle virus* (CPMMV) subjected to serial passages in two experiments. **A.** Recombination events detected in ORF1 of the first experiment in more than three sequences and which coincide with the position of the sites involved in changing the phenotype. **B.** Recombination event detected in ORF1 in the second experiment and which coincides with the position of the sites involved in changing the phenotype. Recombinants blocks are shown in dark gray, and were identified in RDP program v. 4.7. The position of the sites associated with the phenotype changes are indicated by small dots in the upper part of the figure.

SUPPLEMENTARY FIGURE LEGENDS

S1 Fig. Two-dimensional plot representing the percent sequence identities among the isolates from passages P1, P3 and P6 in the first experiment after the serial inoculations of the original CPMMV:BR:GO:10:5 isolate. Nucleotide sequence identities are above the diagonal. **A.** ORF1, RNA-dependent RNA polymerase (RdRp); **B, C** and **D.** ORFs 2-4, triple gene block (TGB1, TGB2 and TGB3, respectively); **E.** ORF5, coat protein (CP); **F.** ORF6, nucleic acid binding protein (NABP)**.**

S2 Fig. Two-dimensional plot representing the percent sequence identities among the isolates from passages BP1 and BP2 in the second experiment after the serial inoculations of the CPMMV:BR:GO:10:5 isolate from local lesion from *N. benthamiana*. Nucleotide sequence identities are above the diagonal. **A.** ORF1, RNA-dependent RNA polymerase (RdRp); **B, C** and **D.** ORFs 2-4, triple gene block (TGB1, TGB2 and TGB3, respectively); **E.** ORF5, coat protein (CP); **F.** ORF6, nucleic acid binding protein (NABP).

S3 Fig. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across ORFs 2, 3, 4, 5 and 6 of CPMMV sequences from serial inoculations in soybean plants, ten sequences from P1 (dark gray line), 30 sequences from P3 (light gray line) and 30 sequences from P6 (dotted line). **A.** ORF 2, triple gene block (TGB 1); **B.** ORF 3 (TGB 2); **C.** ORF4 (TGB 3); **D.**ORF5, coat protein (CP); **E.** ORF6, nucleic acid binding protein (NABP).

S4 Fig. Average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across the ORF2, 3, 4, 5 and 6 of CPMMV sequences from

1014 the second experiment obtained after serial inoculations of the CPMMV:BR:GO:10:5
1015 isolate from local lesions in *N. benthamiana*. Ten sequences from P1 (dark gray line) and
1016 30 sequences from BP2 (light gray line). **A.** ORF 2, triple gene block (TGB 1); **B.** ORF 3
1017 (TGB 2); **C.** ORF4 (TGB 3); **D.**ORF5, coat protein (CP); **E.** ORF6, nucleic acid binding
1018 protein (NABP).
1019
1020 **S5 Fig.** Phylogenetic relationships, based on the 70 sequences of CPMMV isolates from
1021 serial inoculations of P1, P3 and P6 in the first experiment, of individual ORFs (2-6) using
1022 Bayesian inference (implemented in MrBayes 3.1, with 10 million generations). Support for
1023 the nodes is presented as filled circles (posterior probabilities from 0.95 to 1.0). **A.** ORF 2,
1024 triple gene block (TGB1) with model GTR; **B-C.** ORF3-4 (TGB2-3) with selection of
1025 model HKY. **D.** ORF5, coat protein (CP) with model HKY. **E.** ORF6, nucleic acid binding
1026 protein (NABP) with model GTR.
1027
1028 **S6 Fig.** Phylogenetic relationships, based on the 40 sequences of CPMMV isolates from
1029 serial inoculations of BP1 and BP2 in the second experiment, of individual ORFs (2-6)
1030 using Bayesian inference (implemented in MrBayes 3.1, with 10 million generations).
1031 Support for the nodes is presented as filled circles (posterior probabilities from 0.95 to 1.0).
1032 **A.** ORF 2, triple gene block (TGB1) with model HKY; **B.** ORF3 (TGB2) with selection of
1033 model GTR. **C.** ORF4 (TGB3) with model HKY. **D.** ORF5, coat protein (CP) with model
1034 HKY+I. **E.** ORF6, nucleic acid binding protein (NABP) with selection model HKY.
1035
1036 **S7 Fig.** CPMMV ORF1 multivariate statistical clustering analysis of population
1037 subdivision using Discriminant Analysis of Principal Components (DAPC). **A.** DAPC

scatterplot with ellipses representing subpopulations in the first experiment. Dots represent each isolate from each subpopulation: blue, P1; yellow, P3; red, P6. **B.** Discriminant function obtained for the first experiment; discriminant function 1 was chosen. **C.** DAPC scatterplot with ellipses representing subpopulations in the second experiment. Dots represent each isolate from each subpopulation: blue, BP1; red, BP2. **D.** Discriminant function obtained for the bottleneck experiment.

S8 Fig. Phylogenetic relationships, based on the sequences of ORF1 from which recombinants blocks were excluded, using Bayesian inference (implemented in MrBayes 3.1). The recombinant blocks were excluded in RDP program v.5.10. Support for the nodes is presented as filled circles (posterior probabilities from 0.95 to 1.0). **A.** ORF1 tree using 70 sequences from the first experiment with model GTR+I +G and 10 million generations; **B.** ORF1 tree using 40 sequences from the second experiment with model GTR+I and 10 million generations.

Table 1. Descriptors of variability for *Cowpea mild mottle virus* (CPMMV) populations from the first experiment.

Genome region	Inoculation	Number of sequence*	Region length (nt)	*S*[a]	*G+C*	*K*[b]	π[c]	# of mutations	H[d]	Hd[e]	θ-W[f]
ORF1 (REPLICASE)	P1	10	5577	413	0.401	96.956	0.01738	413	10	1	0.02616
	P3	30		662	0.400	76.995	0.01380	760	29	0.998	0.02995
	P6	30		780	0.401	283.168	0.05094	1073	28	0.995	0.03542
ORF2 (TGB1)	P1	10	697	27	0.412	5.400	0.00776	27	7	0.867	0.01371
	P3	30		5	0.411	0.395	0.00057	5	6	0.310	0.00181
	P6	30		7	0.411	0.715	0.00103	7	7	0.510	0.00254
ORF3 (TGB2)	P1	10	321	1	0.420	0.200	0.00062	1	2	0.200	0.00110
	P3	30		6	0.421	0.400	0.00125	6	7	0.366	0.00472
	P6	30		4	0.421	0.267	0.00083	4	5	0.253	0.00315
ORF4 (TGB3)	P1	10	207	1	0.310	0.200	0.00097	1	2	0.200	0.00171
	P3	30		3	0.310	0.262	0.00127	3	4	0.251	0.00366
	P6	30		6	0.310	0.524	0.00253	6	6	0.363	0.00732
ORF5 (CP)	P1	10	867	9	0.426	1.800	0.00208	9	8	0.933	0.00367
	P3	30		14	0.426	1.239	0.00143	14	12	0.777	0.00408
	P6	30		55	0.426	4.453	0.00514	57	13	0.821	0.01601
ORF6 (NABP)	P1	10	342	5	0.418	1.000	0.00292	5	5	0.667	0.00517
	P3	30		7	0.418	0.653	0.00191	7	7	0.464	0.00517
	P6	30		8	0.419	0.772	0.00226	8	5	0.409	0.00590

[a]Total number of segregating sites. [b]Average number of nucleotide differences between sequences. [c]Nucleotide diversity mean (with 95% confidence intervals, using coalescent simulations). [d]Haplotype number. [e]Haplotype diversity. [f]Watterson's estimate of the population-scaled mutation rate based on the total number of segregating sites.

Table 2. Recombination rate (ρ) in populations of CPMMV from the first and bottleneck experiments.

First experiment

Pop.	ρ (CI)	ρ/site (95% CI)	θ/site	ρ/θ
P1	1166.679 (187.653-5048.163)	0.2125 ($3.419x10^{-2}$-0.919)	$2.518x10^{-2}$	8.441
P3	91.698 (62.221-131.29)	$1.75x10^{-2}$ ($1.187x10^{-2}$-$2.505x10^{-2)}$	$2.774x10^{-2}$	0.631
P6	3488.913 (282.606-8693.906)	0.6324 ($5.122x10^{-2}$-1.576)	$2.525\ x10^{-2}$	25.05
All	1237.834 (231.287-3265.553)	0.2244 ($4.192x10^{-2}$-0.5919)	$4.165x10^{-2}$	5.387

Bottleneck experiment

Pop.	ρ (CI)	ρ/site (CI)	θ/site	ρ/θ
BP1	957.927 (19.848-5935.066)	0.188 (3.899x10-3-1.166)	4.791x10-3	39.274
BP2	1635.841	0.318 (2.489x10-2-0.879)	9.623x10-3	33.065
All	9.375 (7.66-11.505)	2.112x10-3 (1.726x10-3-2.592x10-3)	1.478x10-2	0.143

*CI=Confidence intervals; ρ=Recombination rate; θ= mutation rate.

Table 3: Neutrality tests and analysis of selection site by site for ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 of CPMMV isolates from the first experiment implemented in Datamonkey server.

Genome region	Inoculation	Tajima'D	Fu & Li's D*	Fu & Li's F*	SLAC		FEL		IFEL		REL		PARRIS
					N	P	N	P	N	P	N	P	P
	P1	-1.85041*	-2.04262*	-2.25531*	3	-	22	2	1	9	PS	-	-
ORF1	P3	2.33920***	-3.62413**	-3.77908**	4	-	55	7	9	34	-	69	-
	P6	0.17794^{ns}	0.13339^{ns}	0.17250^{ns}	22	2	46	16	35	18	28	136	+
	P1	2.07497***	-2.45320**	-2.66135**	-	-	2	-	-	-	-	-	-
ORF2	P3	-1.87346*	-2.36062^{ns}	-2.57755^{ns}	-	-	1	-	-	1	-	-	-
	P6	-1.75911^{ns}	-0.81656^{ns}	-1.28327^{ns}	-	-	2	-	-	-	PS	-	-
	P1	-1.11173^{ns}	-1.24341^{ns}	-1.34668^{ns}	-	-	-	-	-	-	-	-	-
ORF3	P3	-2.09995*	-3.43598**	-3.53709**	-	-	-	-	-	-	PS	-	-
	P6	-1.88948*	-2.99378*	-3.10061*	-	-	-	-	-	-	PS	-	-
	P1	-1.11173^{ns}	-1.24341^{ns}	-1.34668^{ns}	-	-	-	-	-	-	-	-	-
ORF4	P3	-1.53889^{ns}	-1.47512^{ns}	-1.72941^{ns}	-	-	1	-	-	-	PS	-	-
	P6	-1.86605*	-1.88589^{ns}	-2.18941^{ns}	-	-	2	-	-	-	2	-	-
	P1	-1.88589*	-2.21874**	-2.40603**	-	-	1	-	-	-	-	1	-
ORF5	P3	-2.16704*	-2.63942*	-2.92188*	-	-	2	-	-	-	-	-	-
	P6	-2.58928***	-4.60532**	-4.65377**	1	-	7	-	2	-	1	-	-
	P1	-1.74110*	-2.01007^{ns}	-2.17902^{ns}	-	-	-	-	-	-	-	-	-
ORF6	P3	-1.86290*	-1.51321^{ns}	-1.88745^{ns}	-	-	1	-	-	-	PS	-	-
	P6	-1.87584*	-2.47845^{ns}	-2.68001*	-	-	2	-	-	-	PS	-	-

*P<0.05; **P<0.02(P); ***P< 0.001 Sites under positive selection; (N) sites under negative selection; (PS) purifying selection; (-) no site under selection; (+) sites under positive selection. ^{a,b,c,d,e}Codon-based maximum-likelihood algorithms. ^{a}Single Likelihood Ancestor Counting (SLAC); ^{b}Fixed Effects Likelihood; (FEL); ^{c} Internal Fixed Effects Likelihood (IFEL); ^{d}Random Effects Likelihood (REL) and ^{e}Partitioning for Robust Inference of Selection (PARRIS).

Table 4. Description of sites in ORF1 of CPMMV isolates from both experiments as potential sites involving in phenotype changes.

First Experiment				Bottleneck Experiment			
Site	**Codon**	**Substitution**	**Vi**	**Site**	**Codon**	**Substitution**	**Vi**
1082	344	A-->C	6.24695	648	216	G-->A	7.682213
1096	365*	T-->C	5.364934	819	273	C-->T	7.682213
1119	373*	C-->A	6.24695	957	319	C-->T	4.287483
1122	374*	T-->G	5.364934	1354	452	G-->A	7.682213
1130	377*	G-->C	5.364934	1357	453	A-->G	7.682213
2345	782*	A-->G	7.714678	1358	453	G-->A	7.682213
3225	1075*	A-->C	7.714678	1359	453	A-->T	7.682213
3243	1081	G-->T	6.24695	1360	454	T-->C	7.682213
3246	1082*	A-->G	7.714678	1361	454	C-->T	7.682213
4463	1488*	G-->A	7.714678	1362	454	T-->C	7.682213
4464	1488*	A-->C	11.00037	1627	543	G-->A	7.682213
4465	1489*	C-->T	11.00037	1628	543	A-->G	7.682213
4468	1490	A-->G	11.00037	1629	543	G-->A	7.682213
4470	1490	G-->A	11.00037	1630	544	A-->T	7.682213
4482	1494*	A-->C	11.00037	1631	544	T-->C	7.682213
4504	1502*	T-->C	7.714678	1632	544	C-->A	7.682213
4514	1505	C-->A	11.00037	1633	545	A-->G	7.682213
4524	1508	T-->G	11.00037	1634	545	G-->T	7.682213
4525	1509	G-->A	11.00037	1636	546	T-->C	7.682213
4530	1510*	A-->T	11.00037	1637	546	C-->T	7.682213
4546	1516	C-->A	7.714678	1638	546	T-->G	7.682213
4548	1516	C-->T	11.00037	1639	547	G-->C	7.682213
4565	1522*	T-->C	11.00037	1640	547	C-->T	7.682213
4567	1523*	C-->T	11.00037	1641	547	T-->A	7.682213
4570	1524	G-->A	11.00037	1644	548	A-->T	7.682213
4577	1526	T-->A	11.00037	1646	549	T-->C	7.682213
4578	1526	G-->A	11.00037	1647	549	C-->A	7.682213
				2770	924	C-->A	4.287483
				2771	924	A-->T	7.682213
				2772	924	T-->A	7.682213
				2773	925	A-->C	7.682213
				2774	925	C-->A	7.682213
				2775	925	A-->T	7.682213
				2776	926	T-->G	7.682213
				2777	926	G-->C	7.682213
				2778	926	C-->T	7.682213

*Codons also submitted to positive selection. Vi= evolution index

Fig 1:

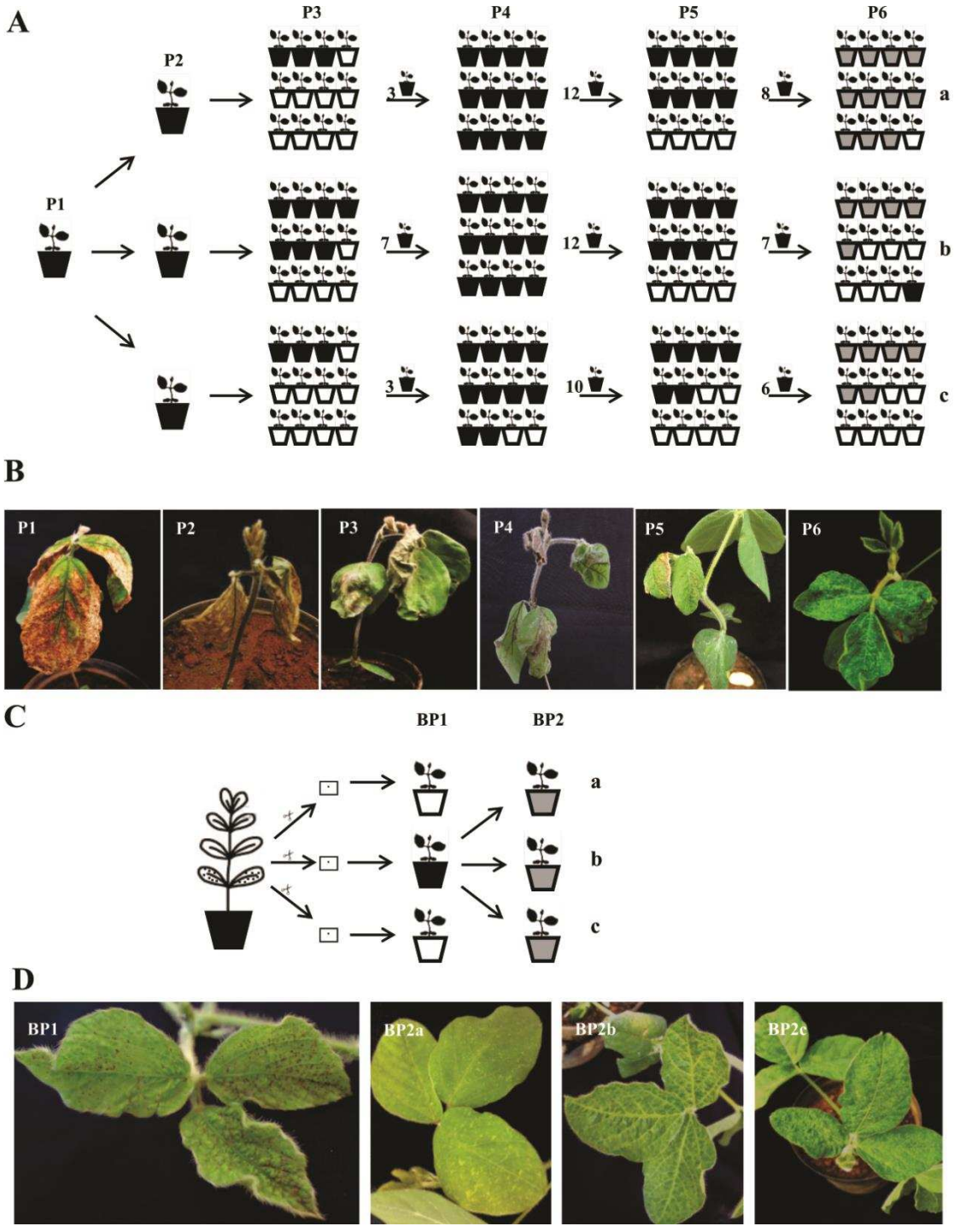

A
P1
P2
P3
P4
P5
P6
3
12
8
a
7
12
7
b
3
10
6
c
B
P1
P2
P3
P4
P5
P6
C
BP1
BP2
a
b
c
D
BP1
BP2a
BP2b
BP2c

Fig 2:

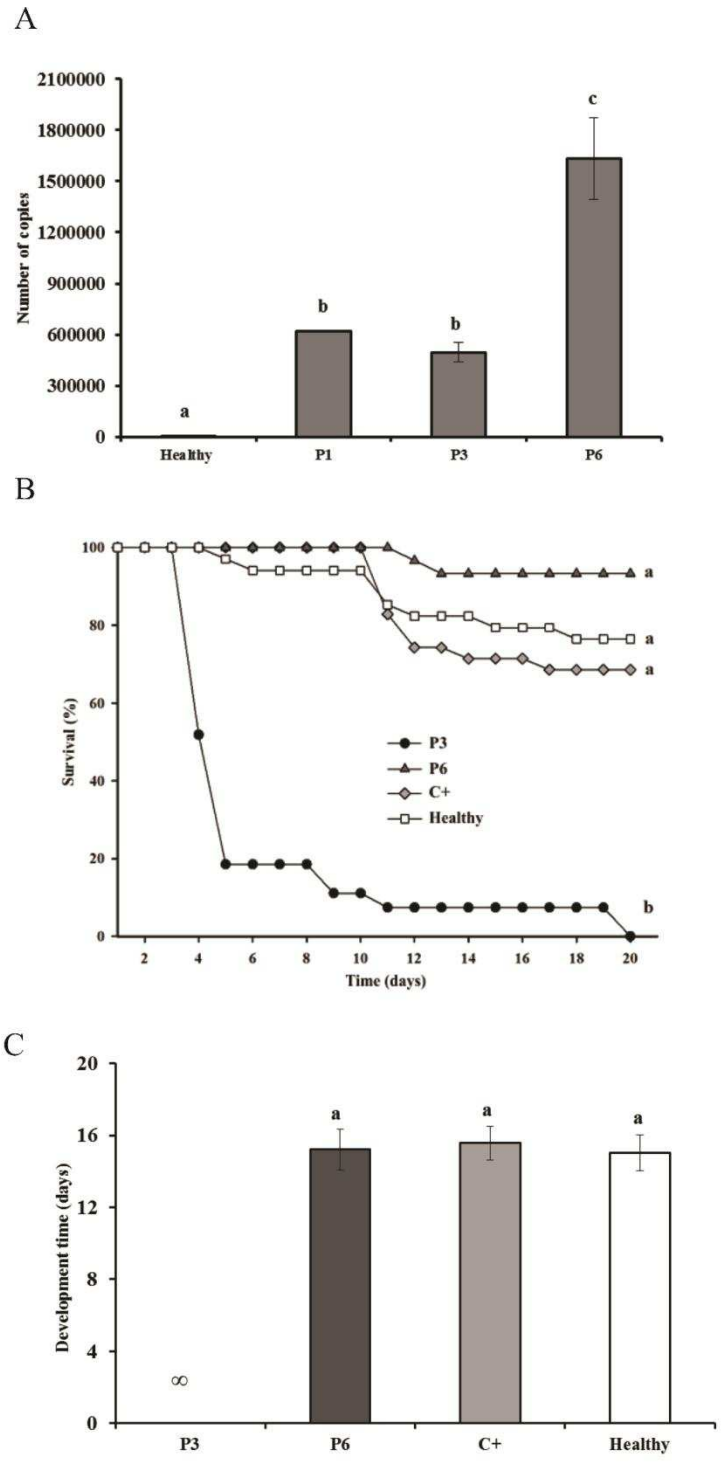

A
2100000
1800000
1500000
1200000
900000
600000
300000
0
Number of copies
a
b
b
c
Healthy
P1
P3
P6
B
100
80
60
40
20
0
Survival (%)
a
a
a
b
P3
P6
C+
Healthy
2
4
6
8
10
12
14
16
18
20
Time (days)
C
20
16
12
8
4
0
Development time (days)
∞
a
a
a
P3
P6
C+
Healthy

Fig 3:

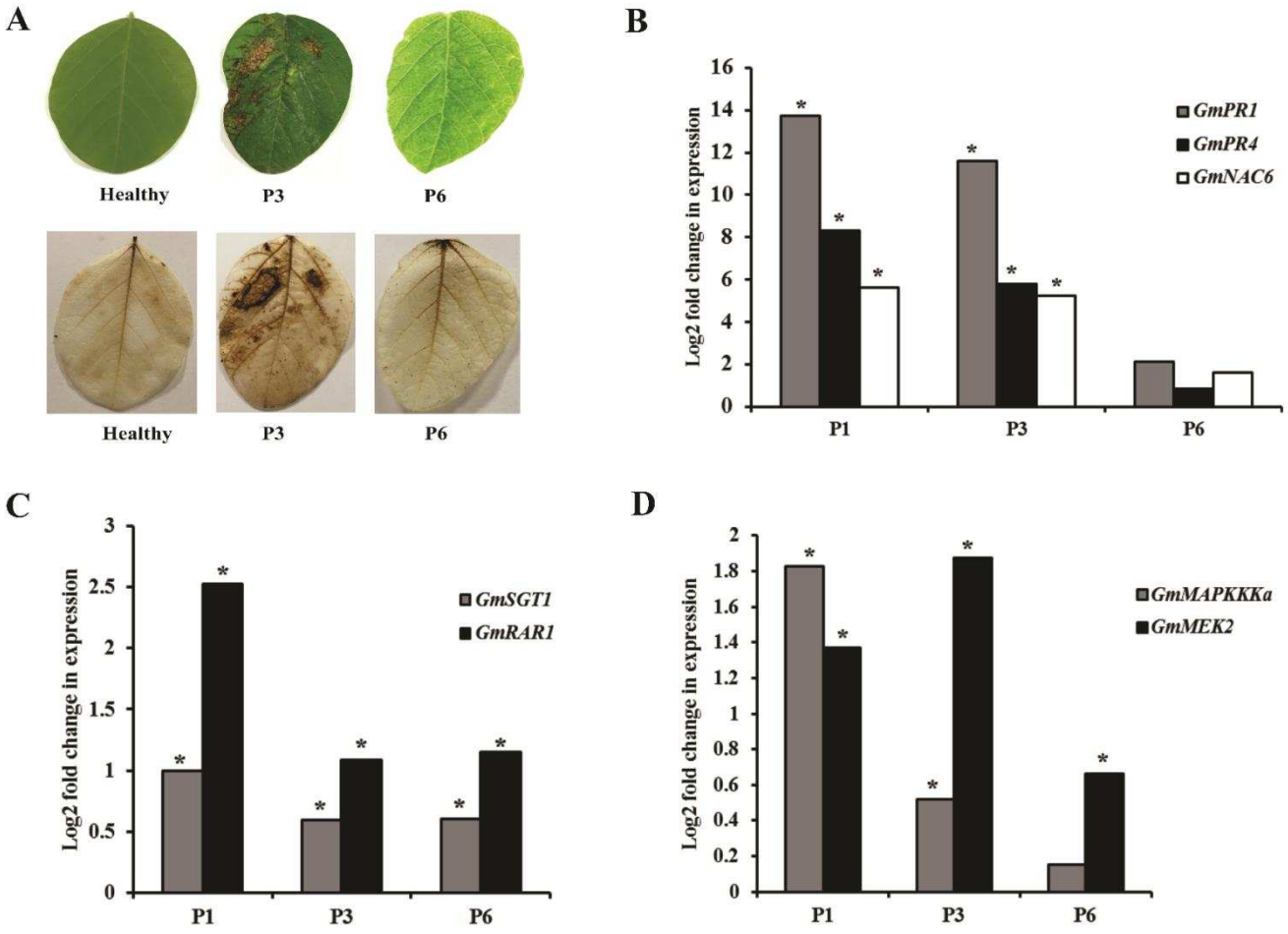
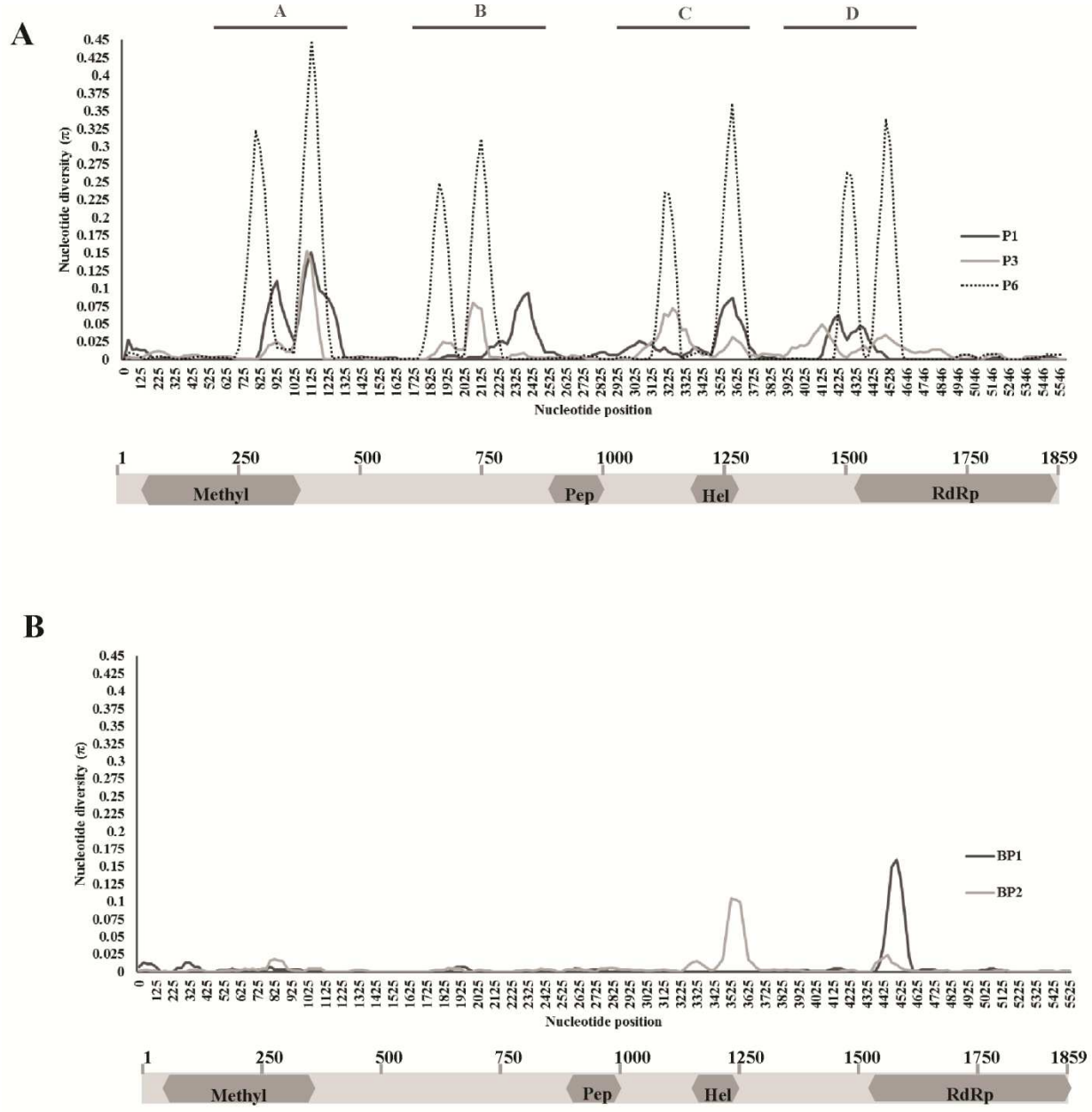
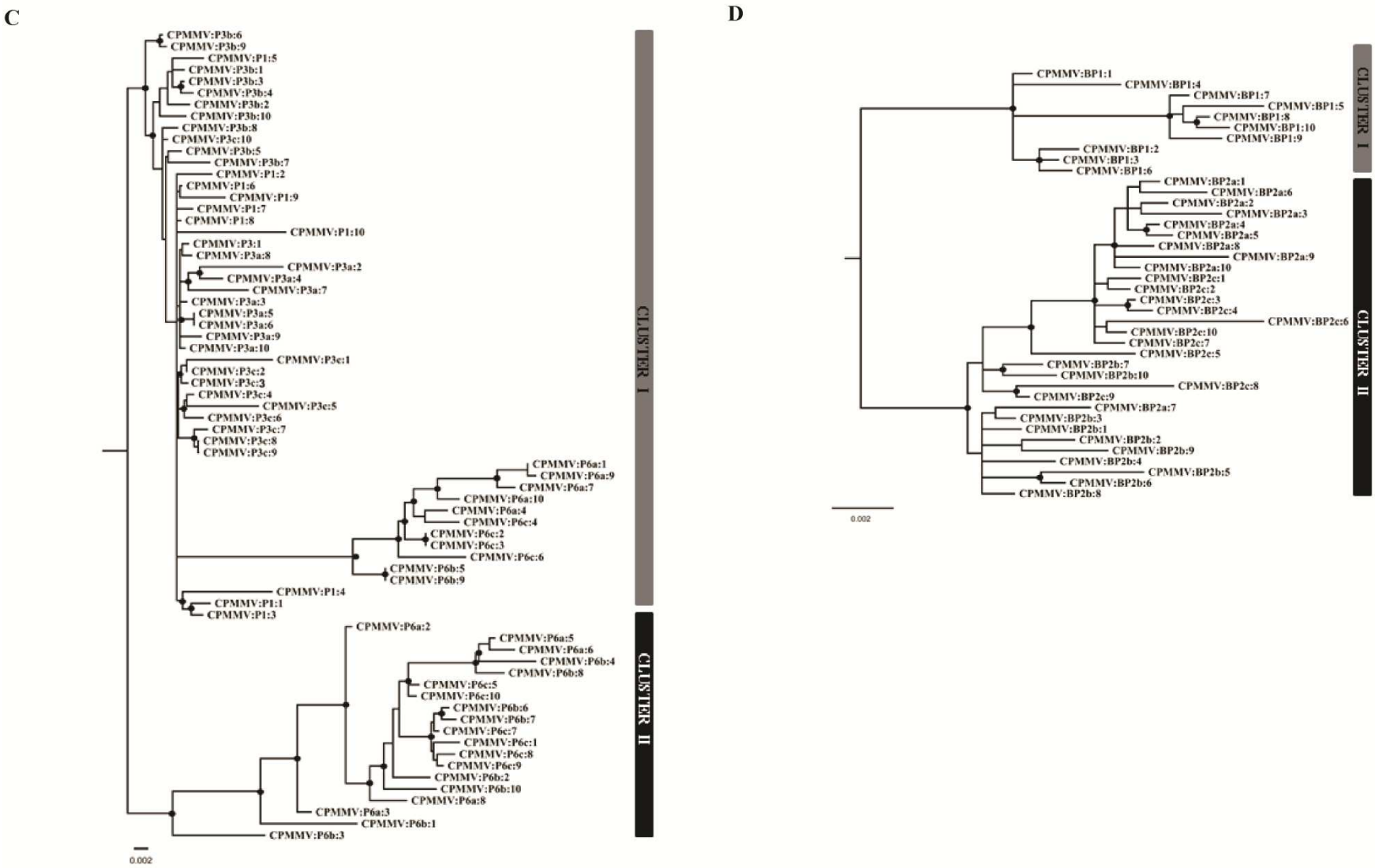
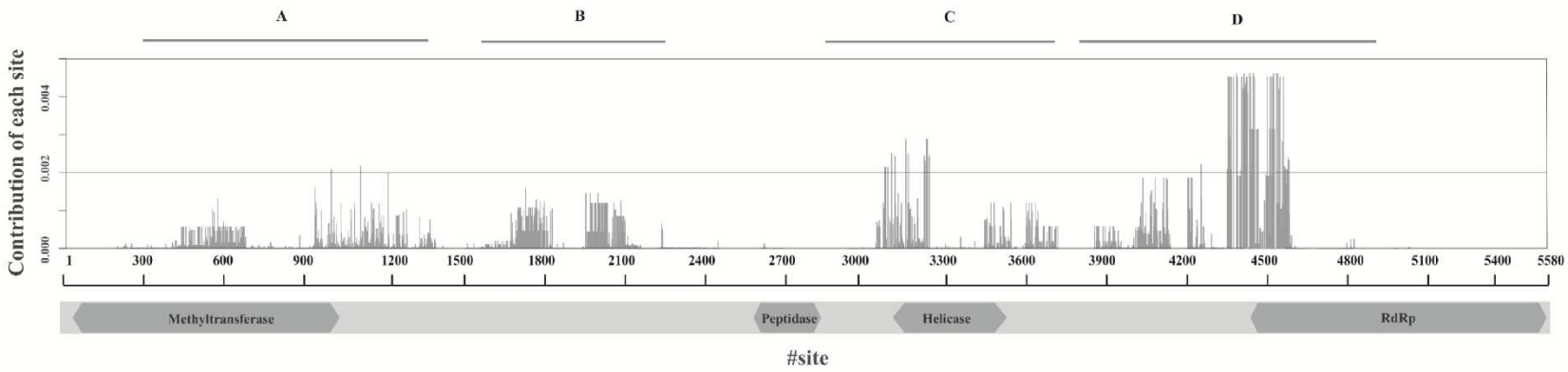

A
Healthy
P3
P6
Healthy
P3
P6
B
Log2 fold change in expression
GmPR1
GmPR4
GmNAC6
P1
P3
P6
C
Log2 fold change in expression
GmSGT1
GmRAR1
P1
P3
P6
D
Log2 fold change in expression
GmMAPKKKa
GmMEK2
P1
P3
P6

Fig 4:


A
A
B
C
D
Nucleotide diversity (π)
Nucleotide position
P1
P3
P6
1
250
500
750
1000
1250
1500
1750
1859
Methyl
Pep
Hel
RdRp
B
Nucleotide diversity (π)
Nucleotide position
BP1
BP2
1
250
500
750
1000
1250
1500
1750
1859
Methyl
Pep
Hel
RdRp


C
CPMMV:P3b:6
CPMMV:P3b:9
CPMMV:P1:5
CPMMV:P3b:1
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:2
CPMMV:P3b:10
CPMMV:P3b:8
CPMMV:P3c:10
CPMMV:P3b:5
CPMMV:P3b:7
CPMMV:P1:2
CPMMV:P1:6
CPMMV:P1:9
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:10
CPMMV:P3:1
CPMMV:P3a:8
CPMMV:P3a:2
CPMMV:P3a:4
CPMMV:P3a:7
CPMMV:P3a:3
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3c:1
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P6a:1
CPMMV:P6a:9
CPMMV:P6a:7
CPMMV:P6a:10
CPMMV:P6a:4
CPMMV:P6c:4
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:6
CPMMV:P6b:5
CPMMV:P6b:9
CPMMV:P1:4
CPMMV:P1:1
CPMMV:P1:3
CPMMV:P6a:2
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6b:4
CPMMV:P6b:8
CPMMV:P6c:5
CPMMV:P6c:10
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6c:7
CPMMV:P6c:1
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6b:2
CPMMV:P6b:10
CPMMV:P6a:8
CPMMV:P6a:3
CPMMV:P6b:1
CPMMV:P6b:3
0.002
CLUSTER I
CLUSTER II
D
CPMMV:BP1:1
CPMMV:BP1:4
CPMMV:BP1:7
CPMMV:BP1:5
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP1:9
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:6
CPMMV:BP2a:1
CPMMV:BP2a:6
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2e:1
CPMMV:BP2c:2
CPMMV:BP2e:3
CPMMV:BP2e:4
CPMMV:BP2c:6
CPMMV:BP2e:10
CPMMV:BP2c:7
CPMMV:BP2e:5
CPMMV:BP2b:7
CPMMV:BP2b:10
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2a:7
CPMMV:BP2b:3
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:9
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:8
0.002
CLUSTER I
CLUSTER II

E


A
B
C
D
Contribution of each site
0.000
0.002
0.004
1
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100
5400
5580
Methyltransferase
Peptidase
Helicase
RdRp
#site

F

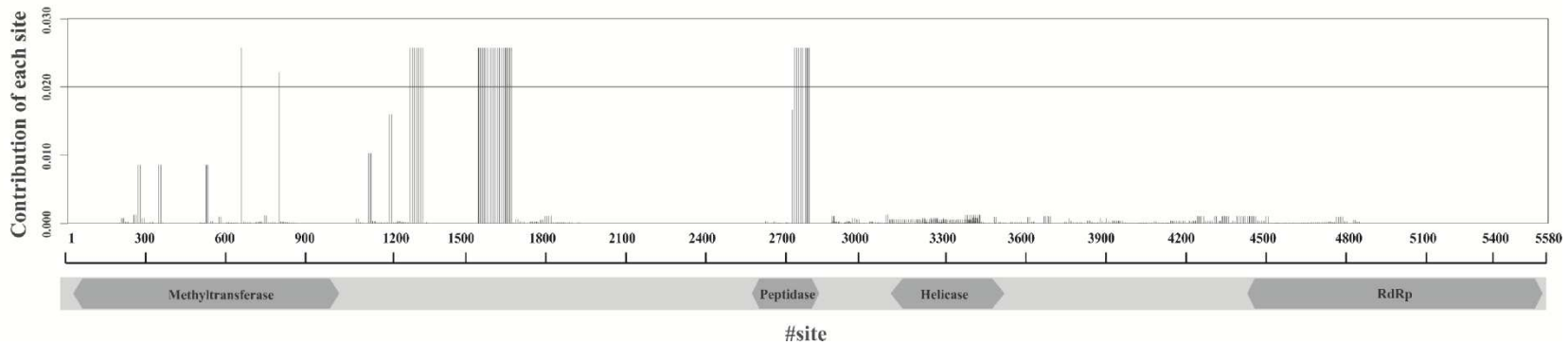
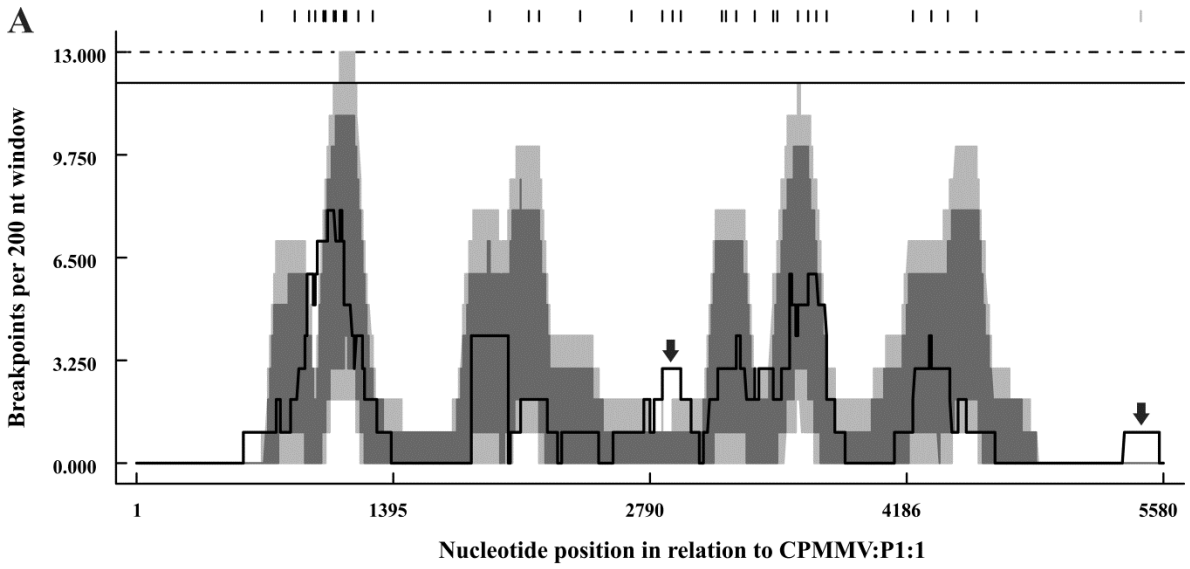
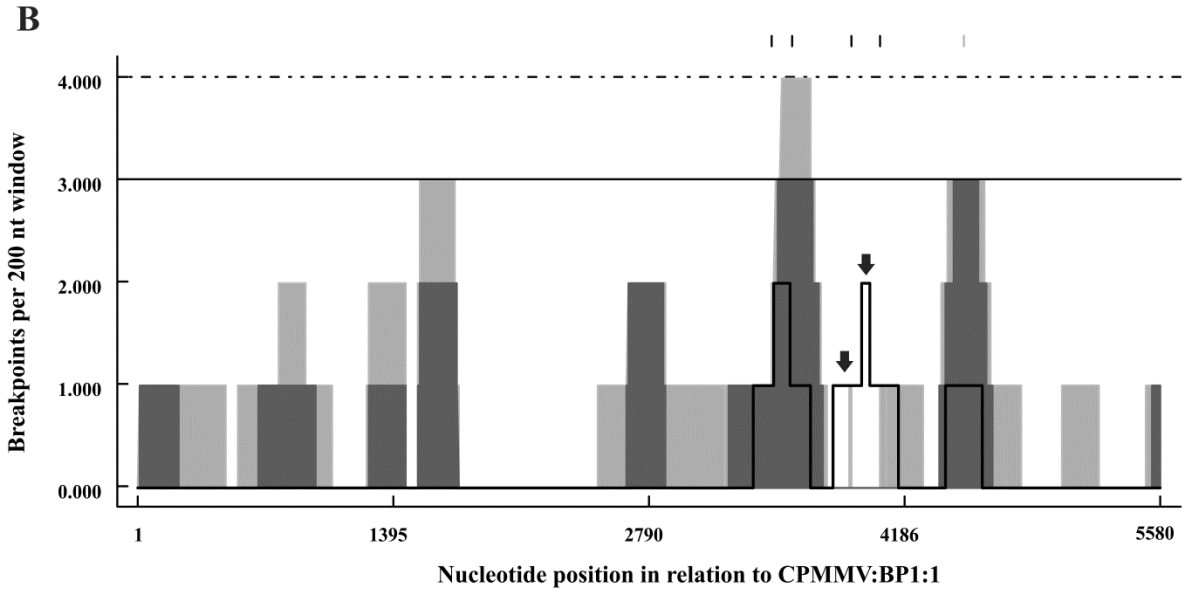

Contribution of each site
0.000
0.010
0.020
0.030
1
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100
5400
5580
Methyltransferase
Peptidase
Helicase
RdRp
#site

Fig 5:

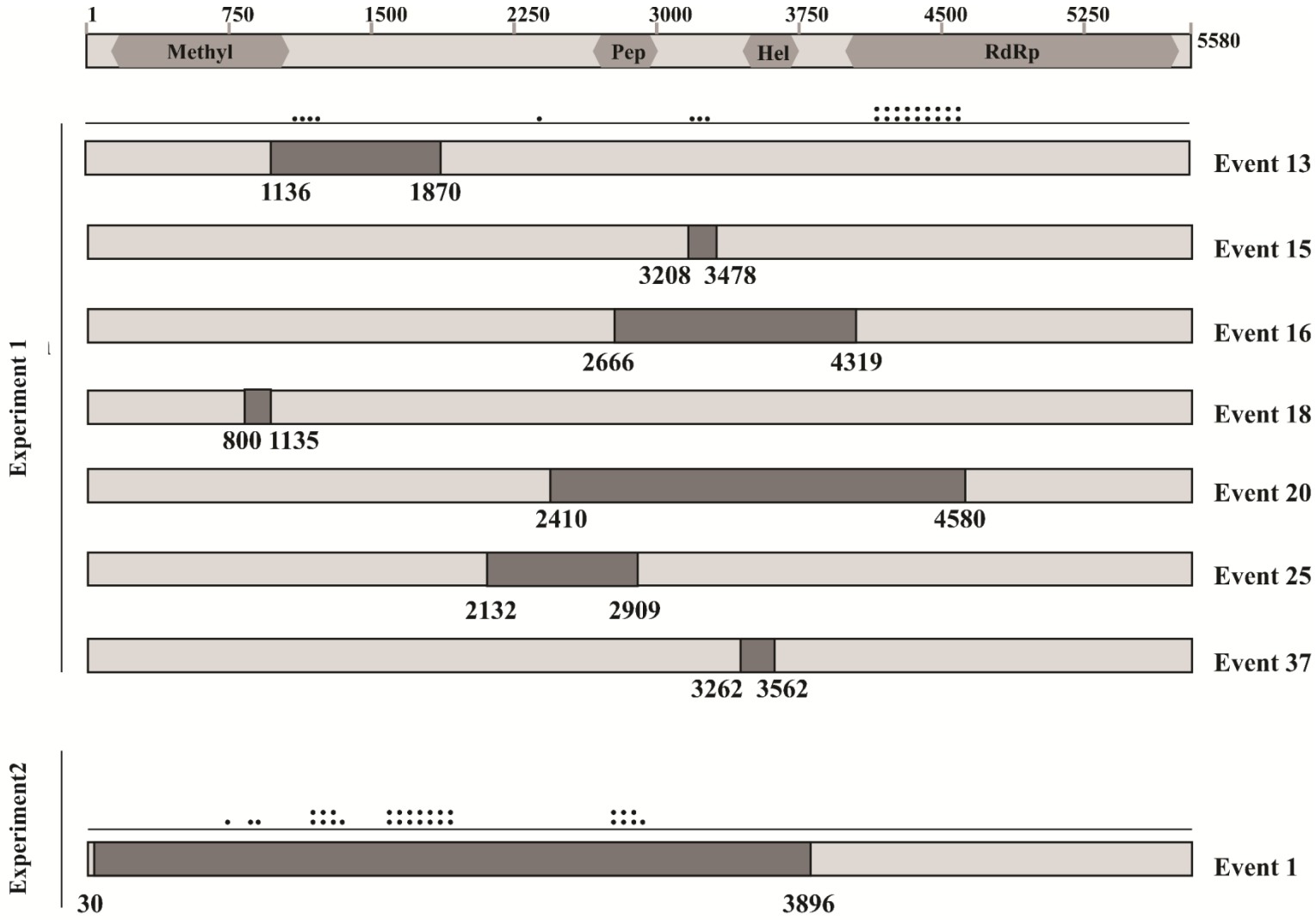


#	Type	Position*	Region	Significance
1	Hot-spot	2811-3055	Pep-IG	Locally
2	Hot-spot	5366-5561	RdRp	Locally



#	Type	Position*	Region	Significance
1	Hot-spot	3797-3881	IG (Hel-RdRp)	Locally
2	Hot-spot	3899-4053	IG (Hel-RdRp)	Locally

Fig 6:


1
750
1500
2250
3000
3750
4500
5250
5580
Methyl
Pep
Hel
RdRp
Experiment 1
Event 13
1136
1870
Event 15
3208
3478
Event 16
2666
4319
Event 18
800
1135
Event 20
2410
4580
Event 25
2132
2909
Event 37
3262
3562
Experiment2
Event 1
30
3896

S1 Table. Descriptors of variability for *Cowpea mild mottle virus* (CPMMV) populations from the bottleneck experiment.

Genome region	Inoculation	Number of isolates*	Region length (nt)	G+C	S^a	K^b	π^c	# of mutations	H^d	Hd^e	θ-W^f
ORF1 (REPLICASE)	BP1	10	5577	0.402	69	24.289	0.00435	69	10	1.000	0.00437
	BP2	30		0.401	207	25.816	0.00463	218	30	1.000	0.00936
ORF2 (TGB1)	BP1	10	697	0.408	6	1.822	0.00262	6	4	0.644	0.00305
	BP2	30		0.407	8	0.595	0.00086	8	9	0.515	0.00290
ORF3 (TGB2)	BP1	10	321	0.421	6	1.200	0.00374	6	6	0.778	0.00661
	BP2	30		0.421	13	0.929	0.00289	13	13	0.00289	0.01022
ORF4 (TGB3)	BP1	10	207	0.310	3	0.600	0.0029	3	4	0.533	0.00512
	BP2	30		0.309	0	0	0	0	1	0	0
ORF5 (CP)	BP1	10	867	0.428	13	3.800	0.00438	13	7	0.933	0.00530
	BP2	30		0.427	17	1.501	0.00173	17	15	0.825	0.00495
ORF6 (NABP)	BP1	10	342	0.428	3	0.600	0.0175	3	4	0.533	0.00310
	BP2	30		0.430	4	0.267	0.00078	4	5	0.253	0.00295

[a]Total number of segregating sites. [b]Average number of nucleotide differences between sequences. [c]Nucleotide diversity mean (with 95% confidence intervals, using coalescent simulations). [d]Haplotype number. [e]Haplotype diversity. [f]Watterson's estimate of the population-scaled mutation rate based on the total number of segregating sites.

S2 Table: Sites involved in subdivision population described in discriminant analysis of principal components DAPC analysis.

Sites involved in subdivision population	
First experiment	**Bottleneck experiment**
1094	648
1119	819
3199	1354
3200	1357
3203	1358
3208	1359
3225	1360
3228	1361
3229	1362
3213	1627
3243	1628
3246	1629
3248	1630
4311	1631
4461	1632
4463	1633
4464	1634
4465	1636
4467	1637
4468	1638
4470	1639
4474	1640
4477	1641
4478	1644
4479	1646
4481	1647
4482	2771
4484	2772
4485	2773
4486	2774
4489	2775
4490	2776
4491	2777
4492	2778
4494	
4496	
4497	
4500	

4501	
4503	
4505	
4508	
4513	
4518	
4520	
4521	
4522	
4524	
4526	
4527	
4528	
4529	
4530	
4541	
4543	
4545	
4546	
4547	
4548	
4562	
4564	
4567	
4568	
4572	
4573	
4574	
4577	
4580	
4581	
4582	
4583	
4584	

*The sites showed in red correspond to sites that were associated with the change of phenotypes in association and localization tests.

S3 table: Recombination detected by RDP in ORF1 of CPMMV populations from the first experiment.

Event	Breakpoints in Alignment		Recombinant Sequence(s)	Parental sequences		Detected by	*P:*value
	Begin	End		Minor	Major		
1	4454*	5483	CPMMV:P3c:1	Unknown	CPMMV:P3c:2	GBMCS**3S**	$1.01xE^{-50}$
2	2957	3358	CPMMV:P3c:5	Unknown	CPMMV:P1:4	RGMCS**3S**	$3.21xE^{-47}$
3	3599	4248*	CPMMV:P3a:7	Unknown	CPMMV:P1:10	**R**GBMCS3S	$2.85xE^{-43}$
4	932	1570*	CPMMV:P1:10	Unknown	CPMMV:P3a:2	**R**GBMCS3S	$1.39xE^{-40}$
5	3178	3644	CPMMV:P6a:3	Unknown	CPMMV:P6a:2	**R**GBMCS3S	$5.84xE^{-28}$
6	3648	4214	CPMMV:P3a:2	Unknown	CPMMV:P6c:3	**R**GBMC3S	$4.53xE^{-36}$
7	1132	4309*	CPMMV:P6b:1	Unknown	CPMMV:P6c:10	RGMCS**3S**	$1.62xE^{-35}$
8	1030	1202	CPMMV:P6a:6	CPMMV:P1:4	CPMMV:P6b:4	**R**GBMCS3S	$1.37xE^{-34}$
9	2240*	4404	CPMMV:P1:4	Unknown	CPMMV:P3c:9	RGBMCS**3S**	$2.65xE^{-33}$
10	1288	2686	CPMMV:P6b:3	CPMMV:P6b:4	CPMMV:P1:2	R**G**BMCS3S	$1.91xE^{-29}$
11	3304*	4324	CPMMV:P6b:3 CPMMV:P6a:3	CPMMV:P1:5 CPMMV:P1:1	CPMMV:P6b:9 CPMMV:P6a:10	RGBMCS**3S**	$1.82xE^{-39}$

12	1922	3592*	CPMMV:P3b:7	CPMMV:P3b:2	CPMMV:P3a:2	RGMCS**3S**	$8.15xE^{-28}$
13	1136	1870*	CPMMV:P6a:2 CPMMV:P1:5 CPMMV:P3b:1 CPMMV:P3b:2 CPMMV:P3b:3 CPMMV:P3b:4 CPMMV:P3b:10 CPMMV:P6a:3 CPMMV:P6a:5 CPMMV:P6a:6 CPMMV:P6a:8 CPMMV:P6b:2 CPMMV:P6b:3 CPMMV:P6b:4 CPMMV:P6b:6 CPMMV:P6b:7 CPMMV:P6b:8 CPMMV:P6b:10 CPMMV:P6c:1	CPMMV:P6c:4 CPMMV:P6a:4	Unknown	**R**GBMCS3S	$5.46xE^{-27}$

			CPMMV:P6c:5				
			CPMMV:P6c:7				
			CPMMV:P6c:8				
			CPMMV:P6c:9				
			CPMMV:P6c:10				
14	742*	1922	CPMMV:P6b:9	CPMMV:P3a:4	Unknown	RGMCS**3S**	$5.83xE^{-26}$
			CPMMV:P6b:5	CPMMV:P1:2	Unknown		
15	3208	3478	CPMMV:P3c:7	CPMMV:P6c:1	CPMMV:P1:1	**R**GBMCS3S	$2.56xE^{-24}$
			CPMMV:P1:1	CPMMV:P6a:2	CPMMV:P1:2		
			CPMMV:P1:2	CPMMV:P6b:2	CPMMV:P1:5		
			CPMMV:P1:5	CPMMV:P6b:6	CPMMV:P1:6		
			CPMMV:P1:6	CPMMV:P6b:10	CPMMV:P1:7		
			CPMMV:P1:7	CPMMV:P6c:10	CPMMV:P1:8		
			CPMMV:P1:8		CPMMV:P1:9		
			CPMMV:P1:9		CPMMV:P3a:1		
			CPMMV:P3a:1		CPMMV:P3a:2		
			CPMMV:P3a:2		CPMMV:P3a:3		
			CPMMV:P3a:3		CPMMV:P3a:4		

CPMMV:P3a:4	CPMMV:P3a:5
CPMMV:P3a:5	CPMMV:P3a:6
CPMMV:P3a:6	CPMMV:P3a:8
CPMMV:P3a:7	CPMMV:P3a:9
CPMMV:P3a:8	CPMMV:P3a:10
CPMMV:P3a:9	CPMMV:P3b:1
CPMMV:P3a:10	CPMMV:P3b:2
CPMMV:P3b:1	CPMMV:P3b:3
CPMMV:P3b:2	CPMMV:P3b:4
CPMMV:P3b:3	CPMMV:P3b:5
CPMMV:P3b:4	CPMMV:P3b:6
CPMMV:P3b:5	CPMMV:P3b:8
CPMMV:P3b:6	CPMMV:P3b:9
CPMMV:P3b:7	CPMMV:P3b:10
CPMMV:P3b:8	CPMMV:P3c:1
CPMMV:P3b:9	CPMMV:P3c:2
CPMMV:P3b:10	CPMMV:P3c:3
CPMMV:P3c:1	CPMMV:P3c:4
CPMMV:P3c:2	CPMMV:P3c:10
CPMMV:P3c:3	CPMMV:P6a:7
CPMMV:P3c:8	CPMMV:P6a:9

			CPMMV:P3c:9		CPMMV:P3c:6		
			CPMMV:P3c:10				
			CPMMV:P6a:1				
			CPMMV:P6a:4				
			CPMMV:P6a:7				
			CPMMV:P6a:9				
			CPMMV:P6a:10				
			CPMMV:P6b:5				
			CPMMV:P6b:9				
			CPMMV:P6c:2				
			CPMMV:P6c:3				
			CPMMV:P6c:4				
			CPMMV:P6c:6				
16	2666*	4319	CPMMV:P1:1	CPMMV:P6c:6	CPMMV:P6c:7	RGBMCS**3S**	$2.89xE^{-38}$
			CPMMV:P1:2		CPMMV:P6b:2		
			CPMMV:P1:3		CPMMV:P6c:1		
			CPMMV:P1:5		CPMMV:P6c:5		
			CPMMV:P1:6		CPMMV:P6c:8		
			CPMMV:P1:7		CPMMV:P6c:9		

CPMMV:P1:8

CPMMV:P1:9

CPMMV:P1:10

CPMMV:P3a:1

CPMMV:P3a:2

CPMMV:P3a:3

CPMMV:P3a:4

CPMMV:P3a:5

CPMMV:P3a:6

CPMMV:P3a:7

CPMMV:P3a:8

CPMMV:P3a:9

CPMMV:P3a:10

CPMMV:P3b:1

CPMMV:P3b:2

CPMMV:P3b:3

CPMMV:P3b:4

CPMMV:P3b:5

CPMMV:P3b:6

CPMMV:P3b:7

CPMMV:P3b:8

CPMMV:P6c:10

			CPMMV:P3b:9				
			CPMMV:P3b:10				
			CPMMV:P3c:1				
			CPMMV:P3c:2				
			CPMMV:P3c:3				
			CPMMV:P3c:4				
			CPMMV:P3c:5				
			CPMMV:P3c:6				
			CPMMV:P3c:7				
			CPMMV:P3c:8				
			CPMMV:P3c:9				
			CPMMV:P3c:10				
17	2189	2665*	CPMMV:P1:2	Unknown	CPMMV:P1:7	RGBMC**3S**	$5.06xE^{-23}$
18	800*	1135*	CPMMV:P3b:8	CPMMV:P6c:6	CPMMV:P6c:7	RGMCS**3S**	$3.67xE^{-28}$
			CPMMV:P1:1		CPMMV:P6b:2		
			CPMMV:P1:3		CPMMV:P6c:1		
			CPMMV:P1:4		CPMMV:P6c:5		
			CPMMV:P1:9		CPMMV:P6c:8		
			CPMMV:P3a:1		CPMMV:P6c:9		

			CPMMV:P3a:2 CPMMV:P3a:3 CPMMV:P3a:6 CPMMV:P3a:7 CPMMV:P3b:7 CPMMV:P3c:1 CPMMV:P3c:2 CPMMV:P3c:4 CPMMV:P3c:5 CPMMV:P3c:6 CPMMV:P3c:7 CPMMV:P3c:8 CPMMV:P3c:9 CPMMV:P6a:1 CPMMV:P6a:7 CPMMV:P6a:9		CPMMV:P6c:10		
19	3748	4588*	CPMMV:P6c:6	Unknown	CPMMV:P6c:3	RGMCS**3S**	$1.45xE^{-20}$
20	2410	4580*	CPMMV:P6a:6 CPMMV:P6a:5	CPMMV:P6a:8 CPMMV:P6a:2 CPMMV:P6b:2	Unknown Unknown Unknown	RGMCS**3S**	$8.69xE^{-18}$

			CPMMV:P6b:8				
21	1066	1134*	CPMMV:P6a:8	Unknown	CPMMV:P6c:5	RGMCS**3S**	$2.44xE^{-18}$
22	742*	1080	CPMMV:P6b:9	CPMMV:P3b:9	CPMMV:P6c:3	RGBMC**3S**	$5.00xE^{-17}$
			CPMMV:P6b:5	CPMMV:P1:2	CPMMV:P6c:2		
23	2240*	2860	CPMMV:P1:4	Unknown	CPMMV:P6c:4	**R**GBM3S	$1.02xE^{-16}$
24	1136*	1861	CPMMV:P6a:4	Unknown	CPMMV:P6a:9	RGMCS**3S**	$1.35xE^{-24}$
			CPMMV:P6a:10		CPMMV:P6a:1		
			CPMMV:P6c:2		CPMMV:P6a:7		
			CPMMV:P6c:3				
			CPMMV:P6c:4				
			CPMMV:P6c:6				
25	2132	2909	CPMMV:P6a:9	CPMMV:P6b:4	CPMMV:P6b:6	RGBMC**S**3S	$1.48xE^{-34}$
			CPMMV:P1:1		CPMMV:P6a:2		
					CPMMV:P6a:8		

CPMMV:P1:3

CPMMV:P1:5

CPMMV:P1:6

CPMMV:P1:7

CPMMV:P1:8

CPMMV:P1:9

CPMMV:P1:10

CPMMV:P3a:1

CPMMV:P3a:2

CPMMV:P3a:3

CPMMV:P3a:4

CPMMV:P3a:5

CPMMV:P3a:6

CPMMV:P3a:7

CPMMV:P6b:2

CPMMV:P6b:7

CPMMV:P6b:10

CPMMV:P6c:1

CPMMV:P6c:5

CPMMV:P6c:7

CPMMV:P6c:8

CPMMV:P6c:9

CPMMV:P6c:10

CPMMV:P3a:8

CPMMV:P3a:9

CPMMV:P3a:10

CPMMV:P3b:4

CPMMV:P3b:5

CPMMV:P3b:10

CPMMV:P3c:1

CPMMV:P3c:2

CPMMV:P3c:3

CPMMV:P3c:4

CPMMV:P3c:5

CPMMV:P3c:6

CPMMV:P3c:7

			CPMMV:P3c:8				
			CPMMV:P3c:9				
			CPMMV:P6a:1				
			CPMMV:P6a:4				
			CPMMV:P6b:1				
			CPMMV:P6b:5				
			CPMMV:P6b:9				
			CPMMV:P6c:4				
			CPMMV:P6c:6				
26	4582	4969*	CPMMV:P3c:1	Unknown	CPMMV:P3a:9	RGB**<u>M</u>**C3S	$6.28xE^{-13}$
27	1922	2122*	CPMMV:P3a:2		CPMMV:P3c:10	RGBMCS**<u>3S</u>**	$2.26xE^{-12}$
			CPMMV:P1:5		CPMMV:P3b:1		
			CPMMV:P3a:1		CPMMV:P3b:2		
					CPMMV:P3b:3		

			CPMMV:P3a:3				
			CPMMV:P3a:4	CPMMV:P3b:4			
			CPMMV:P3a:6	CPMMV:P3b:5			
			CPMMV:P3a:8	CPMMV:P3b:6			
			CPMMV:P3a:9	CPMMV:P3b:8			
			CPMMV:P3a:10	CPMMV:P3b:9			
			CPMMV:P3c:4	CPMMV:P3b:10			
			CPMMV:P3c:5				
			CPMMV:P3c:6				
			CPMMV:P3c:7				
28	1028	1120*	CPMMV:P3b:9 CPMMV:P3b:6	CPMMV:P6b:7 CPMMV:P6b:6	Unknown	**R**GBMCS3S	$5.93xE^{-12}$
29	1*	683	CPMMV:P6a:9	CPMMV:P3b:9	Unknown	**R**GBMC3S	$8.75xE^{-12}$
			CPMMV:P6a:1	CPMMV:P1:1			
			CPMMV:P6a:4	CPMMV:P1:2			

CPMMV:P6a:7

CPMMV:P6a:10

CPMMV:P6b:5

CPMMV:P6b:9

CPMMV:P6c:2

CPMMV:P6c:3

CPMMV:P6c:4

CPMMV:P6c:6

CPMMV:P1:3

CPMMV:P1:4

CPMMV:P1:5

CPMMV:P1:6

CPMMV:P1:7

CPMMV:P1:8

CPMMV:P1:9

CPMMV:P1:10

CPMMV:P3a:1

CPMMV:P3a:2

CPMMV:P3a:3

CPMMV:P3a:4

CPMMV:P3a:5

CPMMV:P3a:6

CPMMV:P3a:7

CPMMV:P3a:8

CPMMV:P3a:9

CPMMV:P3a:10

CPMMV:P3b:1

CPMMV:P3b:2

CPMMV:P3b:3

CPMMV:P3b:4

CPMMV:P3b:5

CPMMV:P3b:6

CPMMV:P3b:8

CPMMV:P3b:10

				CPMMV:P3c:2			
				CPMMV:P3c:3			
				CPMMV:P3c:4			
				CPMMV:P3c:5			
				CPMMV:P3c:6			
				CPMMV:P3c:7			
				CPMMV:P3c:8			
				CPMMV:P3c:9			
				CPMMV:P3c:10			
30	3646	4367*	CPMMV:P6b:9 CPMMV:P6b:5	Unknown	CPMMV:P6c:4 CPMMV:P6a:4	RGMCS**3S**	$4.41xE^{-12}$
31	1014	1135*	CPMMV:P6b:8 CPMMV:P6a:5	CPMMV:P6c:10 CPMMV:P6c:5	Unknown Unknown	**R**GMS3S	$1.59xE^{-11}$

32	862	976	CPMMV:P1:9	Unknown	CPMMV:P6b:10	RGMC**3S**	$7.05xE^{-10}$
33	2929*	3456	CPMMV:P6b:4	Unknown	CPMMV:P6b:10	RGMCS**3S**	$2.20xE^{-11}$
34	3021*	4310*	CPMMV:P6c:8	Unknown	CPMMV:P6b:6 CPMMV:P6b:7	RGMC**3S**	$3.21xE^{-08}$
35	4106*	4786*	CPMMV:P6b:10	Unknown	CPMMV:P6b:6	RGMCS**3S**	$6.60xE^{-08}$
36	1862*	1921*	CPMMV:P6a:4 CPMMV:P6c:2 CPMMV:P6c:3 CPMMV:P6c:4	Unknown	CPMMV:P6a:9 CPMMV:P6a:1	RGBMCS**3S**	$2.16xE^{-07}$
37	3262	3562*	CPMMV:P3c:4 CPMMV:P1:2 CPMMV:P1:9 CPMMV:P1:10 CPMMV:P3a:2	Unknown	CPMMV:P1:4	RMCS**3S**	$1.58xE^{-08}$

CPMMV:P3a:4

CPMMV:P3c:6

CPMMV:P6a:1

CPMMV:P6a:4

CPMMV:P6a:7

CPMMV:P6a:9

CPMMV:P6a:10

CPMMV:P6b:3

CPMMV:P6b:5

CPMMV:P6b:9

CPMMV:P6c:2

CPMMV:P6c:3

CPMMV:P6c:4

			CPMMV:P6c:6				
38	3599*	3651	CPMMV:P3a:7	CPMMV:P3a:4	Unknown	RGMC**3S**	$7.85xE^{-07}$

(*) Indicates that the breakpoint could not be precisely pinpointed.
R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3S, 3Seq.
The reported *P-value* is for the program in bold and underlined type ans is the lowest *P-value* calculated for the event.

S4 table: Recombination detected by RDP in ORF1 of CPMMV populations from the second experiment.

Events	Breakpoints in Alignment		Recombinant Sequence(s)	Parental sequences		Detected by	*P*-value
	Begin	End		Minor	Major		
1	3462	3642	CPMMV:BP2c:8	CPMMV:BP1:5	CPMMV:BP2c:6	**R**GMC3S	6.87E-16
			CPMMV:BP2c:7	CPMMV:BP2a:7	CPMMV:BP2a:1		
			CPMMV:BP2b:1	CPMMV:BP2b:2	CPMMV:BP2a:2		
			CPMMV:BP2b:2	CPMMV:BP2b:4	CPMMV:BP2a:3		
			CPMMV:BP2b:4	CPMMV:BP2b:5	CPMMV:BP2a:4		
			CPMMV:BP2b:8	CPMMV:BP2b:6	CPMMV:BP2a:5		
			CPMMV:BP2b:10	CPMMV:BP2b:9	CPMMV:BP2a:6		
			CPMMV:BP2c:9	CPMMV:BP2b:10	CPMMV:BP2a:8		
				CPMMV:BP1:2	CPMMV:BP2a:10		
				CPMMV:BP1:4	CPMMV:BP2c:1		
				CPMMV:BP1:7	CPMMV:BP2c:2		
				CPMMV:BP1:8	CPMMV:BP2c:3		
				CPMMV:BP1:9	CPMMV:BP2c:4		
				CPMMV:BP1:10	CPMMV:BP2c:7		
2	3540	4036	CPMMV:BP2b:5	CPMMV:BP1:5	CPMMV:BP2c:6	**R**GMC3S	$4.63xe^{-15}$

			CPMMV:BP2b:6	CPMMV:BP1:1	CPMMV:BP2a:1		
			CPMMV:BP2b:9	CPMMV:BP1:2	CPMMV:BP2a:2		
				CPMMV:BP1:3	CPMMV:BP2a:3		
				CPMMV:BP1:4	CPMMV:BP2a:4		
				CPMMV:BP1:6	CPMMV:BP2a:5		
				CPMMV:BP1:7	CPMMV:BP2a:6		
				CPMMV:BP1:8	CPMMV:BP2a:8		
				CPMMV:BP1:9	CPMMV:BP2a:10		
				CPMMV:BP1:10	CPMMV:BP2c:1		
					CPMMV:BP2c:2		
					CPMMV:BP2c:3		
					CPMMV:BP2c:4		
					CPMMV:BP2c:7		
					CPMMV:BP2c:10		
					CPMMV:BP2c:6		
3	2958	4014	CPMMV:BP2b:7	CPMMV:BP1:5	CPMMV:BP2c:6	RGMCS**3S**	$1.19xe^{-18}$
			CPMMV:BP2a:1	CPMMV:BP1:1	CPMMV:BP2a:2		
			CPMMV:BP2a:2	CPMMV:BP1:2	CPMMV:BP2a:3		

			CPMMV:BP2a:3	CPMMV:BP1:3	CPMMV:BP2a:4		
			CPMMV:BP2a:4	CPMMV:BP1:4	CPMMV:BP2a:5		
			CPMMV:BP2a:5	CPMMV:BP1:6	CPMMV:BP2a:6		
			CPMMV:BP2a:6	CPMMV:BP1:7	CPMMV:BP2a:8		
			CPMMV:BP2a:8	CPMMV:BP1:8	CPMMV:BP2a:10		
			CPMMV:BP2a:10	CPMMV:BP1:9	CPMMV:BP2c:1		
			CPMMV:BP2b:3	CPMMV:BP1:10	CPMMV:BP2c:2		
			CPMMV:BP2c:1		CPMMV:BP2c:3		
			CPMMV:BP2c:2		CPMMV:BP2c:4		
			CPMMV:BP2c:3		CPMMV:BP2c:7		
			CPMMV:BP2c:4		CPMMV:BP2c:10		
			CPMMV:BP2c:5		CPMMV:BP2c:6		
			CPMMV:BP2c:7				
			CPMMV:BP2c:10				
4	4120	5337	CPMMV:BP1:2	CPMMV:BP2c:6	CPMMV:BP1:5	**R**GBMCS3S	$6.87x\ e^{-16}$
			CPMMV:BP1:1	CPMMV:BP2a:1	CPMMV:BP1:7		
			CPMMV:BP1:3	CPMMV:BP2a:2	CPMMV:BP1:8		
			CPMMV:BP1:4	CPMMV:BP2a:3	CPMMV:BP1:10		
			CPMMV:BP1:6	CPMMV:BP2a:4			

CPMMV:BP2a:5

CPMMV:BP2a:6

CPMMV:BP2a:7

CPMMV:BP2a:8

CPMMV:BP2a:10

CPMMV:BP2b:1

CPMMV:BP2b:2

CPMMV:BP2b:3

CPMMV:BP2b:4

CPMMV:BP2b:5

CPMMV:BP2b:6

CPMMV:BP2b:7

CPMMV:BP2b:8

CPMMV:BP2b:9

CPMMV:BP2b:10

CPMMV:BP2c:1

CPMMV:BP2c:2

CPMMV:BP2c:3

CPMMV:BP2c:4

				CPMMV:BP2c:5			
				CPMMV:BP2c:7			
				CPMMV:BP2c:8			
				CPMMV:BP2c:9			
				CPMMV:BP2c:10			
5	2954	3461	CPMMV:BP2c:8	Unknown	CPMMV:BP2b:9	**R**GMC3S	6.87xe-16
					CPMMV:BP2a:7		
					CPMMV:BP2b:1		
					CPMMV:BP2b:2		
					CPMMV:BP2b:4		
					CPMMV:BP2b:5		
					CPMMV:BP2b:6		
					CPMMV:BP2b:10		
					CPMMV:BP2c:9		
					CPMMV:BP1:10		
6	3592	4508	CPMMV:BP2a:9	CPMMV:BP1:5	Unknown	RGMCS**3S**	$4.36xe^{-11}$
			CPMMV:BP2c:3	CPMMV:BP1:7			
				CPMMV:BP1:8			
				CPMMV:BP1:9			

(*) Indicates that the breakpoint could not be precisely pinpointed.
R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3S, 3Seq.
The reported *P-value* is for the program in bold and underlined type ans is the lowest *P-value* calculated for the event.

S5 Table: Neutrality tests and analysis of selection site by site for ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 of CPMMV isolates from the second experiment implemented in Datamonkey server.

Genome region	Inoculation	Tajima'D	Fu & Li's D*	Fu & Li's F*	SLAC		FEL		IFEL		REL		PARRIS
					N	P	N	P	N	P	N	P	
ORF1	BP1	-0.02052^{ns}	-0.41837^{ns}	-0.36021^{ns}	-	-	4	-	2	1	-	37	-
	BP2	-2.05656*	-3.24263**	-3.36699**	4	-	21	-	-	6	24	25	-
ORF2	BP1	-0.58262^{ns}	0.20368^{ns}	0.01194^{ns}	-	-	4	-	-	-	- PS	-	-
	BP2	-2.14213*	-3.11216*	-3.28967**	-	-	2	-	-	-	- PS	-	-
ORF3	BP1	-1.79631*	-2.08124*	-2.25642*	-	-	-	-	-	-	- PS	-	-
	BP2	-2.3676***	-3.79378**	-3.92533**			2		-	-	- PS	-	-
ORF4	BP1	-1.56222^{ns}	-1.78443^{ns}	-1.93380^{ns}			1	-	-	-	- PS	-	-
	BP2	-1.50738^{ns}	-2.28108^{ns}	-2.38176	-	-	2	-	-	-	PS	-	-
ORF5	BP1	-0.78755^{ns}	-0.26257^{ns}	-0.44208^{ns}	-	-	1	-	-	-	-	-	-
	BP2	-2.22451***	-4.11900**	-4.13376**	-	-	1	-	-	-	- PS	-	-
ORF6	BP1	-1.56222^{ns}	-1.78443^{ns}	-1.93380^{ns}	-	-	1		-	-	- PS	-	-
	BP2	-1.88948*	-2.99378*	-3.10061*	-	-	1	-	-	-	1	2	-

P<0.05; **P<0.02(P); ***P< 0.001 ns-non significant (P) Sites under positive selection; (N) sites under negative selection; (PS) purifying selection; (-) no site under selection;. [a,b,c,d,e]Codon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood; (FEL); [c] Internal Fixed Effects Likelihood (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning for Robust Inference of Selection (PARRIS).

S6 Table: Description of codons under selection for ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 of CPMMV isolates from P1, P3 and P6 in first experiment implemented in Datamonkey server.

Códons in ORF1 under selection in P1							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
323		323	290	323	308		
371			308	325	381		
793			382	337			
			643	375			
			1083	422			
			1182	790			
			1183	793			
			1184	808			
			1186	975			
				1010			
				1030			
				1047			
				1134			
				1178			
				1185			
				1207			
				1214			
				1247			
				1414			
				1454			
				1486			
				1824			

Códons in ORF1 under selection in P3							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
363		319	47	37	381		11
1257		324	66	65	382		80

1380		357	304	131	1065		81
1387		363	308	141	1101		83
		367	369	192	1103		300
		383	370	207	1489		308
		641	373	246	1787		356
		1380	376	306			360
		1384	380	319			364
			381	324			365
			382	363			368
			698	378			369
			782	383			370
			1070	386			373
			1075	548			376
			1082	658			377
			1083	665			379
			1097	666			380
			1100	674			381
			1101	684			382
			1103	696			649
			1121	728			694
			1142	794			707
			1163	832			740
			1199	872			793
			1202	898			899
			1203	1032			1024
			1205	1034			1054
			1210	1048			1058
			1444	1059			1065
			1475	1067			1066
			1479	1104			1071
			1488	1125			1078
			1489	1137			1080
				1138			1081
				1257			1082
				1267			1083
				1306			1088
				1332			1093
				1344			1097
				1368			1099

				1380			1100
				1384			1101
				1387			1103
				1463			1121
				1520			1210
				1524			1330
				1525			1357
				1538			1363
				1540			1364
				1555			1367
				1577			1371
				1579			1376
				1593			1377
				1663			1381
							1383
							1384
							1385
							1393
							1444
							1475
							1489
							1494
							1495
							1499
							1510
							1583
							1722
							1787

Códons in ORF1 under selection in P6							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
350	365	260	281	24	256	260	250
360	1216	350	365	47	281	350	251
394		360	379	260	365	496	252
496		362	389	295	379	623	253
625		378	697	306	624	630	254

633		394	701	320	702	660	256
660		462	702	350	923	699	257
699		496	703	360	1066	716	258
716		600	726	362	1070	1065	259
829		623	1074	378	1074	1071	262
1121		633	1187	390	1083	1077	263
1140		660	1191	394	1199	1078	265
1150		699	1197	462	1216	1081	266
1435		716	1198	496	1436	1150	267
1437		723	1199	600	1493	1419	268
1513		724	1208	625	1501	1422	269
1515		728	1216	633		1425	270
1518		743	1429	634		1431	271
1523		829		660		1434	272
1663		987		699		1435	274
1672		1117		716		1437	276
1695		1121		723		1513	278
		1140		724		1515	279
		1150		743		1518	280
		1434		829		1522	281
		1435		987		1523	282
		1437		1053		1663	283
		1464		1117		1672	284
		1513		1121			285
		1518		1140			286
		1522		1150			287
		1523		1181			336
		1663		1260			361
		1672		1284			363
		1814		1435			364
				1437			365
				1513			366
				1515			367
				1518			368
				1522			369
				1523			370
				1663			371
				1672			372
				1695			373

				1795			374
				1814			376
							377
							379
							380
							381
							382
							383
							384
							385
							386
							387
							389
							391
							392
							393
							396
							397
							468
							492
							494
							527
							577
							605
							606
							608
							611
							618
							619
							629
							638
							642
							645
							647
							679
							683
							700
							702
							703

719
720
721
725
726
729
749
793
852
916
921
923
950
982
990
1066
1070
1074
1083
1111
1129
1141
1161
1184
1187
1191
1196
1197
1199
1200
1205
1207
1208
1211
1215
1357
1413
1418
1436

							1439
							1465
							1473
							1491
							1493
							1495
							1498
							1501
							1502
							1506
							1721
							1835
							1853
							1857

Códons in ORF2 under selection in P1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				36 224			

Códons in ORF2 under selection in P3							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
			159	36			

Códons in ORF2 under selection in P6							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				157 190			

Códons in ORF4 under selection in P3							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				44			

Códons in ORF4 under selection in P6							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				29 36			

Códons in ORF5 under selection in P1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				171			258

Códons in ORF5 under selection in P6							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
237		237		215		237	
		267		224			
				237			
				267			
				272			
				278			
				286			

Códons in ORF6 under selection in P3							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				20			

Códons in ORF6 under selection in P6							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				37			
				106			

(P) Sites under positive selection; (N) sites under negative selection; Codon-based maximum-likelihood algorithms. Single Likelihood Ancestor Counting (SLAC); Fixed Effects Likelihood; (FEL); Internal Fixed Effects Likelihood (IFEL); Random Effects Likelihood (REL) and Partitioning for Robust Inference of Selection (PARRIS).

S7 table: Description of codons under selection for ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 of CPMMV isolates from BP1 and BP2 in second experiment implemented in Datamonkey server.

Códons in ORF1 under selection in BP1							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
		106	1501	30			6
		319		106			14
				319			38
				1521			124
							214
							642
							657
							869
							890
							924
							1256
							1385
							1397
							1491
							1496
							1499
							1501
							1502
							1504
							1505
							1507
							1508

Códons in ORF1 under selection in BP2							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
453			290	273		974	274
974			869	295		1180	450
1035			1178	318		1181	605
1149			1186	453		1182	615
			1190	938		1184	638

			1216	940		1185	694
				944		1198	818
				974		1201	869
				995		1203	937
				1003		1205	1103
				1035		1208	1108
				1119		1214	1112
				1149		1215	1114
				1176		1217	1120
				1198		1219	1175
				1220		1220	1178
				1255		1222	1186
				1283		1223	1192
				1360		1224	1194
				1432		1225	1195
				1494		1226	1204
						1227	1479
						1228	1481
						1229	1491
							1501

Códons in ORF2 under selection in BP1							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
				3			
				117			
				150			
				152			

Códons in ORF2 under selection in BP2							
Algoritms							
SLAC		**IFEL**		**FEL**		**REL**	
N	**P**	**N**	**P**	**N**	**P**	**N**	**P**
				171			
				229			

Códons in ORF3 under selection in BP1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				43 79			

Códons in ORF4 under selection in BP1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				41			

Códons in ORF4 under selection in BP2							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				19 36			

Códons in ORF5 under selection in BP1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				171			

Códons in ORF5 under selection in BP2							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				149			

Códons in ORF6 under selection in BP1							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				83			

Códons in ORF6 under selection in BP2							
Algoritms							
SLAC		IFEL		FEL		REL	
N	P	N	P	N	P	N	P
				53		53	18 60

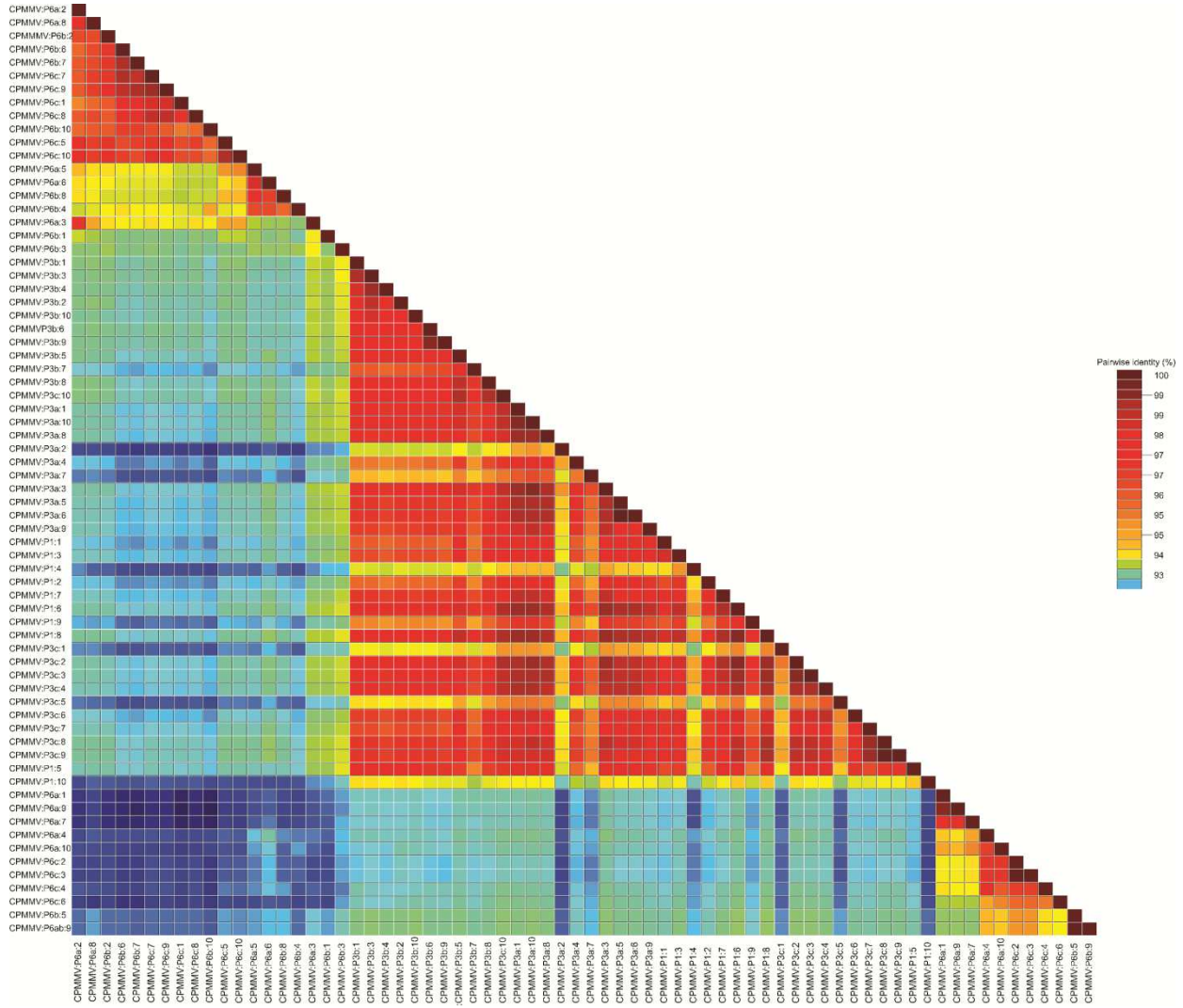
(P) Sites under positive selection; (N) sites under negative selection; Codon-based maximum-likelihood algorithms. Single Likelihood Ancestor Counting (SLAC); Fixed Effects Likelihood; (FEL); Internal Fixed Effects Likelihood (IFEL); Random Effects Likelihood (REL) and Partitioning for Robust Inference of Selection (PARRIS).

S8 Table: Primers used in qRT-PCR viral accumulation and gene expression analyses.

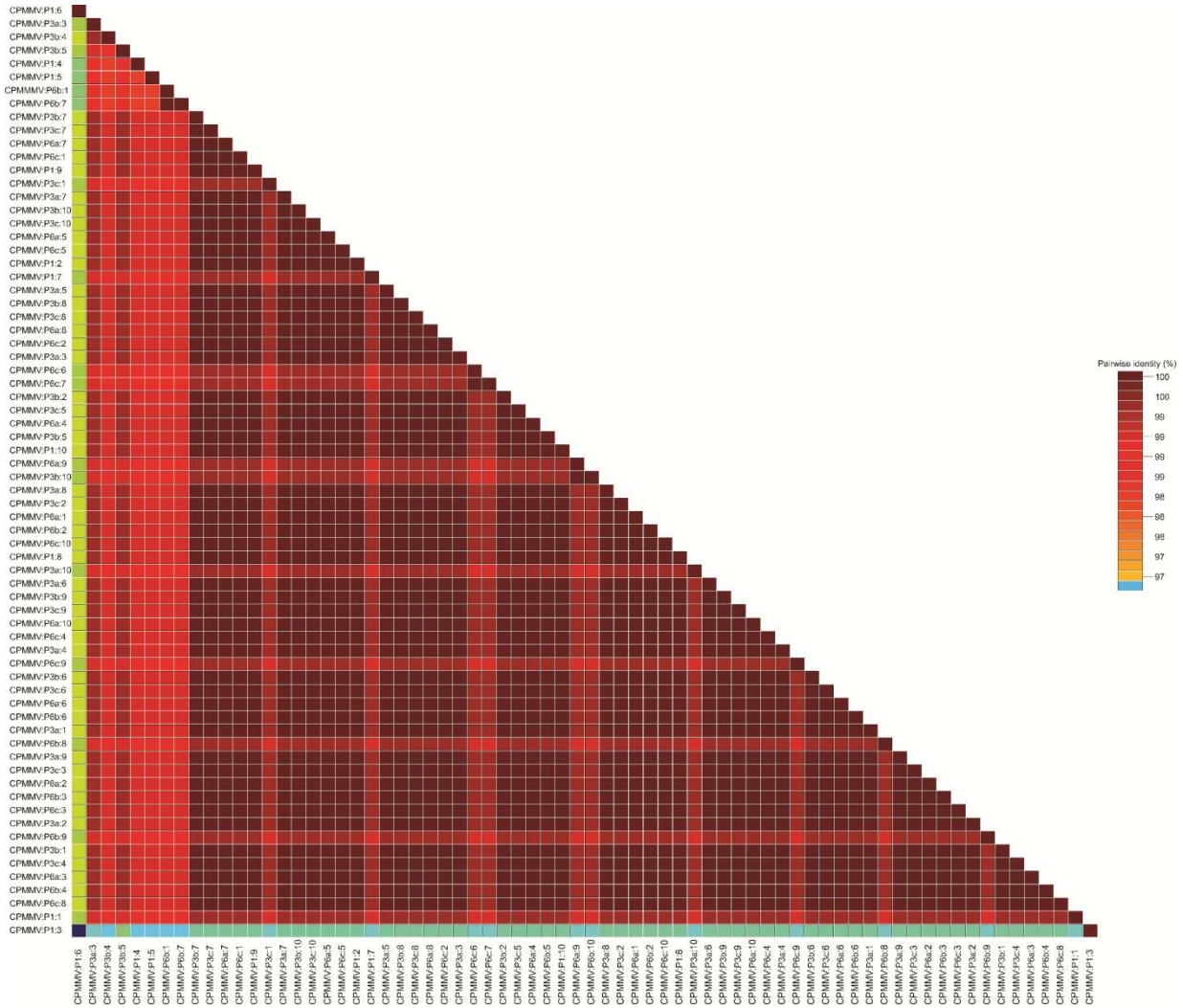
Primer	5’-sequence-3’
GmRAR1F	CCCAGATGTCAGCAGGGTTT
GmRAR1R	GCGGCCTTGTCTCCAATAATAT
GmSGT1F	CATTCACCTCGAACCCAACAA
GmSGT1R	TCAGCAACAGCCTCAGTCAAGT
GmPR1F	AACTATGCTCCCCCTGGCAACTATATTG
GmPR1R	TCTGAAGTGGCTTCTACATCGAAACAA
GmPR4F	TGCTGCACTGATCTACGATTCTC
GmPR4R	TGCGGGTGACAAATACAGGAA
GmNAC6F	CCAACAAAAGCACTTGTGGCA
GmNAC6R	GGACTATTCAACTGACCCAAAAG
GmMAPKKKαF	TCGAAGTGGAAGAAAGGAAAGC
GmMAPKKKαR	TGCGCACAATTGTCCACTATC
GmMEKK2F	ACGATCCGAACGTGGTGAA
GmMEKK2R	CCAGTGACCCTCCGTCCAT
GmHelicaseF	TAACCCTAGCCCCTTCGCCT
GmHelicaseR	GCCTTGTCGTCTTCCTCCTCG
CPMMV CPF	ATAGTGAGATGGCTGATAAACAAAAAC
CPMMV CPR	TTCAGCATCAATGTCTGGAAG

S1 Fig:

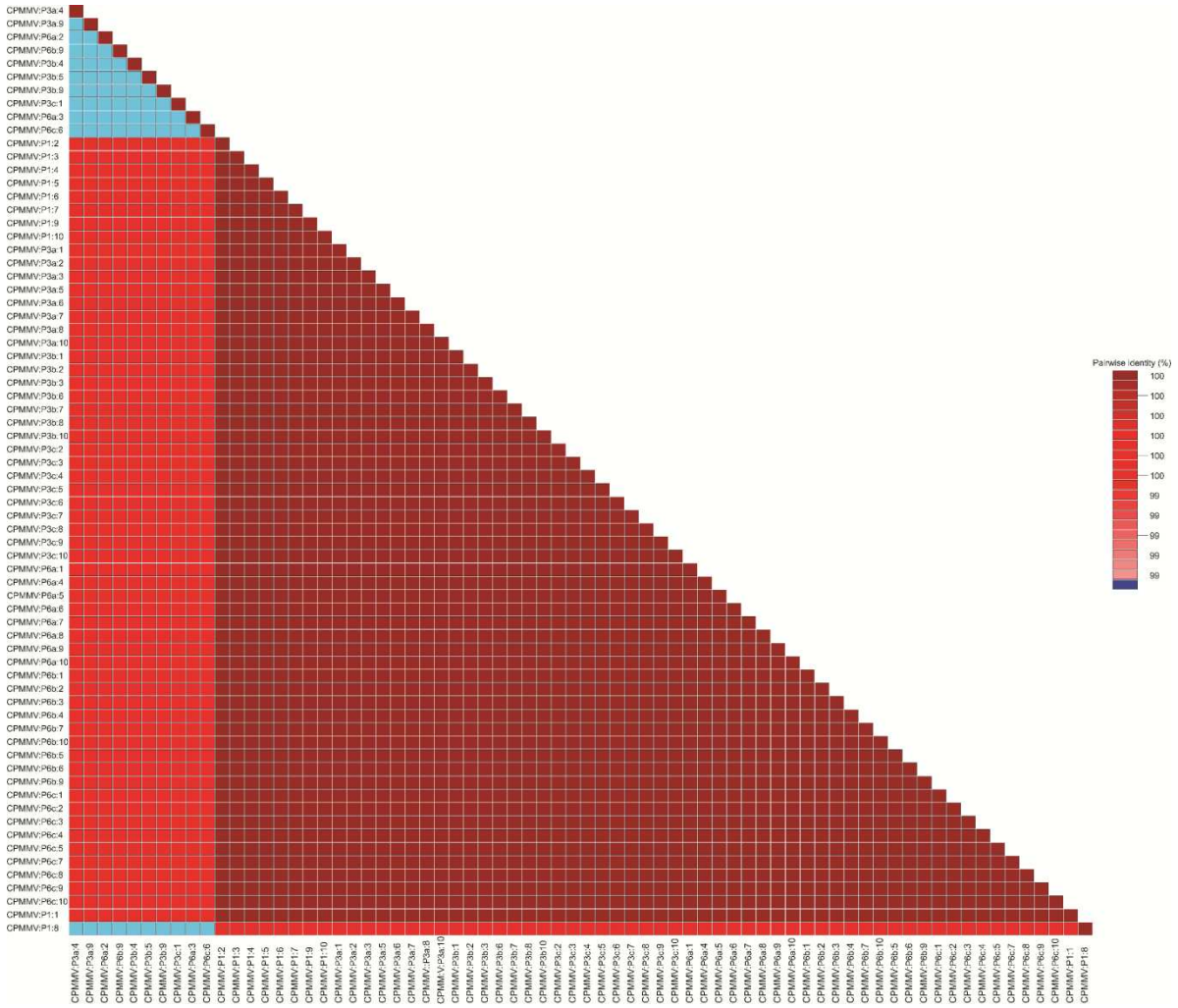
A



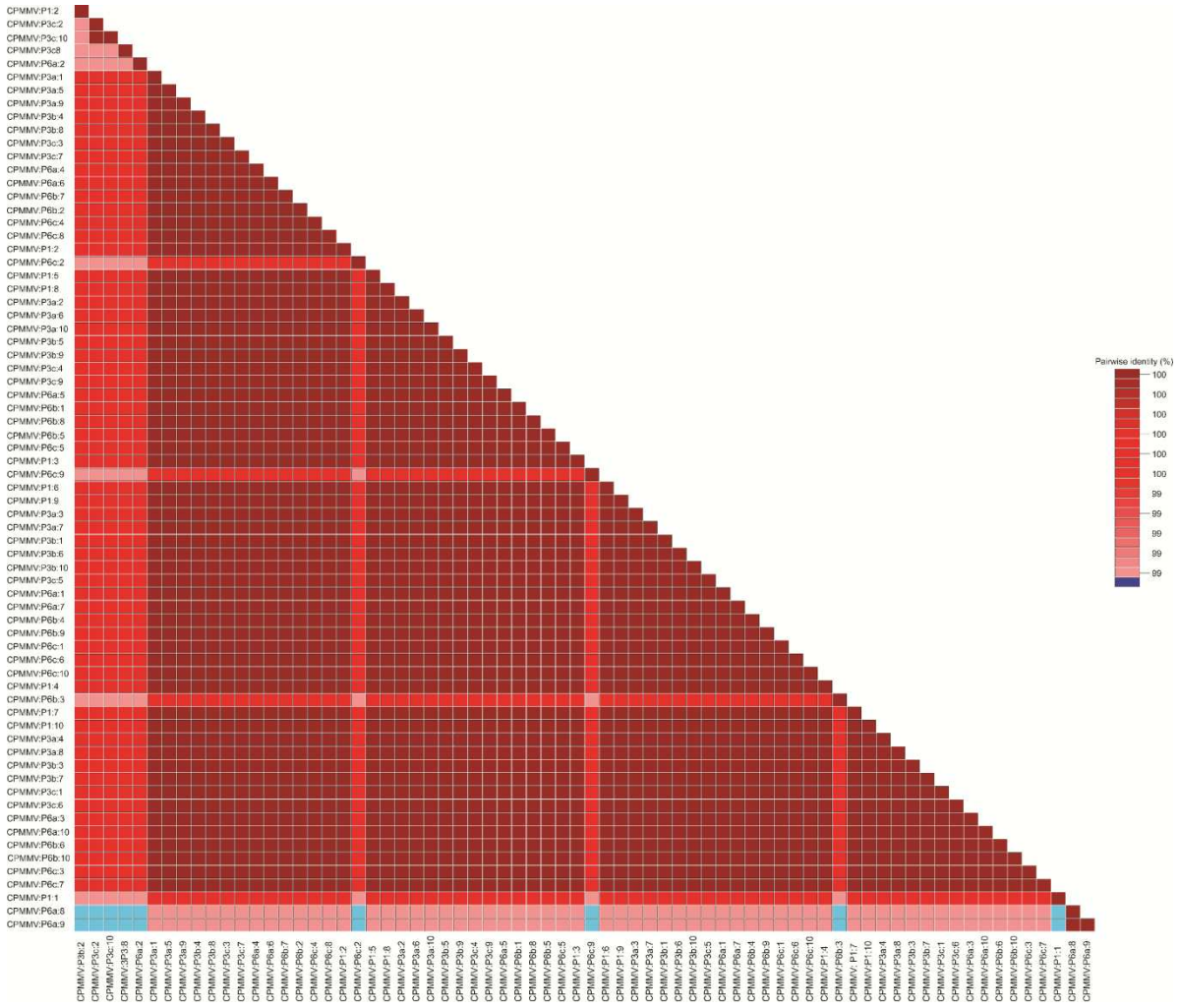
B


CPMMV:P1:6
CPMMV:P3a:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P1:4
CPMMV:P1:5
CPMMMV:P6b:1
CPMMV:P6b:7
CPMMV:P3b:7
CPMMV:P3c:7
CPMMV:P6a:7
CPMMV:P6c:1
CPMMV:P1:9
CPMMV:P3c:1
CPMMV:P3a:7
CPMMV:P3b:10
CPMMV:P3c:10
CPMMV:P6a:5
CPMMV:P6c:5
CPMMV:P1:2
CPMMV:P1:7
CPMMV:P3a:5
CPMMV:P3b:8
CPMMV:P3c:8
CPMMV:P6a:8
CPMMV:P6c:2
CPMMV:P3a:3
CPMMV:P6c:6
CPMMV:P6c:7
CPMMV:P3b:2
CPMMV:P3c:5
CPMMV:P6a:4
CPMMV:P3b:5
CPMMV:P1:10
CPMMV:P6a:9
CPMMV:P3b:10
CPMMV:P3a:8
CPMMV:P3c:2
CPMMV:P6a:1
CPMMV:P6b:2
CPMMV:P6c:10
CPMMV:P1:8
CPMMV:P3a:10
CPMMV:P3a:6
CPMMV:P3b:9
CPMMV:P3c:9
CPMMV:P6a:10
CPMMV:P6c:4
CPMMV:P3a:4
CPMMV:P6c:9
CPMMV:P3b:6
CPMMV:P3c:6
CPMMV:P6a:6
CPMMV:P6b:6
CPMMV:P3a:1
CPMMV:P6b:8
CPMMV:P3a:9
CPMMV:P3c:3
CPMMV:P6a:2
CPMMV:P6b:3
CPMMV:P6c:3
CPMMV:P3a:2
CPMMV:P6b:9
CPMMV:P3b:1
CPMMV:P3c:4
CPMMV:P6a:3
CPMMV:P6b:4
CPMMV:P6c:8
CPMMV:P1:1
CPMMV:P1:3
Pairwise identity (%)
100
100
99
99
99
99
98
98
98
97
97

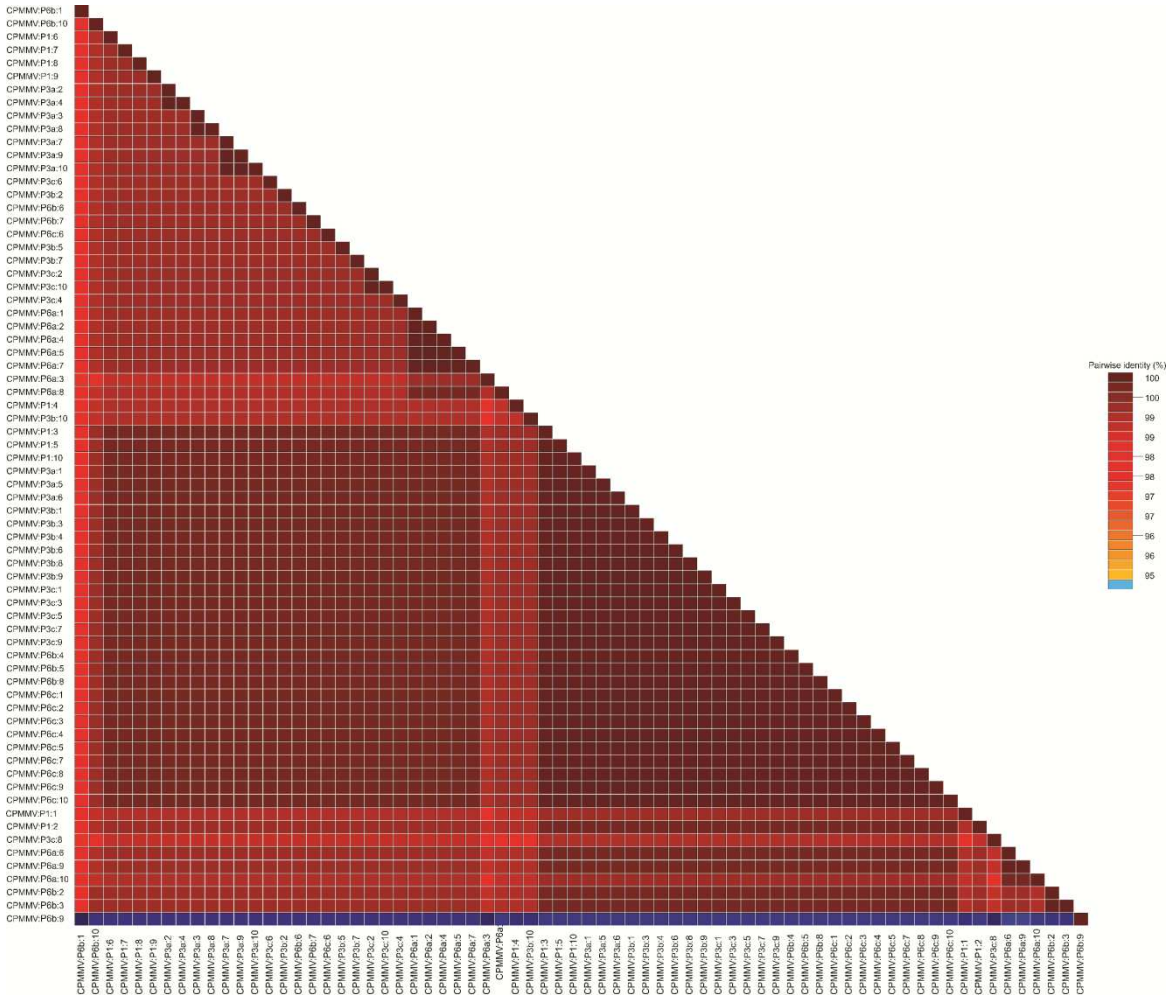
C


CPMMV:P3a:4
CPMMV:P3a:9
CPMMV:P6a:2
CPMMV:P6b:9
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:9
CPMMV:P3c:1
CPMMV:P6a:3
CPMMV:P6c:8
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:2
CPMMV:P3a:3
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:3
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:8
CPMMV:P3b:10
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:10
CPMMV:P6a:1
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6a:7
CPMMV:P6a:8
CPMMV:P6a:9
CPMMV:P6a:10
CPMMV:P6b:1
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:7
CPMMV:P6b:10
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:9
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P1:1
CPMMV:P1:8
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

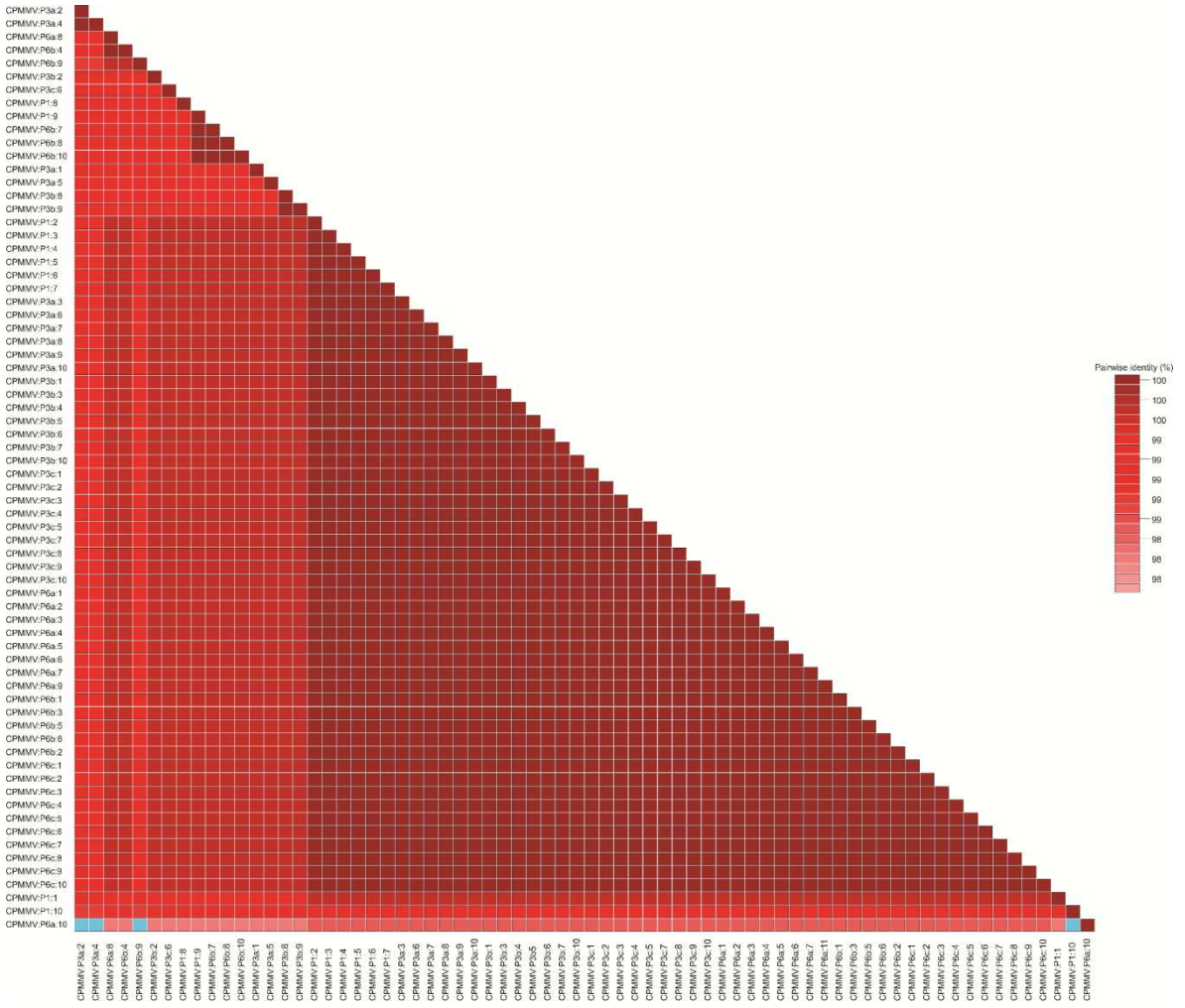
D



E

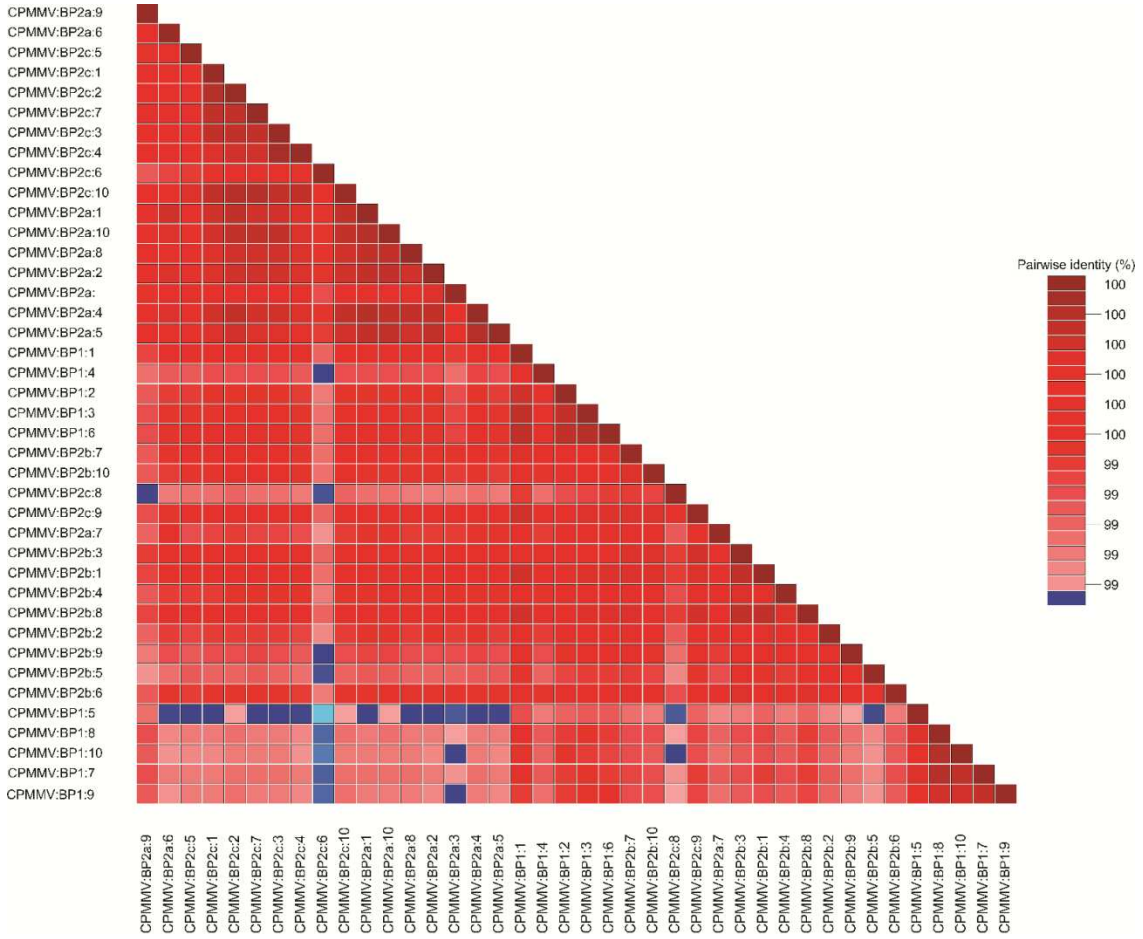

CPMMV:P6b:1
CPMMV:P6b:10
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P3a:2
CPMMV:P3a:4
CPMMV:P3a:3
CPMMV:P3a:8
CPMMV:P3a:7
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3c:6
CPMMV:P3b:2
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6c:6
CPMMV:P3b:5
CPMMV:P3b:7
CPMMV:P3c:2
CPMMV:P3c:10
CPMMV:P3c:4
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:7
CPMMV:P6a:3
CPMMV:P6a:8
CPMMV:P1:4
CPMMV:P3b:10
CPMMV:P1:3
CPMMV:P1:5
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P6b:1
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:6
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P3c:1
CPMMV:P3c:3
CPMMV:P3c:5
CPMMV:P3c:7
CPMMV:P3c:9
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:8
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P1:1
CPMMV:P1:2
CPMMV:P3c:8
CPMMV:P6a:6
CPMMV:P6a:9
CPMMV:P6a:10
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:9
Pairwise identity (%)
100
100
99
99
98
98
97
97
96
96
95

F

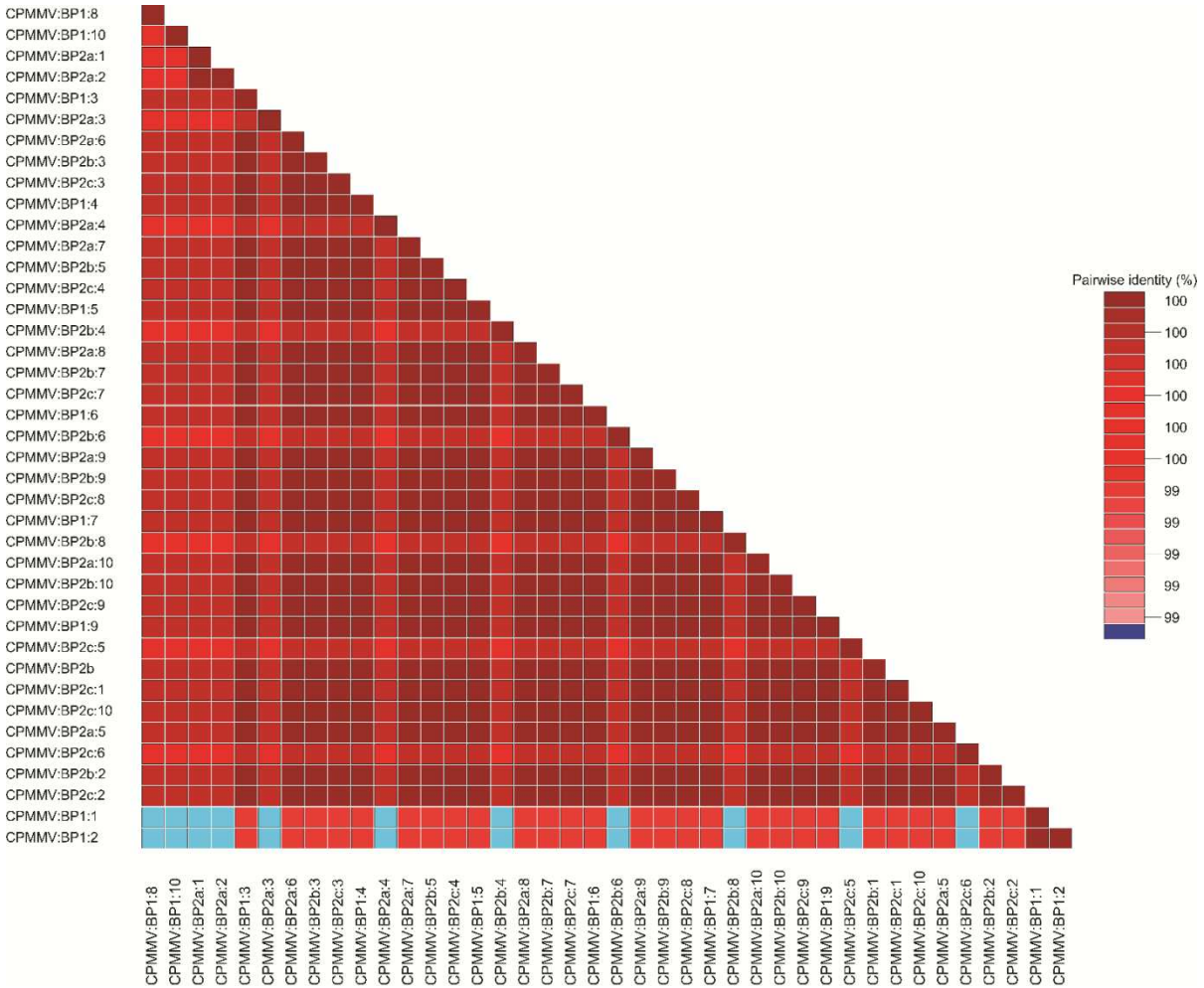

CPMMV:P3a:2
CPMMV:P3a:4
CPMMV:P6a:8
CPMMV:P6b:4
CPMMV:P6b:9
CPMMV:P3b:2
CPMMV:P3c:6
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P6b:7
CPMMV:P6b:8
CPMMV:P6b:10
CPMMV:P3a:1
CPMMV:P3a:5
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P3a:3
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:10
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6a:7
CPMMV:P6a:9
CPMMV:P6b:1
CPMMV:P6b:3
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:2
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:8
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P1:1
CPMMV:P1:10
CPMMV:P6a:10
Pairwise identity (%)
100
100
100
99
99
99
99
99
98
98
98

S2 Fig:

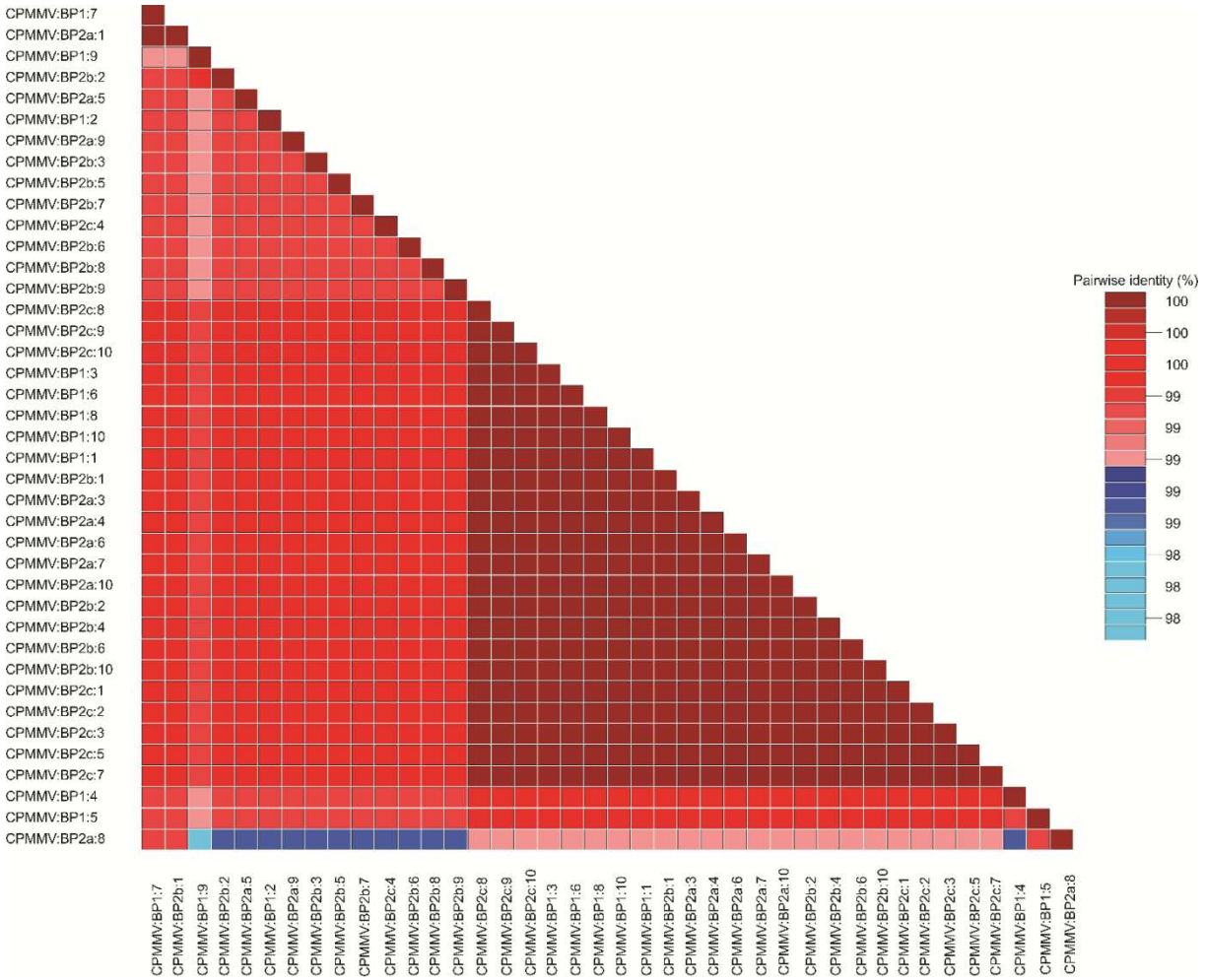
A


CPMMV:BP2a:9
CPMMV:BP2a:6
CPMMV:BP2c:5
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:7
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:6
CPMMV:BP2c:10
CPMMV:BP2a:1
CPMMV:BP2a:10
CPMMV:BP2a:8
CPMMV:BP2a:2
CPMMV:BP2a:
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP1:1
CPMMV:BP1:4
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:6
CPMMV:BP2b:7
CPMMV:BP2b:10
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2a:7
CPMMV:BP2b:3
CPMMV:BP2b:1
CPMMV:BP2b:4
CPMMV:BP2b:8
CPMMV:BP2b:2
CPMMV:BP2b:9
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP1:5
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP1:7
CPMMV:BP1:9
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

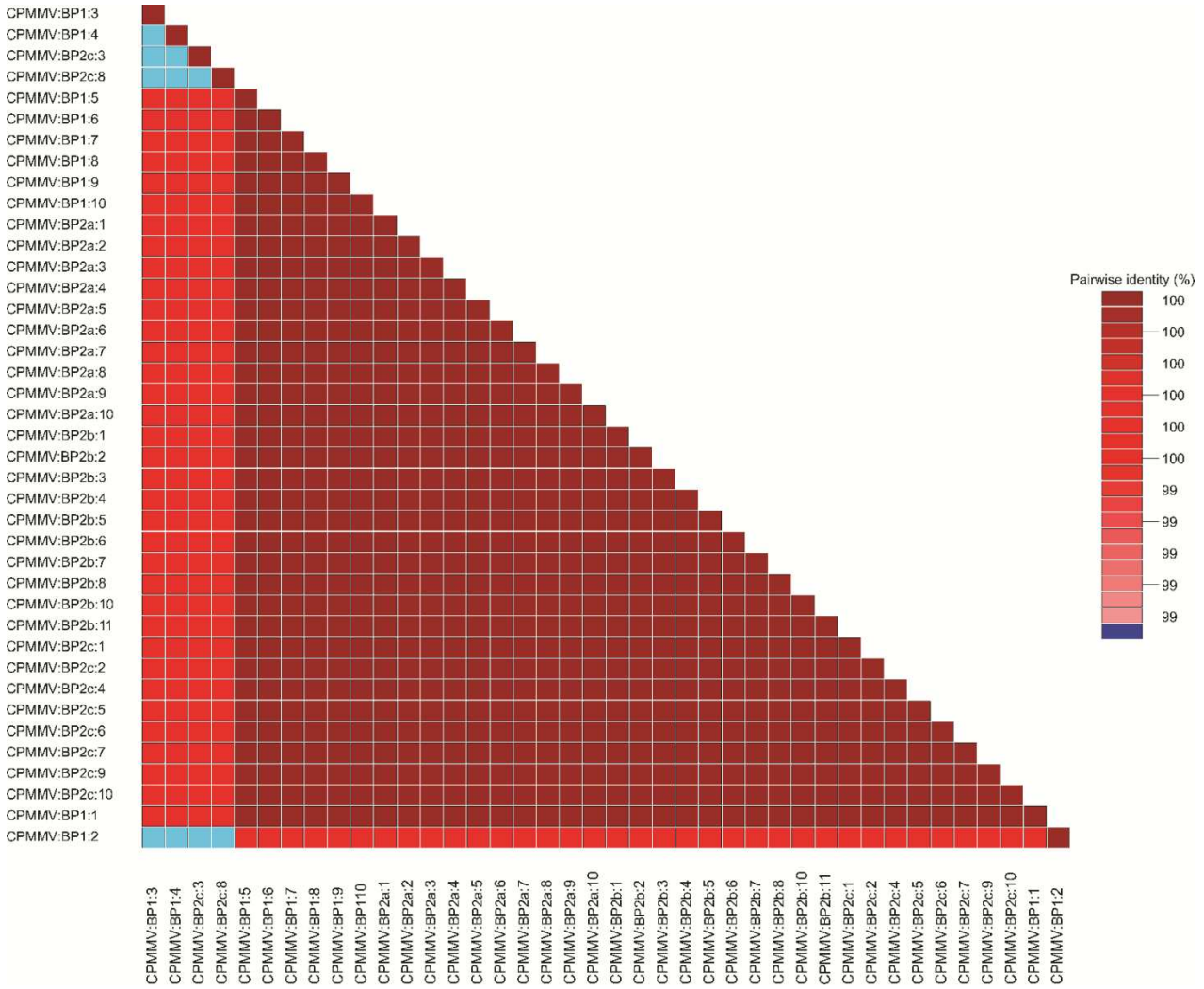
B


CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP1:3
CPMMV:BP2a:3
CPMMV:BP2a:6
CPMMV:BP2b:3
CPMMV:BP2c:3
CPMMV:BP1:4
CPMMV:BP2a:4
CPMMV:BP2a:7
CPMMV:BP2b:5
CPMMV:BP2c:4
CPMMV:BP1:5
CPMMV:BP2b:4
CPMMV:BP2a:8
CPMMV:BP2b:7
CPMMV:BP2c:7
CPMMV:BP1:6
CPMMV:BP2b:6
CPMMV:BP2a:9
CPMMV:BP2b:9
CPMMV:BP2c:8
CPMMV:BP1:7
CPMMV:BP2b:8
CPMMV:BP2a:10
CPMMV:BP2b:10
CPMMV:BP2c:9
CPMMV:BP1:9
CPMMV:BP2c:5
CPMMV:BP2b
CPMMV:BP2c:1
CPMMV:BP2c:10
CPMMV:BP2a:5
CPMMV:BP2c:6
CPMMV:BP2b:2
CPMMV:BP2c:2
CPMMV:BP1:1
CPMMV:BP1:2
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

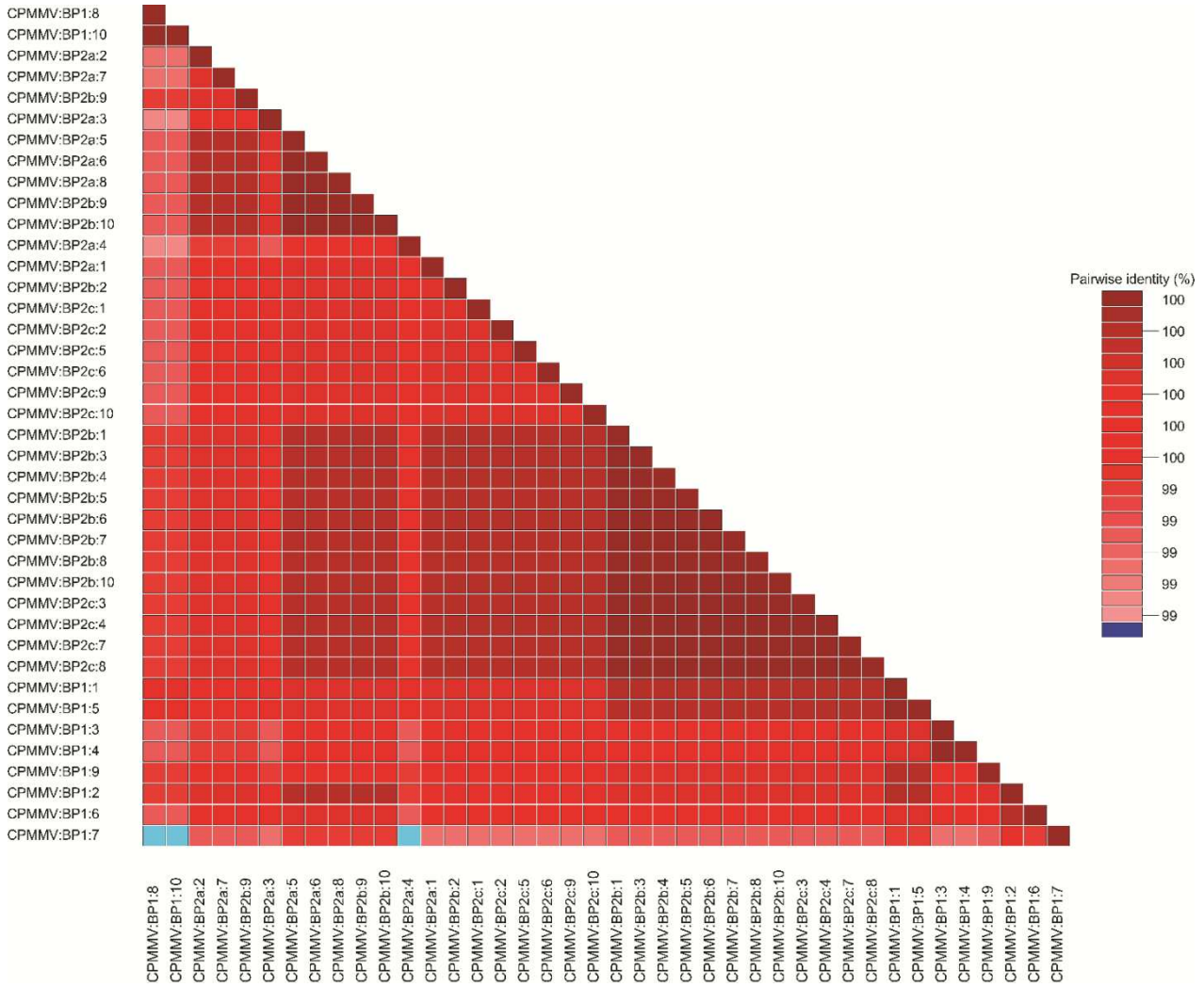
C


CPMMV:BP1:7
CPMMV:BP2a:1
CPMMV:BP1:9
CPMMV:BP2b:2
CPMMV:BP2a:5
CPMMV:BP1:2
CPMMV:BP2a:9
CPMMV:BP2b:3
CPMMV:BP2b:5
CPMMV:BP2b:7
CPMMV:BP2c:4
CPMMV:BP2b:6
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP1:3
CPMMV:BP1:6
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP1:1
CPMMV:BP2b:1
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:10
CPMMV:BP2b:2
CPMMV:BP2b:4
CPMMV:BP2b:6
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:5
CPMMV:BP2c:7
CPMMV:BP1:4
CPMMV:BP1:5
CPMMV:BP2a:8
Pairwise identity (%)
100
100
100
99
99
99
99
99
99
98
98
98

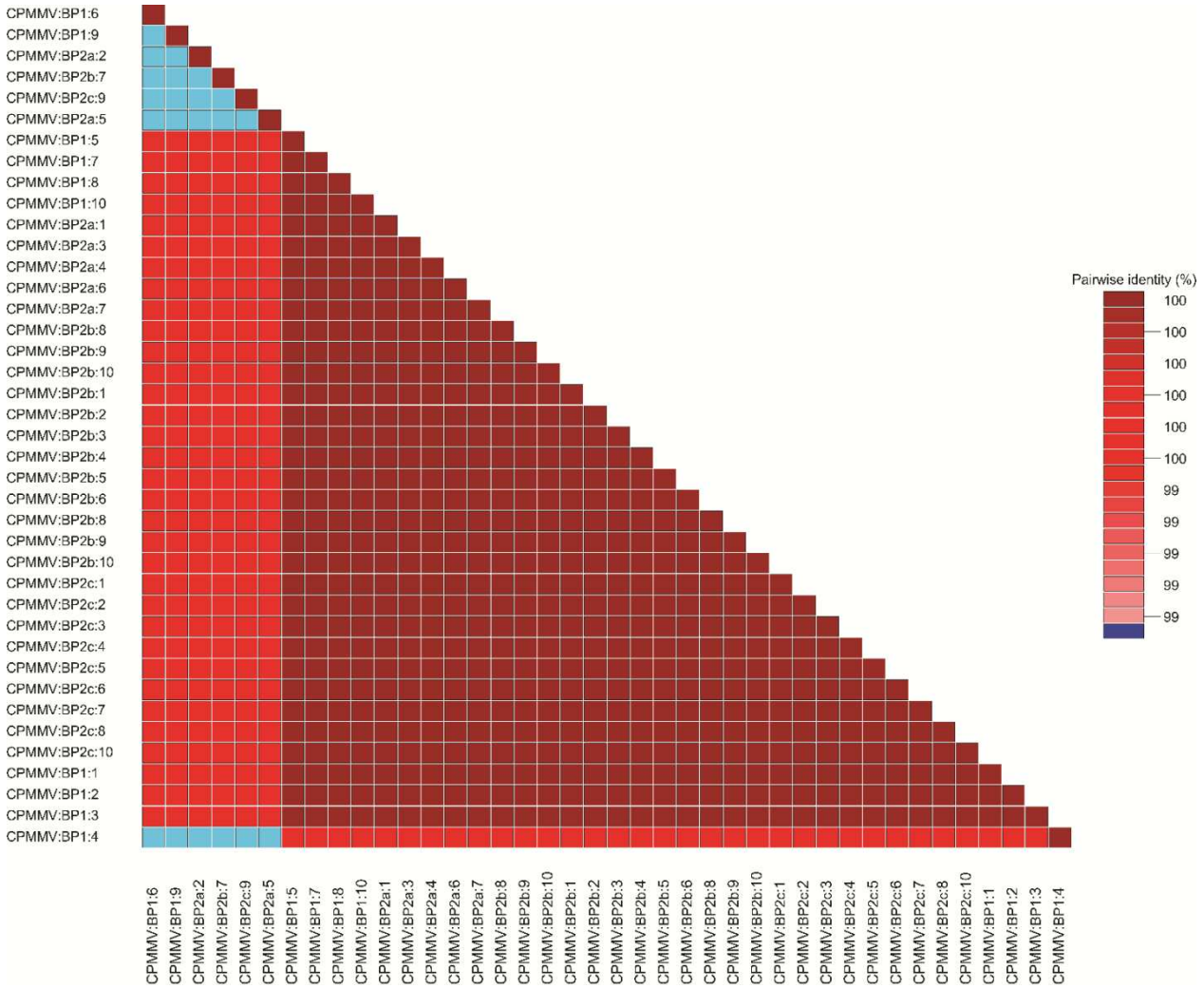
D


CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP2c:3
CPMMV:BP2c:8
CPMMV:BP1:5
CPMMV:BP1:6
CPMMV:BP1:7
CPMMV:BP1:8
CPMMV:BP1:9
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:10
CPMMV:BP2b:11
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP1:1
CPMMV:BP1:2
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

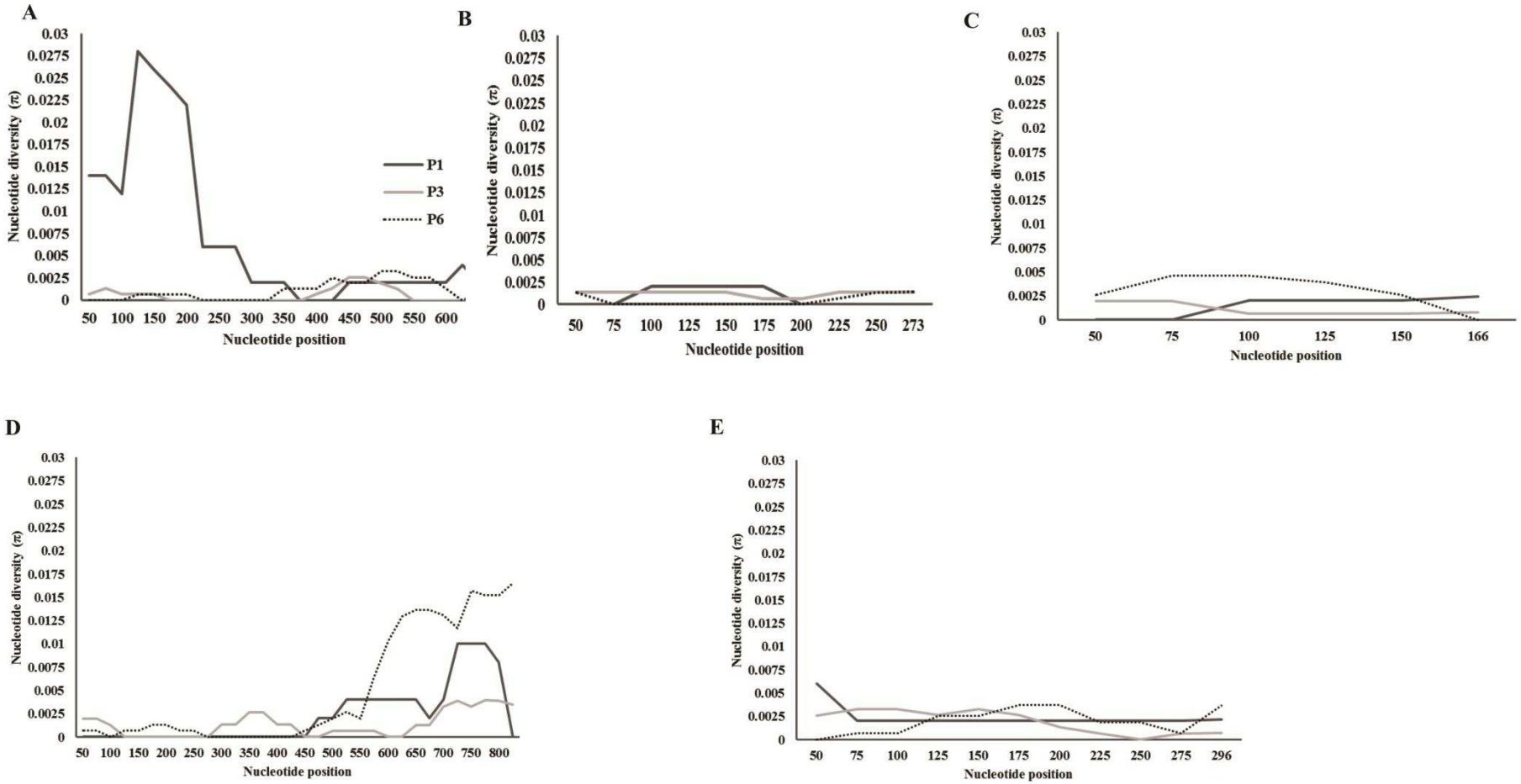
E


CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP2a:2
CPMMV:BP2a:7
CPMMV:BP2b:9
CPMMV:BP2a:3
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2a:4
CPMMV:BP2a:1
CPMMV:BP2b:2
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP2b:1
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:10
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP1:1
CPMMV:BP1:5
CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP1:9
CPMMV:BP1:2
CPMMV:BP1:6
CPMMV:BP1:7
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

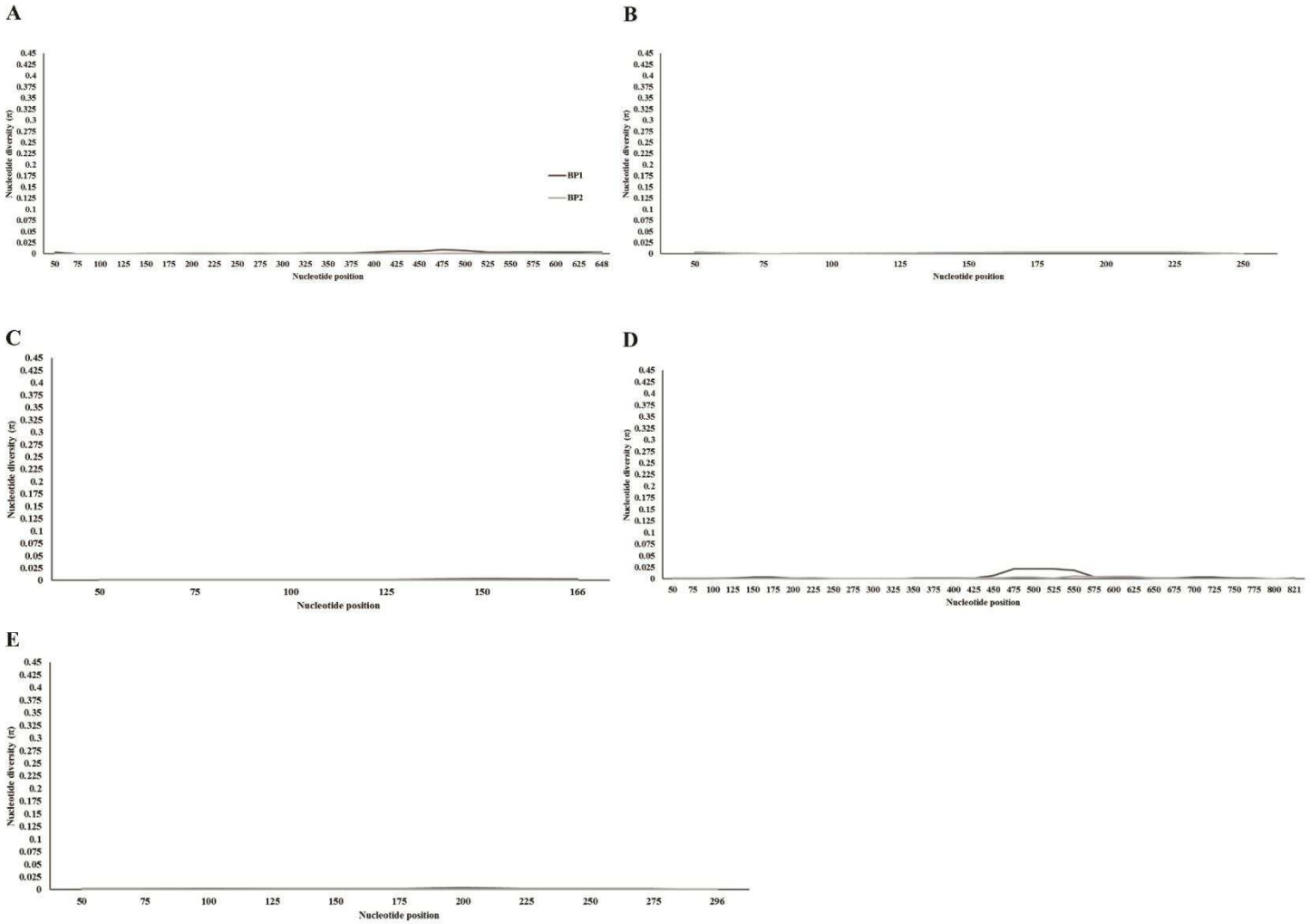
F


CPMMV:BP1:6
CPMMV:BP1:9
CPMMV:BP2a:2
CPMMV:BP2b:7
CPMMV:BP2c:9
CPMMV:BP2a:5
CPMMV:BP1:5
CPMMV:BP1:7
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:10
CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:4
Pairwise identity (%)
100
100
100
100
100
100
99
99
99
99
99

S3 Fig:

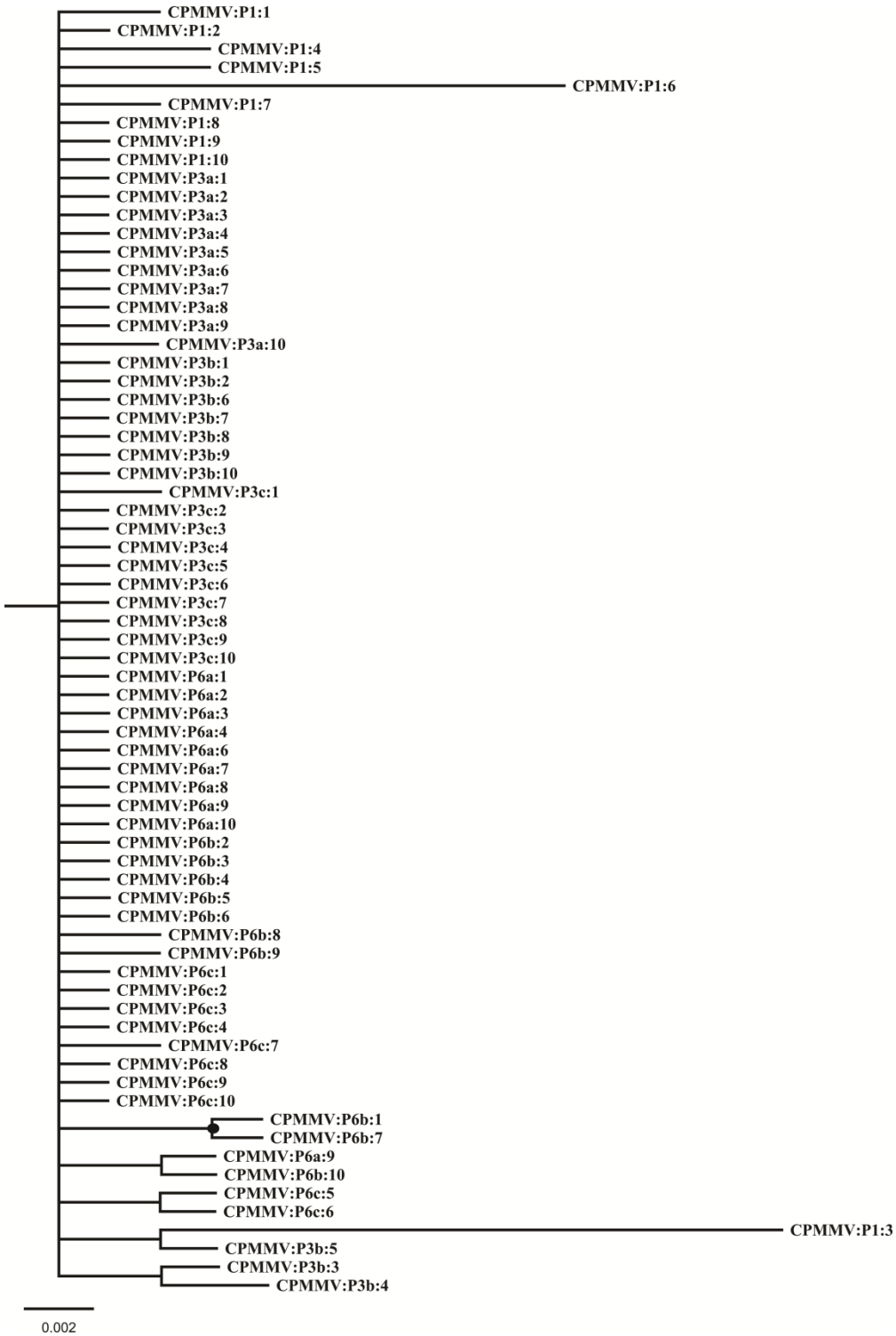

A
B
C
D
E
Nucleotide diversity (π)
Nucleotide position
P1
P3
P6
0.03
0.0275
0.025
0.0225
0.02
0.0175
0.015
0.0125
0.01
0.0075
0.005
0.0025
0
50 100 150 200 250 300 350 400 450 500 550 600
50 75 100 125 150 175 200 225 250 273
50 75 100 125 150 166
50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800
50 75 100 125 150 175 200 225 250 275 296

S4 Fig:

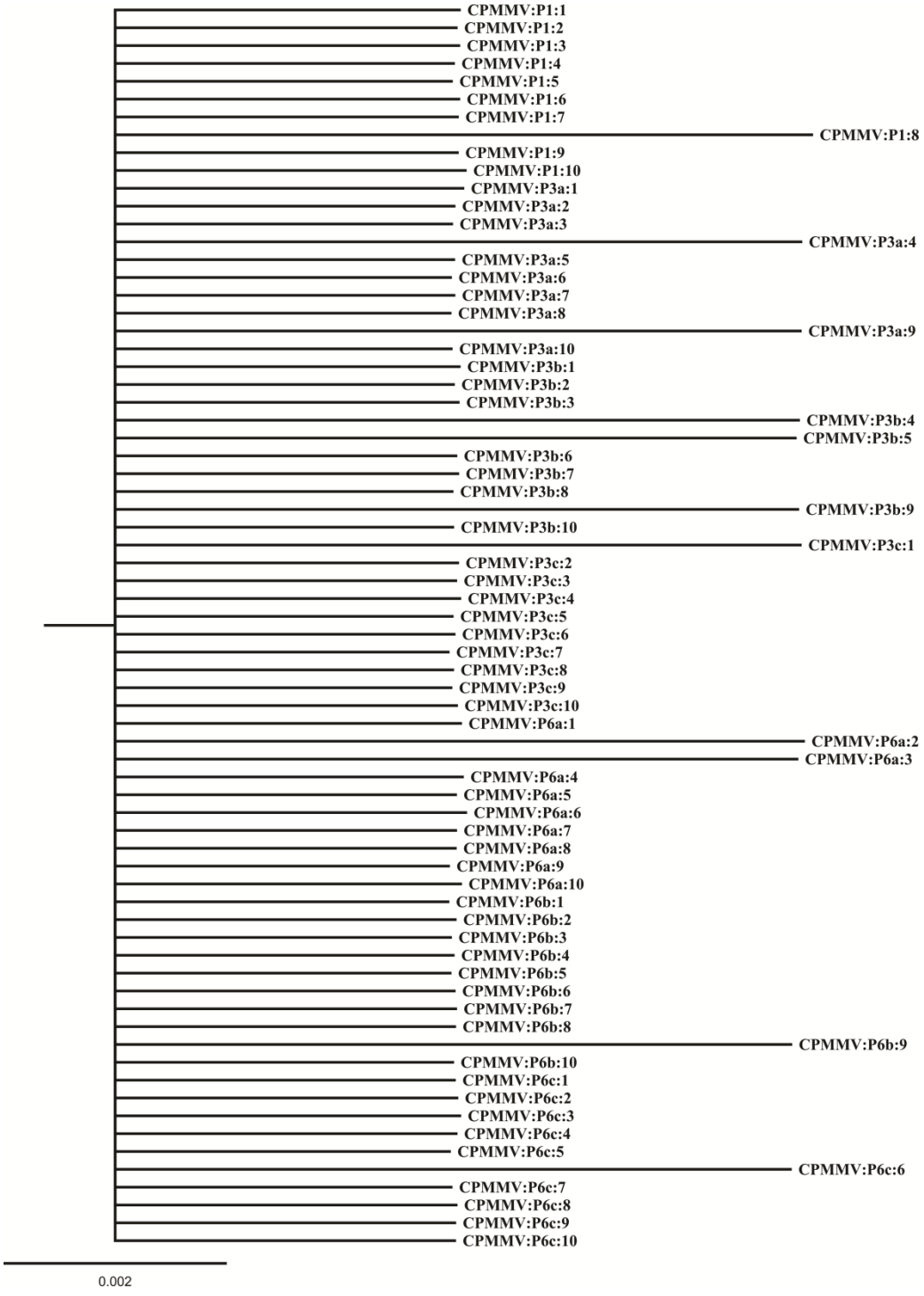


S5 Fig:

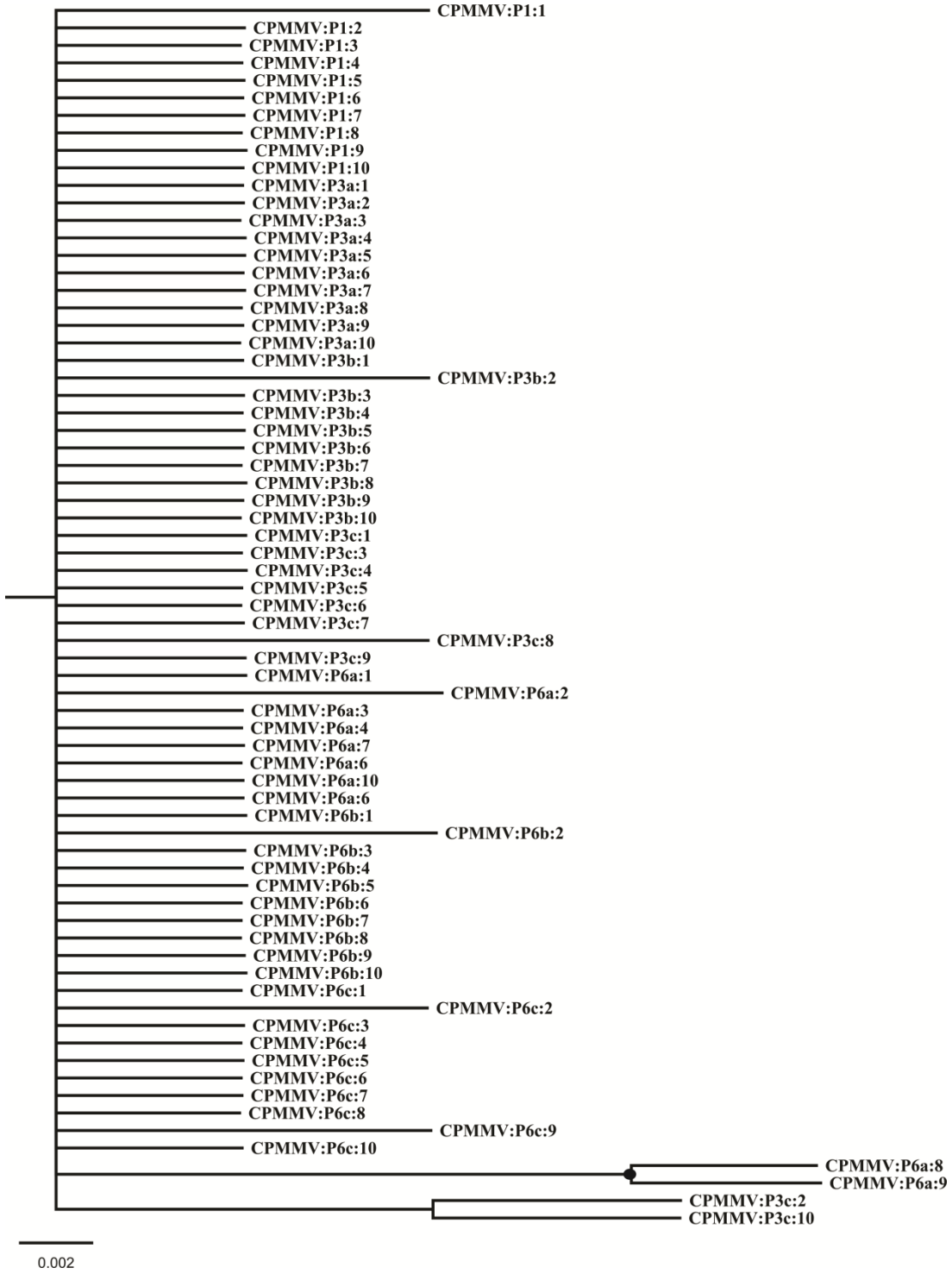
A


CPMMV:P1:1
CPMMV:P1:2
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:2
CPMMV:P3a:3
CPMMV:P3a:4
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:10
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:6
CPMMV:P6a:7
CPMMV:P6a:8
CPMMV:P6a:9
CPMMV:P6a:10
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:8
CPMMV:P6b:9
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P6b:1
CPMMV:P6b:7
CPMMV:P6a:9
CPMMV:P6b:10
CPMMV:P6c:5
CPMMV:P6c:6
CPMMV:P1:3
CPMMV:P3b:5
CPMMV:P3b:3
CPMMV:P3b:4
0.002

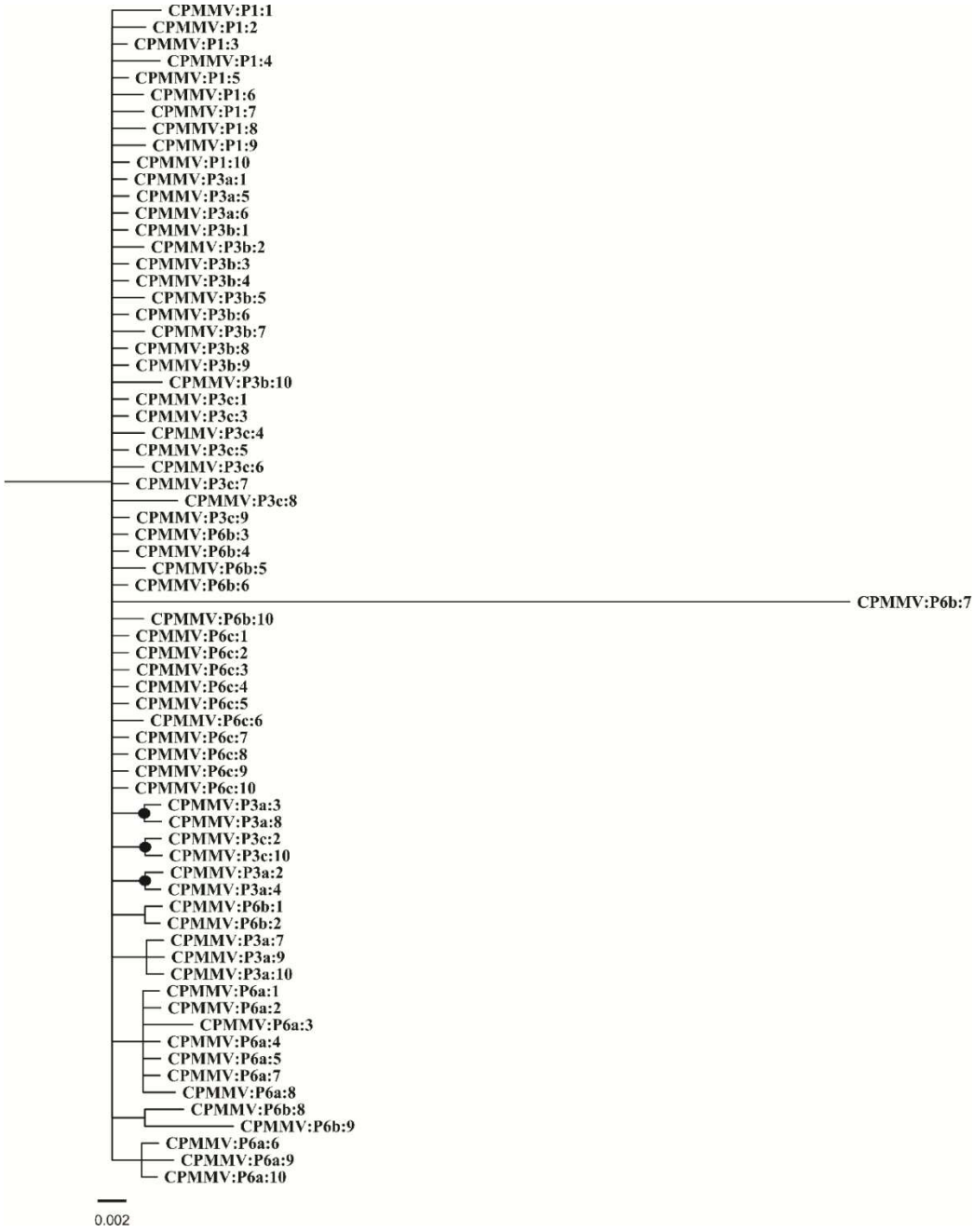
B


CPMMV:P1:1
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:2
CPMMV:P3a:3
CPMMV:P3a:4
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:10
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6a:7
CPMMV:P6a:8
CPMMV:P6a:9
CPMMV:P6a:10
CPMMV:P6b:1
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6b:8
CPMMV:P6b:9
CPMMV:P6b:10
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:6
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
0.002

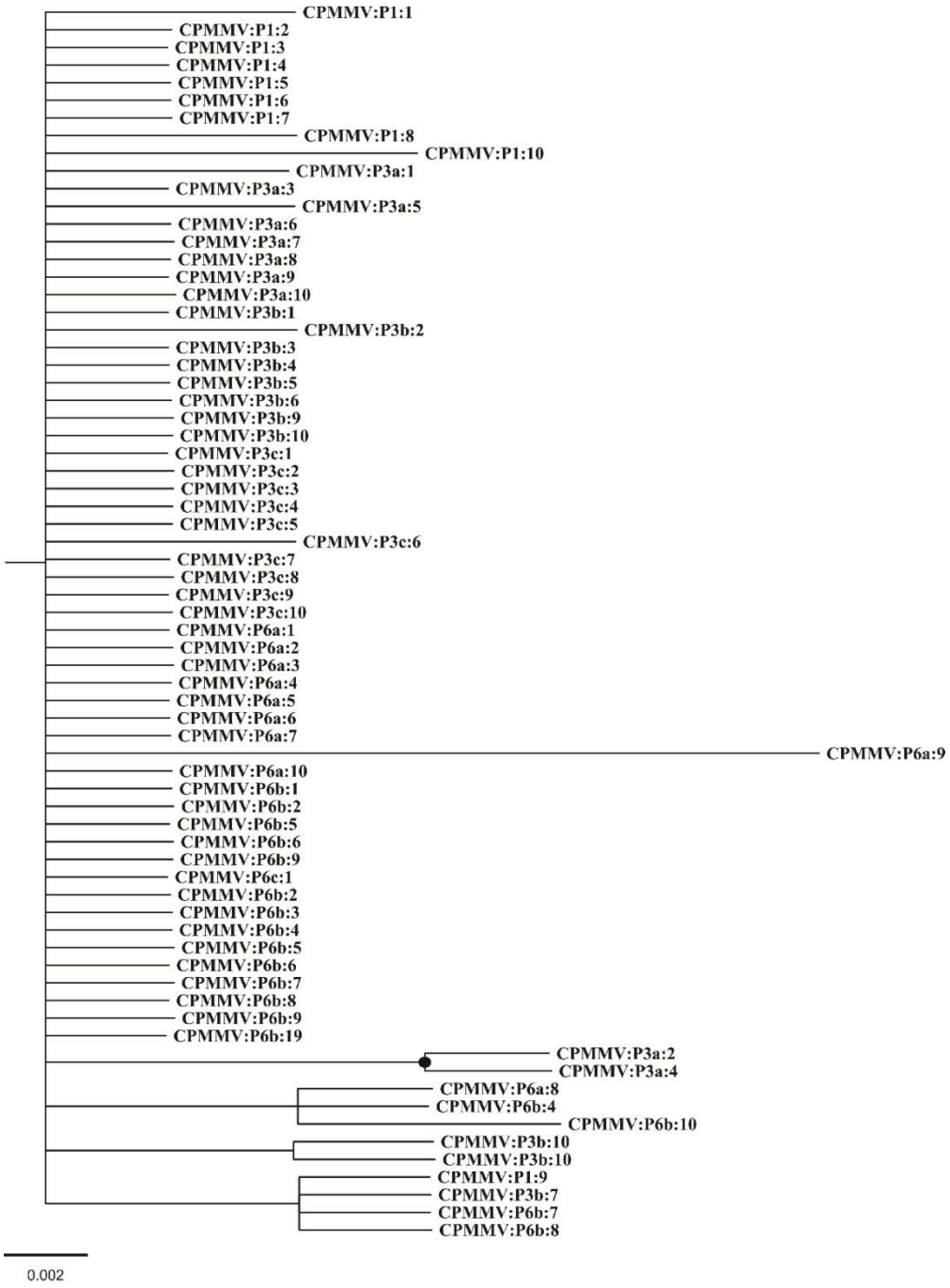
C


CPMMV:P1:1
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:2
CPMMV:P3a:3
CPMMV:P3a:4
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:7
CPMMV:P6a:6
CPMMV:P6a:10
CPMMV:P6a:6
CPMMV:P6b:1
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6b:8
CPMMV:P6b:9
CPMMV:P6b:10
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:6
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P6a:8
CPMMV:P6a:9
CPMMV:P3c:2
CPMMV:P3c:10
0.002

D

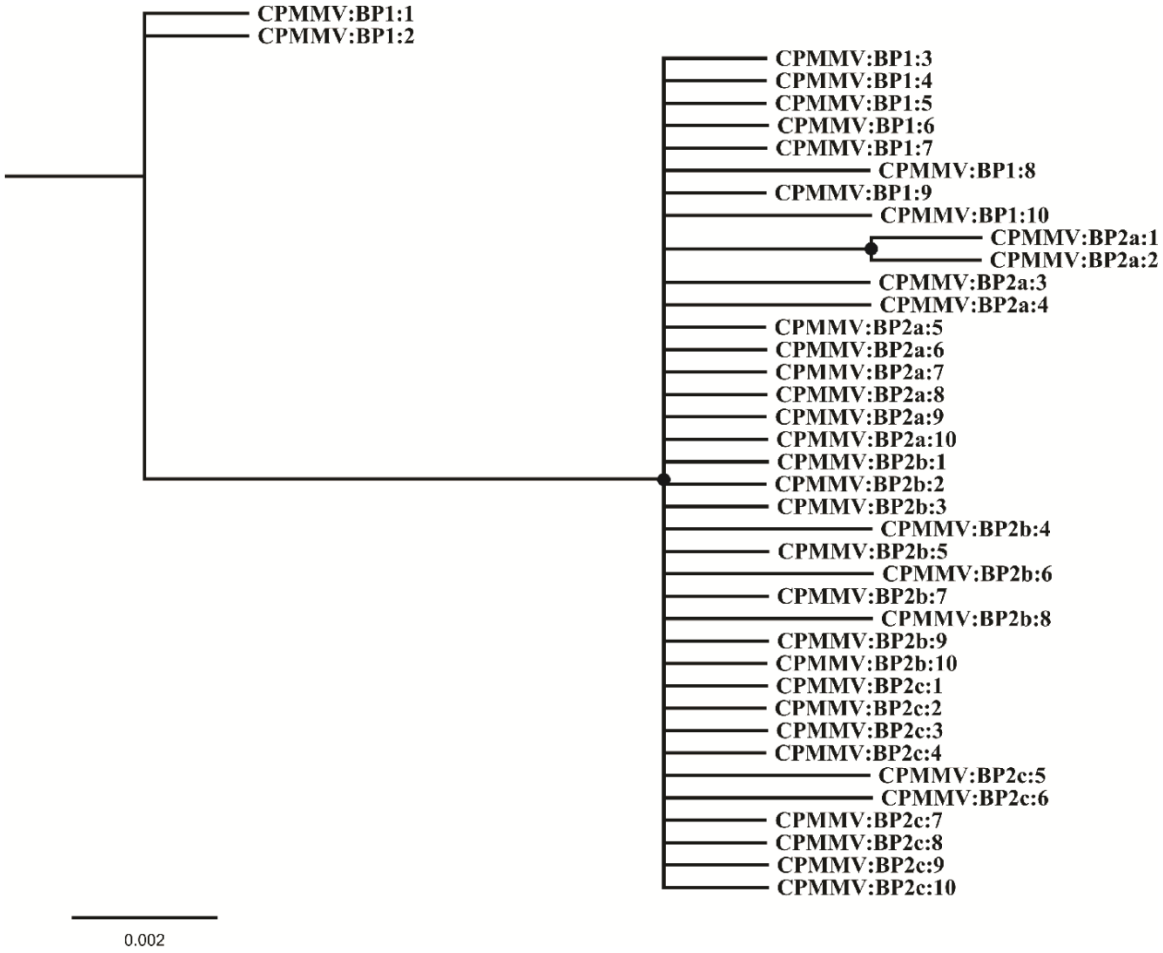

CPMMV:P1:1
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:6
CPMMV:P3b:7
CPMMV:P3b:8
CPMMV:P3b:9
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6b:10
CPMMV:P6c:1
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6c:4
CPMMV:P6c:5
CPMMV:P6c:6
CPMMV:P6c:7
CPMMV:P6c:8
CPMMV:P6c:9
CPMMV:P6c:10
CPMMV:P3a:3
CPMMV:P3a:8
CPMMV:P3c:2
CPMMV:P3c:10
CPMMV:P3a:2
CPMMV:P3a:4
CPMMV:P6b:1
CPMMV:P6b:2
CPMMV:P3a:7
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:7
CPMMV:P6a:8
CPMMV:P6b:8
CPMMV:P6b:9
CPMMV:P6a:6
CPMMV:P6a:9
CPMMV:P6a:10
0.002

E

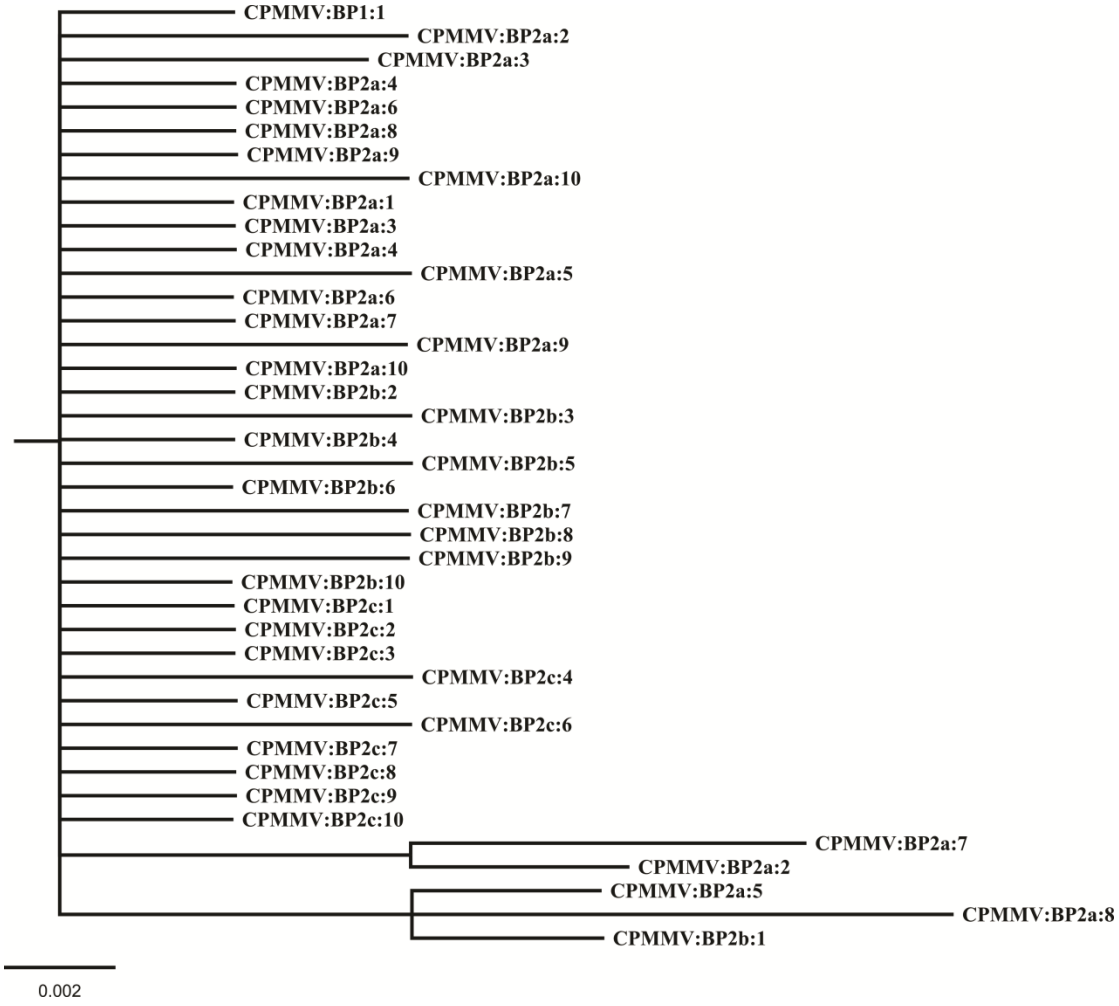

CPMMV:P1:1
CPMMV:P1:2
CPMMV:P1:3
CPMMV:P1:4
CPMMV:P1:5
CPMMV:P1:6
CPMMV:P1:7
CPMMV:P1:8
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:3
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3a:10
CPMMV:P3b:1
CPMMV:P3b:2
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:5
CPMMV:P3b:6
CPMMV:P3b:9
CPMMV:P3b:10
CPMMV:P3c:1
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:10
CPMMV:P6a:1
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:4
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6a:7
CPMMV:P6a:9
CPMMV:P6a:10
CPMMV:P6b:1
CPMMV:P6b:2
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:9
CPMMV:P6c:1
CPMMV:P6b:2
CPMMV:P6b:3
CPMMV:P6b:4
CPMMV:P6b:5
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6b:8
CPMMV:P6b:9
CPMMV:P6b:19
CPMMV:P3a:2
CPMMV:P3a:4
CPMMV:P6a:8
CPMMV:P6b:4
CPMMV:P6b:10
CPMMV:P3b:10
CPMMV:P3b:10
CPMMV:P1:9
CPMMV:P3b:7
CPMMV:P6b:7
CPMMV:P6b:8
0.002

S6 Fig:

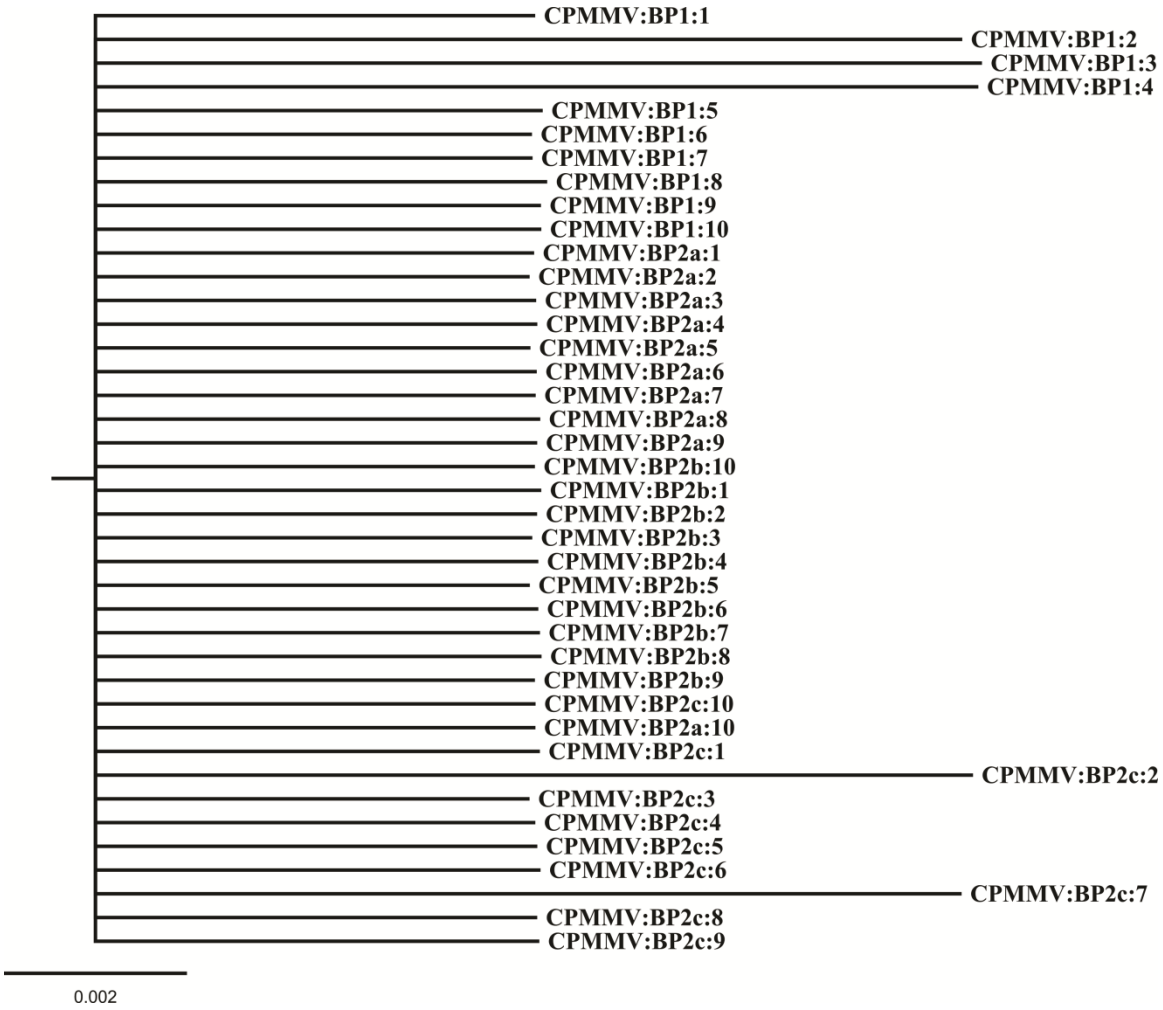
A


CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP1:5
CPMMV:BP1:6
CPMMV:BP1:7
CPMMV:BP1:8
CPMMV:BP1:9
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
0.002

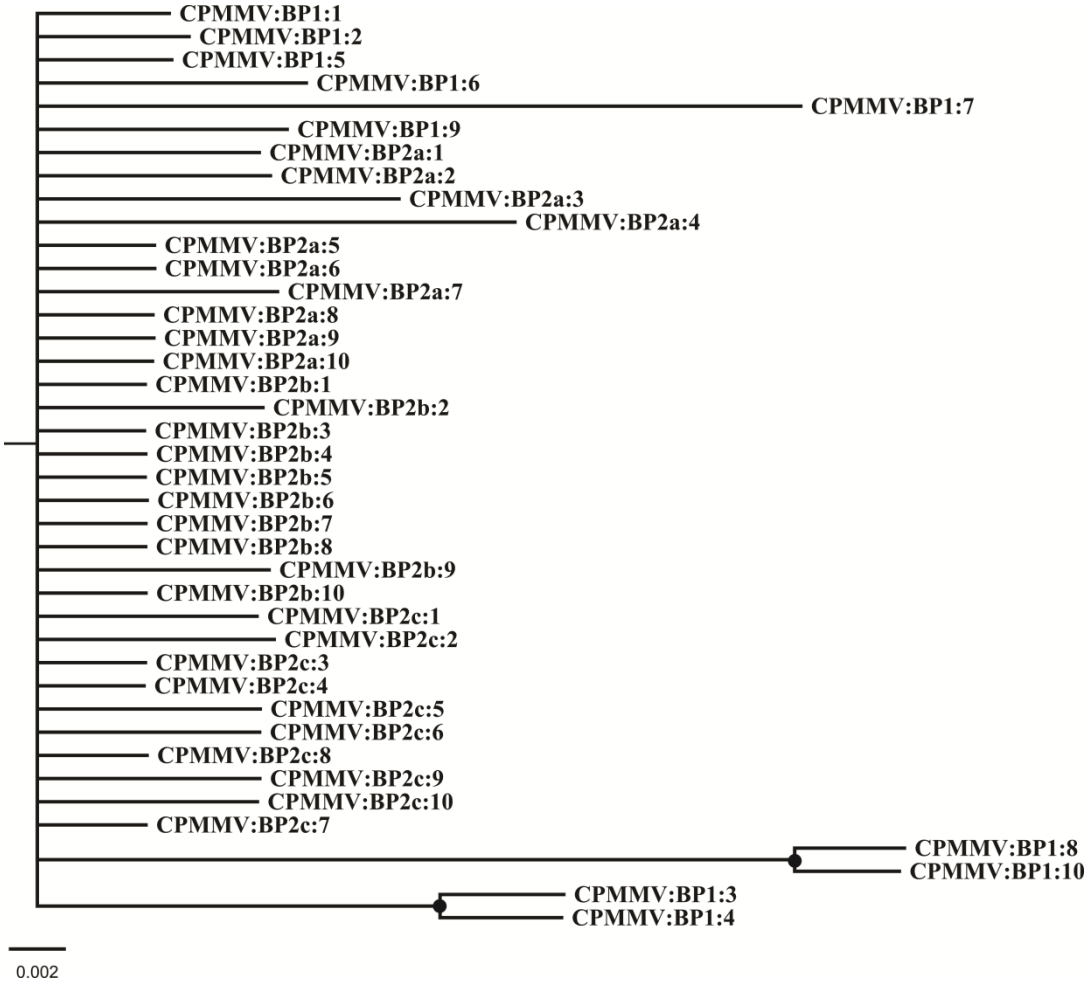
B


CPMMV:BP1:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:6
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2a:1
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP2a:7
CPMMV:BP2a:2
CPMMV:BP2a:5
CPMMV:BP2a:8
CPMMV:BP2b:1
0.002

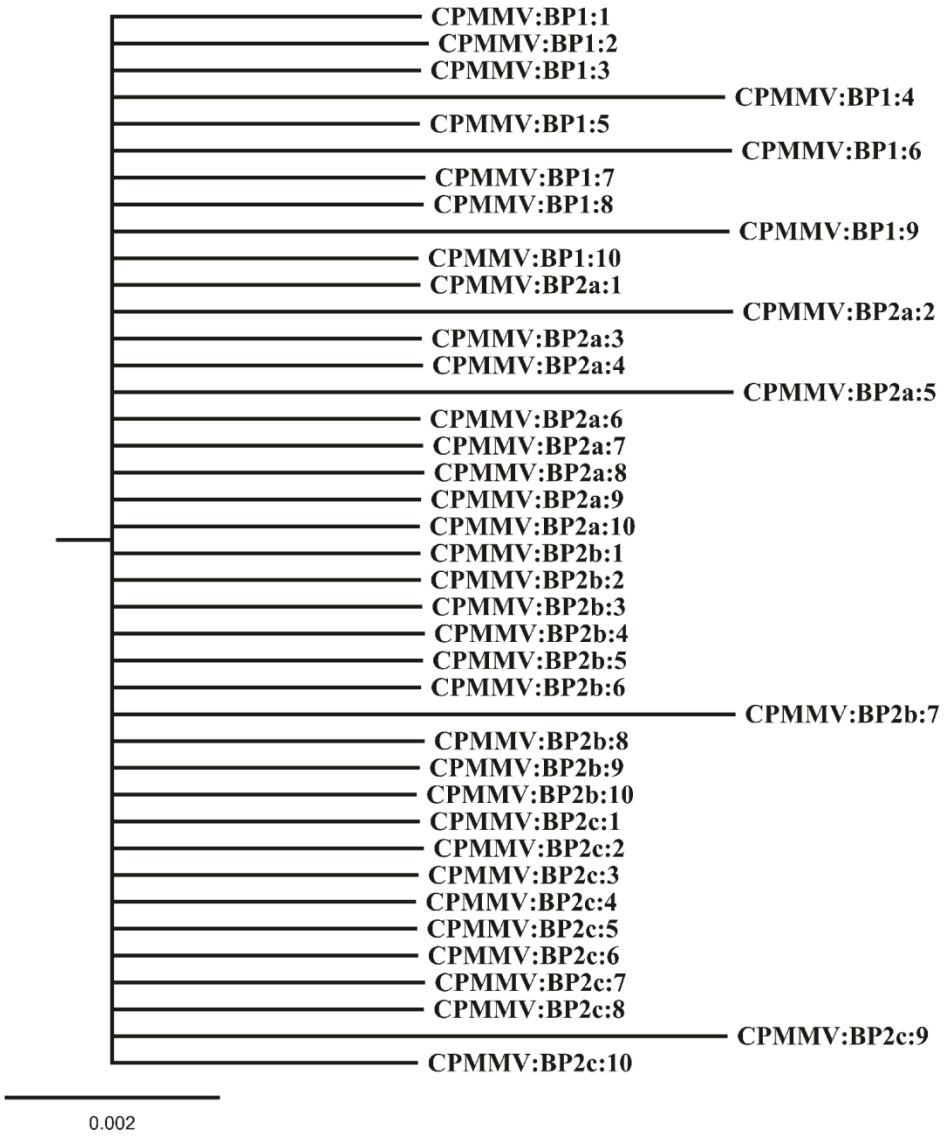
C


CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP1:5
CPMMV:BP1:6
CPMMV:BP1:7
CPMMV:BP1:8
CPMMV:BP1:9
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2b:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2c:10
CPMMV:BP2a:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:9
0.002

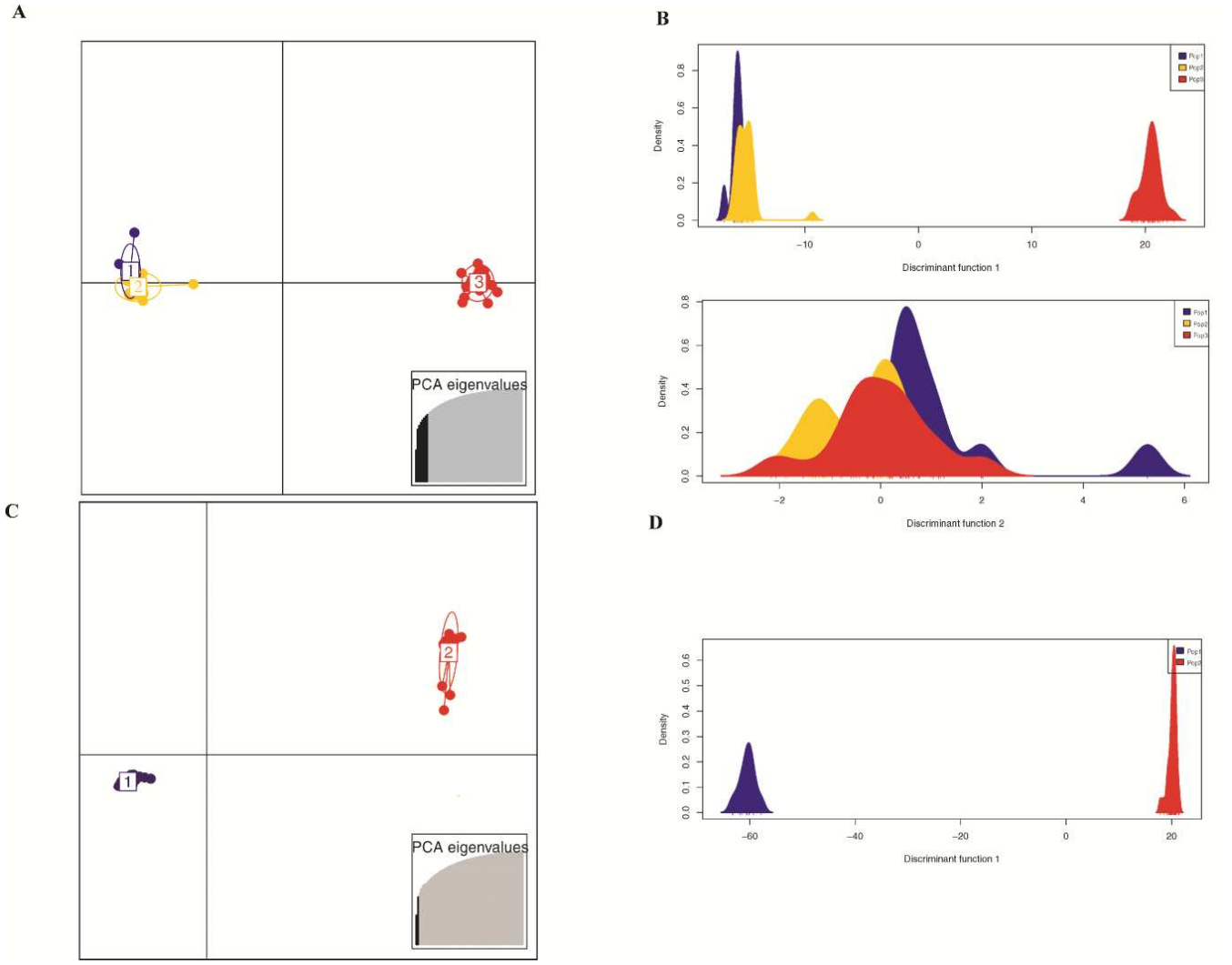
D


CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:5
CPMMV:BP1:6
CPMMV:BP1:7
CPMMV:BP1:9
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP2c:7
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP1:3
CPMMV:BP1:4
0.002

E


CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP1:5
CPMMV:BP1:6
CPMMV:BP1:7
CPMMV:BP1:8
CPMMV:BP1:9
CPMMV:BP1:10
CPMMV:BP2a:1
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2a:8
CPMMV:BP2a:9
CPMMV:BP2a:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:7
CPMMV:BP2b:8
CPMMV:BP2b:9
CPMMV:BP2b:10
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
0.002

S7 Fig:


A
B
C
D
PCA eigenvalues
PCA eigenvalues
Density
Discriminant function 1
Density
Discriminant function 2
Density
Discriminant function 1

S8 Fig:


A
CPMMV:P1:5
CPMMV:P3c:4
CPMMV:P3c:5
CPMMV:P3c:6
CPMMV:P3c:7
CPMMV:P3b:1
CPMMV:P3b:3
CPMMV:P3b:4
CPMMV:P3b:2
CPMMV:P6a:2
CPMMV:P6a:3
CPMMV:P6a:8
CPMMV:P6b:6
CPMMV:P6b:7
CPMMV:P6c:1
CPMMV:P6c:8
CPMMV:P6c:7
CPMMV:P6c:9
CPMMV:P6b:1
CPMMV:P6c:5
CPMMV:P6c:10
CPMMV:P6b:2
CPMMV:P6b:10
CPMMV:P6a:5
CPMMV:P6a:6
CPMMV:P6b:8
CPMMV:P6b:4
CPMMV:P6b:3
CPMMV:P3b:10
CPMMV:P3b:8
CPMMV:P3b:6
CPMMV:P3b:9
CPMMV:P3c:10
CPMMV:P3b:5
CPMMV:P1:2
CPMMV:P1:7
CPMMV:P1:3
CPMMV:P1:6
CPMMV:P1:9
CPMMV:P1:10
CPMMV:P3a:1
CPMMV:P3a:10
CPMMV:P3a:2
CPMMV:P3a:3
CPMMV:P3a:4
CPMMV:P3a:5
CPMMV:P3a:6
CPMMV:P3a:7
CPMMV:P3a:8
CPMMV:P3a:9
CPMMV:P3c:1
CPMMV:P3c:8
CPMMV:P3c:9
CPMMV:P3c:2
CPMMV:P3c:3
CPMMV:P1:8
CPMMV:P3b:7
CPMMV:P6a:1
CPMMV:P6a:9
CPMMV:P6a:7
CPMMV:P6a:4
CPMMV:P6b:5
CPMMV:P6b:9
CPMMV:P6c:4
CPMMV:P6c:6
CPMMV:P6c:2
CPMMV:P6c:3
CPMMV:P6a:10
CPMMV:P1:1
CPMMV:P1:4
0.002
CLUSTER I
CLUSTER II
B
CPMMV:BP2a:9
CPMMV:BP1:1
CPMMV:BP1:2
CPMMV:BP1:3
CPMMV:BP1:4
CPMMV:BP1:6
CPMMV:BP2a:1
CPMMV:BP2a:8
CPMMV:BP2a:10
CPMMV:BP2b:1
CPMMV:BP2b:2
CPMMV:BP2b:3
CPMMV:BP2b:4
CPMMV:BP2b:8
CPMMV:BP2c:1
CPMMV:BP2c:2
CPMMV:BP2c:5
CPMMV:BP2c:6
CPMMV:BP2c:7
CPMMV:BP2c:8
CPMMV:BP2c:9
CPMMV:BP2c:10
CPMMV:BP2a:4
CPMMV:BP2a:5
CPMMV:BP2c:3
CPMMV:BP2c:4
CPMMV:BP2a:6
CPMMV:BP2a:7
CPMMV:BP2b:7
CPMMV:BP2b:10
CPMMV:BP2b:5
CPMMV:BP2b:6
CPMMV:BP2b:9
CPMMV:BP2a:2
CPMMV:BP2a:3
CPMMV:BP1:7
CPMMV:BP1:5
CPMMV:BP1:8
CPMMV:BP1:10
CPMMV:BP1:9
0.002
CLUSTER I
CLUSTER II

CHAPTER 3

The proteins early-responsive to dehydration stress (GmERD4) and inositol methyltransferase (GmIMT) are recruited during the replication of a betaflexivirus in soybean.

Zanardo, L. G., Alves, M. S., Queiroz, S. S., Bruckner, F. P., Silva, F. N., Zerbini, F. M., Carvalho, C. M. The proteins early-responsive to dehydration stress protein (GmERD4) and an inositol methyltransferase (GmIMT) are recruited during the replication of a betaflexivirus in soybean.

Title: The proteins early-responsive to dehydration stress (GmERD4) and an inositol methyltransferase (GmIMT) are recruited during the replication of a betaflexivirus in soybean.

Larissa G. Zanardo; Murilo S. Alves; Sarah S. Queiroz ; Fernanda Prieto Bruckner; Fábio N. Silva; F. Murilo Zerbini; Claudine M. Carvalho.

ABSTRACT

Identification of host factors is an important step to understanding virus infection. We sought to identify soybean candidate factor able to interact with the RNA-dependent RNA polymerase (RdRp) domain of *Cowpea mild mottle virus* (CPMMV) replicase. With two-hybrid screens using a soybean cDNA library, we identified two such candidates: CSC1-like protein ERD4 (Early-responsive to dehydration stress protein-GmERD4) and an Inositol Methyltransferase (GmIMT). We confirmed the interaction using a BiFC assay and evaluated the effects of these proteins on viral infection using soybean protoplasts. Subcellular localizations were predicted using web tools, and were confirmed in localization assays in *Nicotiana benthamiana.* These showed that GmERD4 is localized in the endoplasmic reticulum (ER) and GmIMT in the cytoplasm, while the RdRp domain induced punctate structures in ER and cell membrane. *GmERD4* and *GmIMT* were induced at 3 and 7 days post-inoculation (dpi) in infected plants and the overexpression of *GmERD4* improved viral accumulation at 7 dpi. Both soybean proteins have nucleic acid-binding domains, which, associated with their additional domains, indicates that GmERD4 may be conferring stability to nascent viral RNA, while GmIMT would be involved in viral methylation to avoid degradation of viral RNA. Together, our results suggest that both

GmERD4 and GmIMT are required for CPMMV replication. This is the first study to show the interaction of host proteins with the RdRp domain of a betaflexivirus.

INTRODUCTION

Among the processes of viral infection, replication is central. Through replication, viral genetic material is copied and progeny genomes are generated. The majority of viruses that infect plants have RNA as their nucleic acid. This can be single-stranded RNA (ssRNA), either positive (+) or negative (-), or double-stranded RNA (dsRNA) and can be composed of one or several segments (King *et al.*, 2011; Laliberte & Sanfacon, 2010; Saxena & Lomonossoff, 2014). There is considerable variation in genome organization and virion morphology, though the replication cycle of (+) ssRNA viruses are usually conserved (Ahlquist, 2006). Genome replication involves the copying of the (+) RNA into a complementary (−) RNA strand, that is used for the generation of multiple (+) RNA copies by RNA-dependent RNA-polymerase (RdRp), the core enzyme involved in catalysis of nucleotide polymerization (Laliberte & Sanfacon, 2010).

RNA viruses have small genome and thus encode a limited number of viral proteins. Nevertheless, they are able to establish complex interactions with their host's components, and this capacity can be attributed to the multifunctional nature of their proteins (Vidalain & Tangy, 2010). For RNA viruses in particular, multiple linear motifs of few amino acids allow their proteins to establish interactions with different host factors, facilitation viral replication, movement and spread and overcoming host defense responses (Vidalain & Tangy, 2010). For viral replication, the participation of several factors, of both viral and

host origin, that collectively form replication complexes, is necessary (Laliberte & Sanfacon, 2010). Beyond this, viral infection induces a set of alterations in cell organelles (the formation of spherules, vesicles and multivesicular bodies) that are required for viral genome replication or for virus cell-to-cell movement (Barajas *et al.*, 2009; Rochon *et al.*, 2009; Wang *et al.*, 2005). Positive-strand RNA viruses that infect mammals and plants are models used to explore the endoplasmic reticulum (ER) membrane during viral infections. In this case, the ER would act as an anchor for the formation of a structure harboring the replication complex (viral RNA, host and viral proteins) (Verchot, 2016).

Soybean stem necrosis is one of the most important viral diseases of soybean in Brazil (Almeida, 2008; Almeida *et al.*, 2003). Infected plants show variable symptoms ranging from mosaic and vein clearing to systemic necrosis (Zanardo *et al.*, 2014a; Zanardo *et al.*, 2014b). The causal agent *Cowpea mild mottle virus* (CPMMV) is an emerging virus in soybean plants in Brazil and is still very little studied. It is a member of the genus *Carlavirus* in family *Betaflexiviridae* (Adams & Kreuze, 2016; King *et al.*, 2011). It is a (+) ssRNA with a cap structure [m7GpppG] linked to the 5´ terminus and is polyadenylated at the 3′end, with ~8.2 kb (Menzel *et al.*, 2010). To date, the genomes of nine CPMMV isolates have been completely sequenced (Menzel *et al.*, 2010; Rosario *et al.*, 2014; Zanardo *et al.*, 2014a). The genomic organization includes six open reading frames (ORFs): ORF1 encodes a putative replicase protein, with four conserved motifs: methyltransferase, C23 peptidase, RNA helicase and RNA-dependent RNA polymerase (RdRp). ORFs 2, 3 and 4 encode the triple gene block proteins (TGB1-3p), which are essential for virus movement. ORF 5 encodes the coat protein (CP). ORF 6 encodes a cysteine-rich protein

(CRP) with nucleic acid-binding activity (NABP) (King *et al.*, 2011; Menzel *et al.*, 2010; Zanardo *et al.*, 2014a).

The understanding of the *Betaflexivirus* family of viruses is limited. It is assumed that viral replication is cytoplasmic (Adams *et al.*, 2012), and possibly associated with ER, as with many (+) ssRNA viruses (Bamunusinghe *et al.*, 2009; Chen & Ahlquist, 2000; Dos Reis *et al.*, 2002; Suhy *et al.*, 2000). For the betaflexivirus *Grapevine rupestris stem pitting-associated virus* (GRSPaV – genus *Foveavirus*), it has been shown that the methyltransferase domain of the replicase polyprotein was able to induce punctate bodies in *Nicotiana benthamiana* leaf cells and tobacco protoplasts, which were distributed next to the ER network. Bioinformatic analyses and fractionation analyses of subcellular membranes have shown that the replicase is able to associate with cell membranes, though the nature of the membrane is unknown (Prosser *et al.*, 2015). No other replicase domain for the viruses belonging to this family has been studied.

In addition to determining the functions of each viral replicase domain, it is also important to identify the interactions between these domains and host factors, as only the involvement of viral proteins in betaflexivirus replication process is known (King *et al.*, 2011). We expect that a better understanding of the mechanisms of *Betaflexiviridae* replication will be important scientifically and economically for all crops infected by this group of viruses. All discoveries will help in the development of effective control methods for the virus. We therefore sought to identify host factors able to interact with the RdRp domain of betaflexivirus CPMMV, to study its involvement in viral replication, and determine intracellular localization of RdRp domain and host proteins.

MATERIAL AND METHODS

Viral isolate

The isolate CPMMV:BR:MG:09:2 collected in Capinópolis-MG in 2009, and previously described both molecularly and in terms of its host range Zanardo *et al* (2014a) was used for all experiments. This isolate has been maintained in soybean plants cv. CD206 by successive sap inoculations using inoculation buffer [potassium phosphate buffer (0.2 M, pH 7.2) with 1% sodium bisulfite].

Plant material

Soybean plants cv. CD206 and *Nicotiana benthamiana* were used in all experiments. They were planted and kept in a greenhouse with average daily temperatures of 26 ± 2°C. Viral inoculations in soybean were conducted after 14 days (VC stage) or five days (exclusive to protoplast experiment) of development as described above. *Nicotiana benthamiana* plants were kept in the greenhouse for 28 days and then transferred to growth chambers with controlled temperature (22°C) and photoperiod (14h/10h light/dark). The plants were maintained under these conditions for three days and were then agroinoculated.

RNA extraction and RT-PCR

Total RNA for all experiments was extracted from 100 mg of leaf tissue of soybean plants cv. CD206, infected or uninfected, using the RNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. Reverse transcription (RT) for all experiments was performed using Superscript III Reverse Transcriptase (Invitrogen), according to the manufacturer's protocol, using 500 ng of total RNA and the oligonucleotide ORF6R (for

viral amplifications) (Zanardo *et al.*, 2014a) or oligo(dT) (for host amplifications) (Table S1). PCR amplifications were carried out using Platinum High Fidelity DNA polymerase (Invitrogen), and consisted of 35 cycles of the following profile: 94°C for 1 min, annealing at the temperatures listed in Table S1, elongation at 72°C for 1–2 min, depending on the expected size of the fragment to be amplified, and a final extension step at 72°C for 10 min. The primers and RT-PCR amplifications conditions are listed in Table S1. All amplified PCR products for cloning were gel-purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare).

Yeast two-hybrid library construction

RNA isolation and amplification

The library was prepared from soybean leaf samples at 0, 3, 7 and 14 days post-inoculation (dpi). Leaves of three soybean plants for each time point were collected and pooled for RNA extraction. Total RNA extraction was conducted using the RNAeasy Plant Mini kit (Qiagen) and then submitted to purification using the NucleoTrap mRNA kit (Macherey-Nagel) for messenger RNA (mRNA) enrichment. For cDNA synthesis, 0.126 µg of mRNA and CDSIII primer in SMART ('Switching Mechanism At 5' end of RNA Template') technology (Clontech) were used. The cDNA (3µL) was submitted to a long-distance PCR using the Advantage 2 Polymerase Mix (Clontech) and the "ds cDNA" generated were purified in CHROMA SPIN TE-400 Columns (Clontech). All procedures followed the manufacturer's recommendations.

Yeast transformation

Competent cells of *Saccharomyces cerevisiae* strain Y187 (MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, gal4Δ, met–, gal80Δ, MEL1**,** URA3::GAL1$_{UAS}$ -GAL1$_{TATA}$-lacZ) (Clontec) were prepared based on Yeastmaker TM Yeast Transformation System 2 User Manual. Approximately 7 µg of ds cDNA of infected soybeans plants and 3 µg of linearized pGADT7-Rec were co-transformed into Y187 yeast cells to construct the Soybean cDNA library in yeast two-hybrid vector according to the Make Your Own "Mate & Plate™" Library System User Manual (Clontech). Transformants were plated on SD/-Leu (synthetically defined medium lacking leucine) medium and grown at 30°C for 7 days. After this period the plates were washed with YPDA medium supplemented with 25% glycerol and stored in Eppendorfs at -80 ° C.

Yeast two-hybrid screening

To obtain proteins from the soybean cDNA library able to interact with the RdRp domain (bait) of CPMMV replicase, this bait was cloned into the NdeI and EcoRI sites of the vector pGBKT7, originating the construct pGBKT7-RdRp, which expresses the RdRp domain as a fusion to the yeast GAL4 DNA-binding domain (BD). The construction was used for transformation of *Saccharomyces cerevisiae* strain AH109 (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2:: GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2, URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ) (Clontech). Before screening, the bait protein was tested for autoactivity and toxicity in AH109 yeast, according to the manufacturer's recommendations. A concentrated culture of AH109 yeast (transformed with pGBKT7-RdRp) and 1 mL of the soybean cDNA library in yeast strain Y187 were mixed for mating

according to the Matchmaker™ Gold Yeast Two-Hybrid protocol (Clontech). Nineteen 150 mm plates with SD/-Leu -Trip –His (synthetically defined medium lacking leucine, tryptophan and histidine) and supplemented with 20 mM of 3AT (3-amino-1,2,4-triazole) were used to screen the clones after mating for 7 days at 30°C. All colonies that appeared on the plates were individually transferred to new plates with SD/-Leu -Trip -His supplemented with 50 mM of 3AT and grown at 30°C for 10 days. All colonies capable of growing in this medium were selected and recovered in SD/-Leu-Trip medium. The yeast colonies recovered were subjected to plasmid DNA extraction using the East yeast plasmid isolation kit (Clontech), according to the manufacturer's recommendations, and sequenced using T7promoter and T7terminator universal primers by Macrogen (Seoul, South Korea).

Identification of candidate proteins able to interact with the RdRp domain and sequence analysis

The nucleotide sequences of the candidates for interaction with the RdRp domain were compared in different databases: Phytozome (http://www.phytozome.net/), NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and SoyKb (http://soykb.org/). The candidate proteins were selected based on their functions. The amino acid sequences from selected proteins were aligned and used for phylogenetic analysis in MEGA v.6.06. The trees were constructed using the Neighbor Joining method with 10,000 bootstrap replications. The protein domains were predicted using ORF Finder (www.ncbi.nlm.nih.gov/orffinder/), Pfam (http://pfam.xfam.org/), Interpro (https://www.ebi.ac.uk/interpro/) and UniProtKB (http://www.uniprot.org/uniprot/). The determination of secondary structure was done using PsiPred server (Mcguffin *et al.*, 2000) and the subcellular localization was predicted using

WoLF PSORT (Horton *et al.*, 2007), Yloc (Briesemeister *et al.*, 2010) and TargetP (Emanuelsson *et al.*, 2007) web tools.

Yeast two-hybrid (Y2H) interaction assays

Specific primers for candidate proteins were designed for amplification, digestion and cloning into pGADT7 and pGBKT7 two-hybrid vectors (Table S1). The RdRp domain was also cloned in pGADT7 and was used for the construction of pGBKT7-RdRp, previously described, in all two-hybrid experiments.

Yeast transformation and two-hybrid assays were conducted in accordance with the recommended procedures (Matchmaker Gold Yeast Two-Hybrid System; Yeastmaker Yeast Transformation System 2, Clontech). The AH109 yeast strain was co-transformed with appropriate pGBKT7 and pGADT7 clones and plated in SD/-Leu -Trip medium. Three colonies from SD/-Leu -Trip plates were replica-plated sequentially with increasing stringency of selection and analyzed for growth in SD/-Leu -Trip -His medium containing increasing concentrations of 3AT (5-60mM). The culture was adjusted to an OD_{600} of 0.2. The plates were incubated at 30° C for 7 days. The bait plasmid (pGBKT7-53 or pGBKT7-Lam, Clontech) was co-transformed into AH109 yeast with the prey plasmid (pGADT7-T, Clontech) to serve as positive and negative controls, respectively, according to the manufacturer's recommendations.

β-Galactosidase interaction assay

The β -galactosidase assay was carried out to validate and determine the strength of protein–protein interactions. The substrate 5-bromo-4-cloro-3-indoxil-β-D-galactopyranoside (X-gal, in Colony membrane assay) and the ortho-nitrophenyl-b-galactopyranoside (ONPG, Sigma Aldrich) were used in two independent experiments. For colony membrane assay the yeast colonies with appropriate constructions were transferred from selective plates on to nylon membrane and grown for 48 hrs. Then, the membrane was immersed in liquid nitrogen for 5 min. The membrane was thawed at room temperature and overlaid with X-gal buffer [Z buffer (0.1 M Na_2HPO_4, 0.05M NaH_2PO_4, 0.01 MKCl and 0.002 M $MgSO_4$, pH=7), 0.98M X-gal (5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside) and 0.27% 2-Mercaptoethanol. The filters were incubated overnight at room temperature and until the appearance of the blue color in the positive control. This assay was carried out as described in the yeast protocols handbook (Clontech) with modifications. For the ONPG assay, the yeast cells transformed with appropriate pGBKT7 and pGADT7 constructions were inoculated into SD/-Leu -Trip liquid media and grown at 30° C with vigorous shaking until the OD_{600} reached 0.6–1. Cells (5 ml) were pelleted and β -galactosidase assay was carried out as described in the yeast protocols handbook (Clontech). The β -galactosidase activity was estimated using the formula $1000 \times (OD_{420}/t \times V \times OD_{600})$, wherein OD_{420} represents the product absorbance, OD_{600} cell density, t, time of incubation in min and V, volume of cells in ml. Means were compared using Student's *t* test with $p<0.05$.

All β-galactosidase assays were performed in triplicate using constructs in both the GAL4 activating and DNA binding domain fusion proteins. The bait plasmid (pGBKT7-53 or pGBKT7-Lam, Clontech) and the prey plasmid (pGADT7-T, Clontech) served as positive and negative controls, respectively.

Subcellular localization and Bimolecular Fluorescence Complementation (BiFC) assay

The binary expression vector PK7FWG2 (GFP C-terminal fusion) and the Bimolecular Fluorescence Complementation (BiFC) vectors pSITE-nEYFP-N1 and pSITE-cEYFP-N1 (for split YFP N-terminal/C-terminal fragment expression) were used in experiments of localization and BiFC, respectively. For localization and BiFC assays, the RdRp domain of CPMMV replicase, *GmERD4* and *Gm*IMT were amplified, using the primers described in Table S1, and cloned into pDONR201 vector (Invitrogen) using the gateway recombination system with Gateway BP Clonase^{TM} II Enzime mix (Invitrogen), resulting in pDONR-RdRp, pDONR-ERD4 and pDON-IMT constructs, respectively. The constructs were used to transfer the respective fragment to PK7FWG2, pSITE-nEYFP-C1 and pSITE-cEYFP-C1 vectors using Gateway LR Clonase Enzyme mix (Invitrogen). All procedures of cloning were done according to the manufacturer's recommendations.

For sub-cellular localization, *Agrobacterium tumemfaciens* (GV3101) cells were transformed with the previous constructs in pK7FWG2 vectors. Cultures containing the pK7FWG2 derivatives (pK7FWG2-Ø, pK7FWG2-RdRp, pK7FWG2-ERD4 and pK7FWG2- IMT) were resuspended and diluted to an OD_{600} of 0.2 before leaf infiltration. The cell suspensions were incubated at room temperature for 2 to 4 h and then used to infiltrate 5- to 6-week-old *N. benthamiana* leaves with infiltration medium (10 mM MES,

pH 5.6, 10 mM MgCl2, 200 mM acetosyringone). For co-localization experiments, the construct SP-TagRFP-HDEL (ER marker fused with RFP) and the silencing suppressor P19 were co-agroinfiltrated in equal amounts (final OD_{600} adjusted to 0,2 for each culture) with the constructs in pK7FWG2 vector.

For the BiFC assay, *A. tumemfaciens* (GV3101) cultures, previously transformed with the constructs in pSITE-nEYFP-N1 and pSITE-cEYFP-N1(Martin *et al.*, 2009) (pSITE-nEYFP-N1-Ø, pSITE-nEYFP-N1-RdRp, pSITE-nEYFP-N1-ERD4, pSITE-nEYFP-N1-IMT, pSITE-cEYFP-N1-Ø, pSITE-cEYFP-N1-RdRp, pSITE-cEYFP-N1-ERD4 and pSITE-cEYFP-N1-IMT) were grown and adjusted to an OD_{600} of 1.0, with infiltration medium, prior to *N. benthamiana* leaf infiltration. The silencing suppressor P19 was co-agroinfiltrated in equal amounts (final OD_{600} adjusted to 1.0 for each culture) with pSITEs constructs.

All confocal images were acquired 72 h after agroinfiltration using a 40X oil immersion objective under a laser confocal scanning microscope (LSM 510 META, Carl Zeiss, and Oberkochen, Germany). GFP was excited using Argon laser at 488nm and then captured at 550 nm. RFP was excited at 561nm and captured between 585-610. YFP was excited at 514 nm and the emitted light captured between 530–600 nm. The images were processed with ZEN2.1 software (Carl Zeiss).

Protoplast obtention and transfection for gene overexpression

Protoplasts were obtained from leaves of soybean plants at five dpi with CPMMV and from healthy soybean plants (buffer inoculated, control plants). Briefly, leaves from infected or control plants were cut in pieces and incubated with enzyme solution [1.0% cellulose (from *Aspergillus niger*), 0.5% cellobiase (from *A. niger*), 0.2% pectolyase (from *A. japonicas*), 0.6M mannitol, 0.02M MES (pH 5.5)] during 6 hours under agitation (30 rpm) in the dark. The cell mix was filtered in sieve (65 µM) and centrifuged at 200×g for 10min. The pellet was carefully washed twice with wash buffer [0.6M mannitol and 0.02M MES (pH 5.5)] and electroporation buffer [25 mM HEPES, 0.02M KCl, 0.03M $MgCl_2$, 0.6M mannitol (pH 7.2)]. The supernatant was discarded and the pellet was resuspended in 5mL of electroporation buffer and kept it on ice by 30 minutes before electroporation. The protoplasts of infected or control plants were used for electroporation (25µF and 2500V) using 40 µg of salmon sperm and 10 µg of plasmidial DNA of pK7FWG2-Ø, pK7FWG2-ERD4 or pK7FWG2-IMT. This was incubated in culture plates with MS medium [0.6 M mannitol, 0.23% MS salt and 0.09M sucrose (pH 5.7) at 22° C in dark and recovered after 36 hours. The protoplasts collected were used for RNA extraction and qRT-PCR, as described below.

Real-Time PCR analyses

Real-time PCR analyses were conducted to evaluate the expression of *GmERD4* and *GmIMT* genes and viral accumulation (based on the CPMMV CP gene) in soybean plants and protoplasts. In soybean plants, this was done at 0, 3, 7 and 14 dpi using three plants for each time period and using as the control, soybean plants mock-inoculated with potassium phosphate buffer (0.2 M, pH 7.2) with 1% sodium bisulfite. In soybean protoplasts, gene

299 expression analysis was done exclusively at 7 dpi (electroporations of the pGBKT7
300 construct were done at 5 dpi), protoplasts from healthy soybean were used as the control.
301 Total RNA of protoplasts or leaf tissue was extracted using the RNeasy Plant Mini Kit
302 (Qiagen) according to the manufacturer´s instructions. RNA quality was checked by
303 electrophoresis in agarose gel and quantified by spectrophotometry using a NanoDrop
304 (Thermo Scientific). RNA extraction and the first-strand cDNA were synthesized described
305 above.
306
307 Relative quantification was performed in individual tubes for *GmERD4* and *GmIMT* genes
308 using primers described in Table S1. Cycling conditions were 10 min for 94°C, 40 cycles of
309 15s at 94°C and 1 min at 60°C. All samples were submitted to thermal denaturation to
310 determine their dissociation curve (melting) and to verify the specificity of amplifications.
311 The expression of each gene was tested in three replicates for each time and repetition.
312 Initial tests were performed on serial dilutions of $10^{0,}$ 10^{-1}, 10^{-2} and 10^{-3} of cDNA and the
313 dilution with the best efficiency, according to the $E=10^{-1}$/slope equation, was selected. The
314 slope angle is obtained from regression of the amplification reaction. For quantification of
315 gene expression, *GmHelicase* was used as a reference gene to normalize the real-time PCR
316 data. Relative quantification was done using the comparative cycle threshold method
317 (ΔΔCt) (Livak & Schmittgen, 2001) and statistical comparisons were done using Student's
318 *t* test.
319
320 For absolute quantification, RNA from the same plants was used for cDNA synthesis and
321 amplification. This was done using individual tubes and primers specific for ORF5 (CP
322 coding region; Table S1). Cycling conditions were 10 min for 94°C, 40 cycles of 15s at

94°C and 1 min at 60°C. All samples were submitted to thermal denaturation to determine their dissociation curve (melting) and to verify the specificity of amplifications. A standard curve was obtained by regression of Ct values, using increasing amounts of PCR amplified CPMMV CP gene. Viral accumulation was determined by interpolation of the Ct values of each tested sample within the standard curve (Rutledge & Cote, 2003) and the means were compared using Student's *t* test.

All procedures for real-time PCR, including tests, validation and experiments, were conducted according to the recommendations of Applied Biosystems. All quantifications were performed on StepOnePlus instrument (Applied Biosystems) using specific primers, cDNAs and SYBR Green PCR Master Mix (Applied Biosystems).

RESULTS

Isolation of soybean factors interacting with CPMMV RdRp domain using yeast two-hybrid system and protein identification

To determine the host factors that interact with the RdRp domain of CPMMV replicase, the recombinant plasmid pGBKT7-RdRp was used as bait in yeast two-hybrid screens using the soybean cDNA library cloned in a pGADT7 vector as prey. Initially, the bait was tested for autoactivity and toxicity in AH109 yeast and it was shown to be able to grow in SD-Trip medium when transformed with pGBKT7-RdRp and unable to activate *HIS3* (absence of growing in SD-Trip -His medium) (Figure 1A) and *lacZ* (ausence of blue color in membrane test) (Figure 1B) reporter genes. As a result of sequential screening steps, 58 candidate colonies were obtained by screening the soybean cDNA library (data not shown).

Of the 58 clones, 20 were chosen based on their abilities to grow in SD-Leu-Trip-His medium (activation of *HIS3* reporter gene - histidine prototrophy) and to activate the *lacZ* reporter gene (data not shown), and sequenced for determination of their identity.

All the candidates were soybean proteins and two were chosen for further investigations. These candidates showed similarity with CSC1-like protein ERD4 (Early-responsive to dehydration stress protein) (phytozome: Glyma.15G091600.1, NCBI: XM003546023.3 and soyKB: Glyma15g09820.2) and an O-methyltransferase family protein (inositol methyltransferase-IMT) (phytozome: Glyma.06G295700.1, NCBI: XM_003539858.3 and SoyKB: Glyma06g45050.1). They were here denominated as: GmERD4 and GmIMT, respectively.

RdRp domain is able to interact with GmERD4 and GmIMT by yeast two-hybrid

The full-length coding regions of GmERD4 and GmIMT were cloned into the GAL4 binding domain vector pGBKT7 (pGBKT7-GmERD4 or pGBKT7-GmIMT) and activation domain vector pGADT7 (pGADT7-ERD4 or pGADT7-IMT) and then combinations of plasmid expressing bait (in pGBKT7) and prey (in pGADT7) were co-transformed into *Saccharomyces cerevisae* AH109 strain. The AH109 yeast, co-transformed with pGBKT7-p53 (Murine p53 fused to Gal4 DNA binding domain) and pGADT7-T antigen (SV40 large T-antigen fused to Gal4 activation domain) and with pGBKT7-Lam and pGADT7-T, was used as a positive and a negative control, respectively. Transformants were selected on SD/-Leu -Trip -His containing increasing concentrations of 3AT (5-60mM). The positive control (pGBKT7-53+pGADT7-T) and the transformants with RdRp and ERD4 or IMT were able to grow in prototrophic medium (Figure 2A). The growth of transformants was

371 compared with the positive and negative controls, and it was observed that GmERD4 and
372 GmIMT had greater growth than the negative control (pGBKT7-Lam and pGADT7-T), but
373 less than the positive control (pGBKT7-53+pGADT7-T).
374
375 In order to validate and determine the strength of the interactions, β-galactosidase assays
376 were carried out using X-gal and ONPG as substrates. In the assay using X-gal as substrate,
377 the β-galactosidase activity was determined based on appearance of blue coloration in yeast
378 colonies previously grown on nylon membrane for 48 hrs. The membrane was left in
379 contact with the X-gal solution for 8 h and then withdrawn and kept in the dark for a total
380 period of 24 hrs. The appearance of the blue coloration was observed in yeast colonies co-
381 transformed with RdRp domain and GmERD4 or GmIMT, reinforcing once again the
382 occurrence of an interaction (Figure 2B). The intensity of the interaction was further
383 determined by measuring levels of β-galactosidase activity in yeast extracts using ONPG as
384 substrate (Figure 2C). The β-galactosidase activity levels in interactions between RdRp
385 domain and ERD4 or IMT were compared with negative (pGBKT7-Lam+pGADT7-T) and
386 positive (pGBKT7-53+pGADT7-T) controls and showed that the interaction between the
387 RdRp domain and ERD4 and IMT is weaker than the positive control (Figure 2C).
388
389 **The RdRp domain interacts *in vivo* with GmERD4 and GmIMT**
390 In order to examine whether the interaction between RdRp domain and GmERD4 and
391 GmIMT, identified with the yeast two-hybrid system, also occurs in the plant, a
392 bimolecular fluorescence complementation (BiFC) analysis was conducted (Figure 3).
393 Initially, the RdRp domain and GmERD4 and GmIMT were fused to non-fluorescent
394 fragments (nYFP and cYFP) from fluorescent yellow protein (YFP) in pSITE-cEYFP-N1

and pSITE-nEYFP-N1 vectors and the constructs were used to transform *Agrobacterium tumefactions* GV3101. The transformed *A. tumefactions* were used to agroinoculate *N. benthamiana* leaves. As a negative control, the pSITE-cEYFP-N1 and pSITE-nEYFP-N1 vectors were used (Figure 3A). Examination by confocal microscopy revealed a strong fluorescent signal along the cytomembrane of *N. benthamiana* cells for all combinations of constructs (Figure 3), except in the combination pSITE-cEYFP-N1-RdRp+pSITE-nEYFP-N1-IMT (Figure 3E). These results confirmed that the RdRp domain interacts with the soybean proteins GmERD4 and GmIMT in living plants cells.

GmERD4 and GmIMT have a nucleic acid binding domain

Following the verification of the interactions between the RdRp domain and GmERD4 and GmIMT, the characterization of these proteins was performed from the full amplified and cloned sequences. GmERD4 showed 1,545 nucleotides (nt) encoding a protein of 514 amino acid (aa) residues, while GmIMT showed 903 nt encoding a protein of 300 aa. To understand the relations of GmERD4 and GmIMT with proteins from different hosts, phylogenetic analyses were performed using aa sequences from both proteins and NCBI dataset proteins (Figure 4A and B). Both proteins clustered with homologous proteins from soybean (*Glycine max*) (Figure 4A and B).

Secondary structures of ERD4 and IMT in plants were predicted using PsiPred server (Mcguffin *et al.*, 2000) (Figure S1) while the servers ORF Finder, Pfam, Interpro and UniProtKB were used to predict the principal domains present in both proteins. GmERD4 was found to exhibit three principal domains: RSN1_TM, PHM7_cit and RSN1_7TM (Figure 4C). The RSN1_TM domain (5-164 aa region) is usually present in proteins

associated with Golgi transport (late exocytosis). PHM7_cit (224-294 aa region) is a cytosolic domain from integral membrane proteins and a putative domain from phosphate transporter normally situated between RSN1_TM and RSN1_7TM domains. PHM7_cit belongs to RRM-like Clan, which contains protein families related to the RNA recognition motif (RRM). However, not all proteins from these families are RNA-binding proteins. RSN1_7TM (313-514 aa region) is the seven transmembrane domain region of a putative phosphate transporter.

GmIMT sequences analysis showed that this protein has two principal domains: a dimerization domain in the N-terminal region (43-91 aa region), and a S-adenosyl-L-methionine-dependent methyltransferase domain (140-300 aa region) in the C-terminal region (Figure 4D), which utilizes the ubiquitous methyl donor SAM as a cofactor to methylate proteins, small molecules, lipids, and nucleic acids. Interpro identified an additional winged helix-turn-helix DNA-binding domain in the IMT protein.

Individual localization of the RdRp domain, GmERD4 and GmIMT showed that they are in different subcellular compartments in the absence of infection

The BiFC assay, in addition to detecting the interaction between candidate proteins *in planta*, allows identification of the cellular compartment where this interaction occurs. Here, this assay showed that the interaction between the RdRp domain and the soybean proteins probably occurs in association with the cell membrane. Thus, initially the subcellular localizations of RdRp domain, GmERD4 and GmIMT were predicted using WoLF PSORT (Horton *et al.*, 2007), Yloc (Briesemeister *et al.*, 2010) and TargetP (Emanuelsson *et al.*, 2007) web tools.

The RdRp domain was predicted by WoLF PSORT to have a subcellular localization in the plasma membrane, while YLoc determined that this domain is localized in the nucleus (77.5 % certainty). The TargetP tool detected that the RdRp domain is not associated with mitochondria, chloroplast or membranes of the secretory pathway, showing an unkown localization. GmERD4 localization was associated with the ER by WolF PSORT (68.5% certainty) and membranes from the secretory pathway by Yloc (100% certainty) and TargetP (>80% certainty). The analysis of GmIMT localization showed suggested a cytoplasmic localization (WoLF PSORT 65% and Yloc 99% certainties). This was reinforced by TargetP which showed with 100% certainty that GmIMT is not associated with membranes of mitochondria, chloroplasts or the secretory pathway.

Subcellular localizations of RdRp domain, GmERD4 and GmIMT in *N. benthamiana* plants were then examined. For this, the encoding regions were cloned in the vector PK7FWG2, which allowed the fusion to GFP at the C-terminal region, and introduced into plants by Agrobacterium infiltration. Confocal microscopy indicated that the RdRp domain induced the formation of punctate structures around the cell, possibly in the cellular membrane, and apparently is also localized in the nucleus (Figure 5B). GmERD4 was visualized as a network in the cell cytoplasm, very similar to the pattern observed for proteins associated with the endoplasmic reticulum (Figure 5C). The co-localization of GmERD4 with SP-TagRFP-HDEL, an ER marker fused to RFP, showed that GmERD4 is on the ER network (Figure 6C). This was predicted *a priori* by the analytical tools. GmIMT was predicted to be localized in the cell cytoplasm and the localization assay confirmed this prediction (Figure 5D).

Co-localization assays between the RdRp domain and GmIMT with HDEL were also performed, and it was possible to verify that the punctate structures of the RdRp domain are distributed across the ER network (Figure 6B), while GmIMT was not co-localized with HDEL in the ER (Figure 6D).

GmERD4 and _GmIMT_ expression is induced three and seven days after CPMMV infection

To determine the effect of viral infection on *GmERD4* and *GmIMT* expression, soybean plants were inoculated with the CPMMV:BR:MG:09:2 isolate and viral accumulation and *GmERD4* and *GmIMT* expression were determined after 0, 3, 7 and 14 dpi. As expected, viral accumulation increased throughout the viral infection process (Figure 7A), exhibiting greatest viral accumulation at 14 dpi. The expression analysis of *GmERD4* and *GmIMT* in infected soybean plants demonstrated that expression of both proteins is induced at 3 and 7 dpi (Figure 7B and C), and only *GmIMT* is down-regulated at 14 dpi (Figure 7C). This increase in viral accumulation was accompanied by an increase in the induction of symptoms (Figure 7D).

Overexpression of _GmERD4_, but not _GmIMT_, increased the viral accumulation of CPMMV

To evaluate the direct effect of the soybean proteins on CPMMV infection, *GmERD4* and *GmIMT* were overexpressed in soybean protoplasts. For this, the constructs PK7FWG2-ERD4 and PK7FWG2-IMT were electroporated in protoplasts of soybean plants at 5 dpi and the effect with viral infection was evaluated 36 hrs later (7 dpi). This time was chosen

as both proteins were induced at 3 and 7 dpi (Figure 7B and C). Healthy soybean protoplasts and infected protoplasts electroporated with PK7FWG2-*Ø* (empty) were used as controls.

The overexpression of *GmERD4* and *GmIMT* in soybean protoplasts electroporated with the constructs in PK7FWG2 was confirmed (Figure 7E and F) and the viral accumulation analysis showed that the overexpression of *GmERD4* increased CPMMV accumulation (Figure 7G). In contrast, the overexpression of *GmIMT* did not directly affect the viral accumulation (Figure 7G).

DISCUSSION

Cowpea mild mottle virus is an important pathogen that infects soybean across the world (Brown & Rodrigues, 2014; Zanardo *et al.*, 2014a). It belongs to the *Carlavirus* genus and *Betaflexiviridae* family (Adams & Kreuze, 2016), a group of viruses that is little studied. Viral replication, movement and the hosts' defense response to infection are mechanisms still little explored for this group of viruses. So here, we decided to investigate the viral replication of CPMMV, by identifying host proteins capable of interacting with the RdRp domain of viral replicase, and to evaluate their effects on viral replication. Two host proteins (GmERD4 and GmIMT) were identified as capable of interacting with the RdRp domain of the CPMMV viral replicase. Their subcellular localizations were also determined and ER seems to be the important organelle in CPMMV replication. Both *GmERD4* and *GmIMT* genes were induced during viral infection, but only the overexpression of *GmERD4* increased viral accumulation in soybean protoplasts.

Early-responsive to dehydration (ERD) genes have been identified collectively as genes that are rapidly induced by dehydration in *Arabidopsis thaliana*. Genes that encode these proteins have functional roles in plant stress tolerance and in development (Kiyosue *et al.*, 1994). The expression pattern of ERD4 in *Brassica juncea* (*BjERD4*), analyzed under different stress conditions, showed that transcript levels increased with dehydration, sodium chloride, low temperature, heat and abscisic acid and salicylic acid treatments (Rai *et al.*, 2016), all abiotic stresses. Here, we demonstrated that *GmERD4* is also induced by CPMMV infection (a biotic stress) and that GmERD4 was able to interact with the RdRp domain of CPMMV. The predictions for the GmERD4 protein suggested that it is associated with membranes in the secretory pathway (at the ER). In agreement with these predictions GmERD4 was associated with the ER in localization assays. However, BjERD4 (A9LIW2.2), an ERD4 biologically characterized from *Brassica juncea*, showed chloroplastic localization (Rai *et al.*, 2011; Rai *et al.*, 2016). Both GmERD4 and BjERD4 have the same protein domains (RSN1_TM, PHM7_cit and RSN1_7TM), suggesting similarity in functions. However, an analysis shows that GmERD4 and BjERD4 are phylogenetically distant (Figure 4), so they may have different functions and therefore different localizations. Different functions have been associated to ERDs in *A. thaliana*. *AtERD15*, for example, has been reported to be induced rapidly in response to abiotic (drought) and biotic (infection with the nectrotroph bacterial *Erwinia carotovora*) stress stimuli. ERD15-overexpressing plants led to improved resistance to *E. carotovora* by enhancing the salicylic acid-dependent defense pathway, but reduced the ABA sensitivity of Arabidopsis, manifested in decreased drought tolerance and impaired ability of the plants to increase their freezing tolerance in response to this hormone (Kariola *et al.*, 2006). In soybean it is known that *GmERD15* (*Glycine max* ERD15) is induced by osmotic and

endoplasmic reticulum stresses and activates the osmotic and ER stress-induced N-rich protein (NRP)-mediated cell death response (Alves *et al.*, 2011b). GmERD15 leads to the activation of NRP-B which activates GmNAC6 and induces cell death (Alves *et al.*, 2011a).

The function of GmERD4 in CPMMV infection needs to be determined. This protein has a RSN1_TM domain involved in late Golgi exocytosis, which means that this protein is clearly associated with vesicle transport. Besides this, we identified a cytosolic domain PHM7_cit, which contains an RNA recognition motif (RRM) commonly present in RNA-binding proteins (RBP's). RBP proteins influence the structure and interactions of RNA's and play critical roles in their biogenesis, stability, function, transport and cellular localization (Glisovic *et al.*, 2008). Thus, we believe that GmERD4 may associate with the RdRp domain to confer stability to nascent viral RNA, ensuring its protection and correct targeting through cellular vesicular transport. This possibility is reinforced when we associate this probable function with the induction of *GmERD4* after 3 and 7 dpi, initial stages of viral infection, and with the improvement of viral accumulation during *GmERD4* overexpression.

The other protein studied here was the inositol methyltransferase, GmIMT. It has been associated with methylation of *myo*-inositol forming O-methyl inositol (D-ononitol) in plants under abiotic stress (Ahn *et al.*, 2011). The *IMT* gene has been cloned from the halophytic ice plant *Mesembryanthemum crystallinum* and also from the halophytic wild rice variant *Porteresia coarctata* and in both cases their transcription was found to be induced by osmotic stress (Sengupta *et al.*, 2008; Vernon & Bohnert, 1992). The constitutive expression of the *M. crystallinum IMT* gene in tobacco plants was shown to

563 increase salt and drought tolerance (Sheveleva *et al.*, 1997). Transgenic *Arabidopsis* plants
564 over-expressing *GmIMT* from soybean also displayed improved tolerance to dehydration
565 stress and to a lesser extent to high salinity stress (Ahn *et al.*, 2011). We demonstrated here
566 that *GmIMT* was up-regulated in soybean plants infected with CPMMV, suggesting that is
567 also induced by biotic stress.
568
569 The protein GmIMT has an S-adenosylmethionine-dependent methyltransferase domain
570 and a winged helix-turn-helix DNA-binding domain in its structure. Several winged-helix
571 proteins display an exposed patch of hydrophobic residues thought to mediate protein-
572 protein interactions (Gajiwala & Burley, 2000). It is known that HEN1, S-
573 adenosylmethionine-dependent RNA methyltransferase, is involved in methylation of the 3′
574 terminal nucleotides of siRNAs (Yang *et al.*, 2006), protecting them from degradation and
575 oligouridylation (Li *et al.*, 2005). The subsequent incorporation of the siRNAs into an
576 RNA-induced silencing complex leads to the sequence-specific degradation of target RNAs
577 in post-transcriptional RNA silencing. The same domain was found in GmIMT, suggesting
578 that this protein may be involved in methylation of nascent viral RNA or templates by the
579 viral replicase. Recently, it was reported that the coat protein (CP) of the crinivirus
580 *Tomato chlorosis virus*, a known suppressor of silencing, interacts with S-adenosyl-
581 homocysteine hydrolase (SAHH), a plant protein essential for sustaining the methyl
582 cycle and S-adenosylmethionine-dependent methyltransferase activity. The
583 overexpression of SAHH enhanced local RNA silencing by increasing siRNAs
584 accumulation (Canizares *et al.*, 2013). Thus, in addition to using GmIMT for methylation
585 of viral RNA, the interaction between the RdRp domain and GmIMT would also occur to

avoid its methyltransferase action, preventing the mechanism of post-transitional genetic silencing.

GmIMT was found in the cytoplasm, while the RdRp domain was associated with cell membrane and ER and GmERD4 also showed ER localization, in individual subcellular localization assays. However, the interaction between the RdRp domain and GmERD4 and between RdRp and GmIMT, in the BiFC assays, apparently occurred around the cell membrane. We believe that in the presence of the RdRp domain, and so during the viral infection, the subcellular localization of both soybean proteins maybe altered, with the replication complex possibly displaced to the cell membrane. The location of the RdRp domain and GmERD4 at the ER network suggests that ER may be the organelle with which CPMMV associates during viral replication. If so, CPMMV would associate itself with the ER membrane for formation of the replication complex and this complex, containing viral and host proteins, would be displaced to the cell membrane. This has been demonstrated for the potexvirus *Potato virus X* (PVX). PVX TGBp1 protein was responsible for the inducion at the ER of structures denominated as X-bodies, which includes TGBp2 and TGBp3, in granular vesicles, and nonencapsidated viral RNA (Tilsner *et al.*, 2012). The X-bodies acted as viral replication complexes. TGB2p and TGB3p from these X-bodies were able to establish ER-derived membranous caps at plasmodemata (PD) orifices. Thus, these caps harboring the PVX replicase and nonencapsidated vRNA represent PD-anchored viral replication sites. TGB1 additionally mediated the insertion of the viral coat protein into the PD, which interacts with the 5' terminal region of virions, previously recruited by the TGB2/3 complex. This model was denominated as co-replicational, in which the TGB's

609 (movement proteins) are able to compartmentalize the replication complex and direct
610 movement (Tilsner *et al.*, 2013).
611
612 The possible involvement of GmERD4 and GmIMT in viral replication was demonstrated
613 here by the different *in vitro* and *in vivo* interaction assays. Confocal microscopy assays
614 showed that the prediction of subcellular localization was correct and that, although the
615 proteins are located in different subcellular compartments, the interaction with the RdRp
616 domain is possible. Viral infection induced increased expression of the genes of both
617 proteins in the early stages of infection (3 and 7 dpi), a period in which host machinery is
618 rapidly manipulated to favor viral infection and *GmERD4* when overexpressed favored
619 viral accumulation and thus viral replication. This is the first study showing the interaction
620 of host proteins with the RdRp domain of a betaflexivirus.
621
622 **ACKNOWLEDGMENT**
623 We thank Dr. Michael Goodin for providing us with the BiFC vectors and Dr. Jens Tilsner
624 for the construct SP-TagRFP-HDEL. We are also grateful to Nucleus of Microscopy and
625 Microanalysis of UFV for assistance in the accomplishment of confocal microscopy.
626
627 **REFERENCES**
628 Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN,
629 Ryu KH, Saldarelli P, Yoshikawa N Family Betaflexiviridae. pp.920-941 In: KING AMQ,
630 ADAMS MJ, CARSTENS EB & LEFKOWITZ EJ (Eds.) Virus taxonomy. Ninth Report of
631 the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic
632 Press. 2012.
633
634 Adams M, Kreuze J (2016) Revision of family Betaflexiviridae, order Tymovirales. ICTV
635 Ahlquist P (2006) Parallels among positive-strand RNA viruses, reverse-transcribing
636 viruses and double-stranded RNA viruses. Nature Reviews Microbiology 4:371–382.

Ahn C, Park U, Park PB (2011) Increased salt and drought tolerance by D-ononitol production in transgenic Arabidopsis thaliana. Biochemical and Biophysical Research Communications 415:669-674.

Almeida AMR (2008) Viroses da Soja no Brasil: sintomas, etiologia e controle. Série Documentos 306:1-62.

Almeida AMR, Piuga FF, Kitajima EW, Gaspar JO, Valentin N, Benato LC, Marin SRR, Bineck E, Belintani P, Nunes Junior J, Hoffmann L, Meyer MC (2003) Necrose da haste da soja. Série Documentos 221:1-48.

Alves MS, Fontes EP, Fietto LG (2011a) EARLY RESPONSIVE to DEHYDRATION 15, a new transcription factor that integrates stress signaling pathways. Plant Signal Behav 6:1993-1996.

Alves MS, Reis PA, Dadalto SP, Faria JA, Fontes EP, Fietto LG (2011b) A novel transcription factor, ERD15 (Early Responsive to Dehydration 15), connects endoplasmic reticulum stress with an osmotic stress-induced cell death signal. Journal of Biological Chemistry 286:20020-20030.

Bamunusinghe D, Hemenway CL, Nelson RS, Sanderfoot AA, Ye CM, Silva MA, Payton M, Verchot-Lubicz J (2009) Analysis of potato virus X replicase and TGBp3 subcellular locations. Virology 393:272-285.

Barajas D, Jiang Y, Nagy PD (2009) A unique role for the host ESCRT proteins in replication of Tomato bushy stunt virus. PLoS Pathog 5:e1000705.

Briesemeister S, Rahnenfuhrer J, Kohlbacher O (2010) YLoc-an interpretable web server for predicting subcellular localization. Nucleic Acids Research 38:W497-502.

Brown JK, Rodrigues JCV (2014) Recovery Plan *Cowpea mild mottle virus* Carlavirus: *Betaflexiviridae*; order *Tymovirales*. 1-24.

Canizares MC, Lozano-Duran R, Canto T, Bejarano ER, Bisaro DM, Navas-Castillo J, Moriones E (2013) Effects of the crinivirus coat protein-interacting plant protein SAHH on post-transcriptional RNA silencing and its suppression. Mol Plant Microbe Interact 26:1004-1015.

Chen J, Ahlquist P (2000) Brome mosaic virus polymerase-like protein 2a is directed to the endoplasmic reticulum by helicase-like viral protein 1a. Journal of Virology 74:4310-4318.

Dos Reis FA, Golem S, Goregaoker SP, Culver JN (2002) A nuclear localization signal and a membrane association domain contribute to the cellular localization of the Tobacco mosaic virus 126-kDa replicase protein. Virology 301:81-89.

683 Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell
684 using TargetP, SignalP, and related tools. Nat Protoc 2:953–971.
685
686 Gajiwala KS, Burley SK (2000) Winged helix proteins. Current Opinion in Structural
687 Biology 10:110-116.
688
689 Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-
690 transcriptional gene regulation. FEBS Letters 582:1977-1986.
691
692 Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007)
693 WoLF PSORT: protein localization predictor. Nucleic Acids Research 35:W585-587.
694
695 Kariola T, Brader G, Helenius E, Li J, Heino P, Palva ET (2006) EARLY RESPONSIVE
696 TO DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis.
697 Plant Physiol 142:1559-1573.
698
699 King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (Eds.) Virus taxonomy. Ninth Report
700 of the International Comittee on Taxonomy of Viruses. Elsevier Inc. 1272p. 2011.
701
702 Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Cloning of cDNAs for genes that
703 are early-responsive to dehydration stress (ERDs) in Arabidopsis thaliana L.: identification
704 of three ERDs as HSP cognate genes. Plant Molecular Biology 25:791-798.
705
706 Laliberte JF, Sanfacon H (2010) Cellular remodeling during plant virus infection. Annual
707 Review Phytopathology 48:69-91.
708
709 Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from
710 a 3'-end uridylation activity in Arabidopsis. Current Biology 15:1501-1507.
711
712 Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time
713 quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.
714
715 Martin K, Kopperud K, Chakrabarty R, Banerjee R, Brooks R, Goodin MM (2009)
716 Transient expression in Nicotiana benthamiana fluorescent marker lines provides enhanced
717 definition of protein localization, movement and interactions in planta. Plant Journal
718 59:150-162.
719
720 McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction
721 server. Bioinformatics 16:404-405.
722
723 Menzel W, Winter S, Vetten H (2010) Complete nucleotide sequence of the type isolate of
724 *Cowpea mild mottle virus* from Ghana. Archives of Virology 155:2069-2073.
725
726 Prosser SW, Xiao H, Li C, Nelson RS, Meng B (2015) Subcellular localization and
727 membrane association of the replicase protein of grapevine rupestris stem pitting-associated
728 virus, family Betaflexiviridae. Journal of General Virology 96:921-932.
729

730 Rai A, Suprasanna P, D'Souza SF, Kumar V (2011) Membrane topology and predicted
731 RNA-binding function of the 'early responsive to dehydration (ERD4)' plant protein. PLoS
732 One 7:e32658.
733
734 Rai AN, Tamirisa S, Rao KV, Kumar V, Suprasanna P (2016) Brassica RNA binding
735 protein ERD4 is involved in conferring salt, drought tolerance and enhancing plant growth
736 in Arabidopsis. Plant Molecular Biology 90:375-387.
737
738 Rochon D, Singh B, Reade R, Theilmann J, Ghoshal K, Alam SB, Maghodia A (2009) The
739 p33 auxiliary replicase protein of Cucumber necrosis virus targets peroxisomes and
740 infection induces de novo peroxisome formation from the endoplasmic reticulum. Virology
741 452-453:133-142.
742
743 Rosario K, Capobianco H, Ng TF, Breitbart M, Polston JE (2014) RNA viral metagenome
744 of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus
745 in North America. PLoS One 9:e86748.
746
747 Rutledge RG, Cote C (2003) Mathematics of quantitative kinetic PCR and the application
748 of standard curves. Nucleic Acids Research 31:e93.
749
750 Saxena P, Lomonossoff GP (2014) Virus infection cycle events coupled to RNA
751 replication. Annual Review Phytopathology 52:197-212.
752
753 Sengupta S, Patra B, Ray S, Majumder AL (2008) Inositol methyl tranferase from a
754 halophytic wild rice, Porteresia coarctata Roxb. (Tateoka): regulation of pinitol synthesis
755 under abiotic stress. Plant Cell and Environment 31:1442-1459.
756
757 Sheveleva E, Chmara W, Bohnert HJ, Jensen RG (1997) Increased Salt and Drought
758 Tolerance by D-Ononitol Production in Transgenic *Nicotiana tabacum* L. Plant Physiol
759 115:1211-1219.
760
761 Suhy DA, Giddings TH, Jr., Kirkegaard K (2000) Remodeling the endoplasmic reticulum
762 by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-
763 induced vesicles. Journal of Virology 74:8953-8965.
764
765 Tilsner J, Linnik O, Louveaux M, Roberts IM, Chapman SN, Oparka KJ (2013) Replication
766 and trafficking of a plant virus are coupled at the entrances of plasmodesmata. The Journal
767 of Cell Biology 201:981-995.
768
769 Tilsner J, Linnik O, Wright KM, Bell K, Roberts AG, Lacomme C, Santa Cruz S, Oparka
770 KJ (2012) The TGB1 movement protein of Potato virus X reorganizes actin and
771 endomembranes into the X-body, a viral replication factory. Plant Physiol 158:1359-1370.
772
773 Verchot J (2016) How does the stressed out ER find relief during virus infection? Curr
774 Opin Virol 17:74-79.
775

Vernon DM, Bohnert HJ (1992) A novel methyl transferase induced by osmotic stress in the facultative halophyte Mesembryanthemum crystallinum. Embo J 11:2077-2085.

Vidalain PO, Tangy F (2010) Virus-host protein interactions in RNA viruses. Microbes and Infection 12:1134-1143.

Wang X, Lee WM, Watanabe T, Schwartz M, Janda M, Ahlquist P (2005) Brome mosaic virus 1a nucleoside triphosphatase/helicase domain plays crucial roles in recruiting RNA replication templates. Journal of Virology 79:13747-13758.

Yang Z, Ebright YW, Yu B, Chen X (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Research 34:667-675.

Zanardo LG, Silva FN, Bicalho AAC, Urquiza GPC, Lima ATM, Almeida AMR, Zerbini FM, Carvalho CM (2014a) Molecular and biological characterization of *Cowpea mild mottle virus* isolates infecting soybean in Brazil and evidence of recombination. Plant Pathology 63:456-465.

Zanardo LG, Silva FN, Lima AT, Milanesi DF, Castilho-Urquiza GP, Almeida AM, Zerbini FM, Carvalho CM (2014b) Molecular variability of *Cowpea mild mottle virus* infecting soybean in Brazil. Archives of Virology 159:727-737.

FIGURE LEGENDS:

Figure 1: Autoactivation and toxicity assay of pGBKT7-RdRp in AH109 yeast. A. Toxicity analysis and autoactivation activity by monitoring of Trip and His prototrophy in SD-Trip and SD-Trip-His mediums, respectively. The ability to grow in SD-Trip but not in SD-Trip-His demonstrates non-toxicity to growth of AH109 yeast and no autoactivation activity, respectively. **B.** Autoactivation assay in membrane test of pGBKT7-RdRp construct in AH109 yeast based on β-galactisidade activity (*lacZ* reporter gene). The absence of blue coloration in colonies demonstrates the absence of autoactivation. For all assays, pGBKT7-Ø was used as a negative control. pGBKT7 is a GAL4 binding domain vector.

Figure 2: RdRp domain interaction with GmERD4 and GmIMT by yeast two-hybrid. **A.** RdRp domain, GmERD4 and GMIMT were expressed in yeast as a GAL4 activation domain (AD) fusion (pGADT7–RdRp, pGADT7-ERD4 and pGADT7-IMT) and as a GAL4 binding domain (BD) fusion (pGBKT7-RdRp, pGBKT7-ERD4 and pGBKT7-IMT). The interactions between the tested proteins were examined by monitoring His prototrophy in medium containing increasing concentrations of 3AT (5-60mM). (B–C) Interactions between the proteins were also confirmed by monitoring of β-galactosidase activity (*LacZ* reporter gene). **B.** Analysis of β-galactosidase activity on a nylon membrane using the same constructs. **C.** β-galactosidase activity measured in yeast extracts using ONPG as substrate. Means were compared using Student's *t* test. Letters represent statistically significant differences in means ($P < 0.05$) and error bars are standard deviations. The interactions between pGBKT7-p53 (Murine p53) and pGADT7-T antigen (SV40 large T-antigen) and pGBKT7-Lam and pGADT7-T were used as positive and negative controls, respectively.

Figure 3: *In vivo* interaction with RdRp domain of CPMMV replicase using Bimolecular fluorescence complementation (BiFC) assay. A. Confocal fluorescent image of pSITE-nEYFP-N1-Ø and pSITE-cEYFP-N1- Ø used as negative control. **B-C.** Confocal fluorescent image of RdRp domain and GmERD4 interaction in pSITE-nEYFP-N1 and pSITE-cEYFP-N1 in combinations described below the figures. **D-E.** Confocal fluorescent image of RdRp domain and GmIMT interaction in pSITE-nEYFP-N1 and pSITE-cEYFP-N1 in combinations described below the figures. The fluorescence yellow fluorescent protein (YFP) images were acquired using *Nicotiana benthamiana* leaves co-expressing the construct in the presence of p19 suppressor 72 h after co-infiltration. Scale bars, 10 μm.

Figure 4: Characterization of GmERD4 and GmIMT. A-B. Phylogenetic analysis using GmERD4 and GmIMT amino acid sequences from soybean cv. CD206 (black circle) and sequences of different proteins hosts from NCBI database. The analysis were done in MEGA v.6.06 using Neighbor Joining method with 10,000 bootstrap replications. **C-D.** Schematic domain representation of GmERD4 and GmIMT found using the Pfam web tool. **C.** Domains found in GmERD4: RSN1_TM domain (5-164 aa region, in green) associated with Golgi transport - late exocytosis); PHM7_cit (224-294 aa region, in red) is a cytosolic domain from whole membrane proteins, which contain an RNA recognition motif (RRM) and RSN1_7TM (313-514 aa region, in blue). **D.** Domains found in GmIMT: a dimerization domain in N-terminus region (43-91 aa region, in green) and a S-adenosyl-L-methionine-dependent methyltransferase domain (140-300 aa region, in red).

Figure 5: Subcellular localization of RdRp domain, GmERD4 and GmIMT proteins in *Nicotiana benthamiana* plants. A-D. Confocal visualization of construction in vector pK7FWG2 (Green Fluorescent Protein, GFP C-terminal) described above the images in *Nicotiana benthamiana* leaves after 72 h of agroinfiltration. Scale bar, 10 μ m. **A.** Empty vector pK7FWG2-Ø. **B**. RdRp domain of CPMMV replicase (pK7FWG2-RdRp). **C.** GmERD4 (pK7FWG2- GmERD4) and **D.** GmIMT (pK7FWG2- GmIMT).

Figure 6: Co-localization assay of RdRp domain, GmERD4 and GmIMT proteins with HDEL in *Nicotiana benthamiana* plants. Confocal visualization of construct in vector pK7FWG2 (GFP C-terminal) described above the images in *Nicotiana benthamiana* leaves after 72 h of agroinfiltration in combination with SP-TagRFP-HDEL construct

858 expressing HDEL (ER marker) with Red Fluorescent Protein (RFP). Scale bar, 10 µ m. **A.**
859 SP-TagRFP-HDEL. **B**. RdRp domain of CPMMV replicase (pK7FWG2-RdRp) and SP-
860 TagRFP-HDEL . **C.** GmERD4 (pK7FWG2- GmERD4) and SP-TagRFP-HDEL **D.** GmIMT
861 (pK7FWG2- GmIMT) and SP-TagRFP-HDEL.
862
863 **Figure 7: Expression analysis of GmERD4 and GmIMT in plants and protoplasts**
864 **infected by *Cowpea mild mottle virus* (CPMMV). A.** CPMMV accumulation in soybean
865 plants cv. CD 206 after 0, 3, 7 and 14 days post inoculation (dpi). **B-C.** Relative expression
866 of GmERD4 and GmIMT in soybean plants cv. CD206 after 0, 3, 7 and 14 dpi. **D.**
867 Symptoms induced by CPMMV after 0, 3, 7 and 14 dpi in soybean cv. CD206. Only at 14
868 dpi were typical symptoms of infection observed (mosaic and vein clearing). **E-F.** Relative
869 expression in healthy (H) and infected (I) soybean protoplasts transfected with pK7FWG2-
870 Ø, pK7FWG2-ERD4 or pK7FWG2-IMT constructs **G.** CPMMV accumulation in healthy
871 (H) and infected (I) soybean protoplasts at 7 dpi overexpressing *GmERD4* and *GmIMT*. All
872 analyses were done using three repetitions and compared using Student's *t* test at $p<0.05$.
873 (ns) not significant. *significant.
874
875 **Figure S1: Prediction of secondary structure of the proteins GmERD4 and GmIMT.**
876 **A-B.** Secondary structure predicted in PsiPred server using amino acid sequences of
877 GmERD4 and GmIMT from soybean cv. CD206. **A.** GmERD4 prediction. **B.** GmIMT
878 prediction.

Figure 1:

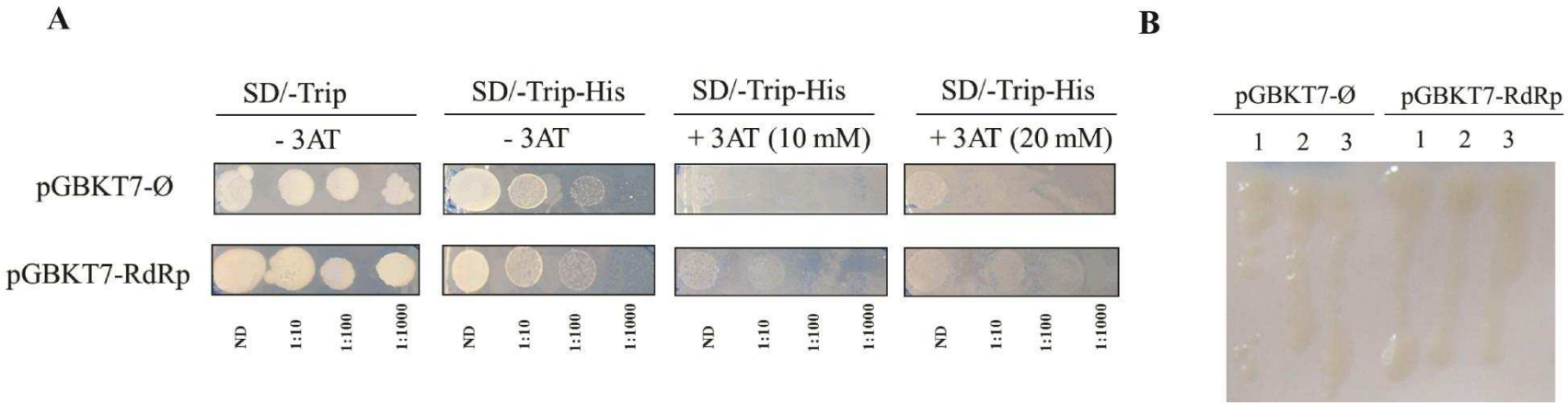
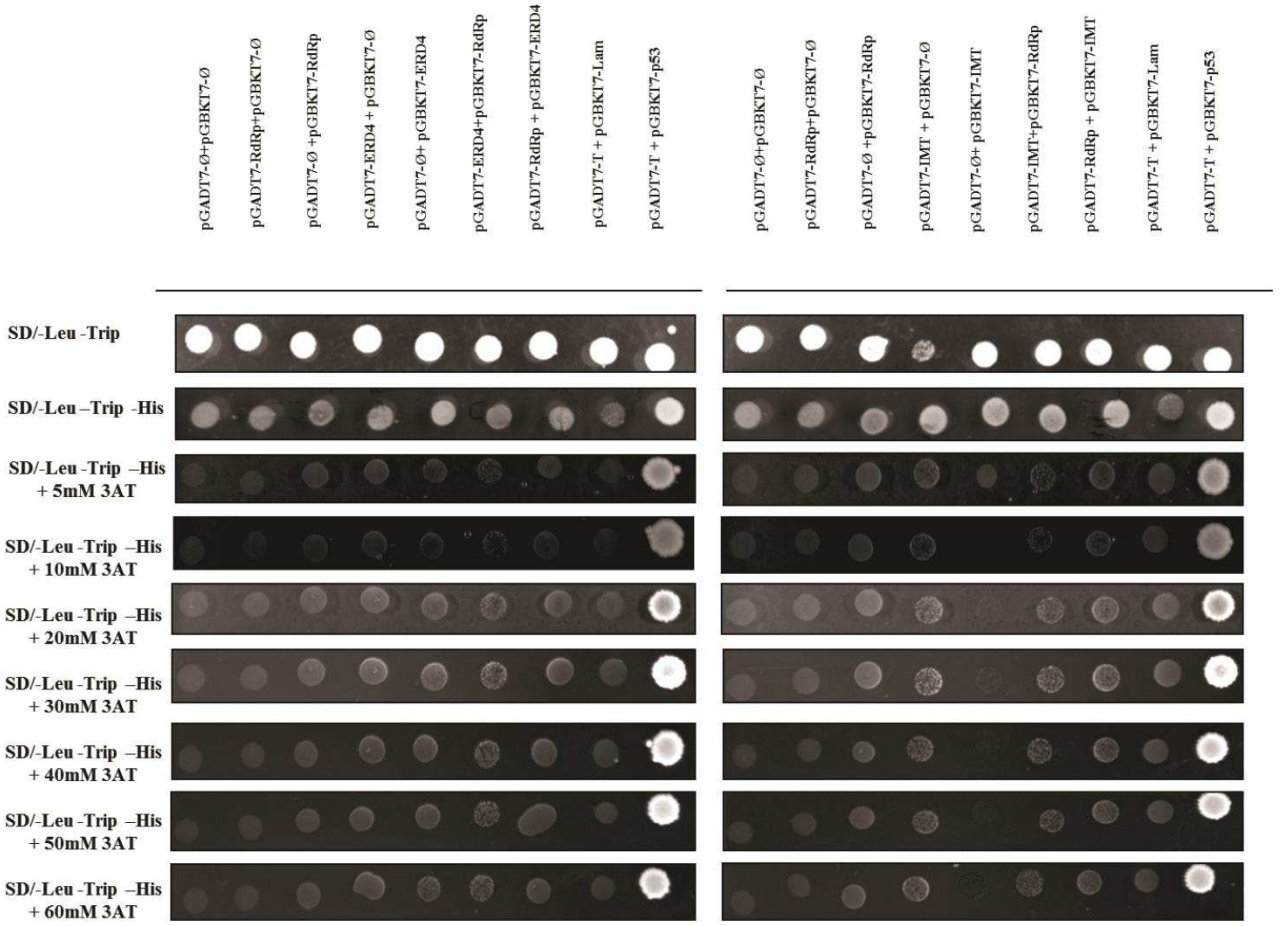

A
SD/-Trip
- 3AT
SD/-Trip-His
- 3AT
SD/-Trip-His
+ 3AT (10 mM)
SD/-Trip-His
+ 3AT (20 mM)
pGBKT7-Ø
pGBKT7-RdRp
ND
1:10
1:100
1:1000
B
pGBKT7-Ø
pGBKT7-RdRp
1 2 3
1 2 3

Figure 2:

A


pGADT7-Ø+pGBKT7-Ø
pGADT7-RdRp+pGBKT7-Ø
pGADT7-Ø +pGBKT7-RdRp
pGADT7-ERD4 + pGBKT7-Ø
pGADT7-Ø+ pGBKT7-ERD4
pGADT7-ERD4+pGBKT7-RdRp
pGADT7-RdRp + pGBKT7-ERD4
pGADT7-T + pGBKT7-Lam
pGADT7-T + pGBKT7-p53
pGADT7-Ø+pGBKT7-Ø
pGADT7-RdRp+pGBKT7-Ø
pGADT7-Ø +pGBKT7-RdRp
pGADT7-IMT + pGBKT7-Ø
pGADT7-Ø+ pGBKT7-IMT
pGADT7-IMT+pGBKT7-RdRp
pGADT7-RdRp + pGBKT7-IMT
pGADT7-T + pGBKT7-Lam
pGADT7-T + pGBKT7-p53
SD/-Leu -Trip
SD/-Leu –Trip -His
SD/-Leu -Trip –His + 5mM 3AT
SD/-Leu -Trip –His + 10mM 3AT
SD/-Leu -Trip –His + 20mM 3AT
SD/-Leu -Trip –His + 30mM 3AT
SD/-Leu -Trip –His + 40mM 3AT
SD/-Leu -Trip –His + 50mM 3AT
SD/-Leu -Trip –His + 60mM 3AT

B


pGADT7-Ø+pGBKT7-Ø
pGADT7-RdRp+pGBKT7-Ø
pGADT7-Ø +pGBKT7-RdRp
pGADT7-ERD4 + pGBKT7-Ø
pGADT7-Ø+ pGBKT7-ERD4
pGADT7-ERD4+pGBKT7-RdRp
pGADT7-RdRp + pGBKT7-ERD4
pGADT7-T + pGBKT7-Lam
pGADT7-T + pGBKT7-p53
ERD4
pGADT7-Ø+pGBKT7-Ø
pGADT7-RdRp+pGBKT7-Ø
pGADT7-Ø +pGBKT7-RdRp
pGADT7-IMT + pGBKT7-Ø
pGADT7-Ø+ pGBKT7-IMT
pGADT7-IMT+pGBKT7-RdRp
pGADT7-RdRp + pGBKT7-IMT
pGADT7-T + pGBKT7-Lam
pGADT7-T + pGBKT7-p53
IMT

C

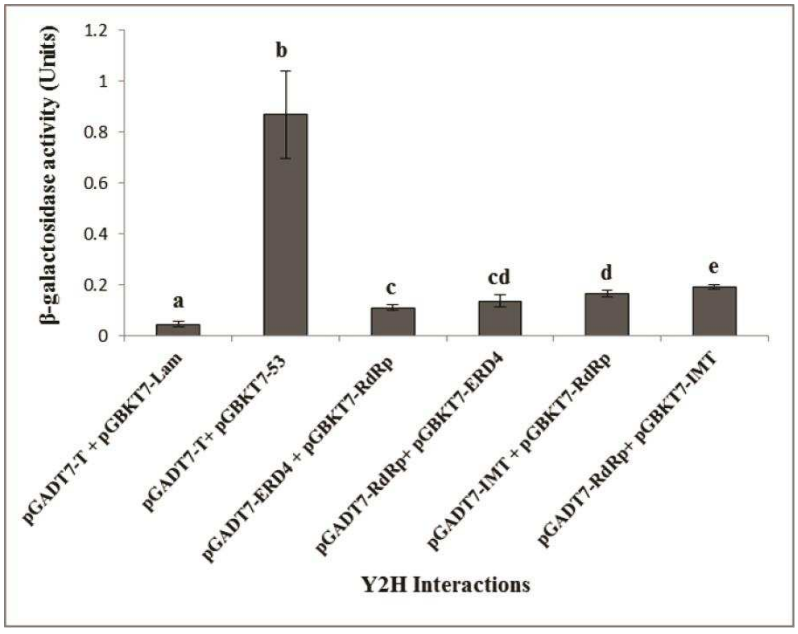

β-galactosidase activity (Units)
1.2
1
0.8
0.6
0.4
0.2
0
a
b
c
cd
d
e
pGADT7-T + pGBKT7-Lam
pGADT7-T+ pGBKT7-53
pGADT7-ERD4 + pGBKT7-RdRp
pGADT7-RdRp+ pGBKT7-ERD4
pGADT7-IMT + pGBKT7-RdRp
pGADT7-RdRp+ pGBKT7-IMT
Y2H Interactions

Figure 3:

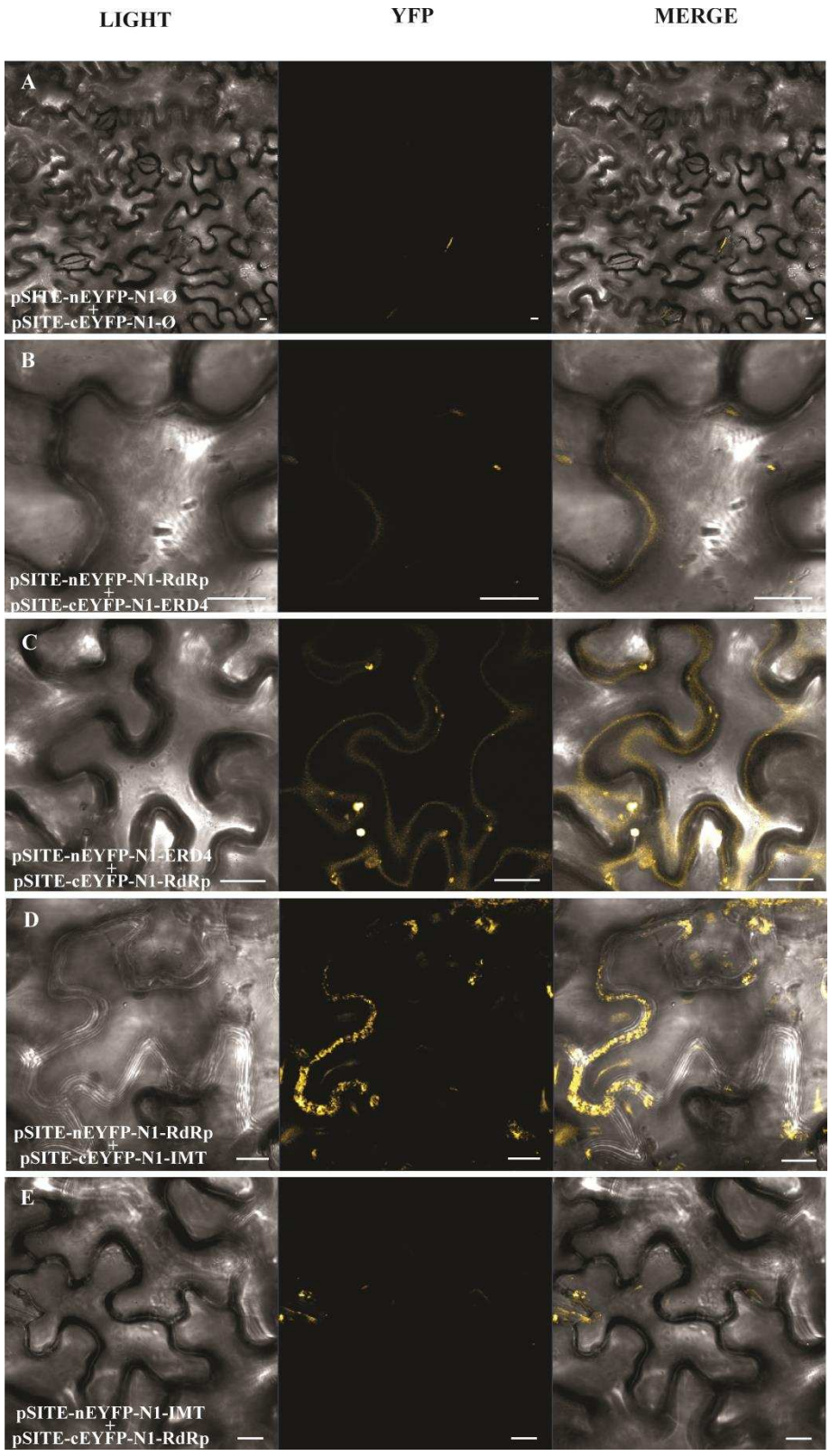

LIGHT
YFP
MERGE
A
pSITE-nEYFP-N1-Ø
+
pSITE-cEYFP-N1-Ø
B
pSITE-nEYFP-N1-RdRp
+
pSITE-cEYFP-N1-ERD4
C
pSITE-nEYFP-N1-ERD4
+
pSITE-cEYFP-N1-RdRp
D
pSITE-nEYFP-N1-RdRp
+
pSITE-cEYFP-N1-IMT
E
pSITE-nEYFP-N1-IMT
+
pSITE-cEYFP-N1-RdRp

Figure 4:

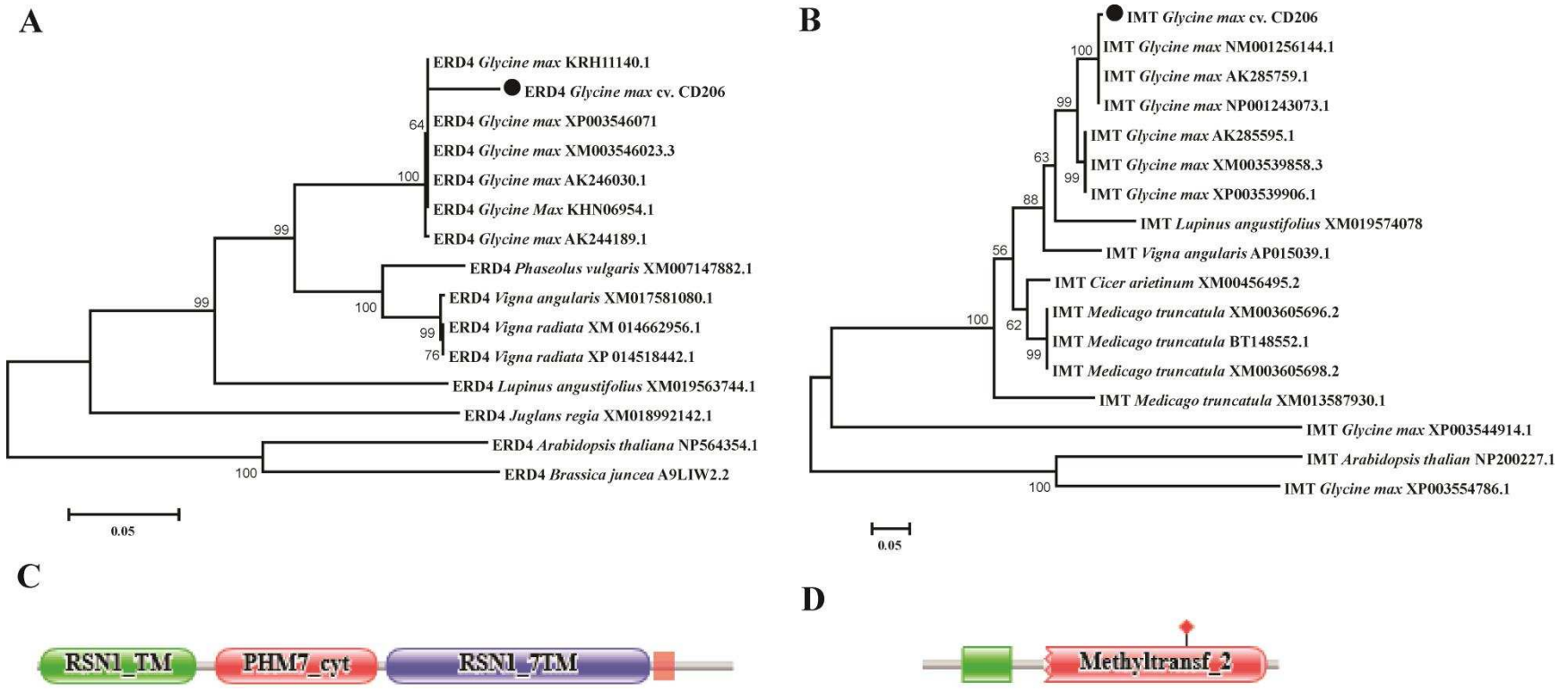

A
ERD4 Glycine max KRH11140.1
ERD4 Glycine max cv. CD206
ERD4 Glycine max XP003546071
ERD4 Glycine max XM003546023.3
ERD4 Glycine max AK246030.1
ERD4 Glycine Max KHN06954.1
ERD4 Glycine max AK244189.1
ERD4 Phaseolus vulgaris XM007147882.1
ERD4 Vigna angularis XM017581080.1
ERD4 Vigna radiata XM 014662956.1
ERD4 Vigna radiata XP 014518442.1
ERD4 Lupinus angustifolius XM019563744.1
ERD4 Juglans regia XM018992142.1
ERD4 Arabidopsis thaliana NP564354.1
ERD4 Brassica juncea A9LIW2.2
64
100
99
100
99
76
99
100
0.05
B
IMT Glycine max cv. CD206
IMT Glycine max NM001256144.1
IMT Glycine max AK285759.1
IMT Glycine max NP001243073.1
IMT Glycine max AK285595.1
IMT Glycine max XM003539858.3
IMT Glycine max XP003539906.1
IMT Lupinus angustifolius XM019574078
IMT Vigna angularis AP015039.1
IMT Cicer arietinum XM00456495.2
IMT Medicago truncatula XM003605696.2
IMT Medicago truncatula BT148552.1
IMT Medicago truncatula XM003605698.2
IMT Medicago truncatula XM013587930.1
IMT Glycine max XP003544914.1
IMT Arabidopsis thalian NP200227.1
IMT Glycine max XP003554786.1
100
99
63
99
88
56
100
62
99
100
0.05
C
RSN1_TM
PHM7_cyt
RSN1_7TM
D
Methyltransf_2

Figure 5:

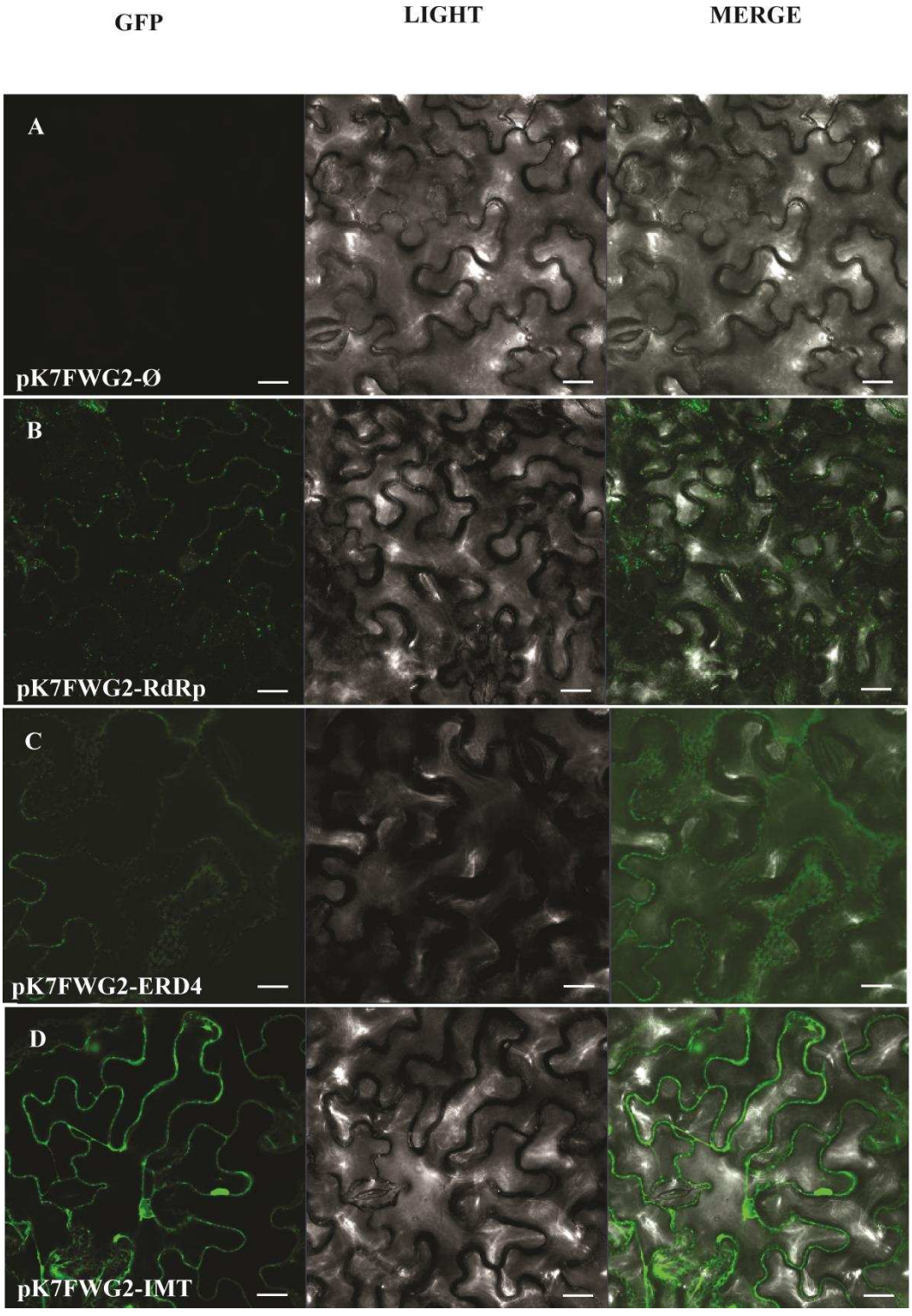

GFP
LIGHT
MERGE
A
pK7FWG2-Ø
B
pK7FWG2-RdRp
C
pK7FWG2-ERD4
D
pK7FWG2-IMT

Figure 6:

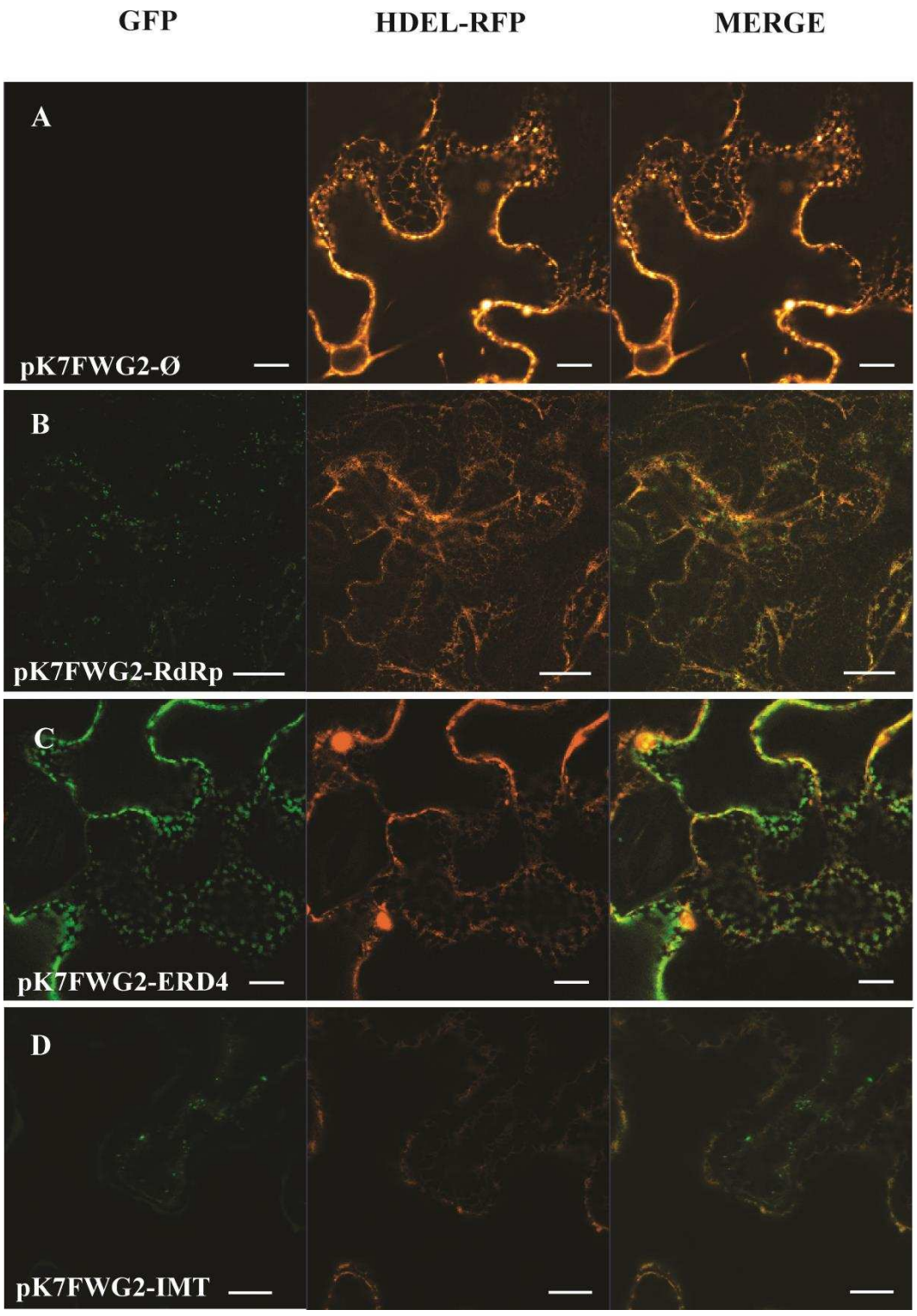

GFP
HDEL-RFP
MERGE
A
pK7FWG2-Ø
B
pK7FWG2-RdRp
C
pK7FWG2-ERD4
D
pK7FWG2-IMT

Figure 7:

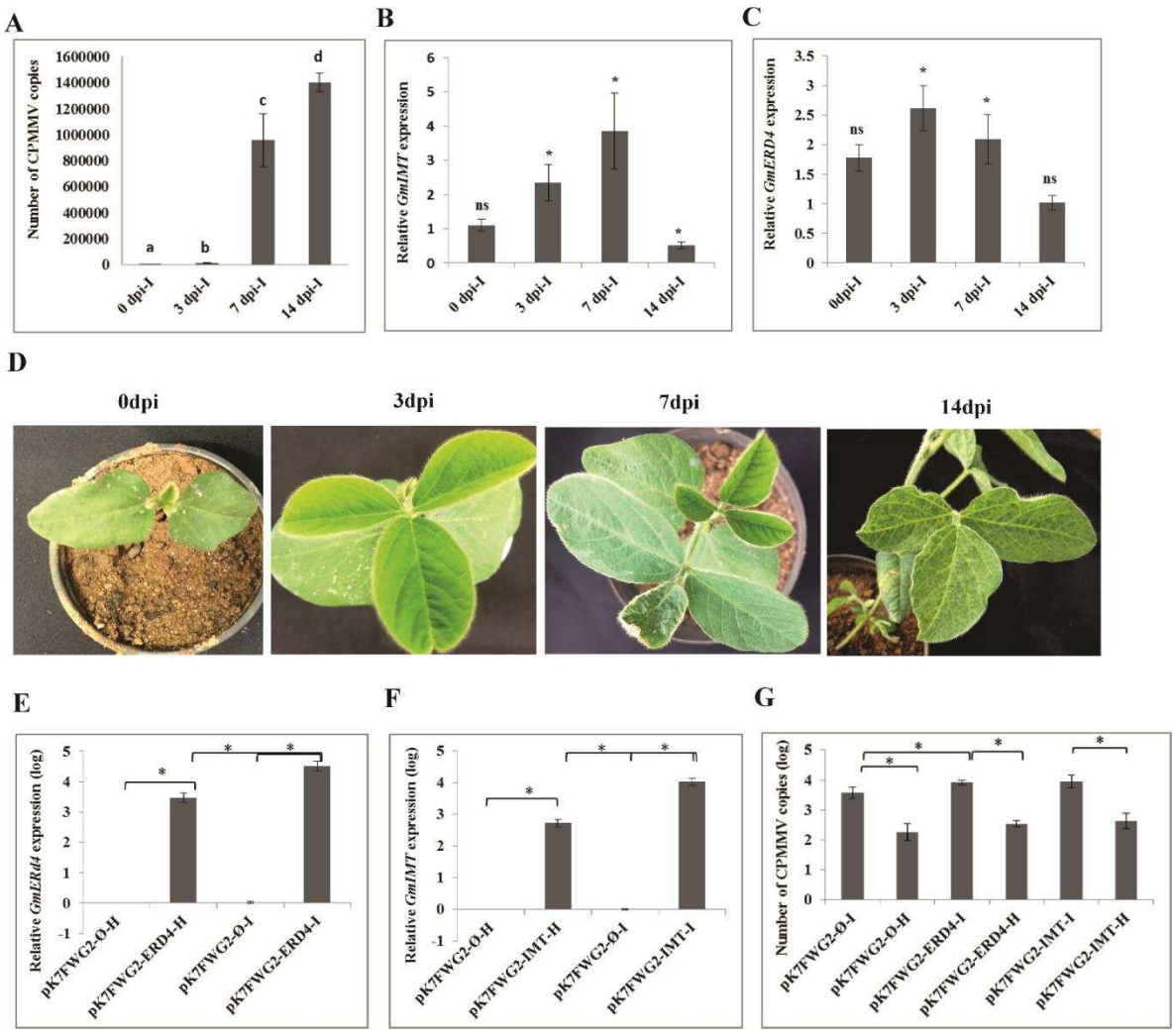
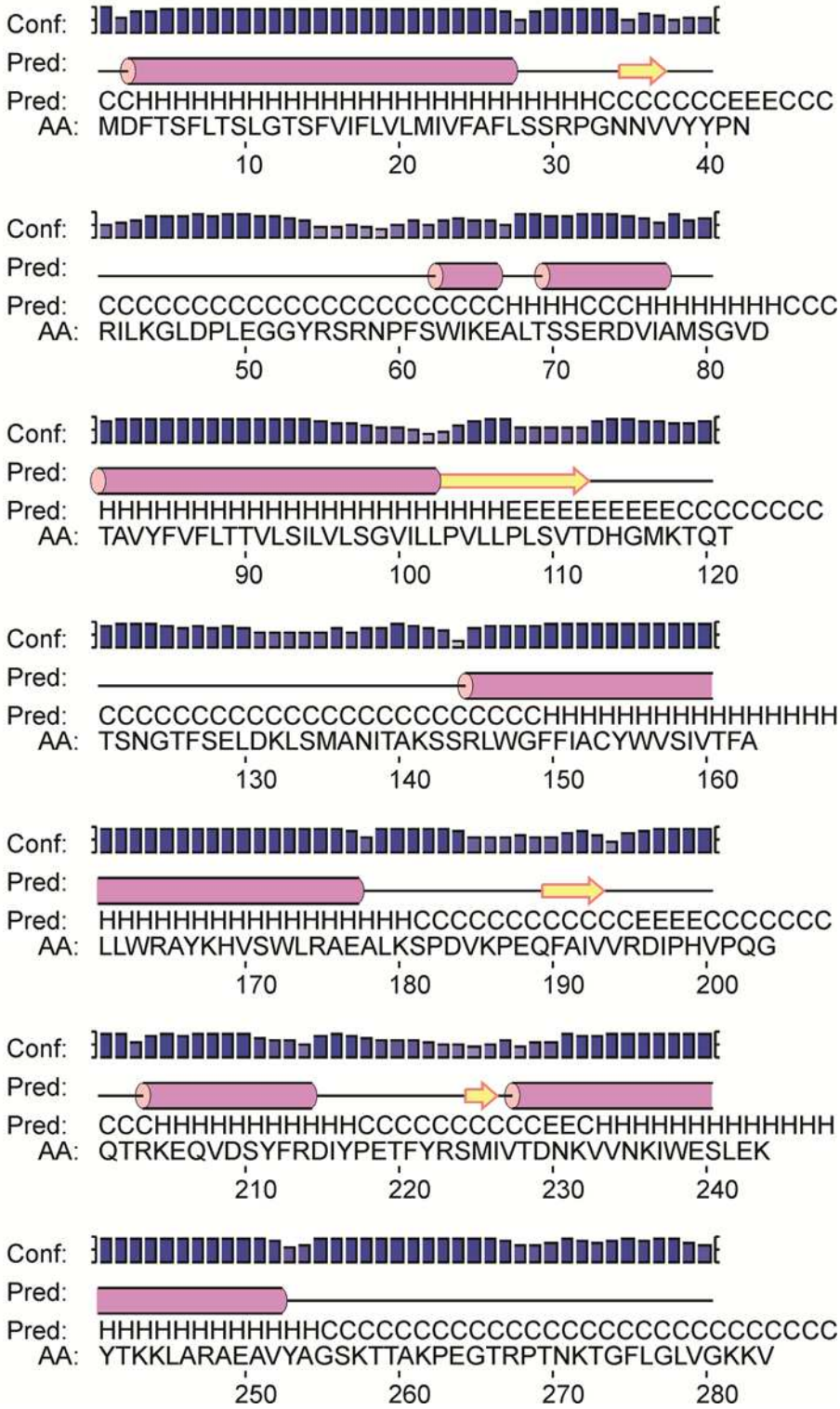
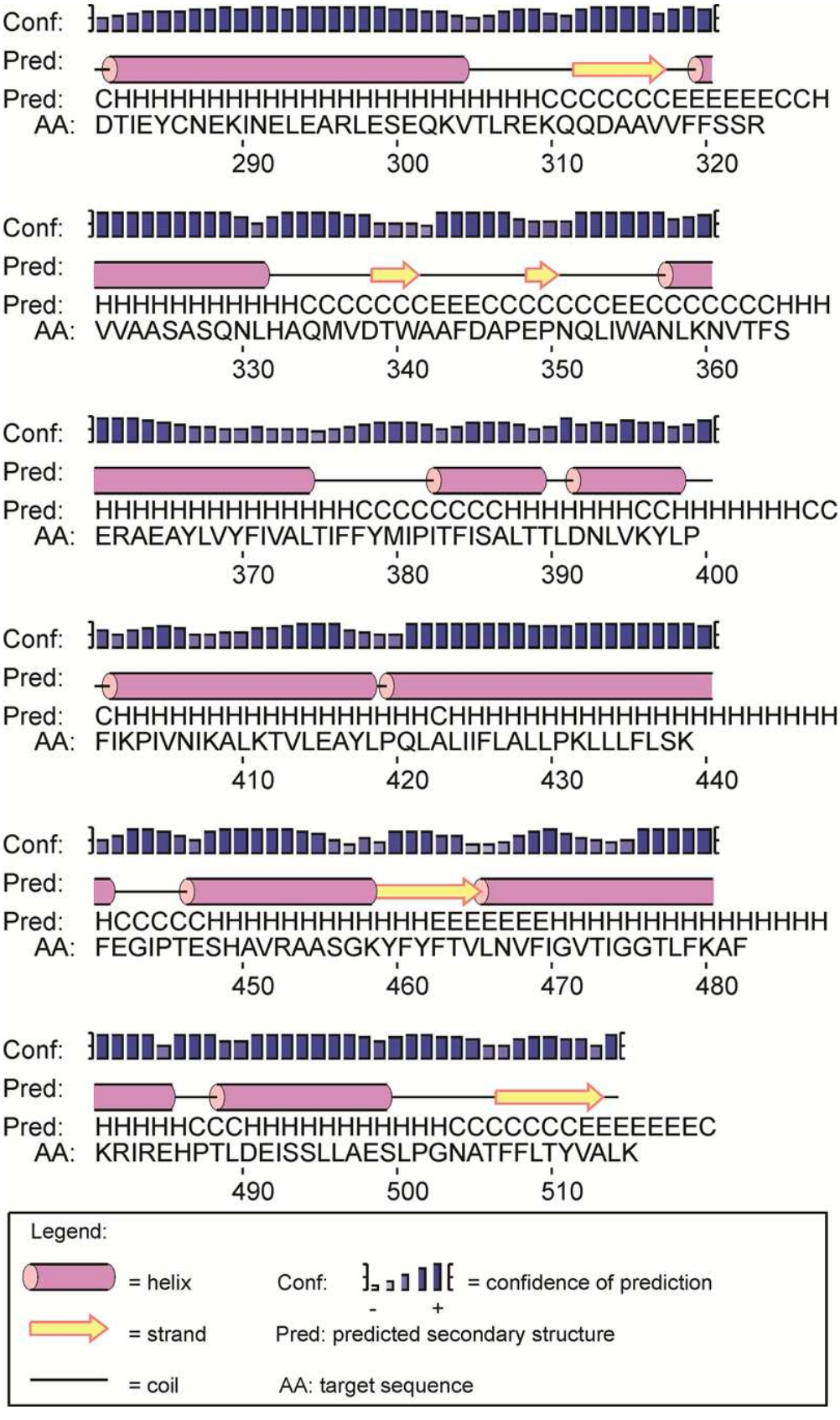

A
Number of CPMMV copies
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
a
b
c
d
0 dpi-I
3 dpi-I
7 dpi-I
14 dpi-I
B
Relative GmIMT expression
ns
*
*
*
0 dpi-I
3 dpi-I
7 dpi-I
14 dpi-I
C
Relative GmERD4 expression
ns
*
*
ns
0dpi-I
3 dpi-I
7 dpi-I
14 dpi-I
D
0dpi
3dpi
7dpi
14dpi
E
Relative GmERd4 expression (log)
pK7FWG2-Ø-H
pK7FWG2-ERD4-H
pK7FWG2-Ø-I
pK7FWG2-ERD4-I
F
Relative GmIMT expression (log)
pK7FWG2-Ø-H
pK7FWG2-IMT-H
pK7FWG2-Ø-I
pK7FWG2-IMT-I
G
Number of CPMMV copies (log)
pK7FWG2-Ø-I
pK7FWG2-Ø-H
pK7FWG2-ERD4-I
pK7FWG2-ERD4-H
pK7FWG2-IMT-I
pK7FWG2-IMT-H

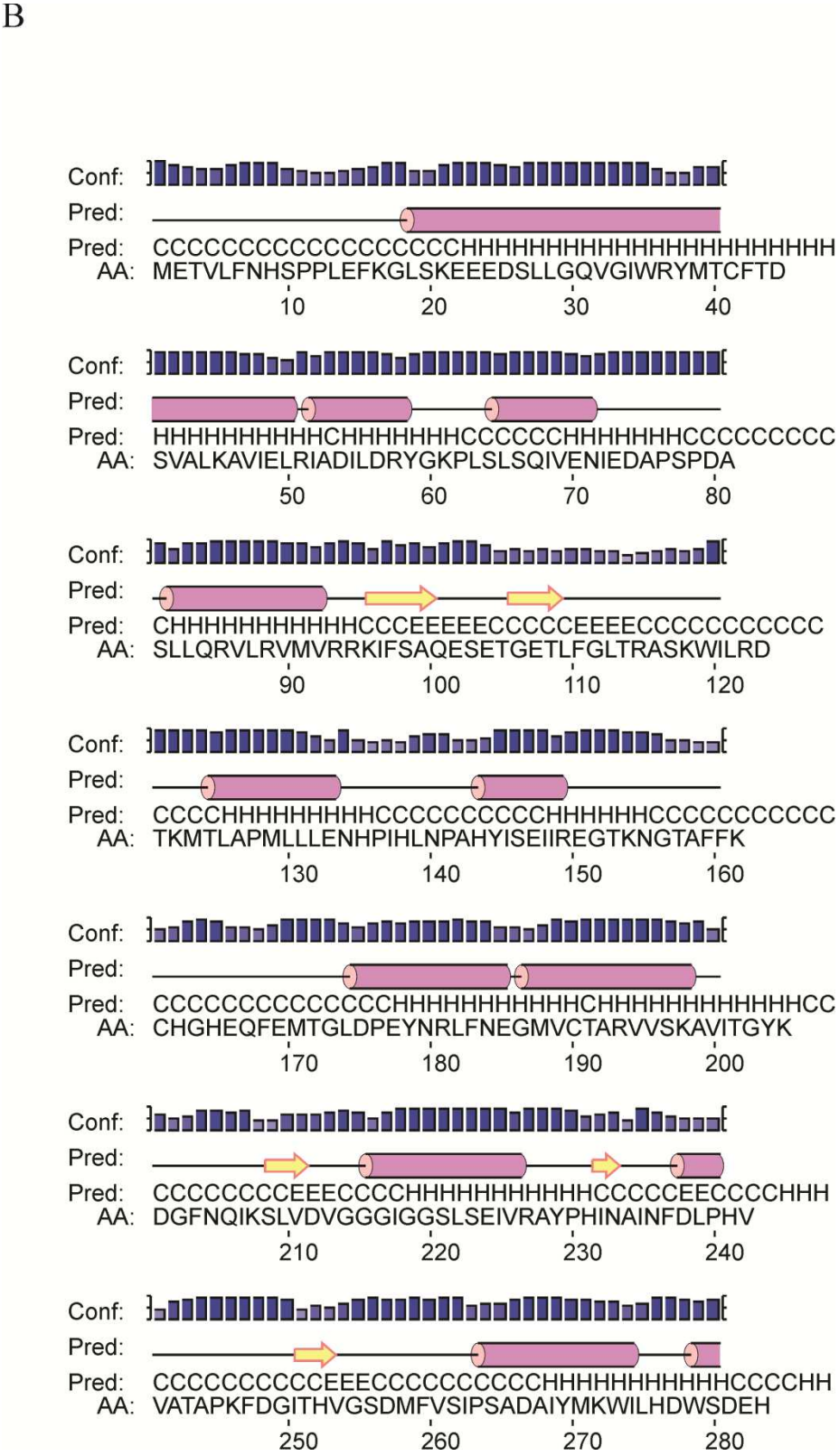
Figure S1:

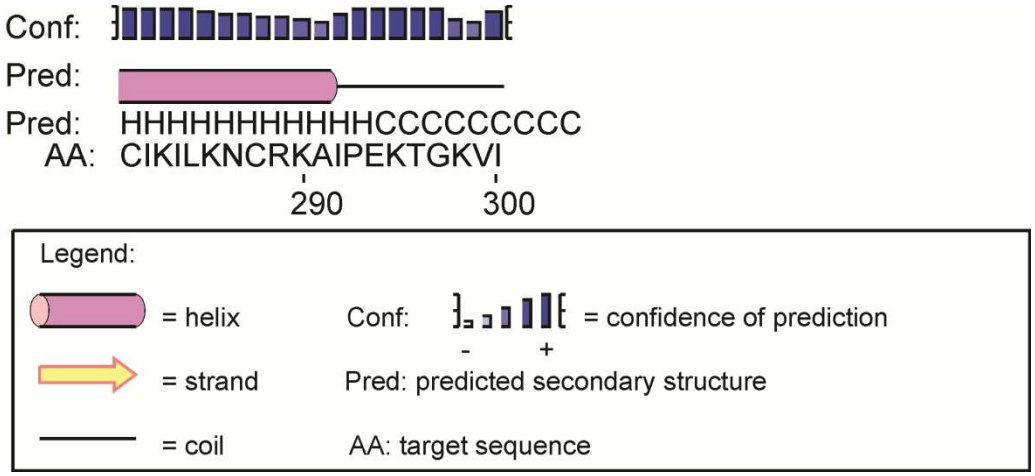
A


Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCEEECCC
AA: MDFTSFLTSLGTSFVIFLVLMIVFAFLSSRPGNNVVYYPN
10 20 30 40
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCHHHHCCCHHHHHHHHCCC
AA: RILKGLDPLEGGYRSRNPFSWIKEALTSSERDVIAMSGVD
50 60 70 80
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHEEEEEEEEECCCCCCCC
AA: TAVYFVFLTTVLSILVLSGVILLPVLLPLSVTDHGMKTQT
90 100 110 120
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHH
AA: TSNGTFSELDKLSMANITAKSSRLWGFFIACYWVSIVTFA
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHCCCCCCCCCCCCEEEECCCCCC
AA: LLWRAYKHVSWLRAEALKSPDVKPEQFAIVVRDIPHVPQG
170 180 190 200
Conf:
Pred:
Pred: CCCHHHHHHHHHHCCCCCCCCCCEECHHHHHHHHHHHH
AA: QTRKEQVDSYFRDIYPETFYRSMIVTDNKVVNKIWESLEK
210 220 230 240
Conf:
Pred:
Pred: HHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCC
AA: YTKKLARAEAVYAGSKTTAKPEGTRPTNKTGFLGLVGKKV
250 260 270 280


Conf:
Pred:
Pred: CHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCEEEEEECCH
AA: DTIEYCNEKINELEARLESEQKVTLREKQQDAAVVFFSSR
290 300 310 320
Conf:
Pred:
Pred: HHHHHHHHHHHCCCCCCCEEECCCCCCCEECCCCCCCHHH
AA: VVAASASQNLHAQMVDTWAAFDAPEPNQLIWANLKNVTFS
330 340 350 360
Conf:
Pred:
Pred: HHHHHHHHHHHHHHCCCCCCCCHHHHHHHCCHHHHHHHCC
AA: ERAEAYLVYFIVALTIFFYMIPITFISALTTLDNLVKYLP
370 380 390 400
Conf:
Pred:
Pred: CHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHH
AA: FIKPIVNIKALKTVLEAYLPQLALIIFLALLPKLLLFLSK
410 420 430 440
Conf:
Pred:
Pred: HCCCCCHHHHHHHHHHHHEEEEEEEHHHHHHHHHHHHHH
AA: FEGIPTESHAVRAASGKYFYFTVLNVFIGVTIGGTLFKAF
450 460 470 480
Conf:
Pred:
Pred: HHHHHCCCHHHHHHHHHHHCCCCCCCEEEEEEEC
AA: KRIREHPTLDEISSLLAESLPGNATFFLTYVALK
490 500 510
Legend:
= helix
Conf:
= confidence of prediction
- +
= strand
Pred: predicted secondary structure
= coil
AA: target sequence

B


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHH
AA: METVLFNHSPPLEFKGLSKEEEDSLLGQVGIWRYMTCFTD
10 20 30 40
Conf:
Pred:
Pred: HHHHHHHHHHCHHHHHHHCCCCCCHHHHHHHCCCCCCCCC
AA: SVALKAVIELRIADILDRYGKPLSLSQIVENIEDAPSPDA
50 60 70 80
Conf:
Pred:
Pred: CHHHHHHHHHHHCCCEEEEECCCCCEEEECCCCCCCCCCC
AA: SLLQRVLRVMVRRKIFSAQESETGETLFGLTRASKWILRD
90 100 110 120
Conf:
Pred:
Pred: CCCCHHHHHHHHHCCCCCCCCCCHHHHHHCCCCCCCCCCC
AA: TKMTLAPMLLLENHPIHLNPAHYISEIIREGTKNGTAFFK
130 140 150 160
Conf:
Pred:
Pred: CCCCCCCCCCCCCCHHHHHHHHHHHHCHHHHHHHHHHHCC
AA: CHGHEQFEMTGLDPEYNRLFNEGMVCTARVVSKAVITGYK
170 180 190 200
Conf:
Pred:
Pred: CCCCCCCCEEECCCCHHHHHHHHHHHHCCCCCEECCCCHHH
AA: DGFNQIKSLVDVGGGIGGSLSEIVRAYPHINAINFDLPHV
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCCCCEEECCCCCCCCCCCCHHHHHHHHHHHHCCCCHH
AA: VATAPKFDGITHVGSDMFVSIPSADAIYMKWILHDWSDEH
250 260 270 280


Conf:
Pred:
Pred: HHHHHHHHHHHCCCCCCCCC
AA: CIKILKNCRKAIPEKTGKVI
290
300
Legend:
= helix
= strand
= coil
Conf:
= confidence of prediction
-
+
Pred: predicted secondary structure
AA: target sequence

CONCLUSÕES GERAIS

- O *Cowpea mild mottle virus* (CPMMV) é um patógeno emergente no Brasil e no mundo, tendo aumentado consideravelmente sua incidência, alterado sua virulência e expandido sua gama de hospedeiros;
- O CPMMV é capaz de induzir necrose sistêmica em soja, a qual apresenta características de resposta hipersensitiva (HR);
- Inoculações sucessivas de um mesmo isolado viral de CPMMV levam a alterações na sua virulência (um isolado antes causador de necrose pode passar a causar sintomas de mosaico), essa mudança leva ao aumento da adaptabilidade do vírus (acúmulo viral) e do vetor;
- A ORF1 (replicase viral) é a região determinante de sintomas do CPMMV;
- A ORF1 é um "hot spot" de recombinação. A recombinação é o mecanismo evolutivo envolvido na alteração do padrão de sintomas, visto que sítios presentes nos blocos recombinantes foram associados à mudança no fenótipo da infecção viral;
- O domínio RNA polimerase dependente de RNA (RdRp) da replicase viral codificada pela ORF1 é capaz de interagir *in vitro* e *in vivo* com as proteínas: ERD4 tipo CSC1 (GmERD4) e inositol metiltransferase (GmIMT);
- GmERD4 localizou-se no retículo endoplasmático (RE), GmIMT apresentou localização citoplasmática enquanto o domínio RdRp induziu estruturas pontuais na membrana plasmática e sobre a rede do RE;

- *GmERD4* e *GmIMT* são induzidas após 3 e 7 dias da inoculação com o CPMMV em plantas de soja cv. CD206;
- A superexpressão de *GmERD4* em protoplastos de soja é capaz de induzir o aumento do acúmulo viral de CPMMV;
- A replicase viral é uma proteína multifuncional.