DIOGO FELIPE MILANESI

ANÁLISE DA DIVERSIDADE DE ISOLADOS DE *Cowpea mild mottle virus* EM CULTIVARES DE FEIJOEIRO CONVENCIONAIS E TRANSGÊNICAS RESISTENTES AO *Bean golden mosaic virus*

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

VIÇOSA
MINAS GERAIS – BRASIL
2017

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

T
M631a
2017

Milanesi, Diogo Felipe, 1992-
Análise da diversidade de isolados de *Cowpea mild mottle virus e*m cultivares de feijoeiro convencionais e transgênicas resistentes ao *Bean golden mosaic virus* / Diogo Felipe Milanesi. – Viçosa, MG, 2017.
vii, 47f. : il. (algumas color.) ; 29 cm.

Orientador: Claudine Márcia de Carvalho.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Vírus de planta. 2. Feijão. 3. Plantas transgênicas. 4. Variabilidade genética . 5. *Cowpea mild mottle virus.* *6. Cowpea mild mottle virus*. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Fitopatologia. II. Título.

CDD 22. ed. 632.32

DIOGO FELIPE MILANESI

ANÁLISE DA DIVERSIDADE DE ISOLADOS DE *Cowpea mild mottle virus* EM CULTIVARES DE FEIJOEIRO CONVENCIONAIS E TRANSGÊNICAS RESISTENTES AO *Bean golden mosaic virus*

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

APROVADA: 23 de fevereiro de 2017

Murilo Siqueira Alves

Francisco Murilo Zerbini Júnior

Claudine Márcia Carvalho
(Orientadora)

AGRADECIMENTOS

A Deus, por me abençoar todos os dias.

Aos meus pais, meus irmãos e meus padrinhos, que são meus maiores exemplos de vida e que sempre estão ao meu lado, me ensinando a importância da família.

Aos meus primos e ao restante da família. Que nós possamos continuar compartilhando bons momentos sempre.

À Mara pelo carinho, apoio, paciência e a força tão valiosa em tantos momentos juntos.

À Universidade Federal de Viçosa, onde adquiri muito crescimento intelectual e profissional.

Ao Departamento de Fitopatologia, professores e caros colegas de estudos.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro para realização do curso.

À professora Claudine Márcia Carvalho pela confiança, atenção, ensinamentos e orientação na realização dos trabalhos durante todos esses anos.

A todos os colegas do Laboratório de Virologia Vegetal Molecular, por compartilhar tantos ensinamentos, além das brincadeiras que animam todos os dias, sou eternamente grato a todos. Especialmente à Larissa, que alegremente me auxilia a tantos anos.

SUMÁRIO

RESUMO

MILANESI, Diogo Felipe, M.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Análise da diversidade de isolados de *Cowpea mild mottle virus* em cultivares de feijoeiro convencionais e transgênicas resistentes ao *Bean golden mosaic virus*.** Orientadora: Claudine Márcia Carvalho.

A cultura do feijoeiro comum no Brasil, além do imenso valor que representa na cadeia econômica e para milhares de agricultores no país, é fundamental devido à contribuição que possui na segurança alimentar da população. Cultivares com evento de resistência ao *Bean golden mosaic virus* (begomovirus), vírus responsável por causar uma das doenças que mais afeta a produtividade da cultura no país, foram desenvolvidas após vários anos de pesquisas. Infecções durante testes em campo desses materiais por outro vírus, o *Cowpea mild mottle virus* (carlavirus), gerou novas preocupações tanto aos pesquisadores envolvidos no projeto do feijoeiro resistente ao mosaico dourado quanto aos produtores que aguardavam a liberação comercial dessas cultivares. Apesar de alguns trabalhos já terem sido desenvolvidos a fim de se avaliar os prejuízos produtivos que o CPMMV causa sobre as isolinhas transgênicas de feijoeiro, assim como sua distribuição, nenhum conhecimento se tem sobre a diversidade desse vírus em feijão comum ou transgênico no Brasil, e poucos trabalhos dessa natureza são encontrados na literatura até hoje. Nesse trabalho, buscou-se avaliar a variabilidade de populações do CPMMV para cada uma de quinze cultivares de feijoeiro comum, sendo dez transgênicas (resistentes ao BGMV) e cinco convencionais, em um campo experimental com ocorrência e transmissão natural do CPMMV. Também foram quantificados os níveis virais em cada cultivar a partir de três repetições. Para cada uma das quinze plantas representando 15 diferentes genótipos de feijoeiro comum, o genoma completo de cinco isolados de CPMMV foi sequenciado pela montagem de sequenciamentos de blocos de PCR. Diferenças foram encontradas na variabilidade dos cinco isolados de CPMMV em plantas transgênicas e em plantas convencionais. Os valores dos descritores de variabilidade π, S, K e Θ foram geralmente maiores nos grupos de isolados de plantas transgênicas. Isso se repetiu para todas as ORF's virais analisadas. As ORF's 2, 3 e 4 foram as que tiveram a maior diversidade registrada, enquanto que a diferenças entre os grupos já citados foi mais perceptível nas regiões das ORF's 2, 5 e 6. Eventos de recombinação foram encontrados na ORF 1 viral, quase sempre ocorrendo em isolados de plantas transgênicas, assim como alguns na ORF 2

e 6. Analisando as sequencias da ORF 1, nota-se que os cinco isolados de cada planta se agrupam e tendem a formar clados próximos a grupos de isolados de genótipos hospedeiros similares, o que pode decorrer da interação entre a replicase viral e a planta. Para a região 3' do genoma, houve a separação do conjunto de 75 isolados em dois grupos de variantes. A identidade nucleotídica par a par entre isolados de grupos distintos variou entre 75 e 85%. Pelos testes de seleção, existe evidência significativa de que vária populações virais estão sobre processo de seleção não neutra. O acúmulo viral não teve diferença significativa entre plantas transgênicas e convencionais. A quantificação também não revelou diferenças em níveis virais em plantas transgênicas originadas de retrocruzamentos com a cultivar Pérola em comparação aos níveis naquelas retrocruzadas com a cultivar BRS Pontal. Os resultados desse trabalho reforçam resultados anteriores de que dois grupos de estirpes de CPMMV estão distribuídos pelas regiões produtoras brasileiras, provavelmente pela presença em plantas daninhas (onde a variabilidade desse vírus nunca foi analisada) e em hospedeiros cultivados como o próprio feijoeiro. Também comprova a alta variabilidade desse vírus de RNA, principalmente nas novas cultivares de feijoeiro resistente ao mosaico dourado por transgenia. É provável que a presença de BGMV nas cultivares convencionais e consequentemente a infecção mista dos dois vírus tenha algum efeito sobre os valores de variabilidade apresentados nesse estudo. Os mecanismos moleculares dessa interação, porém, não são conhecidos. Os resultados apresentados e o fato de que hospedeiros não cultivados estão distribuídos por grandes áreas de produção e que estes podem atuar como reservatório viral, além da grande distribuição da mosca branca pelo Brasil, fazem com que novos trabalhos com esse patógeno sejam de extrema importância.

ABSTRACT

MILANESI, Diogo Felipe, M.Sc., Universidade Federal de Viçosa, February, 2017. **Genetic diversity analysis of *Cowpea mild mottle virus* isolates in conventional and transgenic common bean cultivars resistant to *Bean golden mosaic virus*.** Advisor: Claudine Márcia Carvalho.

The common bean crop in Brazil, besides its economic importance, represents a major source of what is daily consumed by Brazilian population in terms of proteins and carbohydrates, contributing to food security. Cultivars with a transgenic resistance event to *Bean golden mosaic virus* (begomovirus), a virus that causes one of the most important diseases of common bean, were developed after many years of research. The release of these cultivars immune to BGMV is undergoing difficulties because of the re-emergence of *Cowpea mild mottle virus* (carlavirus) in common bean, which has raised some concerns for the researchers and the growers. Although works to access the damage potential into different genotypes of these resistant isolines and to investigate the virus distribution are being reported, no study is found evaluating CPMMV molecular characteristics and diversity in transgenic as well as conventional common bean cultivars in Brazil. In fact, there are very few studies of this kind globally. The objective of this work was to evaluate the variability on CPMMV populations from each of the fifteen common bean cultivars, ten transgenic and resistant to BGMV, and five conventional cultivars, from a field experiment with natural CPMMV transmission by whitefly. CPMMV was also quantified on the three plant replicates of each genotype. Five CPMMV isolates were completely sequenced on all fifteen plants with different genotypes, providing 75 full virus genomes after assembly of PCR sequence blocks. Differences in variability were found between those groups of isolates from transgenic plants to those from conventional ones. With the π, S, K, and Θ-W descriptors, we detected a considerable higher CPMMV variability within transgenic plants in comparison to the virus variability within conventional cultivars in most of the cases. This was the case for all analyzed ORF's. The ORF's 2, 3 and 4 were the ones with the highest variability in the genome; at ORF's 2, 5, and 6, the differences in variability mentioned above are most discernible. Recombination events between isolates happening at the ORF 1 region were detected, as well as at ORF 2 and at ORF 6. Mostly of these were between isolates from transgenic plants. The phylogenetic analysis with ORF 1 sequences of all seventy-five isolates reveals the formation of groups

based on host genotypes, and that these groups are most likely grouping near a cluster of isolates from a similar host plant genotype. These could be the result of the direct and specific interactions needed between the viral replicase and the plant. The results of phylogenetic analysis and sequence comparisons with the 3' region of the viral genome (ORF 2-6), divided the 75 isolates of this study into two groups of CPMMV variants. The pairwise nucleotide differences between isolates from distinct groups ranged from 75 to 85%. The selection tests at some ORF's give significant evidence that some populations are evolving under a non-random process. The viral accumulation on conventional cultivars did not differ statistically to the accumulation at transgenic plants. In addition, there is no evidence of differences between CPMMV levels at transgenic cultivars that have Pérola as the reccurent parent to those that have the BRS Pontal. The results from this work corroborate with previous studies that indicate the existence of two CPMMV strains naturally distributed in Brazilian production areas. It also confirms the expected high variability potential of this RNA virus; the high variability registered on the newly developed BGMV-resistant transgenic common bean cultivars is also troublesome. The presence of BGMV in mixed infections with CPMMV at conventional cultivars is probably influencing the results of CPMMV variability, but the molecular properties of this interaction is still unknown. These results, in addition to the fact that non-cultivated host plants are distributed along major production areas and may act as viral reservoirs and the known widespread of whiteflies in growing regions of Brazil, make further studies with this pathogen of fundamental importance.

GENETIC DIVERSITY ANALYSIS OF *Cowpea mild mottle virus* ISOLATES IN CONVENTIONAL AND TRANSGENIC COMMON BEAN CULTIVARS RESISTANT TO *Bean golden mosaic virus*.

Milanesi, D.F.

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

Abstract

With the importance of common bean crop and the lack of knowledge about molecular variability of *Cowpea mild mottle virus*, we aimed at analyze the intra-host variability of CPMMV on conventional and transgenic *Bean golden mosaic virus*- resistant common bean cultivars. For this, five clones were obtained for each RT-PCR, which were performed at different CPMMV's genome regions, with overlapping fragments. These five clones were sampled for all the different cultivars, sequenced and the analysis performed with sequences from the same plant. Fifteen different common bean genotypes from a field experiment were used. Viral accumulation was also measured by qRT-PCR on all cultivars in three replicates. In most cases, CPMMV populations from transgenic plants showed higher levels of variability in relation to those from conventional cultivars, at all genomic regions. The higher differences were verified on ORF 2, 5, and 6. The triple gene block (TGB) is the genomic region where populations presented higher nucleotide variability (π up to 0.15). At ORF 1, in contrast, isolates had around ten times lower variability levels. In contrast, nineteen recombination events were reported at this region, and only one at ORF 2 and one at ORF 6. No correlation between CPMMV accumulation and its variability was found. The virus accumulation values varied considerably across replicates, and no statistically significant difference in viral accumulation between genotypes could be detected. No difference between CPMMV levels on transgenic and conventional plants were found either. In general, the higher prevalence of different CPMMV variants in the transgenic lines and perhaps the lack of infection with BGMV were responsible for an increased variability assessed on the transgenic genotypes, and the latter will need further investigation.

Introduction

Common bean (*Phaseolus vulgaris* L., Fabaceae) is one of the most important leguminous crops, in particular as it accounts for a substantial proportion of the proteins and carbohydrates consumed daily by the poorest populations of the globe [7]. In Brazil, this bean is consumed on a day-to-day basis in such quantities that even though the country stands as the third highest producer in the world [17], it usually imports a part of its demand – notably during years of production shortfalls [32].

Among the pathogens of this crop is *Cowpea mild mottle virus* (CPMMV). This virus belongs to the genus Carlavirus, family *Betaflexiviridae* - subfamily Quinvirinae - [2], and has been found in Brazil causing angular mosaic symptoms in common bean since at least 1979 [12]. The first description of this virus was from Ghana in 1973 on cowpea (*Vigna unguiculata* L. Walp.) [8] and the same isolate was completely sequenced nearly forty years later [29]. CPMMV is a monopartite, positive single strand RNA virus with a genome of ca. 8,200 nucleotides, containing six Open Reading Frames (ORF's) [29]. The first ORF encodes the viral RNA dependent RNA polymerase (RdRp), followed by the movement proteins encoded in the triple gene block (TGB; ORF 2-4), the coat protein (CP; ORF5) and a nucleic acid binding protein (NABP) in ORF6. The genome has a [m7GpppG] cap structure linked to its 5' terminus, and a 3' poly(A) tail [2, 26]. Carlavirus virions are flexuous filaments of 610-700 nm in length and 12-15 nm in diameter with helical symmetry [1].

CPMMV is transmitted by the whitefly *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in a non-persistent manner [3, 22, 28, 30]. This insect has become a major problem for Brazilian agriculture, especially because of its role in spreading viruses in a number of field crops [11]. The introduction of different cryptic species that diverge in host range and aggressiveness compared to the *Bemisia tabaci* "A" biotype is one of the explanations for this [25]. As recently reported, the whitefly formerly labelled the "B" biotype of *B. tabaci* – but now defined as the species Middle East-Asia Minor 1 [14] – has become the most prevalent in Brazil and is responsible for losses in important agricultural crops [23, 24]. Among the dozens of different viruses that they are able to transmit, some are responsible for causing what are considered emerging diseases on a global scale [31], and among these is CPMMV and the diseases that it causes.

Reports of infection and of losses [20] by CPMMV in *P. vulgaris* have not been significant when compared with other viral species, most notably *Bean golden mosaic virus* (BGMV - *Begomovirus*) and *Bean common mosaic virus* (BCMV - *Potyvirus*). The former has received considerably more attention from researchers due to its great damage potential. As a direct result of this, the Brazilian Agricultural Research Corporation (Embrapa) developed and has been performing agronomic tests on BGMV-resistant transgenic *P. vulgaris* cultivars [5]. During this process, the appearance of viral symptoms during the field trials has become a problem for the release of this technology into the market. The serological and molecular tests have now confirmed that the wide variety of symptoms, that included crinkled leaves and enations visible as necrotic veins, resulted from CPMMV infections [33].

In Brazil, the occurrence of an important disease of soybean (*Glycine max* (L.) Merr; Fabaceae), soybean stem necrosis, that appeared in the 2001/2002 growing season and in some instances caused total losses on areas cultivated with susceptible varieties, was the reason why this virus gained attention and became the subject of more detailed investigation [3]. Since then, dozens of Brazilian isolates have been partially sequenced and the genetic studies have provided some information about the diversity and evolution of this virus. The variety of symptoms that CPMMV induces on soybeans has also been described and a possible division into two strains was proposed [3, 39, 40] . In addition, its presence was confirmed in all major soybean-producing regions in Brazil, with exception of Rio Grande do Sul. A recent study that addresses the re-emergence of CPMMV (especially on transgenic cultivars) [33] also reports the presence of this virus on associated weeds considered alternative hosts. Therefore, it is reasonable to think that this virus is circulating through important agricultural areas on soybean, beans, and other host crops that are still being reported [10], via non-cultivated plants nearby.

As an RNA virus, CPMMV is expected to exhibit a high genetic variability [15, 18, 19]. This variability is usually driven not only by high mutation rates, but also by recombination [21]. In Brazilian isolates of CPMMV, recombination seems to play an important role in its evolution [39, 40]. Studies of genetic diversity of *Cowpea mild mottle virus* are still uncommon and with the first full-length genome being sequenced only in 2010, the majority of studies published until now focus mainly on records of the virus in new countries (or regions) and hosts [4, 6, 35, 38].

In summary, the lack of information about this virus in the newly developed cultivars of common bean resistant to BGMV, and the importance of assessing the genetic changes that may be occurring in CPMMV, with resultant threats to important crops in Brazil, makes additional work in this area vital. Therefore, this study aims to evaluate and characterize the intra-host diversity of CPMMV in *Phaseolus vulgaris* from a field experiment with fifteen common bean cultivars (ten transgenic and resistant to BGMV, and five conventional ones), submitted to natural whitefly transmission. CPMMV quantification from three plant replicates with these genotypes was also carried out using real-time PCR.

Material and methods

Field experiment

Samples were collected from a randomized block design field experiment consisting of 15 different common bean genotypes - ten transgenic lines (with resistance to BGMV) and five conventional lines (Table 1). This trial was conducted at Embrapa Arroz e Feijão (Santo Antônio de Goiás, GO) during the summer/autumn of 2015, in three replicated blocks, with no use of insecticides. At the end of the vegetative state and the beginning of flowering, approximately three leaves were collected from each of the three plant replicates of every genotype described above and stored at -80°C until further usage.

RT-PCR and cloning

Frozen leaves of all fifteen plants from the first replicate representing all genotypes were macerated in liquid nitrogen and 100 mg of leaf tissue submitted to total RNA extraction using the PureLink® RNA Mini Kit (Ambion™) following the manufacturer's instructions. Reverse transcription was performed with Superscript III (Invitrogen) using 500 ng of RNA and the CPMMV-specific primer "ORF6R" (5' - TAAAACCAGGAAATAAC - 3'). PCR amplifications were then carried out with Platinum *Taq* DNA Polymerase (Invitrogen) and forward and reverse primers as previously described in [39], generating eight fragments covering the entire CPMMV genome. All PCR cycles were according to [39]. The generated amplicons of corresponding base pair lengths were purified from agarose gels after electrophoresis with illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare), ligated into pGEM-T Easy Vector (Promega) and transformed into *Escherichia coli* DH5α

cells by electroporation. The plasmid DNA was purified from transformed bacterial cultures using the illustra plasmidPrep Mini Spin Kit (GE Healthcare). Five clones from each of the fifteen plants were sequenced in both directions using universal primers (M13F/M13R) at Macrogen (Seoul, South Korea).

Sequence analyses and phylogeny

Contigs were assembled from the overlapping forward and reverse sequences using DNA BASER Sequence Assembler v.4.36 (Heracle Biosoft). Contigs were then submitted to Blastn searches to confirm that they had similarity with CPMMV. The online program ORF Finder https://www.ncbi.nlm.nih.gov/orffinder/ was used to confirm that the ORFs were not truncated. The sequences from each ORF were aligned by the translated amino acids using the MUSCLE [16] module in MEGA v. 6.06 and the pairwise comparisons between nucleotides (and between the amino acids) done with Sequence Demarcation Tool (SDT v.1.2) http://web.cbio.uct.ac.za/. To investigate viral sequence polymorphism within populations and to compare them, we used DNAsp v.5.10 [36]. Variables evaluated were: (i) the total number of segregating sites (s); (ii) mean nucleotide differences between sequences (k); (iii) nucleotide diversity (π); (iv) number of mutations; (v) number of haplotypes (h); (vi) haplotype diversity (Hd) and (vii) Watterson's estimator for the population-scaled mutation rate based on the total number of segregating sites (θ-W). To calculate the π statistic based on nucleotide position on each ORF, a sliding window of 100 bases with step sizes of 50 bases (ORF's 1, 2, 3, and 5) and 25 bases (ORF's 4, 6), was computed for all fifteen populations.

Phylogenetic trees were constructed for each of the six ORF's individually (ORF 1-6). Phylogenetic relationships were inferred using Bayesian inference (BI) in MrBayes v.3.2.6 [34], at CIPRES Science Gateway V. 3.3 portal (https://www.phylo.org/portal2/home.action). Two runs with four Markov Chain Monte Carlo (MCMC) simulations were conducted simultaneously using 10 million generations; Burn-in was set at 25% from the resulting trees. Trees were visualized with FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) and midpoint rooted.

Selection tests

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 [36].

Gene and site-specific selection were also measured for each viral protein and population in both experiments. Detection of sites under negative and positive selection in the genes was performed 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://datamonkey.org/dataupload.php) 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.

Recombination analyses

The Recombination Detection Program (RDP v.4.80) [27] was used for detection of potential recombinant sequences, their likely parental sequences, and the localization of possible recombination break points, only for the sequences generated here, and for each ORF individually. The recombination-detecting methods RDP (R), Genecon (G), Maximum chi square (M), Bootscan (B), Sister scan (S), Chimaera (C) and 3 SEQ (3S) were selected and only when three or more methods from all seven detected the same recombination event the result was considered valid. Default settings and a multiple comparison-corrected P-value cutoff of 0.05 were used. In addition, a density plot with statistical permutation test to detect significant potential breakpoint hot- and cold-spots, with 10,000 permutations, window size of 200, step size of 100, and confidence intervals (CI) of 95 and 99%, was constructed. This allowed determining whether the breakpoint distribution was significantly non-random.

qRT-PCR

The absolute quantification from three plant replicates of the fifteen genotypes in Table 1 was performed in individual tubes using the specific CPMMV primers CPF (5'-ATAGTGAGATGGCTGATAAACAAAAAC-3') and CPR (5'-TTCAGCATCAATGT CTGGAAG-3'), specific for ORF5 (CP coding region). The cDNAs were diluted accordingly and used in qPCR reactions in a StepOnePlus (Applied Biosystems) instrument with SYBER® Green PCR master mix (Applied Biosystems) and primers. The reactions consisted of 10 minutes at 94°C followed by 40 cycles of 15 seconds at 94°C and 1 min at 60°C. All samples were submitted to thermal denaturation so that their dissociation curve (melting) could be determined and the specificity of the amplification reactions checked. The standard curve was obtained using increasing amounts of amplicon of CPMMV CP gene. The significance of the differences between medians was determined using the Kruskal-Wallis nonparametric test with Dunn's multiple comparison test in GraphPad Prism v.5.01 (GraphPad Software, San Diego California). Significance (*P*) values < 0.05 indicate statistically significant differences between the compared medians for any specified comparison.

Results

CPMMV sequences and initial characterization

The RT-PCR reactions using primers designed for the CPMMV CP region allowed the confirmation that all fifteen common bean plants from replicate 1 were infected. For each plant, eight RT-PCR amplifications were carried out to sequence the full virus genome. Five of these fragments were assembled in order to obtain the complete ORF 1 sequence of each isolate. The initial comparisons using Blastn and SDT for all ORF sequences revealed nucleotide (nt) identity values ranging from 75-100 % (Figure S1). It was verified that the genomes had the six typical carlavirus ORF's [2, 29]. The proteins followed the expected values of amino acid lengths: 1859aa (ORF1), 231aa (ORF2), 106aa (ORF3), 68aa (ORF4), 288aa (ORF5), and 113aa (ORF6). There were isolates with slightly different predicted proteins sizes. One isolate in treatment 3 had 1,851aa, one from treatment 6 had 1,774aa, and one isolate from treatment 10 and another from treatment 4 had 1,858 amino acids in the

RdRp, codified by ORF 1. ORF 6 also had differences in some isolates. Variants with NABP's of 103, 108, and 110 amino acids were sequenced.

CPMMV genetic diversity is higher in transgenic genotypes

The genetic variability of CPMMV was evaluated for each coding region and by groups of five isolates from each plant genotype. In the six ORF's, the intra-host mean nucleotide diversity (π) of CPMMV from the transgenic bean genotypes was consistently higher than that from the groups of isolates in conventional plants (Table S1), especially for ORF's 2, 5, and 6. Although differences up to 100-fold were found, not all transgenic plants had groups of CPMMV isolates with high π values. Occasionally, at some genomic regions, CPMMV diversity from a conventional plant was intermediate or equally high to that seen in the five isolates from a transgenic plant. The values of π from each plant group of isolates at ORF 1 were below 0.015, which is approximately ten times lower than the values at other genomic regions (Table S1). Values around 0.15 were found for isolates from transgenic cultivars at ORF 3. The virus nucleotide diversity was analyzed in the length of the genomic coding regions (ORF 1-6), also separating isolates by cultivars (Figure 1 and data not shown for ORF 1). At all these regions, the transgenic plants (treatments 1-10) contained the groups of CPMMV isolates with the highest π values (Figure 1). Analyzing the other intra-host polymorphism descriptors and comparing them for all genotypes, it stands out that there is a separation between the values for isolates from conventional and transgenic cultivars (Table S1). The CPMMV isolates from transgenic bean showed higher numbers of segregating sites than isolates from conventional genotypes (Table S1). The mutation rate (θ-W) in isolates from transgenic beans was also higher (10^{-2}-10^{-1}) than isolates from conventional ones (10^{-3}-10^{-1}). At ORF 1 the smallest values of θ-W were obtained. These results suggest that the transgenic plants allowed higher genetic variation for CPMMV populations in relation to non-transgenic plants, although the presence of more plants with both types of variants of CPMMV in transgenic genotypes may have been the responsible for an increased variability in the regions of ORF2-6 in these plants.

ORF1 is the region with higher number of recombination events in the genome

As recombination may represent an important factor influencing diversity, specially haplotype diversity, recombination analysis were performed using all CPMMV sequences from this study, separated by ORF (Table S2). For ORF 1, the analysis resulted in evidence of nineteen recombination events with support of three methods at least. Of these, only one of the recombinants had any identified parental sequence (major or minor) isolated from a different plant than the one from which the recombinant clone was isolated. In this case, the recombinant was a clone from treatment 8 (125-Nil Pontal) that has a minor parent also from treatment 8 and a major parent from treatment 4 (10-5-6-1-2-Nil Pérola), both transgenic genotypes (Table S2). In all the other eighteen cases, the major and the minor parents (when both were identified), or only one of them (when the other was unknown), were isolated from the same plant as the recombinant. Although the experiment had 5/15 (33 %) conventional plants, only 3/19 (15.8 %) of the events where detected from these genotypes, two from the same treatment (#15 – CNFP 15882). There was also one recombination event at ORF 2 and one at ORF 6. In both cases, isolates from different transgenic plants were detected as major and minor parents (Table S2). Although CPMMV shows lower genetic polymorphism at ORF1 when compared to the other regions of the genome, it is the region with the highest number of recombination events. The recombination density plot for ORF 1 (data not shown) suggests the existence of three regions with high recombination rates. Most of them, however, co-localize with the overlapping regions from different RT-PCRs used to assemble the complete ORF 1 sequence, which may represent an artificial factor that creates this specific regions. Outside ORF 1, no evidence of regions with higher recombination events rate are supported.

ORF1 phylogenetic tree shows clustering based on cultivars

Phylogenetic analyses were conducted for each coding region (ORF's 1-6). In the ORF 1 phylogenetic tree, with just one exception, all five CPMMV clones from the same plant clustered together, and formed a different clade for each plant, suggesting that isolates from each plant represent different CPMMV populations. In the only divergent case, three clones from treatment 1 formed one clade and the other two formed another clade nearby (Figure 2a). In the trees generated for the other ORF's, a separation into two distinct groups

is evident (Figure 2 b-f), which was not formed by genotype neither transgenic or conventional plant differences, and with an nucleotide variability that is close to the ones found in previous studies that propose the existence of two CPMMV strains (see below). The clades and supported by a posterior probability of 1.

Two different groups of CPMMV variants identified at the ORF 2-6 region

The formation of two separate groups in the identity matrixes (Figure S1 b-f) was capable of differentiating the isolates (in ORF 2-6) exactly the way that they are divided in their phylogenetic trees. Thirty-one isolates diverged into this second group both in the ORF 3 and the ORF 4 coding regions. For ORF's 5 and 6, this group (group 2) has seven isolates. With the amino acid identity matrix of the CP we cannot separate the isolates in any way since the protein is extremely conserved among them, with values higher than 97% (data not shown). The sequences from ORF's 3 and 4, and ORF's 5 and 6, were obtained from exactly the same clones (the PCR fragments for this region contain more than one ORF). At ORF 2, nineteen isolates formed group 2 (Figure S1b).

The NABP of the majority of isolates from CPMMV is normally composed of 113 amino acids. However, in the distinct group that contains seven isolates, six have NABP's with 103aa; the other (ORF6.112.6) is the recombinant isolate and contains 113aa. Inside group 2, the nucleotide similarity among isolates is around 100% at this region; between groups, the similarities are around 80%. The recombinant isolate, although similar to isolates in group 2, has identity values of approximately 85 % to the isolates of group 1 and of 95 % with the ones from group 2 (Figure S1f). This recombinant isolate forms an exclusive clade in the ORF 6 phylogenetic tree (Figure 2f), but that cluster with the group 2 clade, where the ORF6.110.1 isolate, its minor parent, is located. None of this occurs for this isolate at ORF 5 (as might be expected since they come from the same clone), which gives more support to a recombination event. There is something similar with the recombinant isolate ORF2.102.1. In this case, however, a higher similarity to the major parent is evident in both the ORF 2 phylogenetic tree and the identity matrix (Figure 2b and S1b). Another isolate – ORF2.109.1 – separates in the Bayesian inferred tree from the other members of group 1, and could be the result of recombination, but the event was only supported by two methods (data not shown).

Other differences in the NABP include two clones from treatment 1 that have early stop codons, presenting 110aa (ORF6.103.2 and ORF6.103.9) and four from treatment 14 that have 108aa (ORF6.114.1-ORF6.114.5). Besides from clustering in the phylogenetic trees of ORF 5 and 6, these sequences are very similar to the other 68 sequences from group 1 (Figure 2 e-f).

Selection is not responsible for reducing variability, but some codons are under positive selection

In order to verify possible selection sites in the CPMMV genome for each plant genotype, and compare between conventional and transgenic ones, the selection analysis on CPMMV isolates from each plant and on all ORF's were performed and are summarized in Table S3. In general, based on Tajima's D, and Fu and Li's D and F tests (independent of the ORF analyzed), the populations are not under selection (Table S3). This suggests that most of the variability observed in the different genomic regions is not reduced by selection. There are some deviations from neutrality at populations for some ORF's. The tests showed statistical significant values for isolates at ORF 1 (populations from treatments 12 and 13); ORF2 (populations from treatments 7, 10, and 15); ORF3 (treatments 2, 3, 5, 6, 7, and 10); ORF4 (treatments 3, 6, 7, and 10); ORF 5 and 6 (on treatment 12). Analysis of selection site-by-site for all ORF's of CPMMV isolates are also summarized in Table S3. Positive selection sites were detected by REL and PARRIS on groups of isolates at ORF 1. For ORF's 2-6 the REL, FEL and IFEL methods detected a few sites under positive selection. SLAC, the most conservative test, did not detected any such sites. In contrast, all methods detected a variety of sites under negative selection for several isolates through the genome. A high correlation between variability and number of sites under negative selection detected is noticeable when comparing Table S1 and Table S3. Furthermore, in plants where the variability descriptors displayed values highly inconsistent with the others (frequently conventional cultivars) at a specific ORF, the same pattern manifested in the selection analysis.

Accumulation of CPMMV is independent of the host genotype

CPMMV accumulation was determined in the fifteen bean cultivars on all experiment replicates. The objective was to evaluate the effect of the host genotype on CPMMV viral

accumulation. The number of CPMMV copies in the three plants from each treatment varied considerably. Because of the low number of replicates (3), and the fact that we cannot assume a Gaussian distribution of the data, we used a Kruskal-Wallis with Dunns' post-hoc test and no significant difference of medians between genotypes ($P > 0.05$) was detected (Figure 3). Transformations with log and other functions gave a better fit of the data to the normal distribution but there was no differences of means by ANOVA with the Tukey post-hoc test with a significance of 0.05. A boxplot with the number of CPMMV copies in transgenic versus conventional plants is presented in Figure S2b. Using the unpaired t test to compare between both groups means did not result in a significant difference ($P > 0.05$), nor could it support a difference between levels of CPMMV in the transgenic cultivars that have Pérola as the recurrent parent (Nil Pérola – treatment 1-4) against the ones that have the BRS Pontal (Nil Pontal – treatment 5-10 (Figure S2c). When only the first two replicates were analyzed, the median CPMMV number of copies at both groups of transgenic plants were almost identical (Figure S2c).

Discussion

CPMMV was described for the first time in Brazil by Costa et al. (1983). After this, the virus only became a problem in 2000 when it was reported infecting soybean in fields from different Brazilian states. The symptoms it caused were diverse, and the virus attracted attention because of the severe symptoms (stem and leaf necrosis), causing high production losses. Initially, the virus was a problem only for soybean, but in 2013 it was described as a problem in common bean transgenic lines resistant to BGMV [33]. This re-emergence coincided with the period of field tests for the release of the resistant event into commercial cultivars. The natural occurrence of CPMMV in combination with BGMV makes it difficult to assess the virus' distribution across production areas of Brazil and the yield reduction they can cause. Additionally, the fact that alternative hosts were reported to be infected by CPMMV [33] and therefore to constitute viral reservoirs, makes the study of virus evolution with this pathogen extremely necessary. Here, the diversity of CPMMV within conventional (five cultivars) and transgenic common bean (ten BGMV- resistant isolines) was evaluated using five full genome sequences from each. The analysis demonstrated that the CPMMV

isolates from transgenic genotypes formed highly polymorphic populations and high variability was detected at all genomic regions analyzed.

The high polymorphism of CPMMV from transgenic plants could be a result of the absence of BGMV, which would favor the infection by CPMMV on these cultivars. In contrast, begomovirus was not detected by PCR with universal primers (not showed) and there was no difference in accumulation of CPMMV between these groups. In case a mixed infection occurred, it may be possible that BGMV in conventional common beans will somehow interfere with the variability of CPMMV. Nevertheless, this effect has not been reported in a study with *Plum pox virus* (PPV), for which it was demonstrated that there were no significant differences in the genetic diversity of PPV populations infecting transgenic and conventional plums [9]. In our study, differences in polymorphism were not associated with CPMMV's ability to replicate, since no differences in viral accumulation were observed between transgenic and non-transgenic hosts.

The sequence analysis for ORF's 2-6 allowed the identification of two possible, naturally occurring, groups of CPMMV variants present in the experimental field. The formation of two CPMMV groups in the phylogenetic tree has been reported before [40], and these groups were classified as two different strains. Both of these seem to be distributed throughout almost all growing regions of Brazil. The identities between groups were practically the same (for each ORF) to what was former described, although at ORF 5 the amino acid identity values (not showed) between all isolates sequenced here were above 97 %. A few of our sequences also encoded shorter versions of the NABP protein, at 103, 108, and 110 aa. A variant with 103 aa was also described in soybean isolates [40], demonstrating that the analysis here were able to capture a good degree of variation from CPMMV populations.

ORF 1 from all populations was also analyzed. It was verified that the phylogenetic clustering among the isolates was based on host distribution. ORF 1 is responsible for coding the viral replicase, the first protein translated during viral infections, so these results might suggest that the smaller polymorphism of this protein can be the result of the virus-host interactions required. Besides generating a functional protein that is required for replication, the viral replicase needs to interact with different host factors in a host-specific way. Because of this, the variations at ORF 1 could be host specific. Lower levels of polymorphism were

observed in ORF 1 in addition to some variations in the length of the predicted protein. One sequence (ORF1.112.1) for example, had an early stop codon and a predicted replicase with 1,774 aa. The others do not have some amino acids across the ORF or even late start codons, resulting in predicted shorter proteins of 1,851 and 1,858 amino acids. These variations in the CPMMV replicase length were also observed in isolates from soybean [39]. ORF 1 was also the region where most recombination events were detected (19). Recombination at ORF 1 was previously detected in Brazilian isolates infecting soybean [39]. Together with our results, it appears that this is a common characteristic of CPMMV. In the earlier described case of PPV, no recombination evidence was observed in isolates from transgenic or non-transgenic plum plants.

The selection analysis in CPMMV revealed that in most populations the variation observed in different regions of the CPMMV genome is not being restricted by selection. Only in some cases was the occurrence of selection observed. In the site-by-site analysis of selection, sites under positive and negative selection were found. As previously described in [40], at ORF 5 (CP) a high number of sites under negative selection are detected. At the ORF 6 region of isolates from treatment 1, two clones had a divergent 110 amino acids NABP and there is evidence that they are under positive selection.

The effect of the host genotype on viral accumulation was also explored here. It is noteworthy that the results are not only influenced by the virus replication rates in each genotype but also from the likely different moment of infection, since the plants were subjected to natural whitefly transmission. However, the variation in numbers of copies between replicate 1 and 2 are not high for most of the cultivars (Figure S2a). In some of the transgenic cultivars (treatments 4 and 5), low CPMMV copy numbers was detected in all experiment blocks, even in replicate 2 where the highest infection values were observed across all plants (Figure S2a). Although this could mean they can impair CPMMV replication, the difference in productivity between these two cultivars is significant and favors the one with the higher number of virus copies (CNFCT 16205 – treatment 5) as has been demonstrated recently [37]. For the conventional cultivars, the presence of mixed infection with begomoviruses may represent an important aspect for CPMMV variability and replication and still has not been evaluated experimentally but should give a good insight into the interaction of these two viruses in the field. The treatments 13 and 15 were the

conventional ones that consistently had some of the lowest values of CPMMV copies. The IPR Eldorado (treatment 13) is supposed to be resistant to BGMV and to show reduced symptoms and better yields under high disease pressure conditions that usually heavily affects susceptible cultivars. It was shown that this cultivar is less attractive for *B. tabaci* oviposition [13], but the influence this could have had on CPMMV or even BGMV transmission in this experiment is questionable. IPR Eldorado was the plant with the lowest number of CPMMV copies (in replicate one) (Figure S2a) and with the lowest CPMMV variability values detected throughout the entire CPMMV genome. With an apparently lower level of CPMMV in conventional plants compared to transgenic ones, there is an indication that the mixed infection with BGMV can reduce CPMMV infection ability besides its diversity, but the lack of controlled inoculation and the absence of begomovirus in conventional plants tested in replicate one makes it impossible to confirm this without further experiments.

A recent study with some of the same cultivars used here shows that under high CPMMV incidence, the transgenic cultivars that have Pontal as a recurrent parent have better yields than those that have Pérola, even though the symptoms are more severe in the first group [37]. Our data show no significant difference of virus load between the two cultivars. A conventional cultivar evaluated in that study (BRB 169) had a very low incidence of virus disease (caused by both BGMV and CPMMV) compared to other conventional ones, besides a good yield. In our analysis, this genotype (treatment 14) had high CPMMV levels in all three replicates and low CPMMV variability registered on all ORF's. This could represent greater CPMMV fitness on this cultivar or higher infection capacity under mixed infections, although no significant selection pressure was detected in the isolates from this cultivar. The lower disease incidence on some transgenic cultivars registered in previous studies could be the result of symptomless CPMMV infections. In our work, those plant genotypes also had high CPMMV accumulation. Altogether, we could suggest that no relation between symptom development and virus levels in conventional and transgenic common bean cultivars exists for CPMMV.

The results presented here certainly contribute at increasing our knowledge about the molecular diversity of CPMMV not only in common bean but in the BGMV-resistant transgenic cultivars.

References

1. Adams MJ, Antoniw JF, Bar-Joseph M, Brunt AA, Candresse T, Foster GD, Martelli GP, Milne RG, Zavriev SK, Fauquet CM (2004) The new plant virus family *Flexiviridae* and assessment of molecular criteria for species demarcation. Archives of Virology 149:1045-1060

2. Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, Yoshikawa N (2012) Family *Betaflexiviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, pp 920-941

3. Almeida AMR, Piuga FF, Marin SRR, Kitajima EW, Gaspar JO, Oliveira TGd, Moraes TGd (2005) Detection and partial characterization of a carlavirus causing stem necrosis of soybean in Brazil. Fitopatologia Brasileira 30:191-194

4. Baranwal V, Jain P, Saritha R, Jain R, Gautam N (2015) Detection and partial characterization of cowpea mild mottle virus in mungbean and urdbean by deep sequencing and RT-PCR. Crop Protection 75:77-79

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

6. Brito M, Fernández-Rodríguez T, Garrido MJ, Mejías 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

7. Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.) – model food legumes. Plant and Soil 252:55-128

8. 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

9. Capote N, Pérez-Panadés J, Monzó C, Carbonell E, Urbaneja A, Scorza R, Ravelonandro M, Cambra M (2008) Assessment of the diversity and dynamics of Plum pox virus and aphid populations in transgenic European plums under Mediterranean conditions. Transgenic Research 17:367

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

11. Costa A (1975) Increase in the populational density of *Bemisia Tabaci*, a threat of widespread virus infection of legume crops in brazil. In: Bird, J., 2012. Tropical diseases of legumes. Elsevier.

12. 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

13. Da Silva AG, Boiça Junior AL, S. Farias PR, L. Rodrigues NE, S. De Souza BH, Bottega DB, Chiorato AF (2014) Non-preference for oviposition and antibiosis in bean cultivars to *Bemisia tabaci* biotype B (Hemiptera: Aleyrodidae). Revista Colombiana de Entomología 40:7-14

14. Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P (2014) 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

15. Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. Annual Review of Microbiology 51:151-178

16. Edgar R (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BioMed Central Bioinformatics 5:113

17. FAO (2014) Food and Agriculture Organization of the United Nations. FAOSTAT Statistics Database.

18. Garcia-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annual Review Phytopathology 39:157-186

19. Garcia-Arenal F, Fraile A, Malpica JM (2003) Variation and evolution of plant virus populations. International Microbiology 6:225-232

20. Gaspar J, Beriam L, Alves M, Oliveira A, Costa A (1985) Serological identity of bean angular mosaic and cowpea mild mottle viruses. Fitopatologia Brasileira 10:195-199

21. Holmes EC (2009) The evolutionary genetics of emerging viruses. Annual Review of Ecology, Evolution, and Systematics 40:353–372

22. Jeyanandarajah P, Brunt AA (1993) The natural occurrence, transmission, properties and possible affinities of *Cowpea mild mottle virus*. Journal of Phytopathology 137:148-156

23. Lima LHC, Návia D, Inglis PW, Oliveira MRVd (2000) Survey of Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) biotypes in Brazil using RAPD markers. Genetics and Molecular Biology 23:781-785

24. 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

25. Lourenção AL, Nagai H (1994) Surtos populacionais de *Bemisia tabaci* no estado de São Paulo. Bragantia 53:53-59

26. Martelli GP, Adams MJ, Kreuze JF, Dolja VV (2007) Family *Flexiviridae*: a case study in virion and genome plasticity. Annual Review Phytopathology 45:73-100

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

28. 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

29. 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

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

31. Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S (2011) Emerging virus diseases transmitted by whiteflies. Annual Review of Phytopathology 49:219-248

32. Pachico, D.1989. Trends in world common bean production . 2. ed . In: Schwartz, H.F.; Pastor-Corrales, M.A. (eds.). Bean production problems in the tropics . Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. p. 1-8

33. 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. doi:10.1094/APSFeature-2016-10.

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

35. Rosario K, Capobianco H, Ng TF, Breitbart M, Polston JE (2014) RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America. PLoS One 9:e86748

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

37. Santana MV (2015) Danos do *Cowpea mild mottle virus* (CpMMV) e de moscabranca (*Bemisia tabaci* Genn.) no feijoeiro-comum geneticamente modificado resistente ao *Bean golden mosaic virus*. Escola de Agronomia e Engenharia de Alimentos - EAEA (RG). Universidade Federal de Goiás, p 99

38. Tavasoli M, Shahraeen N, Ghorbani S (2009) Serological and RT-PCR detection of *Cowpea mild mottle Carlavirus* infecting soybean. Journal of General and Molecular Virology 1:7-11

39. Zanardo LG, Silva FN, Bicalho AAC, Urquiza GPC, Lima ATM, Almeida AMR, Zerbini FM, Carvalho CM (2014) Molecular and biological characterization of

Cowpea mild mottle virus isolates infecting soybean in Brazil and evidence of recombination. Plant Pathology 63:456-465

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

Figure legends:

Figure 1: Nucleotide diversity (π) plotted along the sequences' nucleotide position generated by a sliding window with step size of 50 nt at ORF 2 **A**; 50 nt at ORF 3 **B**; 25 nt ORF 4 **C**; 50 nt ORF 5 **D**; and 25 nt ORF6 **E**. The lines represents π variation throughout the ORF sequence. π was evaluated from a population of five CPMMV isolates within each plant. The dashed lines represent values from conventional cultivars and solid lines from transgenic. The lines representing CPMMV variability from transgenic plants are colored based on the recurrent parent that generated the isolines.

Figure 2: Phylogenetic relationships between all 75 sequences generated in this study from each virus ORF. The Bayesian inference implemented in MrBayes v.3.2.6 was used. The phylogenetic trees from ORF 1 to 6 are represented in **A**-**F**. The isolates are identified according to the plant identification number (Table 1), the CPMMV genome region and a different number on each of the five clones from the same plant.

Figure 3: Column bar graph representing CPMMV number of copies (±SE) on each of the ten transgenic treatments T1-T10 and five conventional treatments C11-C15. No significant differences were found on virus accumulation between genotypes.

Supplementary figure S1: Nucleotide pairwise comparisons between all 75 CPMMV sequences from this study divided by genomic regions. **A-F** represent the nucleotide identity matrixes from ORF 1 to 6

Supplementary figure S2: A: bar plot of CPMMV number of copies per treatment separated by the three replicates. **B:** box plot comparing number of CPMMV copies between conventional and transgenic cultivars. Means are represented in boxes with the respective values. The means did not differ statistically by t test ($P>0.05$). **C:** box plot of CPMMV levels on transgenic isolines from Pérola and Pontal recurrent parent with values from three replicates, and **D** only the first two replicates.

Table 1 List of common bean cultivars used on this study

Plant ID	Treatment	Name	Genotype	Cultivar
103	1	1-4-8-2-3-Nil Pérola	CNFCT 16201	Transgenic
109	2	1-10-6-1-2-Nil Pérola	CNFCT 16202	Transgenic
111	3	1-16-10-2-Nil Pérola	CNFCT 16203	Transgenic
101	4	10-5-6-1-2-Nil Pérola	CNFCT 16204	Transgenic
110	5	094-Nil Pontal	CNFCT 16205	Transgenic
112	6	097-Nil Pontal	CNFCT 16206	Transgenic
105	7	107-Nil Pontal	CNFCT 16207	Transgenic
102	8	125-Nil Pontal	CNFCT 16208	Transgenic
113	9	132-Nil Pontal	CNFCT 16209	Transgenic
107	10	138-Nil Pontal	CNFCT 16210	Transgenic
106	11	Pérola	BGF 7384	Conventional
115	12	BRS Pontal	CNFC 7813	Conventional
104	13	IPR Eldorado	IPREldorado	Conventional
114	14	BRB 169	BGF 8419	Conventional
108	15	CNFP 15882	CNFP 15882	Conventional

Figure 1

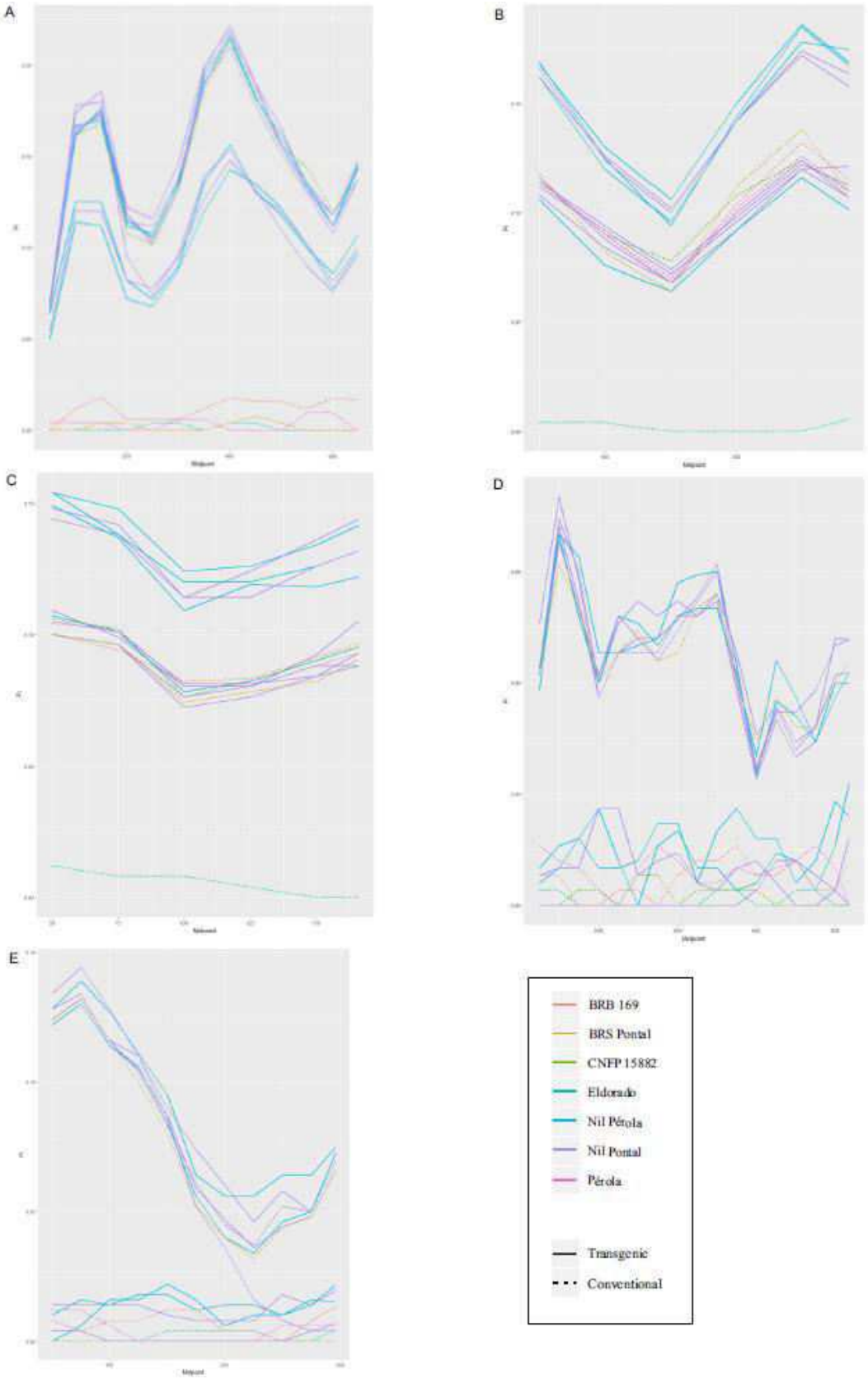
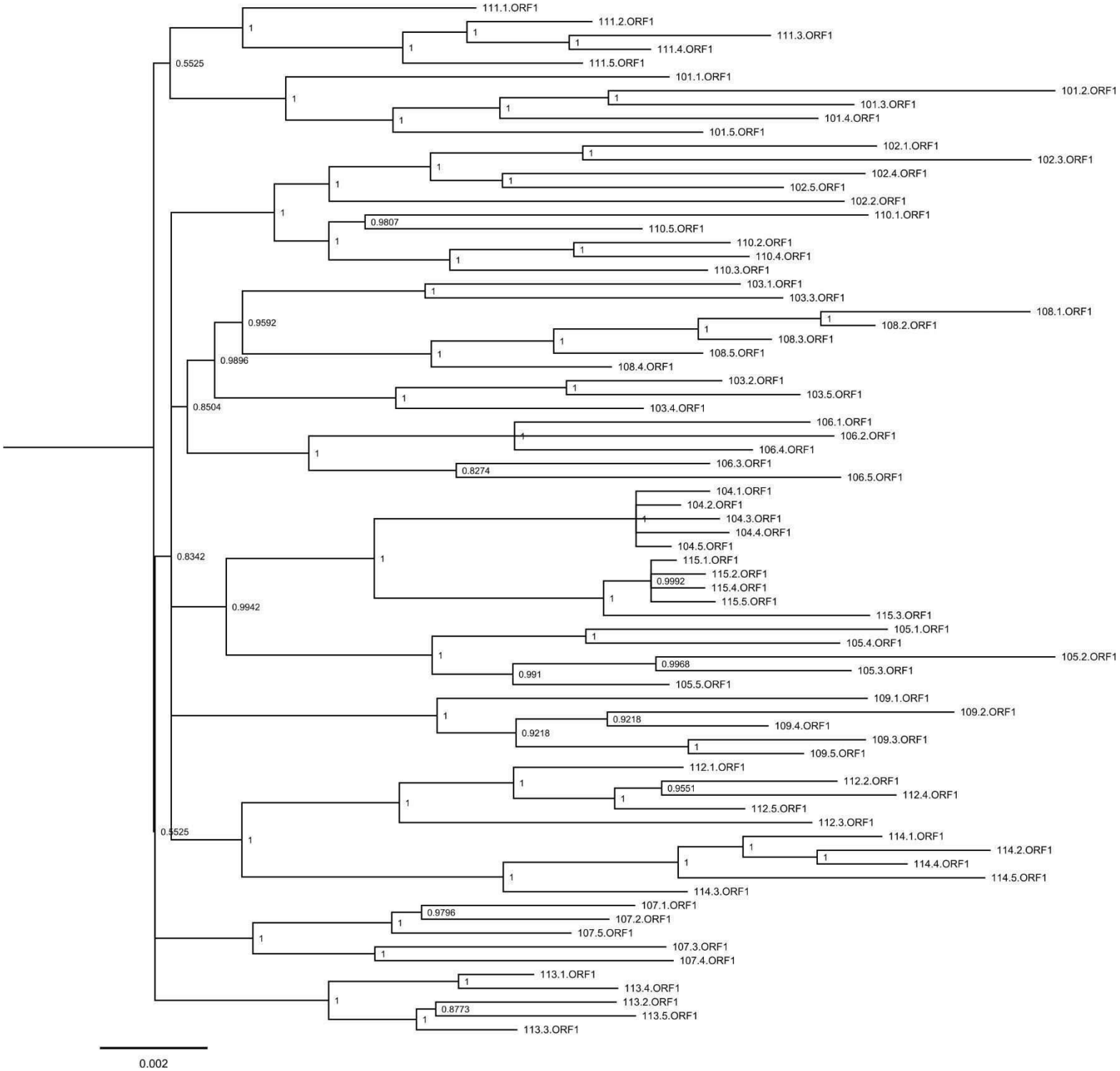

A
B
C
D
E
BRB 169
BRS Pontal
CNFP 15882
Eldorado
Nil Pérola
Nil Pontal
Pérola
Transgenic
Conventional

Figure 2

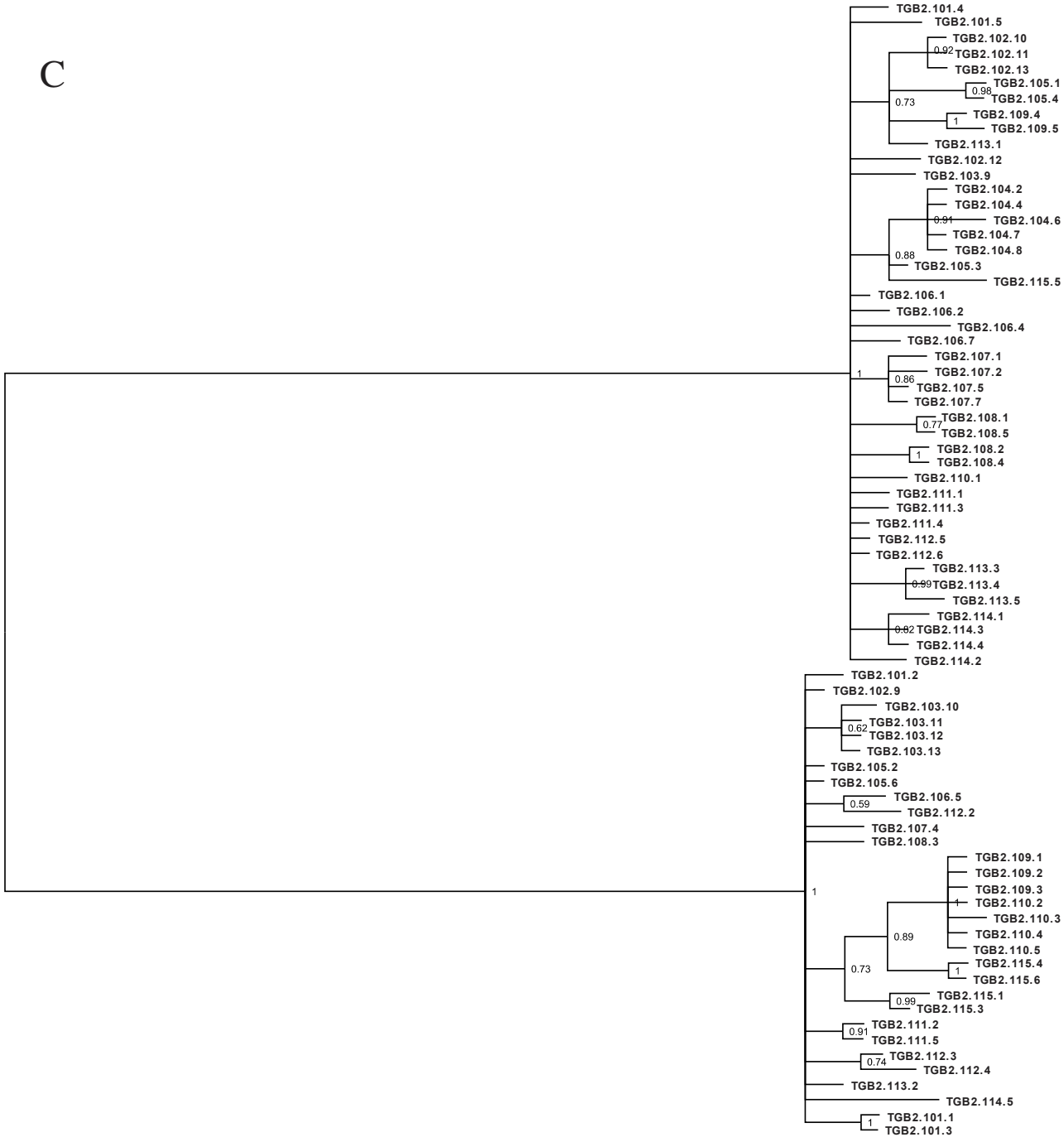
A


111.1.ORF1
111.2.ORF1
111.3.ORF1
111.4.ORF1
111.5.ORF1
0.5525
101.1.ORF1
101.2.ORF1
101.3.ORF1
101.4.ORF1
101.5.ORF1
102.1.ORF1
102.3.ORF1
102.4.ORF1
102.5.ORF1
102.2.ORF1
110.1.ORF1
0.9807
110.5.ORF1
110.2.ORF1
110.4.ORF1
110.3.ORF1
103.1.ORF1
103.3.ORF1
108.1.ORF1
0.9592
108.2.ORF1
108.3.ORF1
108.5.ORF1
0.9896
108.4.ORF1
103.2.ORF1
103.5.ORF1
0.8504
103.4.ORF1
106.1.ORF1
106.2.ORF1
106.4.ORF1
106.3.ORF1
0.8274
106.5.ORF1
104.1.ORF1
104.2.ORF1
104.3.ORF1
104.4.ORF1
104.5.ORF1
0.8342
115.1.ORF1
115.2.ORF1
0.9992
115.4.ORF1
115.5.ORF1
0.9942
115.3.ORF1
105.1.ORF1
105.4.ORF1
105.2.ORF1
0.9968
105.3.ORF1
0.991
105.5.ORF1
109.1.ORF1
109.2.ORF1
0.9218
109.4.ORF1
0.9218
109.3.ORF1
109.5.ORF1
112.1.ORF1
112.2.ORF1
0.9551
112.4.ORF1
112.5.ORF1
112.3.ORF1
0.5525
114.1.ORF1
114.2.ORF1
114.4.ORF1
114.5.ORF1
114.3.ORF1
107.1.ORF1
0.9796
107.2.ORF1
107.5.ORF1
107.3.ORF1
107.4.ORF1
113.1.ORF1
113.4.ORF1
113.2.ORF1
0.8773
113.5.ORF1
113.3.ORF1
0.002

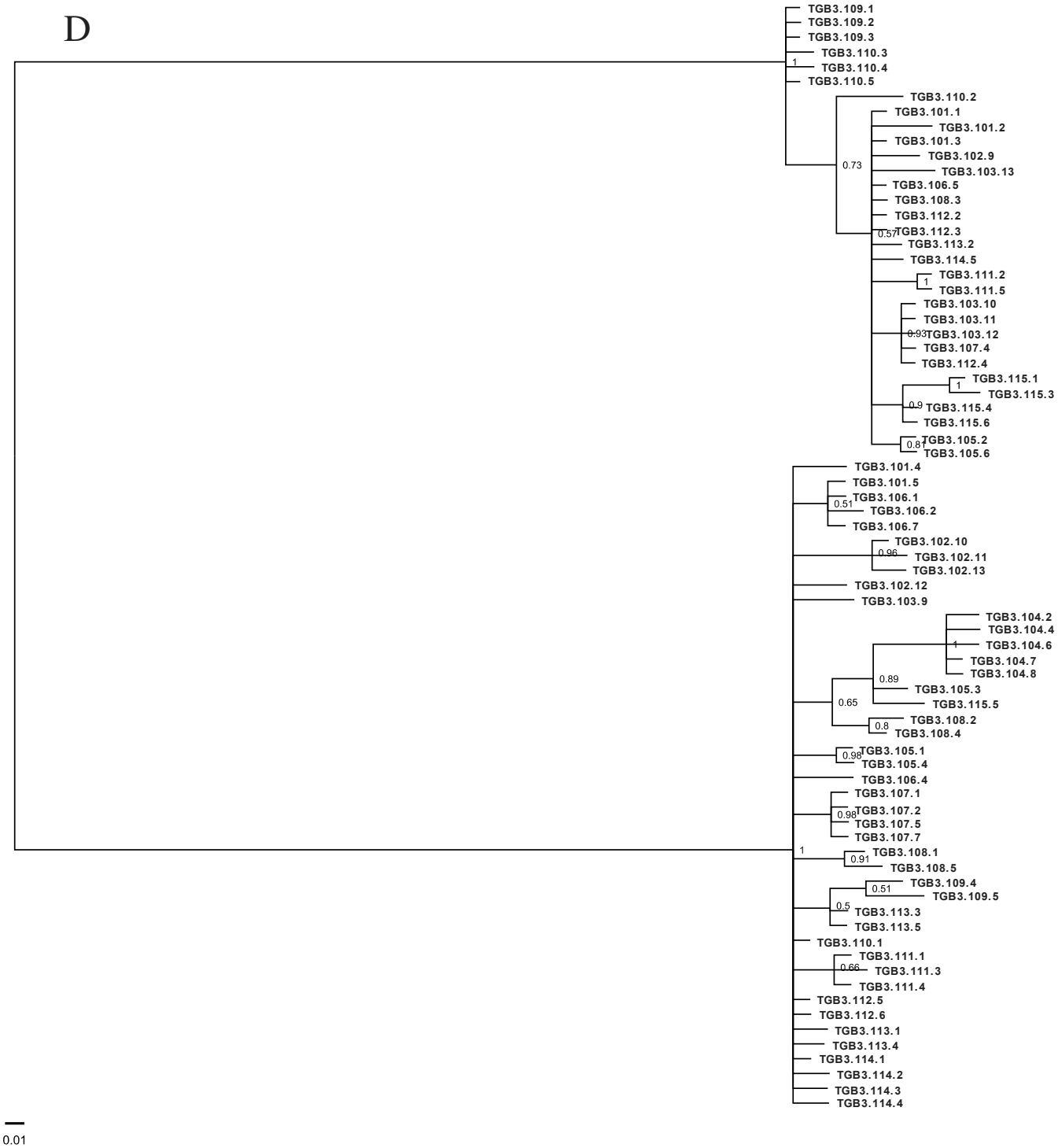
B

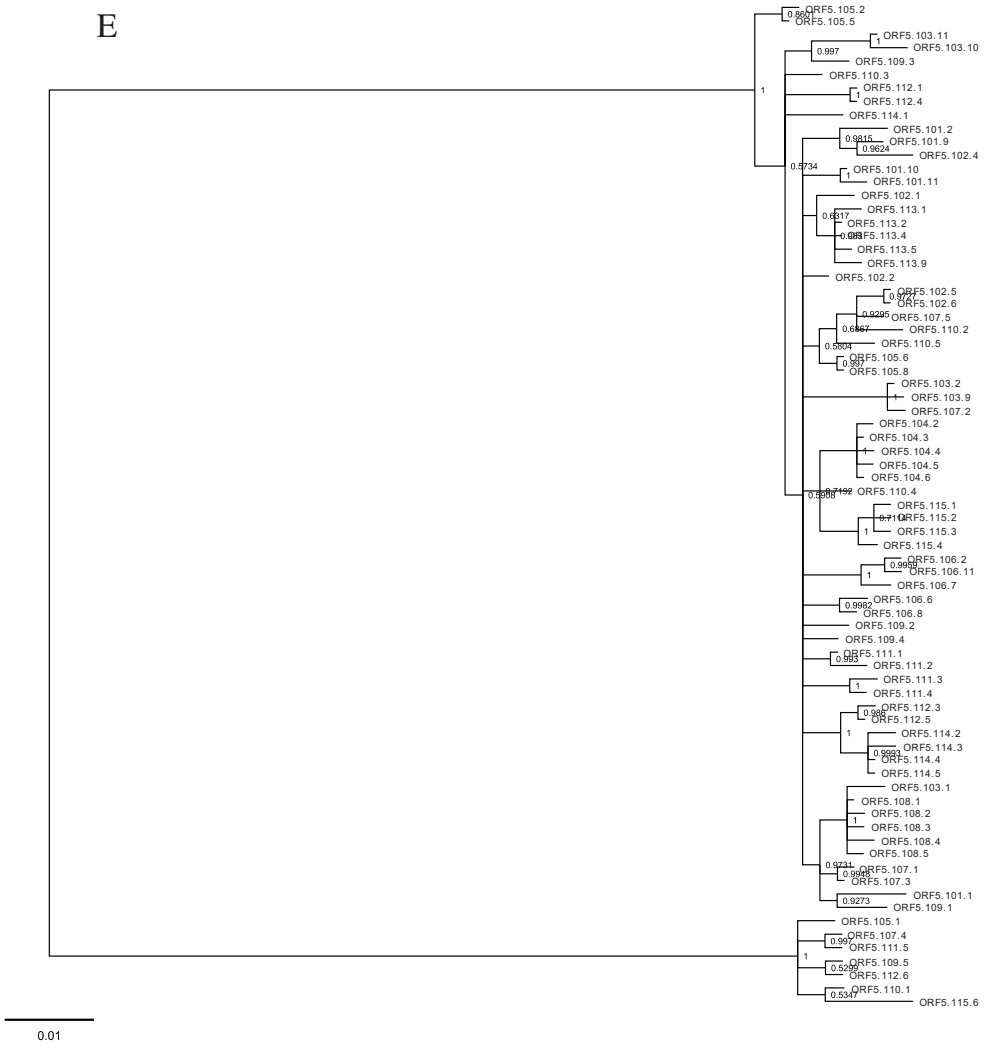

TGB1.101.2
TGB1.101.3
TGB1.101.4
TGB1.105.3
TGB1.105.5
TGB1.107.1
TGB1.107.3
TGB1.110.1
TGB1.110.2
TGB1.111.1
TGB1.111.2
TGB1.113.5
TGB1.112.5
TGB1.102.2
TGB1.103.1
TGB1.108.1
TGB1.108.5
TGB1.109.3
TGB1.112.1
TGB1.102.1
TGB1.109.1
TGB1.101.1
TGB1.101.5
TGB1.106.2
TGB1.106.4
TGB1.106.5
TGB1.107.4
TGB1.108.2
TGB1.108.3
TGB1.108.4
TGB1.111.4
TGB1.114.1
TGB1.114.2
TGB1.114.4
TGB1.112.2
TGB1.112.4
TGB1.103.2
TGB1.103.4
TGB1.103.5
TGB1.113.1
TGB1.113.3
TGB1.113.4
TGB1.103.3
TGB1.105.1
TGB1.105.4
TGB1.109.2
TGB1.109.5
TGB1.114.3
TGB1.114.5
TGB1.112.3
TGB1.107.2
TGB1.107.5
TGB1.110.3
TGB1.102.3
TGB1.102.4
TGB1.102.5
TGB1.109.4
TGB1.106.1
TGB1.106.3
TGB1.111.3
TGB1.111.5
TGB1.104.1
TGB1.104.2
TGB1.104.3
TGB1.104.4
TGB1.104.5
TGB1.105.2
TGB1.115.1
TGB1.115.2
TGB1.115.3
TGB1.115.4
TGB1.115.5
TGB1.110.4
TGB1.110.5
TGB1.113.2
0.01

C


TGB2.101.4
TGB2.101.5
TGB2.102.10
0.92
TGB2.102.11
TGB2.102.13
0.98
TGB2.105.1
0.73
TGB2.105.4
1
TGB2.109.4
TGB2.109.5
TGB2.113.1
TGB2.102.12
TGB2.103.9
TGB2.104.2
TGB2.104.4
0.91
TGB2.104.6
TGB2.104.7
TGB2.104.8
0.88
TGB2.105.3
TGB2.115.5
TGB2.106.1
TGB2.106.2
TGB2.106.4
TGB2.106.7
TGB2.107.1
1
TGB2.107.2
0.86
TGB2.107.5
TGB2.107.7
0.77
TGB2.108.1
TGB2.108.5
1
TGB2.108.2
TGB2.108.4
TGB2.110.1
TGB2.111.1
TGB2.111.3
TGB2.111.4
TGB2.112.5
TGB2.112.6
TGB2.113.3
0.99
TGB2.113.4
TGB2.113.5
TGB2.114.1
0.82
TGB2.114.3
TGB2.114.4
TGB2.114.2
TGB2.101.2
TGB2.102.9
TGB2.103.10
0.62
TGB2.103.11
TGB2.103.12
TGB2.103.13
TGB2.105.2
TGB2.105.6
0.59
TGB2.106.5
TGB2.112.2
TGB2.107.4
TGB2.108.3
1
TGB2.109.1
TGB2.109.2
TGB2.109.3
1
TGB2.110.2
TGB2.110.3
0.89
TGB2.110.4
TGB2.110.5
0.73
1
TGB2.115.4
TGB2.115.6
0.99
TGB2.115.1
TGB2.115.3
0.91
TGB2.111.2
TGB2.111.5
0.74
TGB2.112.3
TGB2.112.4
TGB2.113.2
TGB2.114.5
1
TGB2.101.1
TGB2.101.3
0.2

D


TGB3.109.1
TGB3.109.2
TGB3.109.3
TGB3.110.3
1
TGB3.110.4
TGB3.110.5
TGB3.110.2
TGB3.101.1
TGB3.101.2
TGB3.101.3
TGB3.102.9
0.73
TGB3.103.13
TGB3.106.5
TGB3.108.3
TGB3.112.2
0.57
TGB3.112.3
TGB3.113.2
TGB3.114.5
TGB3.111.2
1
TGB3.111.5
TGB3.103.10
TGB3.103.11
0.93
TGB3.103.12
TGB3.107.4
TGB3.112.4
TGB3.115.1
1
TGB3.115.3
0.9
TGB3.115.4
TGB3.115.6
0.81
TGB3.105.2
TGB3.105.6
TGB3.101.4
TGB3.101.5
TGB3.106.1
0.51
TGB3.106.2
TGB3.106.7
TGB3.102.10
0.96
TGB3.102.11
TGB3.102.13
TGB3.102.12
TGB3.103.9
TGB3.104.2
TGB3.104.4
1
TGB3.104.6
TGB3.104.7
TGB3.104.8
0.89
TGB3.105.3
0.65
TGB3.115.5
0.8
TGB3.108.2
TGB3.108.4
0.98
TGB3.105.1
TGB3.105.4
TGB3.106.4
TGB3.107.1
0.98
TGB3.107.2
TGB3.107.5
TGB3.107.7
1
TGB3.108.1
0.91
TGB3.108.5
TGB3.109.4
0.51
TGB3.109.5
0.5
TGB3.113.3
TGB3.113.5
TGB3.110.1
TGB3.111.1
0.66
TGB3.111.3
TGB3.111.4
TGB3.112.5
TGB3.112.6
TGB3.113.1
TGB3.113.4
TGB3.114.1
TGB3.114.2
TGB3.114.3
TGB3.114.4
0.01


E
ORF5.105.2
ORF5.105.5
ORF5.103.11
ORF5.103.10
ORF5.109.3
ORF5.110.3
ORF5.112.1
ORF5.112.4
ORF5.114.1
ORF5.101.2
ORF5.101.9
ORF5.102.4
ORF5.101.10
ORF5.101.11
ORF5.102.1
ORF5.113.1
ORF5.113.2
ORF5.113.4
ORF5.113.5
ORF5.113.9
ORF5.102.2
ORF5.102.5
ORF5.102.6
ORF5.107.5
ORF5.110.2
ORF5.110.5
ORF5.105.6
ORF5.105.8
ORF5.103.2
ORF5.103.9
ORF5.107.2
ORF5.104.2
ORF5.104.3
ORF5.104.4
ORF5.104.5
ORF5.104.6
ORF5.110.4
ORF5.115.1
ORF5.115.2
ORF5.115.3
ORF5.115.4
ORF5.106.2
ORF5.106.11
ORF5.106.7
ORF5.106.6
ORF5.106.8
ORF5.109.2
ORF5.109.4
ORF5.111.1
ORF5.111.2
ORF5.111.3
ORF5.111.4
ORF5.112.3
ORF5.112.5
ORF5.114.2
ORF5.114.3
ORF5.114.4
ORF5.114.5
ORF5.103.1
ORF5.108.1
ORF5.108.2
ORF5.108.3
ORF5.108.4
ORF5.108.5
ORF5.107.1
ORF5.107.3
ORF5.101.1
ORF5.109.1
ORF5.105.1
ORF5.107.4
ORF5.111.5
ORF5.109.5
ORF5.112.6
ORF5.110.1
ORF5.115.6
0.01

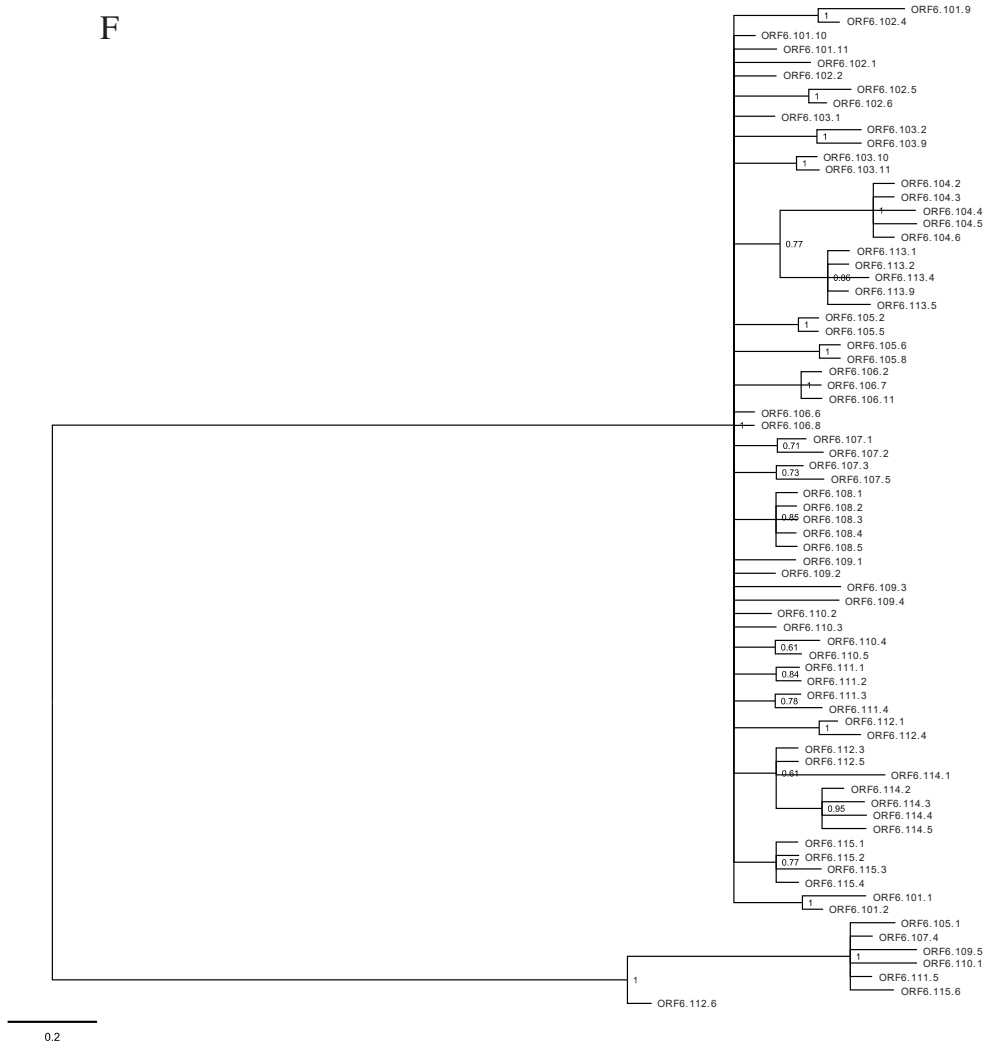
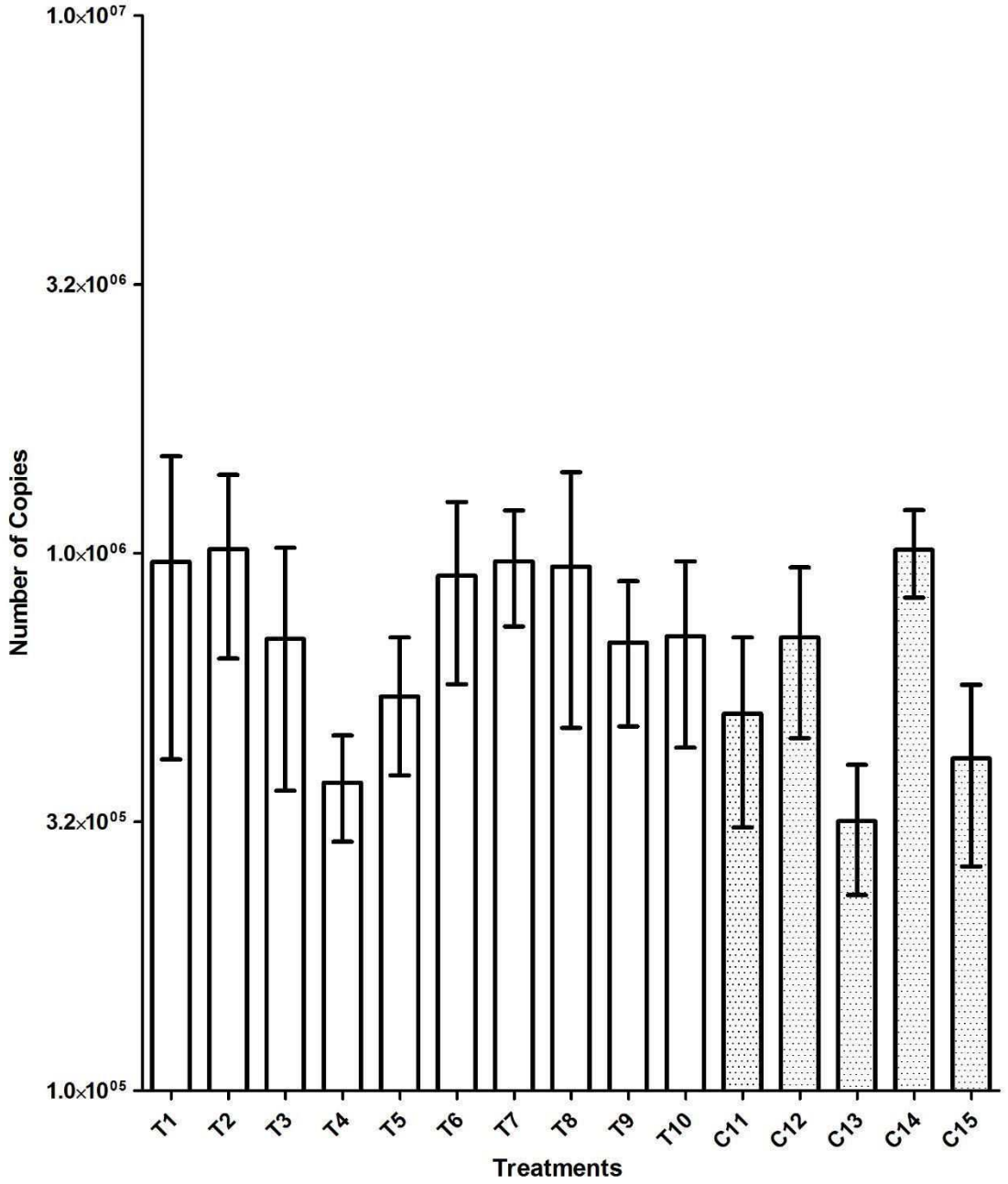

F
ORF6.101.9
ORF6.102.4
ORF6.101.10
ORF6.101.11
ORF6.102.1
ORF6.102.2
ORF6.102.5
ORF6.102.6
ORF6.103.1
ORF6.103.2
ORF6.103.9
ORF6.103.10
ORF6.103.11
ORF6.104.2
ORF6.104.3
ORF6.104.4
ORF6.104.5
ORF6.104.6
ORF6.113.1
ORF6.113.2
ORF6.113.4
ORF6.113.9
ORF6.113.5
ORF6.105.2
ORF6.105.5
ORF6.105.6
ORF6.105.8
ORF6.106.2
ORF6.106.7
ORF6.106.11
ORF6.106.6
ORF6.106.8
ORF6.107.1
ORF6.107.2
ORF6.107.3
ORF6.107.5
ORF6.108.1
ORF6.108.2
ORF6.108.3
ORF6.108.4
ORF6.108.5
ORF6.109.1
ORF6.109.2
ORF6.109.3
ORF6.109.4
ORF6.110.2
ORF6.110.3
ORF6.110.4
ORF6.110.5
ORF6.111.1
ORF6.111.2
ORF6.111.3
ORF6.111.4
ORF6.112.1
ORF6.112.4
ORF6.112.3
ORF6.112.5
ORF6.114.1
ORF6.114.2
ORF6.114.3
ORF6.114.4
ORF6.114.5
ORF6.115.1
ORF6.115.2
ORF6.115.3
ORF6.115.4
ORF6.101.1
ORF6.101.2
ORF6.105.1
ORF6.107.4
ORF6.109.5
ORF6.110.1
ORF6.111.5
ORF6.115.6
ORF6.112.6
0.2

Figure 3


1.0×10^07
3.2×10^06
1.0×10^06
3.2×10^05
1.0×10^05
Number of Copies
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
C11
C12
C13
C14
C15
Treatments

Supplementary Table S1: Variability analysis performed on all CPMMV populations from each genotype. Groups consisting of all isolates from transgenic and all from conventional cultivars, beside all 75 ones, were also analyzed. Conventional cultivars in bold. **A-F** are the values for ORF 1 - 6. **A**

Plant ID	Treatment	Number of Samples	S^a	K^b	π^c	H^d	Hd^e	θ-W^f
103	1	5	134	65	0.01165	5	1	0.01153
109	2	5	120	59	0.01057	5	1	0.01032
111	3	5	87	41	0.00735	5	1	0.00748
101	4	5	162	75.6	0.01356	5	1	0.01394
110	5	5	144	65.2	0.01168	5	1	0.01239
112	6	5	122	53.2	0.00954	5	1	0.0105
105	7	5	127	61.8	0.01108	5	1	0.01092
102	8	5	181	83.8	0.01502	5	1	0.01557
113	9	5	90	40.2	0.0072	5	1	0.00774
107	10	5	124	56.8	0.01018	5	1	0.01067
106	**11**	**5**	**136**	**68.7**	**0.01231**	**5**	**1**	**0.0117**
115	**12**	**5**	**49**	**19.6**	**0.00351**	**5**	**1**	**0.00422**
104	**13**	**5**	**31**	**12.4**	**0.00222**	**5**	**1**	**0.00267**
114	**14**	**5**	**101**	**45.4**	**0.00814**	**5**	**1**	**0.00869**
108	**15**	**5**	**96**	**44.2**	**0.00792**	**5**	**1**	**0.00826**
Conventional		25	463	91.723	0.01644	25	1	0.02197
Transgenic		50	997	96.577	0.01733	50	1	0.03995
All		75	1248	98.627	0.0177	75	1	0.04582

[a] Total number of segregating sites.

[b] Average number of nucleotide differences between sequences.

[c] Nucleotide diversity.

[d] Haplotype number.

[e] Haplotype diversity.

[f] Watterson's estimate of the population mutation rate based on the total number of segregating sites.

B

Plant ID	Treatment	Number of Samples	S[a]	K[b]	π[c]	H[d]	Hd[e]	θ-W[f]
103	1	5	166	68.1	0.09784	5	1	0.11448
109	2	5	170	73.1	0.10503	4	0.9	0.11724
111	3	5	167	99.6	0.1431	5	1	0.11517
101	4	5	168	99.5	0.14296	4	0.9	0.11586
110	5	5	171	101.9	0.14641	5	1	0.11793
112	6	5	169	99.4	0.14282	5	1	0.11655
105	7	5	168	99	0.14224	5	1	0.11586
102	8	5	170	79.1	0.11365	5	1	0.11724
113	9	5	170	69.3	0.09957	5	1	0.11724
107	10	5	169	100.7	0.14468	5	1	0.11655
106	**11**	**5**	**5**	**2.4**	**0.00345**	**4**	**0.9**	**0.00345**
115	**12**	**5**	**3**	**1.2**	**0.00172**	**3**	**0.7**	**0.00207**
104	**13**	**5**	**2**	**0.8**	**0.00115**	**3**	**0.7**	**0.00138**
114	**14**	**5**	**15**	**8**	**0.01149**	**5**	**1**	**0.01034**
108	**15**	**5**	**165**	**98.6**	**0.14167**	**4**	**0.9**	**0.11379**
Conventional		25	180	30.747	0.04418	18	0.967	0.06849
Transgenic		50	221	80.585	0.11578	48	0.998	0.07089
All		75	235	68.752	0.09878	66	0.996	0.06908

[a] Total number of segregating sites.

[b]Average number of nucleotide differences between sequences.

[c]Nucleotide diversity.

[d]Haplotype number.

[e]Haplotype diversity.

[f]Watterson's estimate of the population mutation rate based on the total number of segregating sites.

C

Plant ID	Treatment	Number of Samples	S[a]	K[b]	π[c]	H[d]	Hd[e]	θ-W[f]
103	1	5	73	29.4	0.09159	3	0.7	0.10916
109	2	5	81	48.4	0.15078	3	0.7	0.12112
111	3	5	78	46.4	0.14455	4	0.9	0.11664
101	4	5	78	46.6	0.14517	4	0.9	0.11664
110	5	5	80	32	0.09969	3	0.7	0.11963
112	6	5	77	45.6	0.14206	4	0.9	0.11514
105	7	5	78	46.2	0.14393	3	0.8	0.11664
102	8	5	77	31.4	0.09782	3	0.7	0.11514
113	9	5	80	32.6	0.10156	4	0.9	0.11963
107	10	5	79	31.6	0.09844	4	0.9	0.11813
106	**11**	**5**	**80**	**32.6**	**0.10156**	**5**	**1**	**0.11963**
115	**12**	**5**	**81**	**33.2**	**0.10343**	**4**	**0.9**	**0.12112**
104	**13**	**5**	**2**	**0.8**	**0.00249**	**2**	**0.4**	**0.00299**
114	**14**	**5**	**80**	**32.8**	**0.10218**	**4**	**0.9**	**0.11963**
108	**15**	**5**	**79**	**32.8**	**0.10218**	**3**	**0.8**	**0.11813**
Conventional		25	97	34.74	0.10822	18	0.967	0.08003
Transgenic		50	101	40.624	0.12656	32	0.974	0.07024
All		75	117	39.589	0.12333	49	0.984	0.07457

[a] Total number of segregating sites.

[b]Average number of nucleotide differences between sequences.

[c]Nucleotide diversity.

[d]Haplotype number.

[e]Haplotype diversity.

[f]Watterson's estimate of the population mutation rate based on the total number of segregating sites.

D

Plant ID	Treatment	Number of Samples	S[a]	K[b]	π[c]	H[d]	Hd[e]	θ-W[f]
103	1	5	47	19.7	0.09517	3	0.7	0.10899
109	2	5	45	26.7	0.12899	3	0.7	0.10435
111	3	5	47	28	0.13527	3	0.8	0.10899
101	4	5	50	28.8	0.13913	4	0.9	0.11594
110	5	5	47	19.3	0.09324	5	1	0.10899
112	6	5	46	27.4	0.13237	4	0.9	0.10667
105	7	5	48	28.6	0.13816	3	0.8	0.1113
102	8	5	50	20.7	0.1	5	1	0.11594
113	9	5	45	18.8	0.09082	4	0.9	0.10435
107	10	5	46	18.4	0.08889	2	0.4	0.10667
106	**11**	**5**	**48**	**19.6**	**0.09469**	**4**	**0.9**	**0.1113**
115	**12**	**5**	**45**	**18.6**	**0.08986**	**4**	**0.9**	**0.10435**
104	**13**	**5**	**3**	**1.2**	**0.0058**	**4**	**0.9**	**0.00696**
114	**14**	**5**	**46**	**19.1**	**0.09227**	**5**	**1**	**0.10667**
108	**15**	**5**	**47**	**19.6**	**0.09469**	**5**	**1**	**0.10899**
Conventional		25	64	21.743	0.10504	22	0.99	0.08188
Transgenic		50	76	24.708	0.11936	32	0.975	0.08197
All		75	85	24.339	0.11758	50	0.985	0.08401

[a] Total number of segregating sites.

[b]Average number of nucleotide differences between sequences.

[c]Nucleotide diversity.

[d]Haplotype number.

[e]Haplotype diversity.

[f]Watterson's estimate of the population mutation rate based on the total number of segregating sites.

E

Plant ID	Treatment	Number of Samples	S[a]	K[b]	π[c]	H[d]	Hd[e]	θ-W[f]
103	1	5	26	13.8	0.01592	5	1	0.01439
109	2	5	140	58.7	0.0677	5	1	0.07751
111	3	5	141	58.2	0.06713	5	1	0.07806
101	4	5	22	10.6	0.01223	5	1	0.01218
110	5	5	140	58.2	0.06713	5	1	0.07751
112	6	5	141	58.6	0.06759	4	0.9	0.07806
105	7	5	139	57	0.06574	4	0.9	0.07696
102	8	5	20	9	0.01038	4	0.9	0.01107
113	9	5	5	2	0.00231	4	0.9	0.00277
107	10	5	143	59.6	0.06874	5	1	0.07917
106	**11**	**5**	**16**	**8.4**	**0.00969**	**5**	**1**	**0.00886**
115	**12**	**5**	**141**	**56.6**	**0.06528**	**5**	**1**	**0.07806**
104	**13**	**5**	**3**	**1.2**	**0.00138**	**4**	**0.9**	**0.00166**
114	**14**	**5**	**15**	**6**	**0.00692**	**4**	**0.9**	**0.0083**
108	**15**	**5**	**5**	**2**	**0.00231**	**5**	**1**	**0.00277**
Conventional		25	167	19.64	0.02265	23	0.993	0.05101
Transgenic		50	205	36.714	0.04235	46	0.997	0.05279
All		75	232	31.437	0.03626	69	0.998	0.05474

[a] Total number of segregating sites.

[b]Average number of nucleotide differences between sequences.

[c]Nucleotide diversity.

[d]Haplotype number.

[e]Haplotype diversity.

[f]Watterson's estimate of the population mutation rate based on the total number of segregating sites.

F

Plant ID	Treatment	Number of Samples	S[a]	K[b]	π[c]	H[d]	Hd[e]	θ-W[f]
103	1	5	8	4.2	0.01228	4	0.9	0.01123
109	2	5	77	31.5	0.09238	5	1	0.10839
111	3	5	69	28	0.08211	4	0.9	0.09713
101	4	5	10	4.8	0.01404	5	1	0.01404
110	5	5	69	28.7	0.08416	5	1	0.09713
112	6	5	56	23.4	0.06842	4	0.9	0.0786
105	7	5	73	30.4	0.08889	3	0.8	0.10246
102	8	5	10	4.6	0.01345	5	1	0.01404
113	9	5	2	0.8	0.00234	3	0.7	0.00281
107	10	5	70	28.7	0.08416	5	1	0.09853
106	**11**	**5**	**3**	**1.8**	**0.00526**	**2**	**0.6**	**0.00421**
115	**12**	**5**	**69**	**27.6**	**0.08094**	**3**	**0.7**	**0.09713**
104	**13**	**5**	**2**	**0.8**	**0.00234**	**3**	**0.7**	**0.00281**
114	**14**	**5**	**8**	**3.2**	**0.00936**	**5**	**1**	**0.01123**
108	**15**	**5**	**0**	**0**	**0**	**1**	**0**	**0**
Conventional		25	82	9.35	0.02742	14	0.933	0.06368
Transgenic		50	112	17.882	0.05244	42	0.993	0.07333
All		75	121	15.193	0.04455	55	0.989	0.07259

[a] Total number of segregating sites.

[b]Average number of nucleotide differences between sequences.

[c]Nucleotide diversity.

[d]Haplotype number.

[e]Haplotype diversity.

[f]Watterson's estimate of the population mutation rate based on the total number of segregating sites.

Supplementary Table S2: Recombination events support at least three methods

Overview												
Event number	Found in	Recombinant	Major parent	Minor parent	Detection methods						
					R	G	B	M	C	S	T
1	1	ORF1-102.3	Unknown	102.1	+	+	+	+	+	+	+
2	1	ORF1-105.2	105.3	Unknown	+	-	-	+	+	-	+
3	1	ORF1-106.5	106.2	Unknown	-	-	-	+	-	+	+
4	1	ORF1-102.5	101.2	102.1	+	+	+	+	+	+	-
5	1	ORF1-102.1	102.4	Unknown	-	-	-	+	+	-	+
6	1	ORF1-101.2	101.4	Unknown	+	+	+	+	+	+	+
7	1	ORF1-108.1	108.2	Unknown	+	+	-	+	-	+	+
8	1	ORF1-112.4	112.5	Unknown	-	-	-	+	+	-	+
9	1	ORF1-108.4	108.5	Unknown	-	-	-	+	+	-	+
10	3	ORF1-109.4	109.3	Unknown	-	-	-	+	+	+	-
11	1	ORF1-107.5	107.3	107.2	-	-	-	+	-	+	+
12	1	ORF1-102.4	102.1	102.5	-	-	-	+	-	+	+
13	1	ORF1-101.4	Unknown	101.3	-	-	-	+	-	+	+
14	1	ORF1-103.4	103.1	103.5	-	+	+	+	-	+	+
15	1	ORF1-105.5	105.2	Unknown	-	-	-	+	+	-	+
16	1	ORF1-101.3	101.1	101.5	-	-	+	+	+	+	-
17	2	ORF1-110.5	110.2	Unknown	-	-	-	+	-	+	+
18	1	ORF1-113.3	113.1	113.2	-	-	-	+	-	+	+
19	1	ORF1-103.2	103.5	Unknown	+	-	+	-	-	+	-
20	1	ORF2-102.1	102-3	103-1	+	+	-	+	+	-	+
21	1	ORF6-112.6	112.4	110.1	-	-	-	+	+	+	+

Detection methods and abreviations are cited on the text

Supplementary Table S3: Selection analysis for isolates within each plant genotype at ORF's 1-6 (**A-F)**, using Tajima's D, Fu and Li's D and F, and methods for site-specific selection in the Datamonkey web-server

A

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	0.08049	0.08049	0.0876	0	1	1	37	1	8	0	0	-
109	2	0.18493	0.18493	0.2012	0	8	1	35	0	10	35	16	-
111	3	-0.13816	-0.13816	-0.15013	0	5	1	21	1	1	0	43	-
101	4	-0.21168	-0.21168	-0.23049	0	1	0	39	3	7	0	100	-
110	5	-0.43194	-0.43194	-0.47019	0	2	0	47	2	7	0	89	-
112	6	-0.6965	-0.6965	-0.75779	0	3	1	27	1	1	6	9	+
105	7	0.10488	0.10488	0.11412	0	6	0	39	1	8	42	2	-
102	8	-0.27028	-0.27028	-0.29438	0	2	0	44	1	9	0	111	-
113	9	-0.52734	-0.52734	-0.57309	0	0	0	25	1	3	1	0	-
107	10	-0.4631	-0.43793	-0.48199	0	1	0	46	0	9	1	0	-
106	11	0.34042	0.352	0.3806	0	3	0	44	0	14	2	4	+
115	12	-1.25736	-1.25736*	-1.36155*	0	0	0	6	0	0	0	0	-
104	13	-1.24706	-1.24706*	-1.34447	0	0	0	6	0	0	17	0	-
114	14	-0.48285	-0.48285	-0.525	0	3	0	30	3	0	0	68	-
108	15	0.30997	-0.30997	-0.33695	0	0	0	21	0	4	0	60	-
Conventional		-1.05776	-0.97675	-1.1805	0	31	2	93	3	48	0	83	-
Transgenic		-2.14205*	-2.48128*	-2.80945*	3	118	17	218	24	76	19	95	+
All isolates		-2.22447**	-2.83098*	-3.08581*	3	159	15	289	19	148	23	113	+

(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, $P < 0.05$; **, $P < 0.02$; - analysis could not be performed on some populations.

B

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	-1.18522	-1.15686	-1.26597	0	1	0	58	0	2	0	102	-
109	2	-0.91244	-0.8849	-0.9697	0	2	0	54	1	4	3	0	-
111	3	1.62697	1.68269*	1.82035	0	0	0	57	0	53	1	0	-
101	4	1.564	1.61016	1.74348	0	1	0	58	0	55	0	0	-
110	5	1.57191	1.65313	1.78272	0	0	0	61	0	57	0	0	-
112	6	1.66265	1.68134*	1.82692	0	0	0	62	0	59	0	0	-
105	7	1.73529	1.73529*	1.88969*	0	0	0	59	0	53	0	0	-
102	8	-0.40341	-0.35777	-0.39958	0	2	0	57	2	25	0	0	-
113	9	-1.1868	-1.17751	-1.28436	0	1	0	61	1	1	0	0	-
107	10	1.78408	1.79342**	1.95102**	0	1	0	60	0	60	0	0	-
106	11	0	0	0	0	0	0	1	0	0	0	0	-
115	12	-1.04849	-1.04849	-1.05189	0	0	0	1	0	0	0	0	-
104	13	-0.97256	-0.97256	-0.9544	0	0	0	0	0	0	4	0	-
114	14	0.81235	0.81235	0.86552	0	0	0	9	0	5	0	0	-
108	15	1.86678*	1.86678**	2.03279**	0	0	0	56	0	56	0	0	-
Conventional		-1.52331	1.41126*	0.55225	0	2	0	53	0	51	1	92	-
Transgenic		1.69188	0.39601	1.0617	0	25	0	68	0	49	0	5	-
All isolates		0.91268	0.13958	0.55105	0	29	1	72	3	47	0	19	-

(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, $P < 0.05$; **, $P < 0.02$; - analysis could not be performed on some populations.

C

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	-1.22004	-1.22004	-1.32455	0	0	0	29	0	0	0	0	-
109	2	1.85775*	1.85775**	2.01794**	0	0	0	33	0	0	0	0	-
111	3	1.81511	1.81511**	1.97126**	0	0	0	28	0	0	0	0	-
101	4	1.50649	1.58455	1.70423	0	0	0	29	0	25	0	0	-
110	5	-1.26438	-1.26438**	-1.37332**	0	0	0	32	0	0	0	0	-
112	6	1.65304	1.69356*	1.83046	0	0	0	29	0	0	0	0	-
105	7	1.77459	1.77459**	1.92726*	0	1	0	29	0	0	0	0	-
102	8	-1.14083	-1.14083	-1.2389	0	0	0	29	0	0	0	0	-
113	9	-1.14584	-1.14584	-1.24458	0	0	0	29	0	0	0	0	-
107	10	-1.26424	-1.26424**	-1.37309**	0	0	0	29	0	0	0	0	-
106	11	-1.14584	-1.14584	-1.24458	0	1	0	30	0	0	0	0	-
115	12	-1.1084	-1.1084	-1.20398	0	0	0	35	0	2	0	0	-
104	13	-0.97256	-0.97256	-0.9544	-	-	-	-	-	-	-	-	-
114	14	-1.10633	-1.10633	-1.20166	0	0	0	27	0	0	0	0	-
108	15	-1.10633	-1.06682	-1.16733	0	0	0	29	0	0	0	0	-
Conventional		1.12727	0.73192	1.0095	0	3	0	28	1	27	0	0	-
Transgenic		2.38555*	0.32817	1.31142	0	8	0	25	1	22	0	16	-
All isolates		1.6312	-0.67622	0.3147	0	9	0	27	2	20	2	10	-

(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, $P < 0.05$; **, $P < 0.02$; - analysis could not be performed on some populations.

D

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	-1.09332	-1.06058	-1.15537	0	1	0	15	0	0	0	0	-
109	2	1.57706	1.6112	1.73645	0	0	0	9	0	0	2	9	-
111	3	1.81805	1.81805**	1.96806**	0	0	0	13	0	0	0	0	-
101	4	1.50926	1.50926	1.63458	0	0	0	13	0	13	0	0	-
110	5	-1.22425	-1.19152	-1.29714	0	0	0	12	0	0	0	0	-
112	6	1.81601	1.81601**	1.96551**	0	0	0	12	0	0	0	0	-
105	7	1.82001	1.82001**	1.97051**	0	0	0	14	0	0	0	0	-
102	8	-1.16556	-1.13472	-1.23581	0	0	0	11	0	0	0	0	-
113	9	-1.2566	-1.12292	-1.24451	0	0	0	10	0	0	0	0	-
107	10	-1.25619	-1.25619*	-1.3596*	-	-	-	-	-	-	-	-	-
106	11	-1.12605	-1.12605	-1.21917	0	0	0	12	1	0	0	0	-
115	12	-1.04647	-1.04647	-1.13241	0	0	0	12	0	0	0	0	-
104	13	-1.04849	-1.04849	-1.05189	0	0	0	3	0	0	0	0	-
114	14	-1.15634	-1.12292	-1.2228	0	0	0	14	0	0	2	3	-
108	15	-0.98923	-0.98923	-1.07086	0	0	0	13	0	0	0	0	-
Conventional		0.9496	0.51779	0.76973	0	4	0	16	0	15	0	0	-
Transgenic		1.29298	-0.51253	0.18739	0	2	1	10	2	7	3	13	-
All isolates		0.8988	-1.11029	-0.36265	0	6	1	13	1	15	0	0	-

(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, $P < 0.05$; **, $P < 0.02$; - analysis could not be performed on some populations.

E

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	0.78804	0.78804	0.8477	0	0	0	12	0	2	1	19	-
109	2	-1.19282	-1.1381	-1.25082	0	3	0	81	1	0	0	0	-
111	3	-1.15838	-1.114	-1.22228	0	2	0	81	0	2	0	129	-
101	4	0.02809	0.02809	0.03014	0	0	0	10	0	2	0	0	-
110	5	-1.11286	-1.09052	-1.1919	0	2	0	79	0	2	0	0	-
112	6	-1.0217	-1.0217	-1.1121	0	0	0	89	0	2	0	131	-
105	7	-1.10937	-1.10937	-1.20748	0	0	0	84	1	0	0	0	-
102	8	-0.46204	-0.46204	-0.49502	0	0	0	8	0	0	0	0	-
113	9	-1.12397	-1.12397	-1.15583	0	0	0	1	0	0	0	0	-
107	10	-1.18315	-1.11838	-1.23158	0	0	0	76	0	1	0	129	-
106	11	0.6873	0.6873	0.73329	0	0	0	7	0	4	4	0	-
115	12	-1.24674	-1.24674*	-1.35706*	0	0	0	63	0	0	0	125	-
104	13	-1.04849	-1.04849	-1.05189	0	0	0	0	0	0	0	0	-
114	14	-1.21852	-1.21852	-1.29828	0	0	0	7	0	0	0	0	-
108	15	-1.12397	-1.12397	-1.15583	0	0	0	2	0	0	0	0	-
Conventional		-2.33133**	-3.79517**	-3.91435**	0	7	0	80	0	3	1	263	+
Transgenic		-0.94931	0.12149	-0.33972	0	20	0	91	1	61	0	0	-
All isolates		-1.4167	-0.81823	-1.27889	0	26	1	96	1	68	1	233	-

(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, $P < 0.05$; **, $P < 0.02$; - analysis could not be performed on some populations.

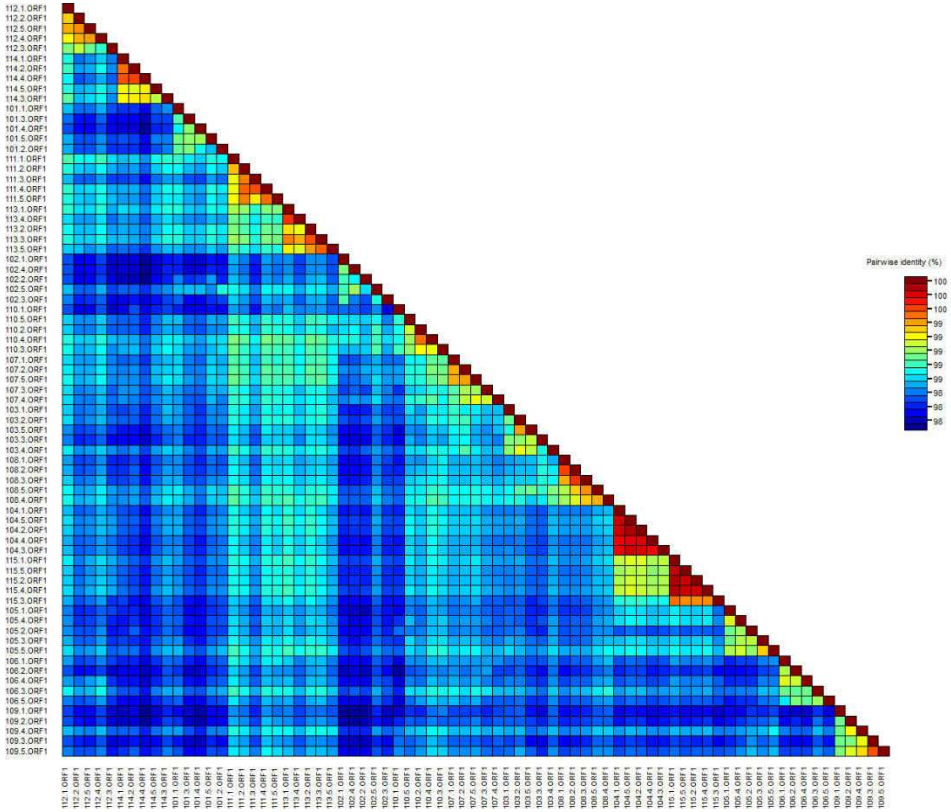
F

					SLAC[a]		FEL[b]		IFEL[c]		REL[d]		PARRIS[e]
Plant ID	Treatment	Tajima's D	Fu and Li's D	Fu and Li's F	PS	NS	PS	NS	PS	NS	PS	NS	PS
103	1	0.66055	0.66055	0.69176	0	0	0	0	0	0	66	17	-
109	2	-1.20332	-1.18306	-1.28924	0	0	1	30	1	0	-	-	-
111	3	-1.17111	-1.17111	-1.27103	0	0	0	28	0	0	0	39	-
101	4	0	0	0	0	0	0	2	0	1	0	0	-
110	5	-1.10494	-1.08239	-1.17973	0	1	1	29	0	3	0	40	-
112	6	-1.09411	-1.03885	-1.13815	0	0	1	24	0	1	-	-	-
105	7	-1.09273	-1.05005	-1.14935	0	0	1	28	0	0	0	41	-
102	8	-0.29817	-0.29817	-0.31445	0	0	0	5	0	2	0	0	-
113	9	-0.97256	-0.97256	-0.9544	0	0	0	1	0	0	1	1	-
107	10	-1.19626	-1.17402	-1.27922	0	0	1	27	1	0	0	40	-
106	11	1.57274	1.57274	1.57783	-	-	-	-	-	-	-	-	-
115	12	-1.2626	-1.2626**	-1.37033**	0	0	0	30	0	0	0	41	-
104	13	-0.97256	-0.97256	-0.9544	0	0	0	0	0	0	3	0	-
114	14	-1.17432	-1.17432	-1.22979	0	0	0	2	0	0	0	2	-
108	15	-	-	-	-	-	-	-	-	-	-	-	-
Conventional		-2.31315**	-3.8898**	-3.98351**	0	2	1	27	1	1	0	44	-
Transgenic		-1.18799	-0.13017	-0.62935	0	4	0	25	1	9	0	0	-
All isolates		-1.52478	-0.52517	-1.11828	0	6	0	24	1	12	0	0	-

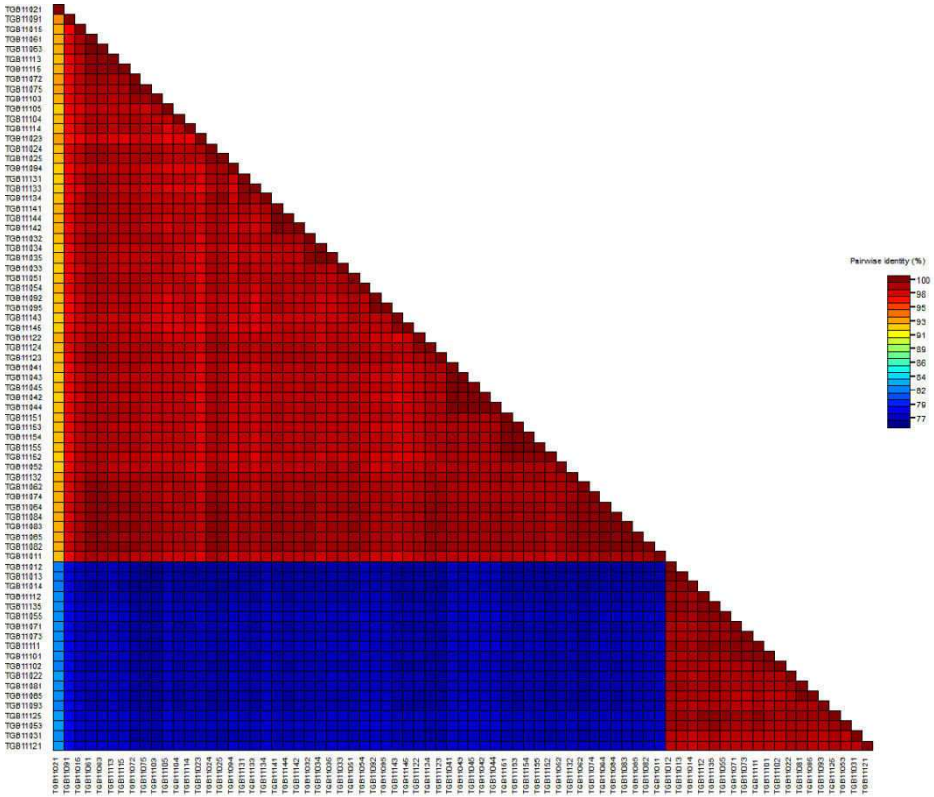
(PS) Sites under positive selection; (NS) sites under negative selection; (-) no site under selection; (+) evidence of sites under positive selection. a,b,c,d,eCodon-based maximum-likelihood algorithms. [a]Single Likelihood Ancestor Counting (SLAC); [b]Fixed Effects Likelihood (FEL); [c]Internal Fixed Effects Likelihoods (IFEL); [d]Random Effects Likelihood (REL) and [e]Partitioning Robust Inference of Selection (PARRIS). *, P < 0.05; **, P < 0.02; - analysis could not be performed on some populations.

Supplementary Figure S1

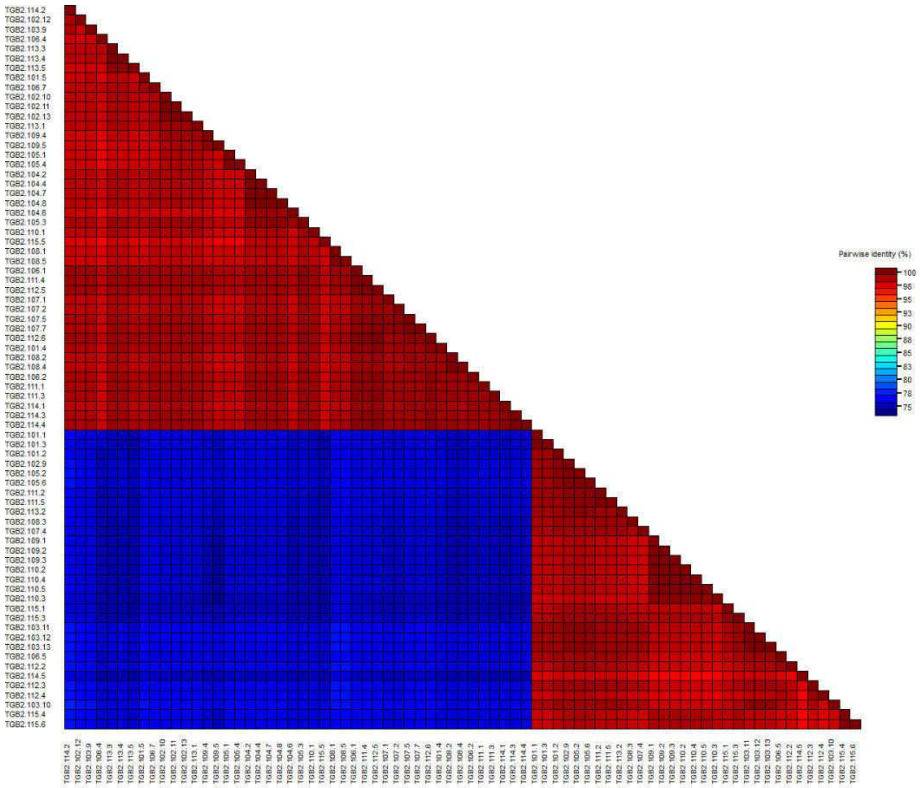
A



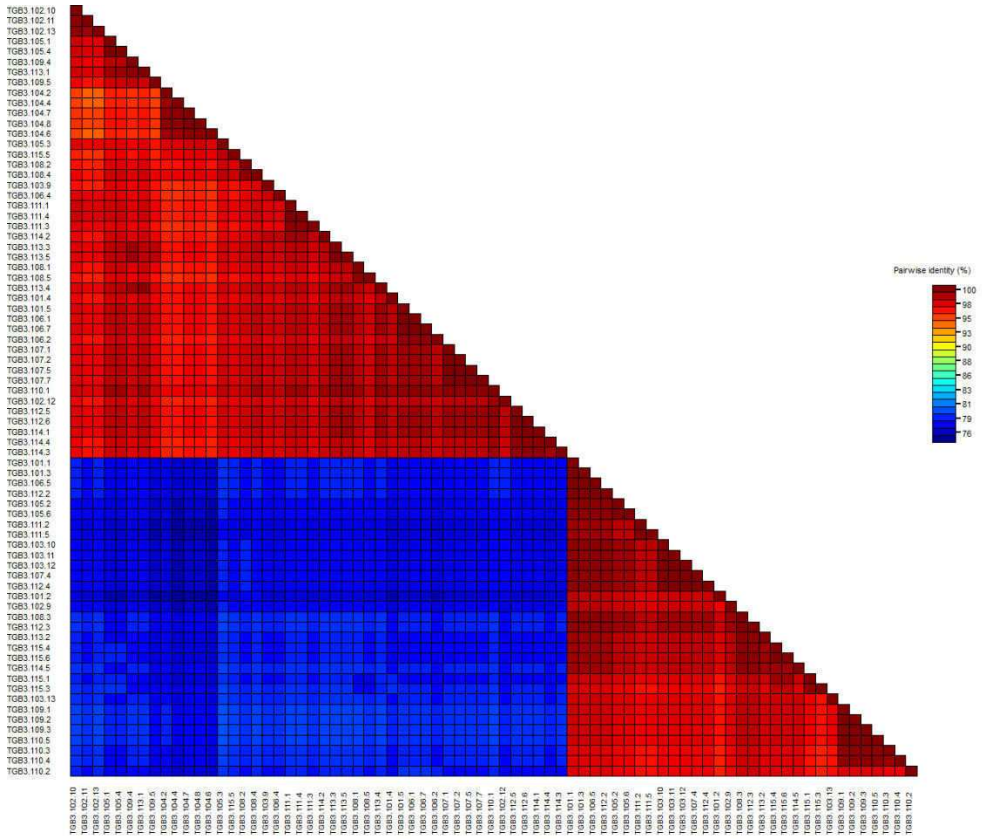
B



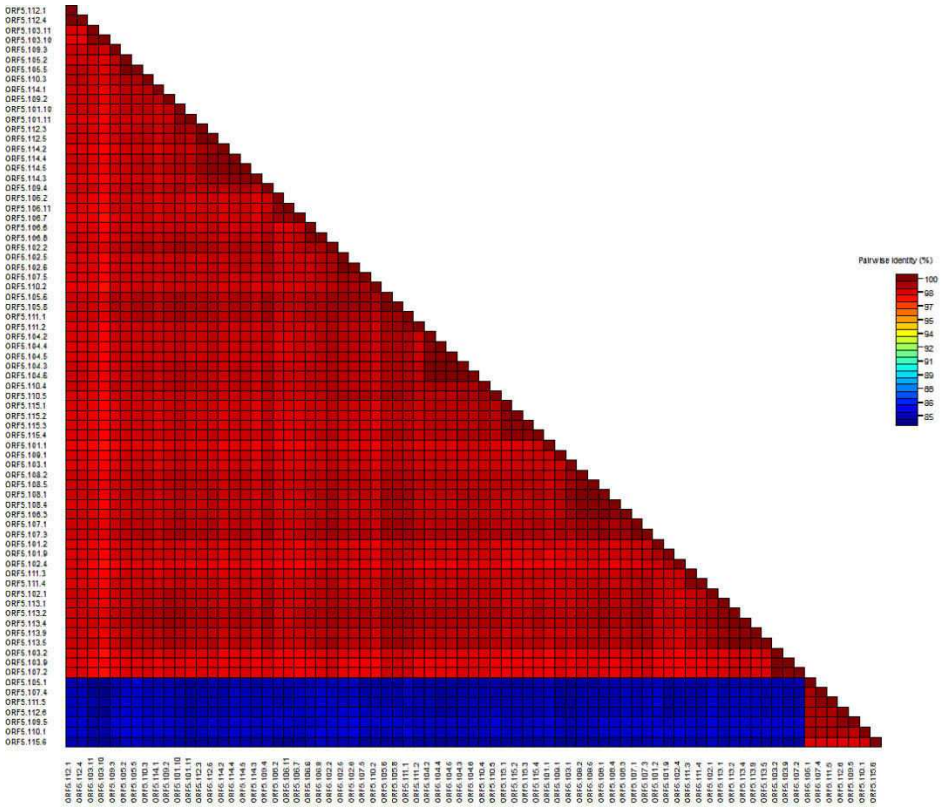
C


Pairwise identity (%)
100
98
95
93
90
88
85
83
80
78
75

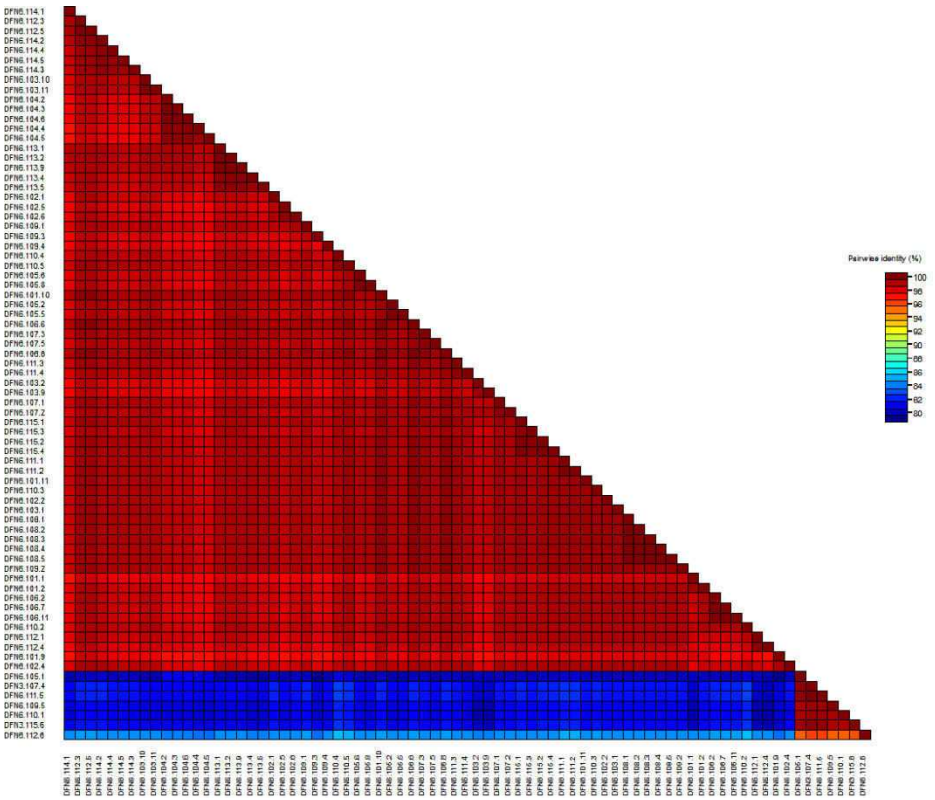
D


Pairwise identity (%)
100
98
95
93
90
88
86
83
81
79
76

E

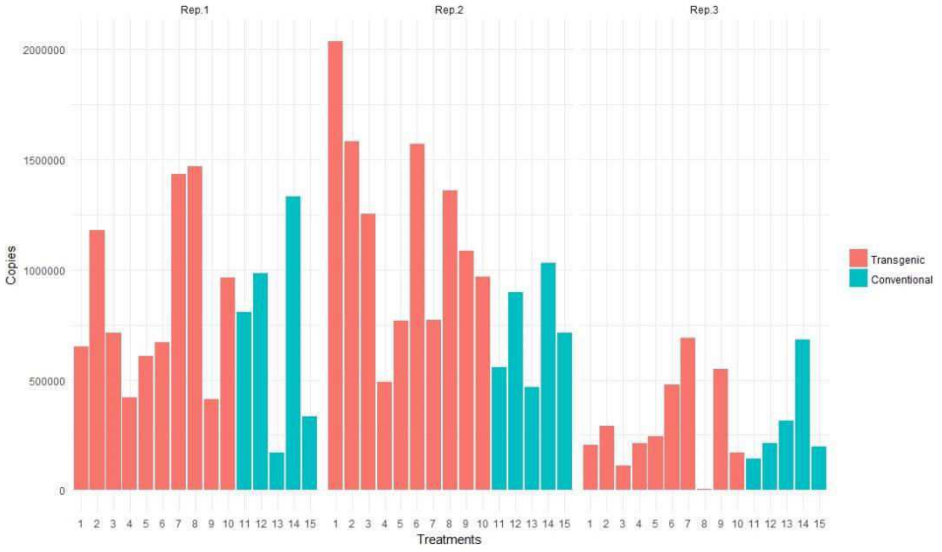

Pairwise Identity (%)
100
98
97
95
94
92
91
89
88
86
85

F

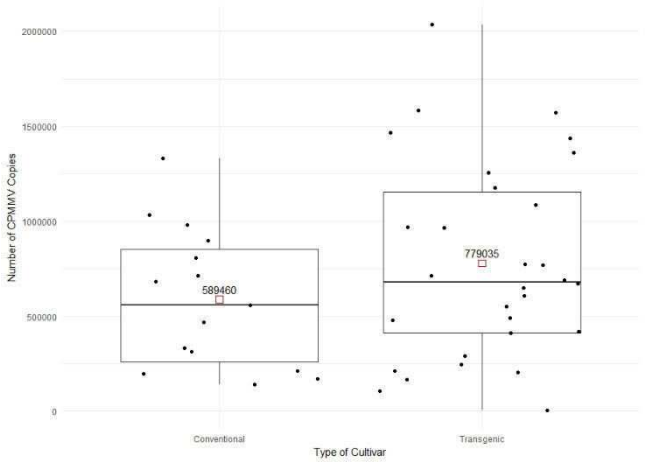

Pairwise Identity (%)
100
96
96
94
92
90
88
86
84
82
80

Supplementary figure S2:

A



B



C