VITOR BATISTA PINTO

MOLECULAR CHARACTERIZATION OF TWO BEGOMOVIRUSES INFECTING Pavonia sp. (MALVACEAE), AND ANALYSIS OF THE INTRA-HOST EVOLUTION OF Tomato severe rugose virus (ToSRV)

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

VIÇOSA
MINAS GERAIS - BRASIL
2015

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

T
P659m
2015

Pinto, Vitor Batista, 1988-
Molecular characterization of two begomoviruses infecting *Pavonia* sp. (Malvaceae), and analysis of the intra-host evolution of *Tomato severe rugose virus* (ToSRV) / Vitor Batista Pinto. – Viçosa, MG, 2015.
x, 65f. : il. (algumas color.) ; 29 cm.

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

1. *Begomovirus*. 2. *Geminiviridae*. 3. Diversidade genética. 4. *Quasispecies*. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 571.9928

VITOR BATISTA PINTO

MOLECULAR CHARACTERIZATION OF TWO BEGOMOVIRUSES INFECTING *Pavonia* sp. (MALVACEAE), AND ANALYSIS OF THE INTRA-HOST EVOLUTION OF *Tomato severe rugose virus* (ToSRV)

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

APROVADA: 29 de julho de 2015.

Fabyano Fonseca e Silva

Poliane Alfenas Zerbini

Francisco Murilo Zerbini Júnior
(Orientador)

Aos meus pais Milton e Eliana,
dedico.

AGRADECIMENTOS

A Deus, por sempre iluminar meus caminhos e permitir mais essa conquista em minha vida;

Aos meus pais, Milton e Eliana, pelo constante apoio, amor, carinho e por sempre acreditarem no meu potencial;

À minha irmã Renata pela amizade e incentivo;

Ao Professor Murilo Zerbini, pela orientação, amizade, paciência e dedicação para realização deste trabalho;

À Universidade Federal de Viçosa (UFV) por proporcionar a realização do mestrado em Genética e Melhoramento;

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro;

Aos funcionários da Pós-Graduação em Genética e Melhoramento pela paciência e disponibilidade todas as vezes nas quais precisei;

Aos amigos da Pós-Graduação em Genética e Melhoramento, especialmente à Isadora, sem você meu mestrado não teria sido o mesmo;

Aos amigos do Laboratório de Virologia Vegetal Molecular, Josiane, André, Hermano, Márcio, Igor, César, Larissa, Letícia, Flávia, Fernanda, Murilo, Talita, Priscila, Patrícia e especialmente à Camila, Angélica e João pela sincera amizade, apoio e momentos de descontração;

Ao Cleysinho e Márcio pela valiosa colaboração na execução deste trabalho;

Ao servidor Pedro M. P. Vidigal e ao Núcleo de Análises de Biomoléculas (NuBioMol) da Universidade Federal de Viçosa pelo suporte na análise dos dados gerados;

Aos familiares que me apoiaram e sempre deram palavras de incentivo que confortavam nos momentos mais difíceis;

Aos amigos Mateus, Mayara, Luiza, Daiana, Luiz Felipe, Sil, Paulo, Daniel, Úrsula, Christiane, Adriana, Marília, Carla, Carol M., Rizia, Gustavo e Rosane que, mesmo distantes, sempre se fazem presentes e a todos os amigos que conquistei nesta etapa, em especial ao Luan, Carol H., Thaís, Rodrigo, Vinícius e Marcone, pelas risadas, histórias, festas e por me aguentarem todo esse tempo em Viçosa;

Aos companheiros de república, Michel, Pedro e Vinícius, pela convivência, compreensão e paciência;

A todos que contribuíram direta ou indiretamente para conclusão deste trabalho.

BIOGRAFIA

VITOR BATISTA PINTO, filho de Milton Antônio Caetano Pinto e Eliana Pereira Batista Pinto, nasceu em Sabará, Minas Gerais, no dia 19 de novembro de 1988.

Realizou o ensino médio e o curso técnico em Química Industrial no Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), sendo concluído em 2006. Graduou-se em Agronomia pela Universidade Federal de Minas Gerais (UFMG), na cidade de Montes Claros, Minas Gerais, em julho de 2013. Ingressou no mês seguinte no curso de pós-graduação em Genética e Melhoramento em nível de Mestrado na Universidade Federal de Viçosa e, em julho de 2015, submeteu-se à defesa de dissertação.

SUMÁRIO

ABSTRACT

PINTO, Vitor Batista, M.Sc., Universidade Federal de Viçosa, July, 2015. **Molecular characterization of two begomoviruses infecting Pavonia sp. (Malvaceae), and analysis of the intra-host evolution of Tomato severe rugose virus (ToSRV)**. Advisor: Francisco Murilo Zerbini Júnior.

The family Geminiviridae is comprised of viruses with a circular, single-stranded DNA genome encapsidated in twinned icosahedral particles. The viruses in the genus Begomovirus are transmitted by the whitefly Bemisia tabaci to dicot plants. Begomoviruses have mutation and nucleotide substitution rates similar to those reported for RNA viruses, and a high frequency of recombination. Due to their rapid evolutionary process, new begomovirus species are often found in the field. This study aimed to perform the molecular characterization of two begomovirus species infecting Pavonia sp. (Malvaceae), and to follow and quantify the evolution of Tomato severe rugose virus (ToSRV) in a cultivated and a non-cultivated host. Two begomoviruses were isolates from Pavonia sp. plants collected in the municipalities of Albuquerque and Corumbá, Mato Groso do Sul, Brazil. Sequence comparisons and phylogenetic analysis showed that these were two novels species, with the typical features of bipartite, New World begomoviruses. The names Pavonia mosaic virus (PavMV) and Pavonia yellow mosaic virus (PavYMV) were proposed for the two new species. In the study to evaluate the evolutionary dynamics of ToSRV, tomato and Nicandra physaloides plants were inoculated via biobalistics with an infectious clone of ToSRV and maintained in a greenhouse. Total DNA was extrated from leaves collected at 30, 75 and 120 days after inoculation, and was sequenced in the Illumina HiSeq 2000 platform. The DNA libraries from each of the two hosts were submitted to quality control analysis with FastQC software. The genome

assembly was performed with the program Geneious using the infectious clone as reference. Both genomic components of ToSRV showed substitution rates similar to those of RNA viruses: 3.06 x 10^{-3} and 2.03 x 10^{-3} sub/site/year for the DNA-A and DNA-B, respectively, in N. physaloides, and 1.38 x 10^{-3} and 8.68 x 10^{-4} sub/site/year for the DNA-A and DNA-B, respectively, in tomato. Substitution rates in the range of those already described for other begomoviruses were found also for the CP, Rep, MP and NSP genes in both hosts. We quantified synonymous and non-synonymous substitutions, transversions and transitions, as well as deletions and insertions in the CP, Rep, MP and NSP genes. A decrease in the number of variable sites was observed during the course of the experiment, with a corresponding increase in the number of identical sites to the reference genome. Suppression of the stop codons of the MP and NSP genes was observed in the N. physaloides libraries, suggesting an adaptive strategy. Determination of Shannon entropy indicated mutation hotspots in the N-terminal region of the Rep gene, the intergenic common region in the DNA-A and DNA-B (CR-A and CR-B, respectively) and the long intergenic region between the MP and NSP genes in the DNA-B (LIR-B). Overall, the results indicate that ToSRV evolves as a quasispecies, with a high degree of genetic variability which could be partly responsible for its prevalence in the field.

RESUMO

PINTO, Vitor Batista, M.Sc., Universidade Federal de Viçosa, julho de 2015. **Caracterização molecular de dois begomovírus infectando Pavonia sp. (Malvaceae), e análise da evolução intra-hospedeiro do Tomato severe rugose virus (ToSRV)**. Orientador: Francisco Murilo Zerbini Júnior.

A família Geminiviridae é composta por vírus com genoma de DNA circular de fita simples, encapsidado em partículas icosaédricas geminadas. Os vírus pertencentes ao gênero Begomovirus são transmitidos pela mosca-branca Bemisia tabaci a plantas dicotiledôneas. Os begomovírus apresentam taxas de mutação e substituição nucleotídica semelhantes às relatadas para vírus de RNA, e uma elevada taxa de recombinação. Devido ao seu rápido processo evolutivo, novas espécies de begomovírus são frequentemente encontradas no campo. Este trabalho teve como objetivos realizar a caracterização molecular de duas espécies de begomovírus infectando Pavonia sp. (Malvaceae), e quantificar a evolução do Tomato severe rugose virus (ToSRV) em um hospedeiro cultivado e um não-cultivado. Dois begomovírus foram isolados de plantas de Pavonia sp. coletadas nos municípios de Albuquerque e Corumbá, Mato Grosso do Sul. Comparação de sequências e análise filogenética indicaram tratar-se de duas novas espécies com as características de begomovírus bissegmentados das Américas. Foram propostos os nomes Pavonia mosaic virus (PavMV) e Pavonia yellow mosaic virus (PavYMV). No estudo para avaliar a dinâmica evolutiva do ToSRV, plantas de tomateiro e Nicandra physaloides foram inoculadas via biobalística com clone infeccioso do ToSRV e mantidas em casa-de-vegetação. DNA total foi extraído de folhas coletadas aos 30, 75 e 120 dias após a inoculação, e submetido a sequenciamento na plataforma Illumina HiSeq 2000. As bibliotecas de DNA foram submetidas a análise de qualidade no software

FastQC. O alinhamento dos reads foi realizado no software Geneious, utilizando o clone infeccioso como sequência referência. Ambos os componentes genômicos do ToSRV apresentam taxa de substituição semelhantes a de vírus de RNA: 3,06 x 10^{-3} e 2,03 x 10^{-3} subst/sítio/ano para o DNA-A e DNA-B, respectivamente, em N. physaloides, e 1,38 x 10^{-3} e 8,68 x 10^{-4} subst/sítio/ano para o DNA-A e DNA-B, respectivamente, em tomateiro. Valores de taxa de substituição similares aos já descritos para begomovírus foram encontrados também para os genes CP, Rep, MP e NSP em ambos os hospedeiros. Foram quantificadas substituições sinônimas e não-sinônimas, transversões e transições, além de deleções e inserções nos genes CP, Rep, MP e NSP. Foi observado um decréscimo no número de variações ao longo do tempo e consequente aumento do número de sítios idênticos ao genoma referência. Nas bibliotecas provenientes de N. physaloides foi observada a supressão dos códons de terminação dos genes MP e NSP em todo decorrer da infecção, sugerindo uma estratégia adaptativa. O cálculo da entropia de Shannon identificou como hotspots de mutação a região N-terminal do gene Rep, as regiões comuns no DNA-A e DNA-B (CR-A e CR-B, respectivamente) e a região intergênica longa entre os genes MP e NSP no DNA-B (LIR-B). Estes dados sugerem que o ToSRV evolui como quasispecies, apresentando elevada variabilidade genética, o que poderia explicar em parte sua prevalência no campo.

INTRODUCTION

Various agricultural systems in different regions of the world have suffered serious economic constraints due to geminivirus infection. Some factors, including the emergence of new viral species/variants, emergence of more efficient vector populations, climate change, changes in production systems, unchecked transit of infected material and introduction of susceptible varieties, individually or together, have contributed to the increased incidence and severity of infection by geminiviruses (Varma and Malathi, 2003).

The Geminiviridae family is comprised of viruses with a circular, single-stranded DNA genome encapsidated in twinned icosahedral particles. Members of this family can infect monocot or dicot plants and are widespread in all tropical and subtropical regions of the world, causing severe diseases in many economically relevant crops such as bean, cassava, cotton, maize, pepper and tomato (Moffat, 1999; Shepherd et al., 2010; Navas-Castillo et al., 2011). Based on the type of insect vector, host range, genomic organization and phylogenetic relationships, the geminiviruses are classified into seven genera: Begomovirus, Becurtovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus and Turncurtovirus (Brown et al., 2012; Varsani et al., 2014). The begomoviruses have one or two genomic components, are transmitted by the whitefly Bemisia tabaci and infect dicot plants.

Begomoviruses can be divided into two groups: Old World (Eastern hemisphere: Europe, Africa, Asia and Oceania) and New World (Western hemisphere: the Americas) (Rybicki, 1994; Padidam et al., 1999a; Paximadis et al., 1999). Most Begomoviruses from the New World have two genomic components known as DNA-A and DNA-B (Fauquet et al., 2005). Tomato leaf deformation virus

(ToLDeV), which has only a single, DNA-A-like component, was the first monopartite begomovirus reported to naturally occur in the New World (Melgarejo et al., 2013). The two components of bipartite begomoviruses do not share significant sequence identity, except for a region with approximately 200 nt known as the common region (CR), which includes the origin of replication (Hanley-Bowdoin et al., 1999). Old World (OW) begomoviruses can be either mono- or bipartite. Monopartite begomoviruses have a genomic component which is homologous to the DNA-A of bipartite begomoviruses (Padidam et al., 1996; Mansoor et al., 2003).

Populations of geminiviruses, including begomoviruses, have a high degree of genetic variability, comparable to viruses with RNA genomes (Ge et al., 2007; Prasanna et al., 2010; Rocha et al., 2013). The main sources of genetic variability in plant viruses are mutation, recombination and pseudo-recombination (Monci et al., 2002; García-Arenal et al., 2003; Seal et al., 2006). Frequent recombination events (Padidam et al., 1999b; Rocha et al., 2013), the occurrence of pseudo-recombination between viruses with bipartite genomes (Gilbertson et al., 1993; Andrade et al., 2006) and high mutation and nucleotide substitution rates (Duffy et al., 2008; Duffy and Holmes, 2009) are the main factors that promote the high variability observed for begomoviruses.

Mutation is the primary mechanism for generation of variability, with natural selection, recombination, genetic drift and gene flow acting to mold the genetic structure of the population (Duffy et al., 2008). The mutation rates and nucleotide substitution rates observed for geminiviruses are similar to those calculated for RNA viruses, despite the expectation that they would be lower since geminiviruses use the host's proof-reading replication machinery, which in theory would increase the fidelity of replication (Duffy et al., 2008). There is evidence that the rapid evolution

of geminiviruses is at least partly directed by mutational processes that act specifically on ssDNA (Harkins et al., 2009). Studies in bacterial and animal systems have indicated that substitutions rates of dsDNA and ssDNA viruses differ significantly (Duffy et al., 2008).

Recombination is a very common event in geminiviruses (Padidam et al., 1999b; Lefeuvre et al., 2009) and seems to contribute greatly to their genetic variability, increasing their evolutionary potential and local adaptation (Harrison and Robinson, 1999; Padidam et al., 1999b; Berrie et al., 2001; Monci et al., 2002; Lima et al., 2013; Rocha et al., 2013). Knowledge on the existence and frequency of recombination in a virus population may further advance the understanding of which genes are interchanged. This information is important, for example, to maximize the durability of genetic resistance, since new recombinant variants may have an increased ability to infect previously resistant genotypes (Monci et al., 2002; Awadalla, 2003; Sattar et al., 2013). Indeed, recombination events have been directly implicated in the emergence of new begomovirus diseases and epidemics (Zhou et al., 1997; Pita et al., 2001), including the devastating epidemic of cassava mosaic in Uganda and neighboring countries, caused by a recombinant isolate of East African cassava mosaic virus (Zhou et al., 1997; Pita et al., 2001), the epidemics of tomato leaf curl disease in Spain, with the emergence of the recombinant species Tomato yellow leaf curl Malaga virus and Tomato yellow leaf curl Axarquia virus (Monci et al., 2002; García-Andrés et al., 2006; García-Andrés et al., 2007a; García-Andrés et al., 2007b), and the epidemics of cotton leaf curl disease in Pakistan caused by a complex of recombinant Cotton leaf curl virus isolates (Zhou et al., 1998; Idris and Brown, 2002).

The existence of two genomic components in many begomoviruses promotes an alternative mechanism for the exchange of genetic material, known as pseudo-recombination (also known as reassortment): the exchange of genomic components between two distinct viruses (more often strains of the same species, but sometimes between closely related species) without the need for intermolecular recombination (Gilbertson et al., 1993; Sung and Coutts, 1995; Andrade et al., 2006; revised by Rojas et al., 2005). The viability of pseudo-recombinants indicates that factors involved in replication and movement can be interchangeable. On the other hand, the often observed asymmetry of reciprocal pseudo-recombinants indicates that this is a complex phenomenon involving interactions among multiple viral and host factors (Hill et al., 1998).

Typically, in genetic variability studies, the complete genome of begomoviruses is cloned from rolling circle amplification (RCA) or PCR products (Briddon et al., 2000; Inoue-Nagata et al., 2004), followed by conventional sequencing. As informative as these studies have been, they are limited by the impossibility of cloning every single variant present in an infected plant - inevitably, only the most prevalent variants will be cloned. Thus, even the most complete studies provide only a rough estimate of the true genetic variability of the viral population, and may grossly underestimate the presence of minor (less fit) variants which could become prevalent after a genetic bottleneck, or after horizontal transfer.

Recently, a study used a new approach, RCA followed by next generation sequencing (NGS), to evaluate the diversity of a population of a begomovirus and associated DNA satellites in naturally infected tomato and okra plants (Idris et al., 2014). The authenticity of the sequences and reproducibility of the approach was validated by comparing the results with those obtained by cloning and Sanger

sequencing. This was the first report of NGS implementation to explore the diversity and identify begomovirus-satellite complexes directly out of naturally infected plants, optimizing the exploration of diversity and populational structure independently of viral abundance (Idris et al., 2014).

NGS was used also to determinate the nucleotide substitution rate of a potyvirus. Dunham et al. (2014) determined the extension and pattern of genetic diversity of Zucchini yellow mosaic virus (ZYMV) by sequencing 23 leaves that grew sequentially along a single Curcubita pepo subsp. texana vine. The authors verified that systemic movement is characterized by sequential bottlenecks in the ZYMV population, although not enough to reduce the population to a single virion since multiple variants were consistently transmitted between the leaves. Moreover, the authors examined the fixation of mutations that resulted in a conformational change in the CI protein, suggesting that these mutations may confer a selective advantage related to systemic movement of the virus in C. pepo.

This study aimed to perform the molecular characterization of two novel begomoviruses infecting non-cultivated plants in Brazil, and follow and quantify the evolution of the begomovirus Tomato severe rugose virus (ToSRV) in a cultivated host (tomato) and in a non-cultivated host (Nicandra physaloides) using NGS.

REFERENCES

ANDRADE, E.C., MANHANI, G.G., ALFENAS, P.F., CALEGARIO, R.F., FONTES, E.P.B., ZERBINI, F.M. 2006. Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from Sida sp., forms pseudorecombinants with begomoviruses from tomato but not from Sida. J Gen Virol 87:3687-3696.

AWADALLA, P. 2003. The evolutionary genomics of pathogen recombination. Nat Rev Genet 4:50-60.

BARBOSA, J.C., BARRETO, S.S., INOUE-NAGATA, A.K., REIS, M.S., FIRMINO, A.C., BERGAMIN, A., REZENDE, J.A.M. 2009. Natural infection of Nicandra physaloides by Tomato severe rugose virus in Brazil. J Gen Plant Pathol 75:440-443.

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

BERRIE, L.C., RYBICKI, E.P., REY, M.E.C. 2001. Complete nucleotide sequence and host range of South African cassava mosaic virus: Further evidence for recombination amongst begomoviruses. J Gen Virol 82:53-58.

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

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

BRIDDON, R.W., MANSOOR, S., BEDFORD, I.D., PINNER, M.S., MARKHAM, P.G. 2000. Clones of cotton leaf curl geminivirus induce symptoms atypical of cotton leaf curl disease. Virus Genes 20:19-26.

BROWN, J.K., FAUQUET, C.M., BRIDDON, R.W., ZERBINI, F.M., MORIONES, E., NAVAS-CASTILLO, J. 2012. Family Geminiviridae. Pages 351-373 in: Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses, A.M.Q. King, M.J. Adams, E.B. Carstens, and E.J. Lefkowitz, eds. Elsevier Academic Press, London, UK.

CASTILLO-URQUIZA, G.P., BESERRA JUNIOR, J.E.A., ALFENAS-ZERBINI, P., VARSANI, A., LIMA, A.T.M., BARROS, D.R., ZERBINI, F.M. 2007. Genetic diversity of begomoviruses infecting tomato in Paty do Alferes, Rio de Janeiro state, Brazil. Virus Rev Res 12:233.

CASTILLO-URQUIZA, G.P., BESERRA JR., J.E.A., BRUCKNER, F.P., LIMA, A.T.M., VARSANI, A., ALFENAS-ZERBINI, P., ZERBINI, F.M. 2008. Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Arch Virol 153:1985-1989.

COTRIM, M.A., KRAUSE-SAKATE, R., NARITA, N., ZERBINI, F.M., PAVAN, M.A. 2007. Diversidade genética de begomovírus em cultivos de tomateiro no Centro-Oeste Paulista. Summa Phytopathol 33:300-303.

DUFFY, S., HOLMES, E.C. 2009. Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. J Gen Virol 90:1539-1547.

DUFFY, S., SHACKELTON, L.A., HOLMES, E.C. 2008. Rates of evolutionary change in viruses: Patterns and determinants. Nat Rev Genet 9:267-276.

DUNHAM, J.P., SIMMONS, H.E., HOLMES, E.C., STEPHENSON, A.G. 2014. Analysis of viral (Zucchini yellow mosaic virus) genetic diversity during systemic movement through a Cucurbita pepo vine. Virus Res 191:172-179.

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

FAUQUET, C.M., MAYO, M.A., MANILOFF, J., DESSELBERGER, U., BALL, L.A., eds. 2005. Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego.

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

FERNANDES, F.R., ALBUQUERQUE, L.C., OLIVEIRA, C.L., CRUZ, A.R.R., ROCHA, W.B., PEREIRA, T.G., NAITO, F.Y.B., DIAS, N.D., NAGATA, T., FARIA, J.C., ZERBINI, F.M., ARAGÃO, F.J.L., INOUE-NAGATA, A.K. 2011. Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting Euphorbia heterophylla plants. Arch Virol 156:2063-2069.

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

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

GARCÍA-ANDRÉS, S., ACCOTTO, G.P., NAVAS-CASTILLO, J., MORIONES, E. 2007a. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. Virology 359:302-312.

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

GARCÍA-ARENAL, F., FRAILE, A., MALPICA, J.M. 2003. Variation and evolution of plant virus populations. Int Microbiol 6:225-232.

GE, L.M., ZHANG, J.T., ZHOU, X.P., LI, H.Y. 2007. Genetic structure and population variability of Tomato yellow leaf curl China virus. J Virol 81:5902-5907.

GILBERTSON, R.L., HIDAYAT, S.H., PAPLOMATAS, E.J., ROJAS, M.R., HOU, Y.-H., MAXWELL, D.P. 1993. Pseudorecombination between infectious cloned

DNA components of tomato mottle and bean dwarf mosaic geminiviruses. J Gen Virol 74:23-31.

HANLEY-BOWDOIN, L., SETTLAGE, S.B., OROZCO, B.M., NAGAR, S., ROBERTSON, D. 1999. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Pl Sci 18:71-106.

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

HARRISON, B.D., ROBINSON, D.J. 1999. Natural genomic and antigenic variation in white-fly transmitted geminiviruses (begomoviruses). Annu Rev Phytopath 39:369-398.

HILL, J.E., STRANDBERG, J.O., HIEBERT, E., LAZAROWITZ, S.G. 1998. Asymmetric infectivity of pseudorecombinants of Cabbage leaf curl virus and Squash leaf curl virus: Implications for bipartite geminivirus evolution and movement. Virology 250:283-292.

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

IDRIS, A., AL-SALEH, M., PIATEK, M., AL-SHAHWAN, I., ALI, S., BROWN, J.K. 2014. Viral metagenomics: Analysis of begomoviruses by Illumina high-throughput sequencing. Viruses 6:1219-1236.

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

INOUE-NAGATA, A.K., ALBUQUERQUE, L.C., ROCHA, W.B., NAGATA, T. 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. J Virol Met 116:209-211.

JOVEL, J., RESKI, G., ROTHENSTEIN, D., RINGEL, M., FRISCHMUTH, T., JESKE, H. 2004. Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. Arch Virol 149:829-841.

LEFEUVRE, P., LETT, J.M., VARSANI, A., MARTIN, D.P. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J Virol 83:2697-2707.

LIMA, A.T.M., SOBRINHO, R.R., GONZALEZ-AGUILERA, J., ROCHA, C.S., SILVA, S.J.C., XAVIER, C.A.D., SILVA, F.N., DUFFY, S., ZERBINI, F.M. 2013. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. J Gen Virol 94:418-431.

MANSOOR, S., BRIDDON, R.W., ZAFAR, Y., STANLEY, J. 2003. Geminivirus disease complexes: An emerging threat. Trends Plant Sci 8:128-134.

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

MELLO, R.N., ALMEIDA, A.M.R., ZERBINI, F.M. 2000. Detection and identification of geminiviruses infecting soybean and associated weeds in Brazil. Fitopatol Bras 25:444.

MELLO, R.N., COTRIM, M.A.A., LOPES, E.F., MOREIRA, A.G., CONTIN, F.S., FONTES, E.P.B., ALMEIDA, A.M.R., ZERBINI, F.M. 2002. Survey of begomoviruses associated with soybean and identification of Sida mottle virus (SiMoV) infecting this crop in Brazil. Virus Rev Res 7(Supplement):157.

MOFFAT, A.S. 1999. Plant pathology - Geminiviruses emerge as serious crop threats. Science 286:1835-1835.

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

NAVAS-CASTILLO, J., FIALLO-OLIVÉ, E., SÁNCHEZ-CAMPOS, S. 2011. Emerging virus diseases transmitted by whiteflies. Annu Rev Phytopath 49:219-248.

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

NOZAKI, D.N., KRAUSE-SAKATE, R., HASEGAWA, J.M., CESAR, M.A., DZIUBA, P.H., PAVAN, M.A. 2005. Ocorrência de Tomato severe rugose virus em pimentão (Capsicum annuum L.) no estado de São Paulo. Fitopatol Bras 30 (Suplemento):S189.

OROZCO, B.M., MILLER, A.B., SETTLAGE, S.B., HANLEY-BOWDOIN, L. 1997. Functional domains of a geminivirus replication protein. J Biol Chem 272:9840-9846.

PADIDAM, M., BEACHY, R.N., FAUQUET, C.M. 1996. The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. Virology 224:390-404.

PADIDAM, M., BEACHY, R.N., FAUQUET, C.M. 1999a. A phage single-stranded DNA (ssDNA) binding protein complements ssDNA accumulation of a geminivirus and interferes with viral movement. J Virol 73:1609-1616.

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

PAXIMADIS, M., IDRIS, A.M., TORRES-JEREZ, I., VILLARREAL, A., REY, M.E.C., BROWN, J.K. 1999. Characterization of tobacco geminiviruses in the Old and New Worlds. Arch Virol 144:703-717.

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

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

PRASANNA, H.C., SINHA, D.P., VERMA, A., SINGH, M., SINGH, B., RAI, M., MARTIN, D.P. 2010. The population genomics of begomoviruses: Global scale population structure and gene flow. Virol J 7:220.

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

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

ROJAS, M.R., HAGEN, C., LUCAS, W.J., GILBERTSON, R.L. 2005. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Annu Rev Phytopath 43:361-394.

RYBICKI, E.P. 1994. A phylogenetic and evolutionary justification for three genera of Geminiviridae. Arch Virol 139:49-77.

SATTAR, M.N., KVARNHEDEN, A., SAEED, M., BRIDDON, R.W. 2013. Cotton leaf curl disease - an emerging threat to cotton production worldwide. J Gen Virol 94:695-710.

SEAL, S.E., VAN DEN BOSCH, F., JEGER, M.J. 2006. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Crit Rev Pl Sci 25:23-46.

SHEPHERD, D.N., MARTIN, D.P., VAN DER WALT, E., DENT, K., VARSANI, A., RYBICKI, E.P. 2010. Maize streak virus: An old and complex 'emerging' pathogen. Mol Plant Pathol 11:1-12.

SILVA, A.K.F., SANTOS, C.D.G., NASCIMENTO, A.K.Q. 2010. Begomovirus transmission from weeds to tomato by the whitefly. Planta Daninha 28:507-514.

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

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

SUNG, Y.K., COUTTS, R.H. 1995. Pseudorecombination and complementation between potato yellow mosaic geminivirus and tomato golden mosaic geminivirus. J Gen Virol 76:2809-2815.

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

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

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

VARMA, A., MALATHI, V.G. 2003. Emerging geminivirus problems: A serious threat to crop production. Ann Appl Biol 142:145-164.

VARSANI, A., NAVAS-CASTILLO, J., MORIONES, E., HERNÁNDEZ-ZEPEDA, C., IDRIS, A., BROWN, J.K., ZERBINI, F.M., MARTIN, D.P. 2014. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. Arch Virol 159:2193-2203.

VOINNET, O., PINTO, Y.M., BAULCOMBE, D.C. 1999. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci USA 96:14147-14152.

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

ZERBINI, F.M., ANDRADE, E.C., BARROS, D.R., FERREIRA, S.S., LIMA, A.T.M., ALFENAS, P.F., MELLO, R.N. 2005. Traditional and novel strategies for geminivirus management in Brazil. Australas Plant Path 34:475-480.

ZHOU, X., LIU, Y., ROBINSON, D.J., HARRISON, B.D. 1998. Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. J Gen Virol 79:915-923.

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

CHAPTER 1

TWO NOVEL BEGOMOVIRUSES INFECTING THE MALVACEOUS WEED *Pavonia* SP. IN BRAZIL

Pinto, V.B., Silva, J.P., Fiallo-Olivé, E., Navas-Castillo, J., Zerbini, F.M. Two novel begomoviruses infecting the malvaceous weed Pavonia sp. in Brazil. Archives of Virology, submitted.

Two novel begomoviruses infecting the malvaceous weed Pavonia sp. in Brazil

Vitor Batista Pinto[1,2], João Paulo Silva[1,2], Elvira Fiallo-Olivé[2,3], Jésus Navas-Castillo[2,3], Francisco Murilo Zerbini[1,2]

[1]Dep. de Fitopatologia, [2]National Institute for Plant-Pest Interactions, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil; [3]Instituto de Hortofruticultura Subtropical y Mediterránea ″La Mayora″, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), 29750 Algarrobo-Costa, Málaga, Spain

Abstract

Begomoviruses are whitefly-transmitted, single-stranded DNA viruses that are often associated with non-cultivated plants. Here, we report the detection of members of two new begomovirus species in Pavonia sp. (Malvaceae). Sequence comparisons and phylogenetic analysis have shown that these novel species are related to New World begomoviruses. The nucleotide sequences of the DNA-A of both viruses had the greatest identity with Abutilon mosaic Bolivia virus (AbMBoV). Based on symptoms observed in the field and considering the host, we proposed the names Pavonia mosaic virus (PavMV) and Pavonia yellow mosaic virus (PavYMV) for these two new begomoviruses.

The genus Begomovirus (family Geminiviridae) includes species that infect dicotyledoneous plants and whose genomes are composed of one or two molecules

of circular, single-stranded DNA [1]. In nature, these viruses are spread by the Bemisia tabaci sibling species group (Hemiptera: Aleyrodidae) and its distribution is related to the spread of the vector [2, 3]. The species B. tabaci Middle East-Asia Minor 1 (MEAM1) is highly efficient in transmitting begomoviruses and has a wide host range [4, 5]. Begomoviruses indigenous to the New World are typically bipartite, except for Tomato leaf deformation virus, recently described in Peru and Ecuador [6]. The two genomic components, referred to as DNA-A and DNA-B, are of similar size (2.5-2.7 kb). The DNA-A contains five open reading frames (ORFs) which encode proteins with functions in viral replication (Rep, Ren), suppression of host defenses (Trap) and particle formation (CP), while the two ORFs in the DNA-B enconde a movement protein (MP) and a nuclear shuttle protein which is also involved in suppression of host defenses (NSP) [7, 8]. Begomoviruses cause severe diseases in economically important crops throughout the tropics and subtropics, and are also frequently associated with non-cultivated plants [5]. Malvaceous plants are one of the largest natural begomovirus reservoirs in the Americas [9-17]. The genus Pavonia, probably the largest within the family Malvaceae, is represented by 250 species [18]. Many species are used as ornamentals and some have potential for the cellulose industry by having cellulose without contaminants [18].

As part of an ongoing effort to assess begomovirus diversity and the emergence of novel species, symptomatic Pavonia sp. plants were collected in the cities of Albuquerque (S19º23'49.7", W57º25'29.5"; sample #51) and Corumbá (S19°11'05.8", W057°31'54.1"; sample #40) in Mato Grosso do Sul state, Brazil, in September 2014 (Suppl. Fig. S1). Total DNA was extracted [19], and the presence of a begomovirus was confirmed by rolling-circle amplification (RCA) [20]. Full-length genomic components were cloned into pBLUESCRIPT KS+ (Stratagene) after

monomerization with the restriction enzymes ApaI, EcoRI, and SacI and were completely sequenced by primer walking (Macrogen, Inc., Seoul, South Korea). The assembly of nucleotide sequences was carried out using Geneious version 8.0.5. Pairwise sequence comparisons were performed using Sequence Demarcation Tool (SDT) v.1.2 [21]. Multiple sequence alignments were obtained using the MUSCLE algorithm implemented in MEGA6 [22]. Phylogenetic trees were constructed using Bayesian inference performed with MrBayes v.3 [23], using the General Time Reversible (GTR) nucleotide substitution model selected by MrModeltest v.2.2 [24] in the Akaike Information Criterion (AIC). The analysis was run for 10 million generations, excluding the first 2,000,000 generations as burn-in. The trees were visualized in FigTree v.1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). Recombination analysis was performed with Recombination Detection Program (RDP) v.4.5.1 [25] using default settings and a Bonferroni-corrected P-value cutoff of 0.05. Only the recombination events detected by more than four of the seven tests implemented in RDP were considered to be reliable.

DNA-A and -B components were cloned for the two Pavonia samples. All four components have the conserved nonanucleotide that contains the origin of replication (5'-TAATATTAC-3'). That the A and B components from each sample are cognate components of a bipartite begomovirus is indicated by their identical iteron sequences (GGTG for the components from sample #40; GGGG for the components from sample #51) and by the digestion of the RCA products from each sample with a 4-base cutter restriction enzyme, which indicates that they were the only two DNA components present in each sample (data not shown). All four DNA components have the typical organization of New World, bipartite begomoviruses, with five ORFs in the DNA-A and two ORFs in the DNA-B. Pairwise sequence

analysis of the DNA-A indicated that the two isolates, named BR-Cor40-14 and BR-Alb51-14, have 87% nucleotide sequence identity with each other and <82% identity with previously described begomoviruses (Table 1). The two DNA-B components are 89% identical and have <79% identity with other begomoviruses. The most closely related species is Abutilon mosaic Bolivia virus (AbMBoV) for both components (Table 1). Based on the begomovirus species demarcation criteria recently updated by the Geminiviridae study group of the ICTV [1], each isolate comprises a new begomovirus species for which we propose the names Pavonia yellow mosaic virus (PavYMV) (BR-Cor40-14) and Pavonia mosaic virus (PavMV) (BR-Alb51-14).

Pairwise sequence comparisons based on the deduced amino acid sequences of each viral protein indicated that the CP and Ren are the most conserved proteins in both species, with 78-85% and 79-89% identity with other begomoviruses, respectively, and Rep and NSP the least conserved, both with <81% identity (Suppl. Tables S2-S7).

Phylogenetic analysis based on both the complete DNA-A and DNA-B sequences indicate that PavMV and PavYMV are most closely related to AbMBoV (Fig. 1; Suppl. Fig. S2).

Recombination events were detected in the genomes of both viruses. The events involved two genomic regions for PavYMV (nt 674-1226 and 1994-2198 in the CP and Rep genes, respectively) and three for PavMV (nt 355-408 in the CP gene; nt 2049-2197 and 2198-2433 in the Rep gene). Malvaceous-infecting begomoviruses from the same phylogenetic clade (AbMBoV, SiBoV) were identified as putative parents. These results indicate that, as for the vast majority of

begomoviruses, recombination is involved in the evolution of both PavMV and PavYMV.

The presence of whitefly-transmitted viruses in non-cultivated malvaceous hosts, causing the so-called "infectious chlorosis", has been reported in Brazil since the 1940's [26]. However, it was only with the recent development of sequence-unbiased detection tools that the incredible species diversity of these viruses started to unravel [14, 17, 27, 28]. It is curious that severe epidemics of begomoviruses in malvaceous crops such as cotton and okra have not been reported in Brazil, unlike in other regions such as the Indian Subcontinent and West Africa [29, 30]. This could be related to the preference of local whitefly populations for other hosts.

Acknowledgments

This work was funded by the Science Without Borders program of CNPq (grant 401838/2013-7 to FMZ and JNC) and by FAPEMIG (grant CAG-APQ-02037-13 to FMZ). JNC was the recipient of a Special Visiting Scientist fellowship under the SWB program. EFO was the recipient of a Visiting Scientist fellowship from FAPEMIG.

References

1. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olive E, Briddon RW, Hernandez-Zepeda C, Idris A, Malathi VG, Martin DP, Rivera-Bustamante R, Ueda S, Varsani A (2015) Revision of Begomovirus taxonomy based on pairwise sequence comparisons. Arch Virol 160:1593-1619
2. Jones DR (2003) Plant viruses transmitted by whiteflies. Eur J Plant Pathol 109:195-219
3. Brown JK (2007) The Bemisia tabaci complex: Genetic and phenotypic variability drives begomovirus spread and virus diversification. APSNet

Features. Available at: http://www.apsnet.org/publications/apsnetfeatures/Pages/BemisiatabaciComplex.aspx.

4. Brown JK, Idris AM (2005) Genetic differentiation of whitefly Bemisia tabaci mitochondrial cytochrome oxidase I, and phylogeographic concordance with the coat protein of the plant virus genus Begomovirus. Ann Entom Soc Am 98:827-837

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

6. Melgarejo TA, Kon T, Rojas MR, Paz-Carrasco L, Zerbini FM, Gilbertson RL (2013) Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. J Virol 87:5397-5413

7. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL (2005) Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Annu Rev Phytopath 43:361-394

8. Brustolini OJ, Machado JP, Condori-Apfata JA, Coco D, Deguchi M, Loriato VA, Pereira WA, Alfenas-Zerbini P, Zerbini FM, Inoue-Nagata AK, Santos AA, Chory J, Silva FF, Fontes EP (2015) Sustained NIK-mediated antiviral signalling confers broad-spectrum tolerance to begomoviruses in cultivated plants. Plant Biotechnol J doi: 10.1111/pbi.12349

9. Costa AS (1955) Studies on Abutilon mosaic in Brazil. J Phytopath 24:97-112

10. Frischmuth T, Engel M, Lauster S, Jeske H (1997) Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, Sida-infecting bipartite geminiviruses in Central America. J Gen Virol 78:2675-2682

11. Hofer P, Engel M, Jeske H, Frischmuth T (1997) Nucleotide sequence of a new bipartite geminivirus isolated from the common weed Sida rhombifolia in Costa Rica. Virology 78:1785-1790

12. Roye ME, McLaughlin WA, Nakhla MK, Maxwell DP (1997) Genetic diversity among geminiviruses associated with the weed species Sida spp., Macroptilium lathyroides and Wissadula amplissima from Jamaica. Plant Dis 81:1251-1258

13. Echemendía AL, Ramos PL, Díaz L, Peral R, Fuentes A, Pujol M, González G (2004) First report of Sida golden yellow vein virus infecting Sida species in Cuba. Plant Pathol 53:234

14. Jovel J, Reski G, Rothenstein D, Ringel M, Frischmuth T, Jeske H (2004) Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. Arch Virol 149:829-841

15. Fiallo-Olivé E, Martinez-Zubiaur Y, Moriones E, Navas-Castillo J (2010) Complete nucleotide sequence of Sida golden mosaic Florida virus and phylogenetic relationships with other begomoviruses infecting malvaceous weeds in the Caribbean. Arch Virol 155:1535-1537

16. Fiallo-Olive E, Navas-Castillo J, Moriones E, Martinez-Zubiaur Y (2012) Begomoviruses infecting weeds in Cuba: Increased host range and a novel virus infecting Sida rhombifolia. Arch Virol 157:141-146

17. Tavares SS, Ramos-Sobrinho R, Gonzalez-Aguilera J, Lima GSA, Assunção IP, Zerbini FM (2012) Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on Sida spp. Planta Daninha 30:305-315

18. Fryxell PA (1999) Pavonia Cavanilles (Malvaceae). Flora Neotropica Monograph 76

19. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull 19:11-15

20. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met 116:209-211

21. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277

22. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729

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

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

25. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003

26. Orlando A, Silberschmidt K (1946) Studies on the natural dissemination of the virus of "infectious chlorosis" of the Malvaceae and its relation with the insect vector Bemisia tabaci Genn. Arq Inst Biol 17:1-36

27. Castillo-Urquiza GP, Beserra Jr. JEA, Bruckner FP, Lima ATM, Varsani A, Alfenas-Zerbini P, Zerbini FM (2008) Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Arch Virol 153:1985-1989

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

29. Sattar MN, Kvarnheden A, Saeed M, Briddon RW (2013) Cotton leaf curl disease - an emerging threat to cotton production worldwide. J Gen Virol 94:695-710

30. Kon T, Rojas MR, Abdourhamane IK, Gilbertson RL (2009) Roles and interactions of begomoviruses and satellite DNAs associated with okra leaf curl disease in Mali, West Africa. J Gen Virol 90:1001-1013

Table 1 - Percent identities between the complete DNA-A (above the diagonal) and DNA-B (below the diagonal) nucleotide sequences of the two new begomoviruses species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) with the most closely related begomoviruses.

	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavYMV
TGMV[a]	--	82	77	78	79	81	76	78	78	79	78	75	76	75
ToSRV	71	--	76	80	80	78	77	79	78	79	78	80	76	75
ToMoLCV	66	68	--	77	77	79	75	76	76	77	75	74	76	76
ToLDV	-[b]	-	-	--	84	80	82	84	83	84	79	79	78	78
ToYSV	68	70	67	-	--	81	80	83	88	89	79	78	78	79
SiMBoV1	70	70	67	-	70	--	77	81	80	82	78	76	79	78
SiBrV	70	69	67	-	69	70	--	80	80	81	75	80	78	75
SimMV	68	76	66	-	73	73	69	--	83	83	76	76	76	76
SiMoV	-	-	-	-	-	-	-	-	--	88	78	76	76	77
SiYNV	-	-	-	-	-	-	-	-	-	--	79	78	78	79
AbMBoV	69	72	67	-	68	73	70	70	-	-	--	76	82	82
BGMV	68	69	67	-	68	69	69	69	-	-	67	--	77	74
PavMV[c]	68	70	66	-	70	72	73	70	-	-	78	68	--	87
PavYMV[d]	69	71	67	-	70	72	73	70	-	-	79	69	89	--

[a]Access numbers as in Fig 1.
[b]Not done (sequences not available)
[c]Isolate BR-Alb51-14
[d]Isolate BR-Cor40-14

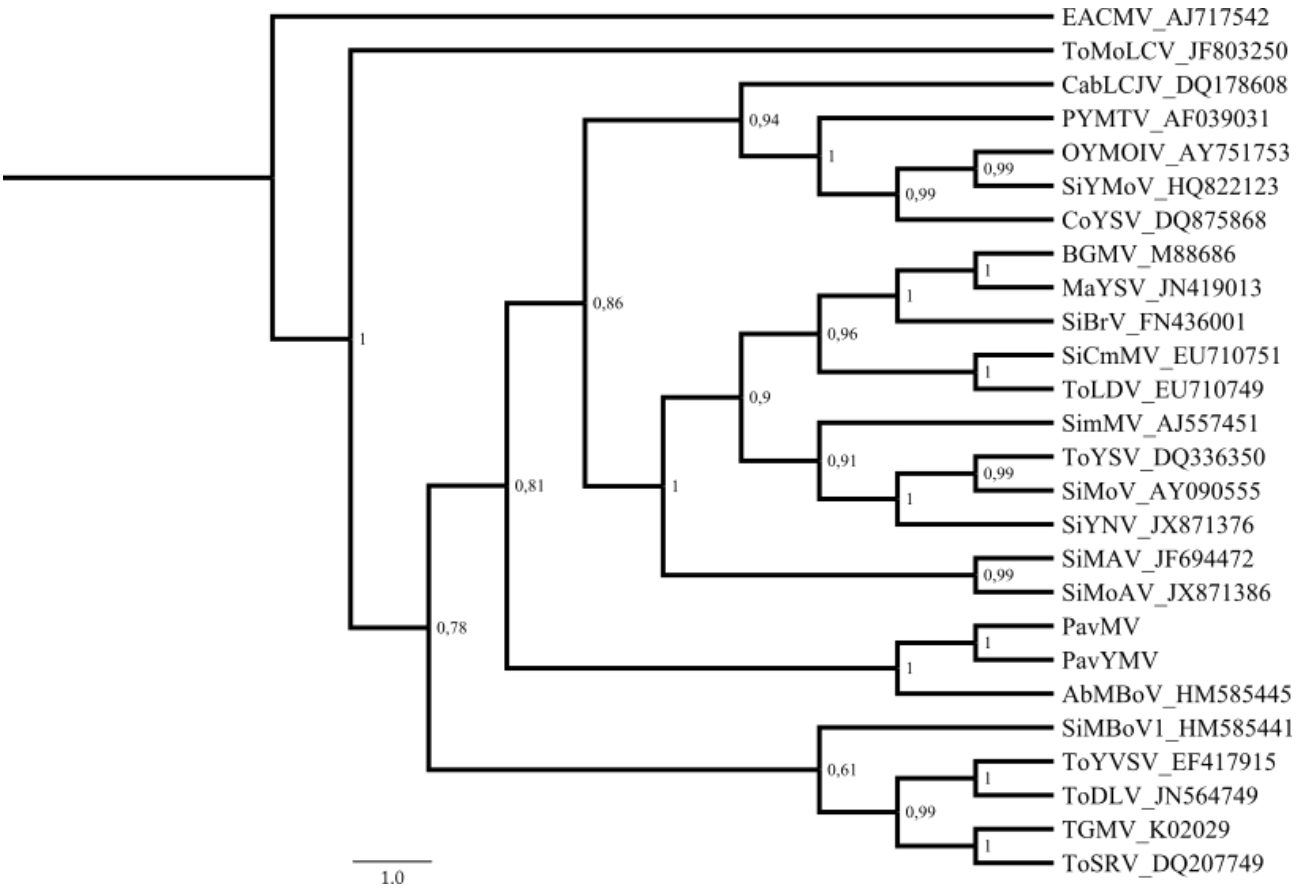


Figure 1 - Phylogenetic tree based on the complete DNA-A sequences of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV), plus additional sequences from New World begomoviruses. The tree was constructed by Bayesian inference using the GTR nucleotide substitution model. Numbers at the nodes indicate Bayesian posterior probabilities. East African cassava mosaic virus (EACMV, an Old World begomovirus) was used as an outgroup. AbMBoV, Abutilon mosaic Bolivia virus; BGMV, Bean golden mosaic virus; CabLCJV, Cabbage leaf curl Jamaica virus; CoYSV, Corchorus yellow spot virus; MaYSV, Macroptilium yellow spot virus; OYMOIV, Okra yellow mottle Iguala virus; PYMTV, Potato yellow mosaic Trinidad virus; SiBrV, Sida Brazil virus; SiCmMV, Sida commom mosaic virus; SiMAV, Sida mosaic Alagoas virus; SiMBoV1, Sida mosaic Bolivia virus 1; SimMV, Sida micrantha mosaic virus; SiMoAV, Sida mottle Alagoas virus; SiMoV, Sida mottle virus; SiYMoV, Sida yellow mottle virus; SiYNV, Sida yellow net virus; TGMV, Tomato golden mosaic virus; ToDLV, Tomato dwarf leaf virus; ToLDV, Tomato leaf distortion virus; ToMoLCV, Tomato mottle leaf curl virus; ToSRV, Tomato severe rugose virus; ToYSV, Tomato yellow spot virus; ToYVSV, Tomato yellow vein streak virus.

Supplementary Figure S1 - Symptoms of mosaic and bright yellow mosaic in the two Pavonia sp. samples collected in Albuquerque (left) and Corumbá (right).

Supplementary Table S1 - Percent identities between the deduced amino acid sequences of the Rep (AC1) protein encoded by the DNA-A of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavMYV
TGMV	--	82	75	79	82	81	73	80	78	79	79	74	76	75
ToSRV		--	75	79	81	80	74	80	78	78	79	79	74	75
ToMoLCV			--	74	76	79	72	75	75	75	74	73	74	76
ToLDV				--	81	79	78	79	78	81	79	79	75	76
ToYSV					--	83	75	80	86	87	79	76	75	77
SiMBoV1						--	75	83	81	85	81	76	77	77
SiBrV							--	75	75	76	74	81	77	72
SimMV								--	78	80	75	76	74	73
SiMoV									--	87	78	74	74	76
SiYNV										--	79	76	76	79
AbMBoV											--	76	81	80
BGMV												--	77	72
PavMV													--	82
PavMYV														--

Supplementary Table S2 - Percent identities between the deduced amino acid sequences of the Trap (AC2) protein encoded by the DNA-A of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	81	79	79	80	86	77	80	82	81	82	77	80	79
ToSRV		--	78	80	82	78	79	81	81	82	80	83	77	78
ToMoLCV			--	78	79	81	77	80	79	78	79	74	81	81
ToLDV				--	86	82	84	88	85	86	82	81	81	81
ToYSV					--	85	83	84	93	92	87	81	83	85
SiMBoV1						--	76	80	84	84	85	78	84	84
SiBrV							--	84	82	82	78	83	79	78
SimMV								--	87	85	81	81	78	81
SiMoV									--	93	84	80	81	83
SiYNV										--	85	82	82	83
AbMBoV											--	79	89	88
BGMV												--	79	79
PavMV													--	96
PavYMV														--

Supplementary Table S3 - Percent identities between the deduced amino acid sequences of the Ren (AC3) protein encoded by the DNA-A of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	83	81	82	82	87	80	81	81	82	81	80	81	81
ToSRV		--	83	82	85	84	82	82	82	84	81	84	81	83
ToMoLCV			--	82	83	82	79	81	80	81	82	78	82	81
ToLDV				--	87	83	85	87	85	86	83	82	81	83
ToYSV					--	84	87	86	90	93	84	83	83	85
SiMBoV1						--	81	81	80	83	83	81	83	83
SiBrV							--	83	84	85	81	83	83	81
SimMV								--	87	87	82	82	80	81
SiMoV									--	93	82	82	80	82
SiYNV										--	84	82	82	84
AbMBoV											--	81	89	88
BGMV												--	81	80
PavMV													--	94
PavYMV														--

Supplementary Table S4 - Percent identities between the deduced amino acid sequences of the AC4 protein encoded by the DNA-A of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	86	72	84	87	82	67	84	81	83	80	67	67	76
ToSRV		--	72	83	85	83	70	86	78	81	78	69	67	77
ToMoLCV			--	77	73	74	68	74	70	76	71	69	66	74
ToLDV				--	87	86	74	88	79	87	80	74	70	78
ToYSV					--	86	71	87	84	88	79	71	69	76
SiMBoV1						--	73	91	79	86	80	72	72	76
SiBrV							--	71	70	72	69	88	77	67
SimMV								--	78	86	77	74	69	75
SiMoV									--	87	77	69	70	72
SiYNV										--	81	76	75	80
AbMBoV											--	67	69	78
BGMV												--	79	66
PavMV													--	70
PavYMV														--

Supplementary Table S5 - Percent identities between the deduced amino acid sequences of the CP (AV1) protein encoded by the DNA-A of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

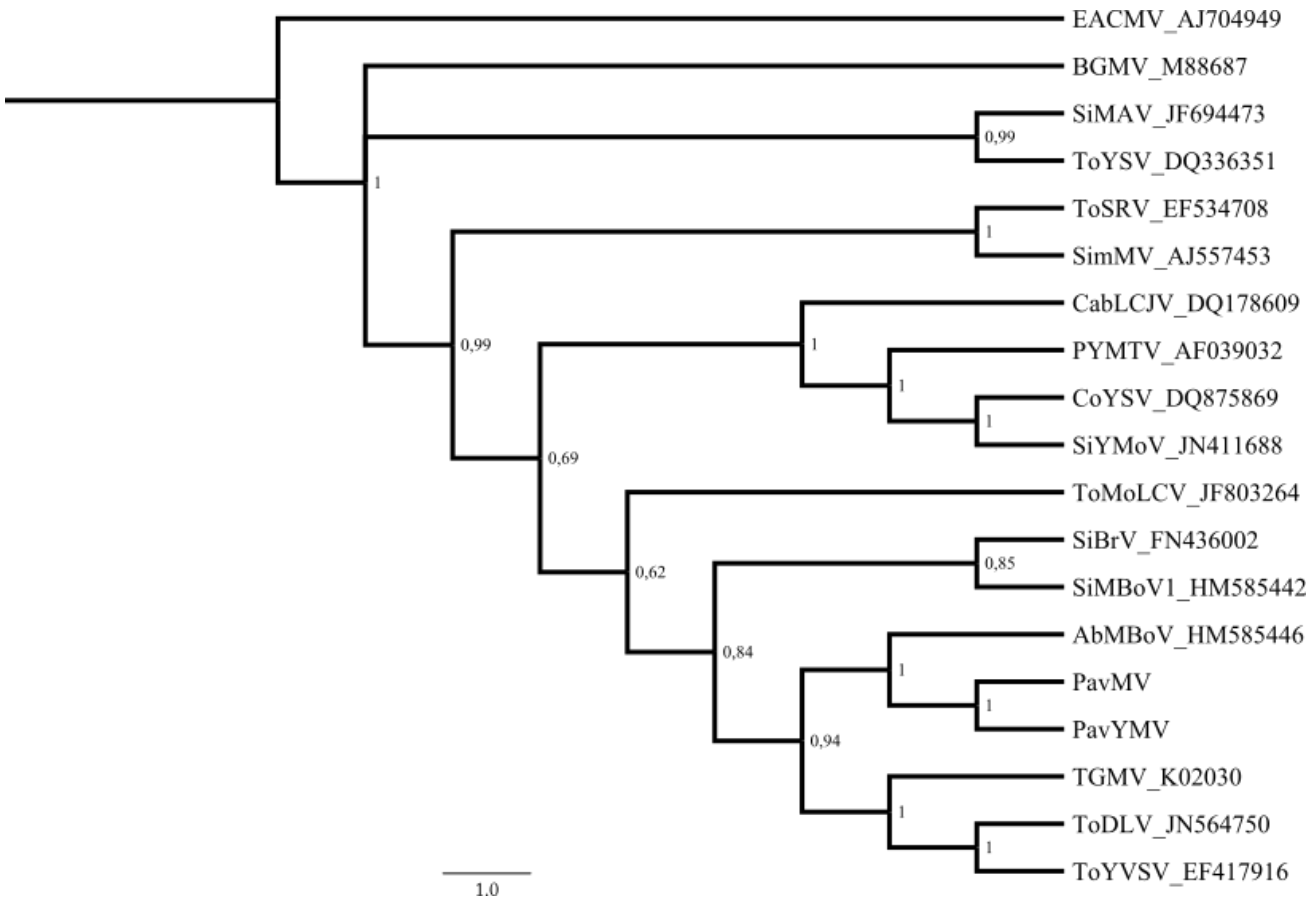
	TGMV	ToSRV	ToMoLCV	ToLDV	ToYSV	SiMBoV1	SiBrV	SimMV	SiMoV	SiYNV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	86	83	84	83	82	85	83	84	84	81	81	80	78
ToSRV		--	82	83	82	81	84	81	82	83	81	82	79	78
ToMoLCV			--	84	83	82	82	82	83	83	81	83	82	81
ToLDV				--	89	84	90	92	92	90	84	83	83	82
ToYSV					--	81	87	89	89	92	83	81	83	83
SiMBoV1						--	84	84	83	82	80	79	84	82
SiBrV							--	90	88	90	81	82	82	81
SimMV								--	91	89	83	81	81	81
SiMoV									--	90	83	82	81	80
SiYNV										--	82	85	83	82
AbMBoV											--	80	82	85
BGMV												--	79	78
PavMV													--	92
PavYMV														--

Supplementary Table S6 - Percent identities between the deduced amino acid sequences of the NSP (BV1) protein encoded by the DNA-B of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToYSV	SiMBoV1	SiBrV	SimMV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	74	71	73	74	73	72	73	73	72	73
ToSRV		--	71	75	75	74	79	74	73	74	75
ToMoLCV			--	72	71	73	71	72	73	71	73
ToYSV				--	75	75	78	77	76	76	77
SiMBoV1					--	73	76	75	73	76	75
SiBrV						--	74	76	74	74	75
SimMV							--	77	74	76	76
AbMBoV								--	73	81	81
BGMV									--	72	73
PavMV										--	90
PavYMV											--

Supplementary Table S7 - Percent identities between the deduced amino acid sequences of the MP (BC1) protein encoded by the DNA-B of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV) and the most closely related viruses.

	TGMV	ToSRV	ToMoLCV	ToYSV	SiMBoV1	SiBrV	SimMV	AbMBoV	BGMV	PavMV	PavYMV
TGMV	--	76	74	76	79	76	76	78	76	79	79
ToSRV		--	73	76	78	74	83	77	75	77	77
ToMoLCV			--	72	75	73	73	75	73	75	75
ToYSV				--	76	74	77	77	76	77	77
SiMBoV1					--	80	79	81	73	81	81
SiBrV						--	76	77	74	78	79
SimMV							--	78	73	77	78
AbMBoV								--	76	88	89
BGMV									--	76	77
PavMV										--	94
PavYMV											--



Supplementary Figure S2 - Phylogenetic tree based on the complete DNA-B sequences of the two new begomovirus species detected in Pavonia sp. (Pavonia mosaic virus, PavMV, and Pavonia yellow mosaic virus, PavYMV), plus additional sequences from New World begomoviruses. The tree was constructed by Bayesian inference using the GTR nucleotide substitution model. Numbers at the nodes indicate Bayesian posterior probabilities. East African cassava mosaic virus (EACMV, an Old World begomovirus) was used as an outgroup. AbMBoV, Abutilon mosaic Bolivia virus; BGMV, Bean golden mosaic virus; CabLCJV, Cabbage leaf curl Jamaica virus; CoYSV, Corchorus yellow spot virus; PYMTV, Potato yellow mosaic Trinidad virus; SiBrV, Sida Brazil virus; SiMAV, Sida mosaic Alagoas virus; SiMBoV1, Sida mosaic Bolivia virus 1; SimMV, Sida micrantha mosaic virus; SiYMoV, Sida yellow mottle virus; TGMV, Tomato golden mosaic virus; ToDLV, Tomato dwarf leaf virus; ToMoLCV, Tomato mottle leaf curl virus; ToSRV, Tomato severe rugose virus; ToYSV, Tomato yellow spot virus; ToYVSV, Tomato yellow vein streak virus.

CHAPTER 2

INTRA-HOST EVOLUTION OF Tomato severe rugose virus (ToSRV)

Abstract

To evaluate and quantify the mutational dynamics of the bipartite begomovirus Tomato severe rugose virus (ToSRV) in a cultivated and a non-cultivated host, plants of tomato and Nicandra physaloides were biolistically inoculated with an infectious clone and the leaves sampled at 30, 75 and 120 days after inoculation. Total DNA was extracted and sequenced in the Illumina HiSeq 2000 platform. The datasets were trimmed with the quality score limit set to 0.01, and the assembly was performed using the infectious clone sequence as reference. We inferred high rates of nucleotide substitution for the two DNA components in both hosts: 3.06×10^{-3} and 2.03×10^{-3} sub/site/year for the DNA-A and DNA-B, respectively, in N. physaloides, and 1.38×10^{-3} and 8.68×10^{-4} sub/site/year the for DNA-A and DNA-B, respectively, in tomato. These values are similar to those estimated for other begomoviruses and for viruses with single-stranded RNA genomes. Strikingly, the number of substitutions decreased over time, with a corresponding increase in the number of identical sites, suggesting the presence of bottlenecks during the systemic infection. In N. physaloides, but not in tomato, stop codon suppression in the MP and NSP genes was detected at the three time points, suggesting an adaptive strategy. Determination of Shannon entropy indicated mutation hotspots in the N-terminal region of Rep gene, the intergenic common region in the DNA-A and DNA-B (CR-A and CR-B, respectively) and the long intergenic region between the MP and NSP genes in the DNA-B (LIR-B). These results indicate that ToSRV evolves as a quasispecies, with a high degree of genetic variability which could be partly responsible for its prevalence in the field.

Introduction

The family Geminiviridae is comprised of viruses with circular, single-stranded DNA genomes encapsidated in twinned icosahedral particles. The family includes seven genera, defined on the basis of host range (monocots or dicots), type of insect vector (leafhoppers, treehoppers or whiteflies), genome organization (mono- or bipartite) and phylogenetic relationships: Begomovirus, Becurtovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus and Turncurtovirus (Brown et al., 2012; Varsani et al., 2014). The genus Begomovirus includes viruses with mono- or bipartite genomes, transmitted by whiteflies (Bemisia tabaci) to dicotyledonous plants. Begomoviruses are widespread in all tropical and subtropical regions of the world, and cause severe diseases in a number of economically relevant crops (Moffat, 1999; Navas-Castillo et al., 2011).

Based on phylogenetic relationships and genome characteristics, begomoviruses can be divided into two groups: Old World (OW; Europe, Africa and Asia) and New World (NW; the Americas) (Rybicki, 1994; Padidam et al., 1999b; Paximadis et al., 1999). The majority of NW begomoviruses are bipartite, with two genomic components named DNA-A and DNA-B. The DNA-A contains genes involved in replication, suppression of host defenses and encapsidation of the viral progeny. The DNA-B contains genes required for intra- and intercelular movement in the plant and suppression of host defenses (reviewed by Rojas et al., 2005).

Populations of geminiviruses, including begomoviruses, have a high degree of genetic variability, equivalent to that observed for viruses with single-stranded RNA genomes (Ge et al., 2007; Duffy et al., 2008; Prasanna et al., 2010; Rocha et al., 2013). Frequent recombination events (Padidam et al., 1999a), the occurrence of

pseudo-recombination between viruses with bipartite genomes (Andrade et al., 2006), and high mutation and nucleotide substitution rates (Duffy et al., 2008; Duffy and Holmes, 2009) are the main factors that promote the high variability observed for begomoviruses.

Brazil is a center of genetic diversity for begomoviruses, with reports of their detection going back to the mid-20th century (Orlando and Silberschmidt, 1946; Costa, 1955; Flores et al., 1960). More recently, a large number of new begomovirus species have been characterized on both cultivated and non-cultivated plants (Ribeiro et al., 2003; Fernandes et al., 2006; Calegario et al., 2007; Ribeiro et al., 2007; Castillo-Urquiza et al., 2008; Fernandes et al., 2008; Fernandes et al., 2009; Silva et al., 2011; Albuquerque et al., 2012; Silva et al., 2012; Tavares et al., 2012). These viruses, all indigenous, emerged after the introduction of the polyphagous Bemisia tabaci Middle East-Asia Minor 1 (MEAM 1, previously know as B. tabaci biotype B) in the mid-1990's (Ribeiro et al., 1998). The characterization of these viruses substantiated their high degree of genetic variability (Silva et al., 2012; Lima et al., 2013; Rocha et al., 2013). Some of these species are widely distributed throughout the country, while others are restricted to certain regions. Tomato severe rugose virus (ToSRV) is the predominant virus on tomato crops in southeastern and midwestern Brazil (Fernandes et al., 2008; González-Aguilera et al., 2012).

Nicandra physaloides is a non-cultivated plant commonly found throughout South America, including Brazil, where it is sometimes used in folk medicine (Agra et al., 1994). Barbosa et al. (2009) found plants of N. physaloides showing symptoms of mosaic and leaf deformation in tomato fields near the city of Sumaré in the state of São Paulo. After cloning and sequencing, the authors identified the presence of

ToSRV, concluding that N. physaloides can serve as a natural reservoir for this virus under field conditions.

The analysis of begomovirus populations found in cultivated and non-cultivated plants indicates that viral populations infecting non-cultivated plants have a higher degree of genetic variability compared to those present in cultivated plants (Lima et al., 2013; Rocha et al., 2013). Rocha et al. (2013) found that the Brazilian populations of begomoviruses are highly recombinant, have a rapid rate of molecular evolution and are structured based on geographical location, which explains the predominance of certain viral species in different regions of the country.

Until recently, all studies of genetic structure and variability of begomovirus populations have been based on cloning and sequencing viral genomes obtained from samples collected in the field. Although these studies have provided valuable information, they are limited by systematic sampling, often carried out in a fragmented way both spatially and temporally, and by the impossibility of sequencing every genomic variant present in a given sample. In this context, the recent advances of mass sequencing technologies (often called next generation sequencing, or NGS) open new possibilities, allowing for example that the viral population in a single plant be sequenced with a coverage of tens or hundreds of thousands of times. Viral evolution could this way be followed in real time.

The wide distribution of ToSRV in the field and the high mutation rate of begomoviruses emphasize the importance of studying the evolution of this species in cultivated and non-cultivated hosts, to understand its evolutionary dynamics and the emergence and establishment of new variants during a short-term viral infection. This study aimed to track, using NGS, the evolution of a population of ToSRV in tomato and N. physaloides.

Material and methods

Viral isolate

Clones corresponding to the DNA-A and DNA-B of the isolate ToSRV-[BR:Pir1:05], obtained by Lima (2007), were used. Both clones correspond to 1.5 copies of the genome including two origins of replication in the same orientation. The infectivity of the clones was confirmed by biolistic inoculation on tomato and Nicotiana benthamiana plants.

Plant material

Tomato seeds (cv. Santa Clara) were sowed in plastic pots containing commercial substrate (Plantmax). Seeds of Nicandra physaloides obtained from a commercial source (AgroCosmos) were submitted to dormancy breaking by incubating in 95% sulfuric acid for 10 minutes and rinsing with running water, and were then planted in plastic cups containing commercial substrate (Plantmax).

Viral inoculation and sample collection

The plants were inoculated via biolistics (Aragão et al., 1996) when they showed the first pair of true leaves. For the inoculation, 10 μg of each genomic component were mixed with tungsten particles (M-10, Bio-Rad). Twenty plants of each species were inoculated. Tomato plants were bombarded at 40 kgf/mmHg, and N. physaloides plants at 50 kgf/mmHg. Two days after bombardment, the plants were transplanted to pots containing a mixture of soil and sand in a 3:1 ratio, and maintained in a greenhouse.

Infection was confirmed by extracting total DNA from systemically-infected leaves (Doyle and Doyle, 1987), followed by rolling circle amplification (Inoue-Nagata et al., 2004). Aliquots of the amplifications were subjected to cleavage with the EcoRI restriction enzyme, which cleaves the ToSRV DNA-A at a single site, and analyzed by agarose gel electrophoresis.

Sample collection started 15 days after inoculation (dai) and was performed at 15-day intervals until 120 dai. Young leaves located in the apical region of each plant were collected and each subsequent collection was held at the subsequent node to the previously collected branch. The process was the same for each of the two plant species. Samples consisted of the leaves from each one of the twenty plants of each host at each time point and were stored at -80°C until DNA extraction.

DNA extraction and sequencing

About 100 mg of leaf tissue from one of the twenty plants from each species was used for genomic DNA extraction. The remaining 19 samples were kept for future use, if necessary. DNA extraction was performed using the ZR Plant kit/Seed MiniPrep DNA (Zymo Research). This kit retains DNA until 40 kpb. Next, the DNA was purified with the DNA Clean & Concentrator-5 kit (Zymo Research), which retains DNA until 23 kbp. After quantification, the DNA samples collected at 30, 75 and 120 dai were sent to Macrogen Inc. (Seoul, South Korea), where they were sequenced in the Illumina HiSeq 2000 platform.

Quality control and filtering

Sequence libraries were received for bioinformatics analysis. Reads were pair-ended with a size of 101 bases each. Quality control was initially performed

using the program FastQC (Andrews, 2010), to check the adapters and the quality distribution of the bases without any prior computer processing of the libraries. Then, the reads were trimmed using Printseq-Lite v. 0.20.4 for the first ten bases, and were filtered for a quality score greater than 30.

Assembly and analysis of reads

The previously trimmed reads, processed to a quality score of 30, were aligned using Geneious v. 8.0.5 assuming up to 14% mismatches. The sequence of the ToSRV clone (2,591 nt for the DNA-A and 2,568 nt for the DNA-B) was used as the references for alignment of the reads and for detection of nucleotide substitutions along the viral genome.

For base polymorphism detection, an algorithm in Geneious was used to find SNPs along the genome and in each of the coding regions in the two genomic components. Through the output of this tool, it was possible to calculate the frequency, to find the replacement site throughout the genome, the type of substitution and the amino acid changed.

The total number of variations found in each analyzed region divided by the total number of nucleotides was used to calculate the nucleotide substitution rate, assuming that 3% of the variations were due to sequencing errors. The substitution rate was estimated for the complete sequences of the DNA-A and DNA-B and for the coding regions of the CP, Rep, MP and NSP genes.

The plot correlating sequence coverage with the number of mismatches was generated in QtiPlot v. 0.9.8.10.

Shannom entropy plots were generated for each library using a script written by Castro (2015).

Results

Sequence quality

Most libraries showed good sequence quality, except for the first time point (30 dai) in tomato. We opted to filter all libraries for a score of Q30, corresponding to an accuracy of 99.99%, and thus obtaining high homogeneity in all reads (Suppl. Fig. S1). The FastQC analysis indicated that all data sets presented noise in the first ten nucleotides. Thus, we chose to trim the first ten nucleotides to prevent bias in the detection of mutations in the viral genome (Suppl. Fig. S2).

Determination of the maximum number of mismatches allowed in the assembly

Due to the bipartite nature of ToSRV, with two genomic components containing an aprox. 200-bp common region, a strategy was necessary to avoid incorporation of reads originating from the DNA-A in the DNA-B assembly, and vice-versa. Assemblies were tested allowing a maximum of 4, 10, 14, 15, 20 or 25% mismatches with no gaps, and the mapping coverage in all positions of the viral genome was compared (Figure 1).

When up to 25% of mismatches were allowed, a considerable increase in coverage at the common region of both genomics components was observed, without a corresponding increase of the coverage along the remaining regions of the genome (Figure 1). This indicated that alignment of reads from different genomic component was taking place at the common region, with an overestimation of mutations. This was also observed when up to 15% and 20% of mismatches were allowed (Figure 1).

On the other hand, when up to 4, 10 or 14% of mismatches were alllowed, only small increases in coverage were observed, and the topology was the same with no significant increase at the common region (Figure 1). For greater reliability in the

analysis, all sequences mapping to the common region and showing divergence from the reference sequence were submitted to a BLASTn search to verify the genomic component to which these reads showed higher identity.

Based on these observations, we decided to perform the genome assembly allowing up to 14% of mismatches.

Determination of nucleotide substitutions and of the substitution rate

After mapping of all reads to the reference sequence of each genomic component, a significantly higher coverage of the DNA-B compared to the DNA-A was observed, together with a reduction of coverage as the viral infection progressed, in both hosts (Table 1). This indicated that the DNA-B was present at a higher concentration during the course of the viral infection, suggesting a higher replication rate for the DNA-B compared to the DNA-A.

Nucleotide substitutions were quantified for the full-length DNA-A and DNA-B as well as the CP, Rep, MP and NSP genes, allowing gaps in relation to the reference genome (to account for insertions or deletions), up to 14% of mismatches, and "seeds" of 18 bases. The parameters included the number of variations that occurred in each ORF analyzed site by site, as well as the number of synonymous and non-synonymous substitutions, the presence of stop codons, deletions and insertions and the percentage of identical sites at each time point (30, 75 and 120 dai) (Figure 2; Suppl. Tables S1-S4).

From the four genes analyzed, CP and Rep (both in the DNA-A) showed a lower number of variations compared to NSP and MP (encoded by the DNA-B). All four genes showed a consistently higher number of non-synonymous substitutions in relation to synonymous substitutions. Samples from N. physaloides showed higher

number of variations when compared to tomato, for both genomic components. Interestingly, mutations that introduced premature stop codons were observed in all four genes, although they were more frequent in the NSP and MP genes (Figure 2; Suppl. Tables S1-S4). The frequency ranged from 0.1% to 0.2% (mean p-value = 9.3 x 10^{-2}) and 0.04% to 0.1% (mean p-value = 8.57 x 10^{-2}) for the MP gene in N. physaloides and tomato, respectively, and 0.1% to 0.3% (mean p-value = 5.52 x 10^{-2}) and 0.1% to 0.2% (mean p-value = 5.17 x 10^{-3}) for the NSP gene in N. physaloides and tomato, respectively.

In the N. physaloides samples, suppression of the regular stop codons was observed in MP gene due to an insertion (TAA→ATA) at 30, 75 and 120 dai, and due to a transversion (TAA→AAA) at 75 and 120 dai; in the NSP gene at 30 dai (TAA→ATA) and 75 dai (TAA→CAA; →TTA; →TAC); and in the CP gene at 30 and 120 dai (TAA→CAA). In tomato, stop codon suppression was observed only in the CP and MP genes, but they were deleterious over time. These mutations were not observed in the Rep gene, in either host.

In both hosts and for all four genes, the number of variations decreased uniformly in the genome as time passed, with a corresponding increase in the number of identical sites (Figure 2; Suppl. Tables S1-S4).

Substitution rates were estimated for the DNA-A and DNA-B and for the same four genes, based on the data for the 120 dai libraries (Table 2). The calculated substitution rates were in the same order of magnitude for the DNA-A and its two encoded genes (CP and Rep) in tomato and N. physaloides. The DNA-B and its two encoded genes (NSP and MP) showed substitution rates one order of magnitude higher in N. physaloides than in tomato.

Shannon entropy

The complexity, or diversity, of a viral population can be measured using Shannon entropy of a sample of genomes. Nucleotide positions will have a higher entropy if they display a high frequency of variation. Shannon entropy was calculated for each position of the DNA-A and DNA-B at 120 dai, in the two hosts (Figures 3 and 4). For the DNA-A, entropy values were generally low (<0.1) in tomato. In N. physaloides, entropy values of approx. 0.4 were observed in the 5'-half of the common region. For the DNA-B, moderate entropy values (approx. 0.3) were observed in the short intergenic region in both tomato and N. physaloides. The high entropy value in 3'-IR in N. physaloides is most likely due to low coverage detected in this region.

Shannon entropy was also calculated for each host/time point to identify mutation hotspots (Figures 5 and 6). In both hosts, the DNA-A regions encompassing the 3'-end of the Trap gene (nt positions 1300-1500) and the middle of Rep (nt positions 1800-2000nt) were identified as mutation hotspots (Figures 5A and 6A). In N. physaloides, the 5'-end of the CR is a strong mutation hotspot (Figure 6A). In the DNA-B, the two intergenic regions (nt positions 1200-1300 and 2300-2600) were identified as hotspots for mutation (Figures 5B and 6B).

Discussion

Our experiments constitute the first attempt to calculate nucleotide substitution rates for a geminivirus using NGS. We used a virus that has been shown to be prevalent in tomato fields in Brazil for the last ten years (Fernandes et al., 2008; González-Aguilera et al., 2012), and inoculated its commercial host (tomato) as well as a host which has been shown to act as a reservoir in the field (Nicandra

physaloides) (Barbosa et al., 2009). Since our primary objective was to calculate substitution rates, a deliberate choice was made to avoid any amplification of the viral DNA (either by RCA or PCR) before sequencing. While this may have contributed to the relatively low coverage obtained (specially for the DNA-A), it adds confidence to the data, in terms of the observed substitutions being the result of natural processes of mutation. A higher coverage could probably have been achieved if an enrichment step were added to the protocol. Actually, the use of size exclusion columns during DNA extraction was an attempt to enrich for low molecular weight DNA (ie, viral DNA rather than plant DNA). However, it is possible that plant DNA was fragmented and therefore was not excluded.

Geminiviruses have mutation frequencies and nucleotide substitution rates which are similar to those observed for RNA viruses, even though they use the host's proofreading replication machinery which should increase the fidelity of replication (Duffy et al., 2008). Under different selection conditions, three isolates of the mastrevirus Maize streak virus (MSV) showed a substitution rate in the order of 10^{-4} to 10^{-5} sub/site/year, suggesting that these isolates evolve in a quasispecies organization due their high sequence heterogeneity (Isnard et al., 1998). Analyzing the genetic variability of the monopartite begomovirus Tomato yellow leaf curl China virus (TYLCCNV), Ge et al. (2007) reported mutation frequencies of 3.5×10^{-4} and 5.3×10^{-4} in Nicotiana benthamiana and tomato, respectively, after 60 days of infection, and also concluded that it evolves as a quasispecies. Using Bayesian coalescence-based analysis, Duffy and Holmes (2008) found substitution rates of 2.88×10^{-4} sub/site/year for Tomato yellow leaf curl virus (TYLCV). By analyzing the nucleotide substitution rate of the bipartite EACMV, assuming that the two genome segments have distinct evolutionary histories, Duffy and Holmes (2009)

found average rates of 1.60 x 10^{-3} and 1.33 x 10^{-4} sub/site/year for the DNA-A and DNA-B, respectively.

Our substitution rates results are compatible with those listed above, and consistent with the EACMV model of the two genomic components evolving at different rates. The ToSRV DNA-A, in both tomato and N. physaloides, displayed substitution rates similar to those reported for other members of the family Geminiviridae. Interestingly, the substitution rate of the DNA-B and its coding regions displayed values that were one order of magnitude higher in N. physaloides compared to tomato. It is noteworthy that the observed values in tomato are similar to the values found for the DNA-B of East Africa cassava mosaic virus (EACMV), the only other bipartite begomovirus for which substitution rates were estimated (Duffy and Holmes, 2009). It has been suggested that viruses infecting non-cultivated hosts may evolve faster than those infecting cultivated hosts (Rocha et al., 2013). It is possible that, by having a large genetic base, N. physaloides may be capable of mounting more efficient defense responses that act specifically on the DNA-B, placing a higher selective pressure and thus leading to the incorporation of mutations until a variant emerges which escapes these defense responses.

In all four coding regions analyzed, nucleotide substitutions which incorporated premature stop codons were verified. The frequency of these substitutions was much higher in the genes encoded by the DNA-B than in those encoded by the DNA-A. However, the frequency decreased over time. These premature termination codons lead to the translation of truncated proteins which will, most likely, be defective. Thus, it is possible that the DNA components harboring these mutations are defective DNAs. By analyzing the degree of heterogeneity in structured quasispecies populations of Hepatitis C virus (HCV) (a virus with a

single-stranded RNA genome), Martell et al. (1992) observed that 10% of the sequences included premature termination codons, indicating that a significant fraction of particles in circulation should contain defective genomes.

Interestingly, suppression of the "normal" stop codons in the MP and NSP genes (and to a lesser extent also in the CP gene) was observed during infection in N. physaloides. Zaccomer et al. (1995) showed that RNA viruses use suppression of termination codons as a gene expression strategy, in addition to divided genomes, subgenomic RNAs, frame-shifting, etc. The possible fixation of a codon for an amino acid in place of the termination codon in the MP and NSP gene during infection in N. physaloides may reflect the necessity of the virus to generate variability so that the two movement proteins can interact efficiently with host factors, facilitating cell-to-cell movement. Regardless, this observation suggests that these variants may present a selective advantage. The ability to rapidly generate diversity has been proposed as one of the reasons that make viruses capable of adapting to new ecological niches (Roossinck, 1997).

A notable observation in our study is that the number of variations decreased during the course of the experiment, suggesting the presence of bottlenecks during the systemic infection. A similar observation was reported in a study where 12 experimental mutants of Cucumber mosaic virus (CMV) were inoculated in tobacco, and a reduction of the number of mutants was observed in successive leaves as a function of distance from the source (Li and Roossinck, 2004). However, in a study which used NGS, an increase of Zucchini yellow mosaic virus (ZYMV) variants was observed in successive leaves of Curcubita pepo (Dunham et al., 2014). The use of NGS may provide greater visibility to the bottleneck phenomenon in a viral population, as minor variants are sampled, giving better support to the analysis.

References

Agra, M., Rocha, E., Formiga, S., Locatelli, E. 1994. Plantas medicinais dos Cariris Velhos, Paraíba. Parte I: Subclasse Asteridae. Rev Bras Farm 75:61-64.

Albuquerque, L.C., Varsani, A., Fernandes, F.R., Pinheiro, B., Martin, D.P., Oliveira Ferreira, P.d.T., Lemos, T.O., Inoue-Nagata, A.K. 2012. Further characterization of tomato-infecting begomoviruses in Brazil. Arch Virol 157:747-752.

Andrade, E.C., Manhani, G.G., Alfenas, P.F., Calegario, R.F., Fontes, E.P.B., Zerbini, F.M. 2006. Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from Sida sp., forms pseudorecombinants with begomoviruses from tomato but not from Sida. J Gen Virol 87:3687-3696.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aragão, F.J.L., Barros, L.M.G., Brasileiro, A.C.M., Ribeiro, S.G., Smith, F.D., Sanford, J.C., Faria, J.C., Rech, E.L. 1996. Inheritance of foreign genes in transgenic bean (Phaseolus vulgaris L.) co-transformed via particle bombardment. Theor Appl Genet 93:142-150.

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

Brown, J.K., Fauquet, C.M., Briddon, R.W., Zerbini, F.M., Moriones, E., Navas-Castillo, J. 2012. Family Geminiviridae. Pages 351-373 in: Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses, A.M.Q. King, M.J. Adams, E.B. Carstens, E.J. Lefkowitz, eds. Elsevier Academic Press, London, UK.

Calegario, R.F., Ferreira, S.S., Andrade, E.C., Zerbini, F.M. 2007. Characterization of Tomato yellow spot virus, (ToYSV), a novel tomato-infecting begomovirus from Brazil. Pesq Agropec Bras 42:1335-1343.

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

Castro, G.M. (2015). Identificação da quasispecies Papaya ringspot virus em uma biblioteca de cDNA de Fevillea cordifolia. M.Sc. Dissertation, Instituto de Biologia, Universidade Estadual de Campinas. Campinas, SP.

Costa, A.S. 1955. Studies on Abutilon mosaic in Brazil. J Phytopath 24:97-112.

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

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

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

Duffy, S., Shackelton, L.A., Holmes, E.C. 2008. Rates of evolutionary change in viruses: Patterns and determinants. Nat Rev Genet 9:267-276.

Dunham, J.P., Simmons, H.E., Holmes, E.C., Stephenson, A.G. 2014. Analysis of viral (Zucchini yellow mosaic virus) genetic diversity during systemic movement through a Cucurbita pepo vine. Virus Res 191:172-179.

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

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

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

Flores, E., Silberschmidt, K., Kramer, M. 1960. Observações de "clorose infecciosa" das malváceas em tomateiros do campo. O Biológico 26:65-69.

García-Arenal, F., Fraile, A., Malpica, J.M. 2003. Variation and evolution of plant virus populations. Int Microbiol 6:225-232.

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

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

Inoue-Nagata, A.K., Albuquerque, L.C., Rocha, W.B., Nagata, T. 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met 116:209-211.

Isnard, M., Granier, M., Frutos, R., Reynaud, B., Peterschmitt, M. 1998. Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. J Gen Virol 79:3091-3099.

Li, H.Y., Roossinck, M.J. 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. J Virol 78:10582-10587.

Lima, A.T.M. 2007. Caracterização de dois begomovírus (Tomato severe rugose virus e Tomato yellow vein streak virus) que infectam tomateiro e obtenção de clones infecciosos. M.Sc. Dissertation, Dep. de Fitopatologia, Universidade Federal de Viçosa. Viçosa, MG.

Lima, A.T.M., Sobrinho, R.R., Gonzalez-Aguilera, J., Rocha, C.S., Silva, S.J.C., Xavier, C.A.D., Silva, F.N., Duffy, S., Zerbini, F.M. 2013. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. J Gen Virol 94:418-431.

Martell, M., Esteban, J.I., Quer, J., Genescà, J., Weiner, A., Esteban, R., Guardia, J., Gomez, J. 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: Quasispecies nature of HCV genome distribution. J Virol 66:3225-3229.

Moffat, A.S. 1999. Geminiviruses emerge as serious crop threats. Science 286:1835-1835.

Monci, F., Sanchez-Campos, S., Navas-Castillo, J., Moriones, E. 2002. A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. Virology 303:317-326.

Navas-Castillo, J., Fiallo-Olivé, E., Sánchez-Campos, S. 2011. Emerging virus diseases transmitted by whiteflies. Annu Rev Phytopath 49:219-248.

Orlando, A., Silberschmidt, K. 1946. Studies on the natural dissemination of the virus of "infectious chlorosis" of the Malvaceae and its relation with the insect vector Bemisia tabaci Genn. Arq Inst Biol 17:1-36.

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

Padidam, M., Beachy, R.N., Fauquet, C.M. 1999b. A phage single-stranded DNA (ssDNA) binding protein complements ssDNA accumulation of a geminivirus and interferes with viral movement. J Virol 73:1609-1616.

Paximadis, M., Idris, A.M., Torres-Jerez, I., Villarreal, A., Rey, M.E.C., Brown, J.K. 1999. Characterization of tobacco geminiviruses in the Old and New Worlds. Arch Virol 144:703-717.

Prasanna, H.C., Sinha, D.P., Verma, A., Singh, M., Singh, B., Rai, M., Martin, D.P. 2010. The population genomics of begomoviruses: Global scale population structure and gene flow. Virol J 7:220.

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

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

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

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

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

Roossinck, M.J. 1997. Mechanisms of plant virus evolution. Annu Rev Phytopath 35:191-209.

Rybicki, E.P. 1994. A phylogenetic and evolutionary justification for three genera of Geminiviridae. Arch Virol 139:49-77.

Seal, S.E., Van den Bosch, F., Jeger, M.J. 2006. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Crit Rev Pl Sci 25:23-46.

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

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

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

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J.K., Zerbini, F.M., Martin, D.P. 2014. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. Arch Virol 159:2193-2203.

Zaccomer, B., Haenni, A.-L., Macaya, G. 1995. The remarkable variety of plant RNA virus genomes. J Gen Virol 76:231-247.

Table 1. Total reads mapped to the Tomato severe rugose vrius (ToSRV) genome in the tomato and Nicandra physaloides libraries at each time point (30, 75 and 120 days after inoculation, dai).

	Tomato		**N. physaloides**	
Dai	DNA-A	DNA-B	DNA-A	DNA-B
30	6.658	44.563	5.176	32.723
75	4.686	32.238	4.271	33.271
120	3.129	20.437	2.665	21.215

Table 2. Substitution rates in the full-length (FL) genomic components (DNA-A and DNA-B) and in the CP, Rep, NSP and MP genes of ToSRV in **(A)**N. physaloides and **(B)** tomato.

A

	FL DNA-A	**CP**	**Rep**	**FL DNA-B**	**NSP**	**MP**
Number of variations	254	33	57	1,333	254	267
Number of variations (corrected)	246.38	32.01	55.29	1293.01	246.38	258.99
Nucleotides sequenced	245,106	64,546	91,860	1,933,133	531,472	700,699
Substitution rate (sub/site/year)	1.01×10^{-3}	5.0×10^{-4}	6.02×10^{-4}	6.69×10^{-4}	4.64×10^{-4}	3.7×10^{-4}
Substitution rate (estimated) (sub/site/year)	3.06×10^{-3}	1.51×10^{-3}	1.83×10^{-3}	2.03×10^{-3}	1.41×10^{-3}	1.12×10^{-3}

B

	FL DNA-A	**CP**	**Rep**	**FL DNA-B**	**NSP**	**MP**
Number of variations	134	26	43	548	144	157
Number of variations (corrected)	129.98	25.22	41.71	531.56	139.68	152.29
Nucleotides sequenced	287,330	74,776	108,764	1,862,335	527,879	699,655
Substitution rate (sub/site/year)	4.52×10^{-4}	3.37×10^{-4}	3.84×10^{-4}	2.85×10^{-4}	2.65×10^{-4}	2.17×10^{-4}
Substitution rate (estimated) (sub/site/year)	1.38×10^{-3}	1.03×10^{-3}	1.17×10^{-3}	8.68×10^{-4}	8.05×10^{-4}	6.62×10^{-4}

Figure legends

Figure 1. Coverage of the viral genome assuming different values of mismatches.

Figure 2. Variability observed in the ToSRV genes (CP, Rep, MP, NSP) after biolistic inoculation of tomato and Nicandra physaloides. **(A)** Total number of variations, **(B)** Percentage of identical sites, **(C)** Total number of synonymous substitutions, **(D)** Total number of non-synonymous substitutions. dai, days after inoculation.

Figure 3. Shannon entropy for the ToSRV DNA-A **(A)** and DNA-B **(B)** in tomato at 120 days after inoculation.

Figure 4. Shannon entropy for the ToSRV DNA-A **(A)** and DNA-B **(B)** in Nicandra physaloides at 120 days after inoculation.

Figure 5. Shannon entropy for the ToSRV DNA-A **(A)** and DNA-B **(B)** at each time point (30, 75 and 120 days after inoculation) in tomato.

Figure 6. Shannon entropy for the ToSRV DNA-A **(A)** and DNA-B **(B)** at each time point (30, 75 and 120 days after inoculation) in Nicandra physaloides.

Supplementary Figure S1. Box-plots of the quality of bases and their positions in the reads for the sample of N. physaloides collected 30 days after inoculation. On the

left, reads R1 (+) and on the right, reads R2 (-). **(A)** Unprocessed for filtering quality, **(B)** processed for filtering quality for score Q30.

Supplementary Figure S2. Sequence content among reads for the sample of N. physaloides with 30 dai. Above, reads R1 (+) without trimming of the first ten nucleotides and below, reads R1 after trimming of the first ten nucleotides.

Figure 1

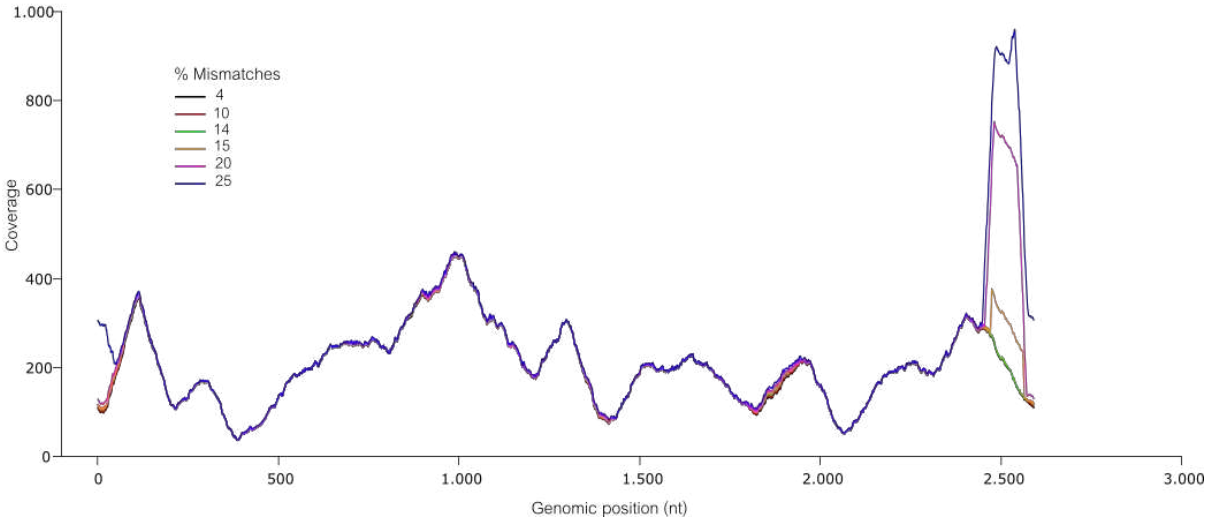

1.000
800
600
400
200
0
Coverage
% Mismatches
4
10
14
15
20
25
0
500
1.000
1.500
2.000
2.500
3.000
Genomic position (nt)

Figure 2

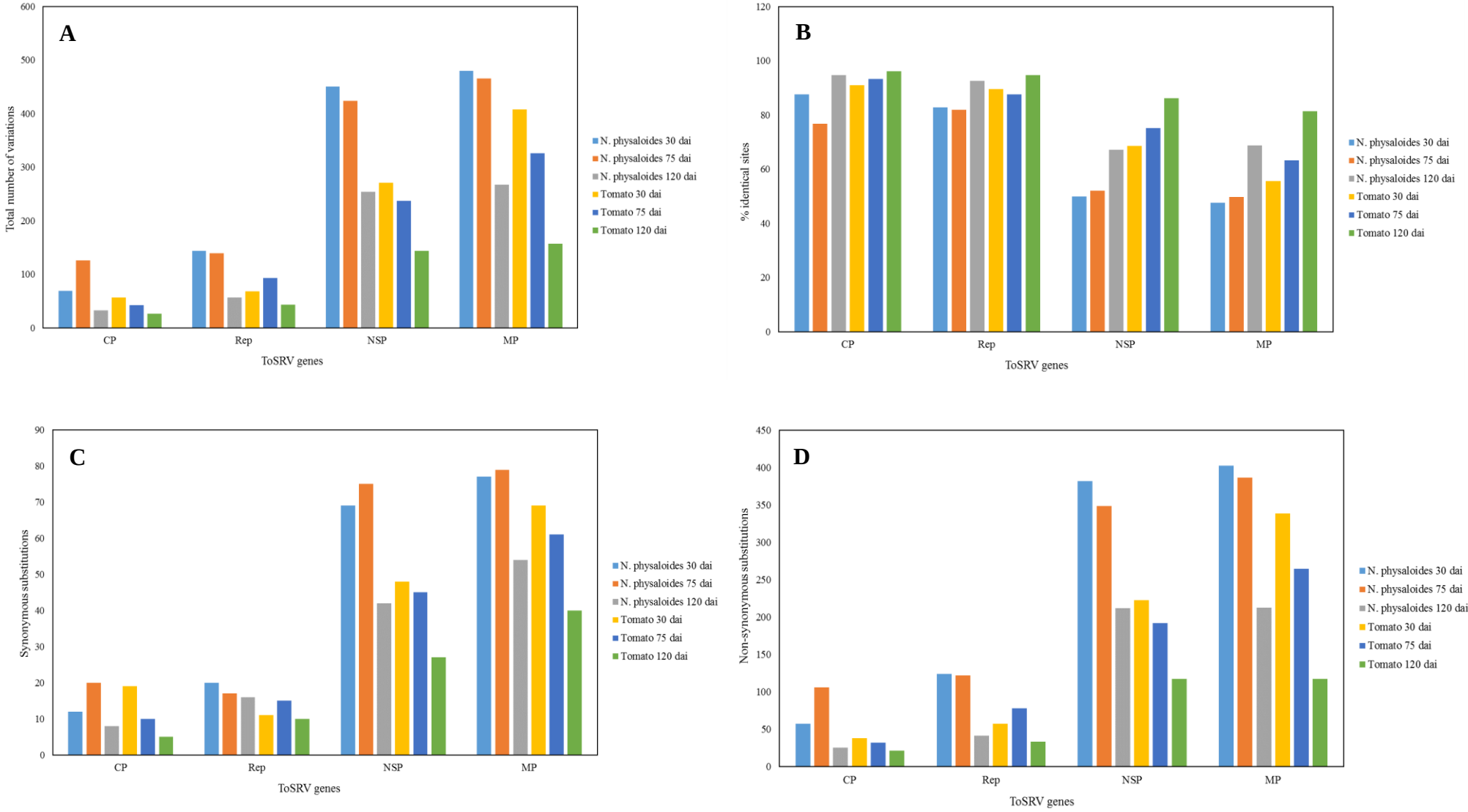
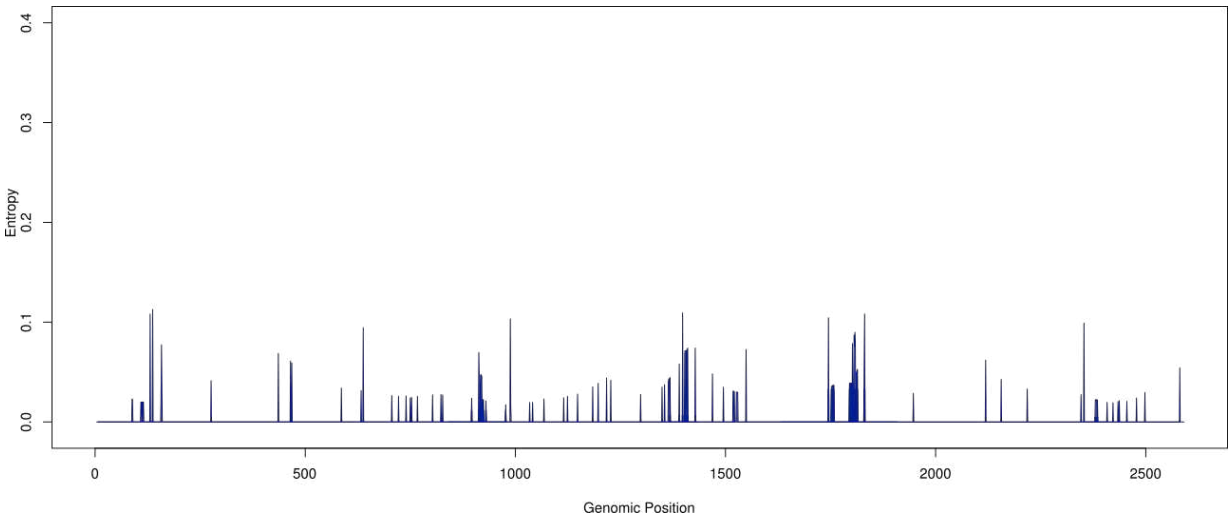

A
Total number of variations
600
500
400
300
200
100
0
CP
Rep
NSP
MP
ToSRV genes
N. physaloides 30 dai
N. physaloides 75 dai
N. physaloides 120 dai
Tomato 30 dai
Tomato 75 dai
Tomato 120 dai
B
% identical sites
120
100
80
60
40
20
0
CP
Rep
NSP
MP
ToSRV genes
N. physaloides 30 dai
N. physaloides 75 dai
N. physaloides 120 dai
Tomato 30 dai
Tomato 75 dai
Tomato 120 dai
C
Synonymous substitutions
90
80
70
60
50
40
30
20
10
0
CP
Rep
NSP
MP
ToSRV genes
N. physaloides 30 dai
N. physaloides 75 dai
N. physaloides 120 dai
Tomato 30 dai
Tomato 75 dai
Tomato 120 dai
D
Non-synonymous substitutions
450
400
350
300
250
200
150
100
50
0
CP
Rep
NSP
MP
ToSRV genes
N. physaloides 30 dai
N. physaloides 75 dai
N. physaloides 120 dai
Tomato 30 dai
Tomato 75 dai
Tomato 120 dai

Figure 3

A


Entropy
0.0
0.1
0.2
0.3
0.4
0
500
1000
1500
2000
2500
Genomic Position

B

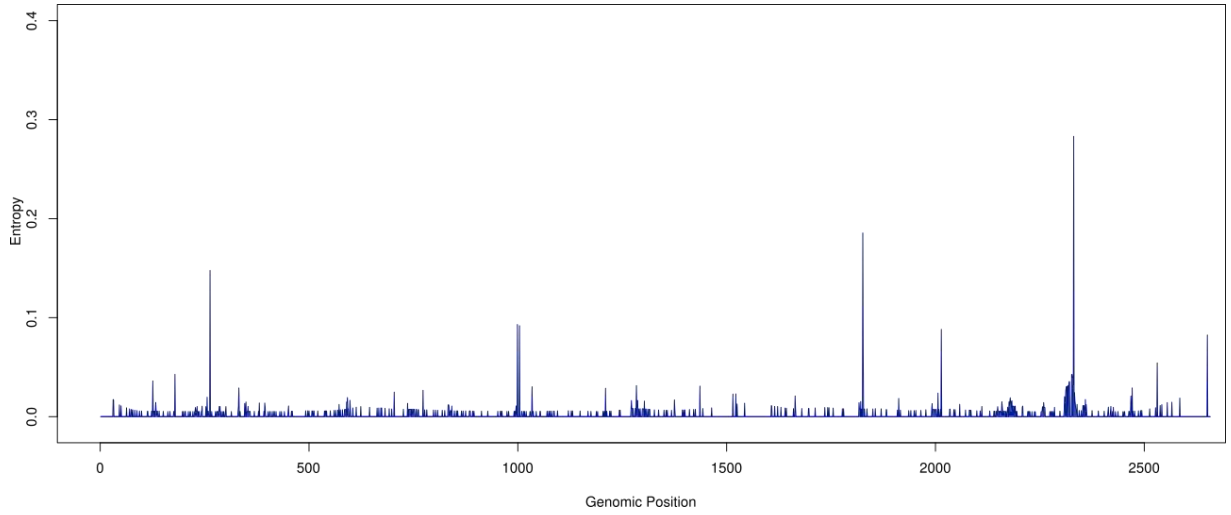
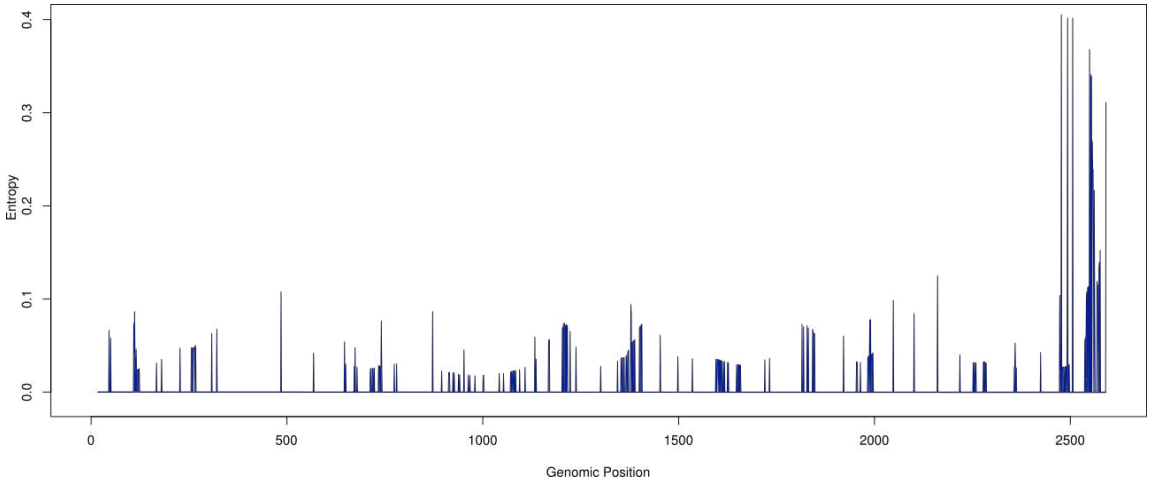

Entropy
0.0
0.1
0.2
0.3
0.4
0
500
1000
1500
2000
2500
Genomic Position

Figure 4

A


Entropy
0.0
0.1
0.2
0.3
0.4
0
500
1000
1500
2000
2500
Genomic Position

B

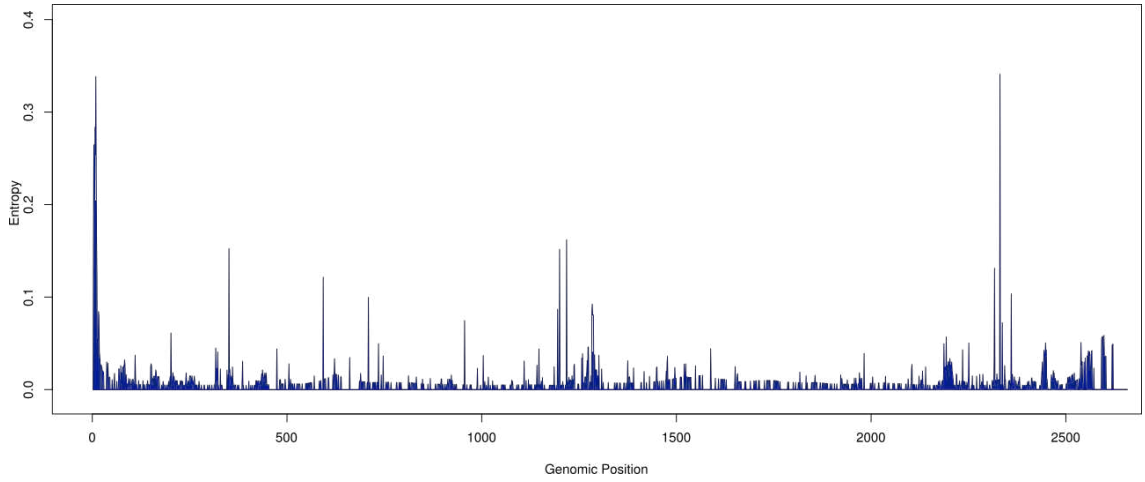
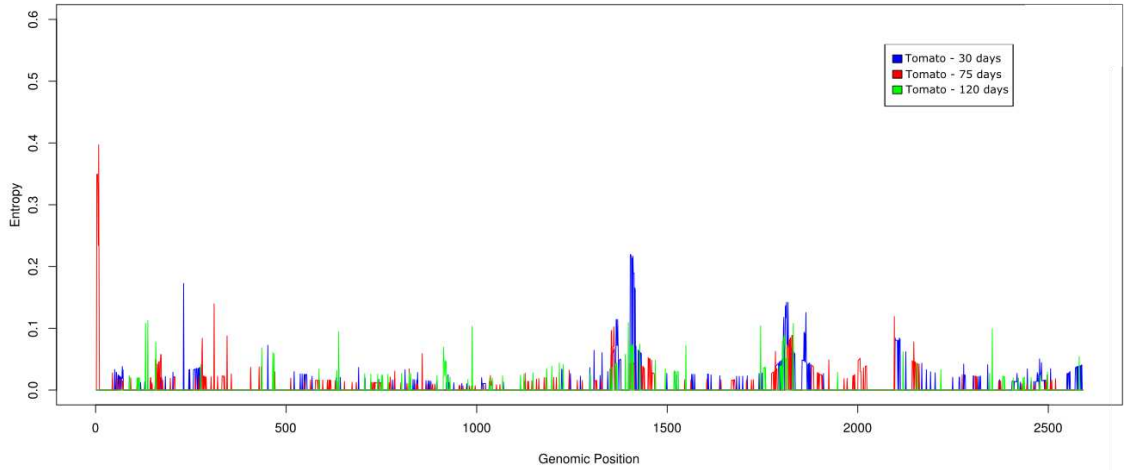

Entropy
0.0
0.1
0.2
0.3
0.4
0
500
1000
1500
2000
2500
Genomic Position

Figure 5

A


Tomato - 30 days
Tomato - 75 days
Tomato - 120 days
Entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
500
1000
1500
2000
2500
Genomic Position

B

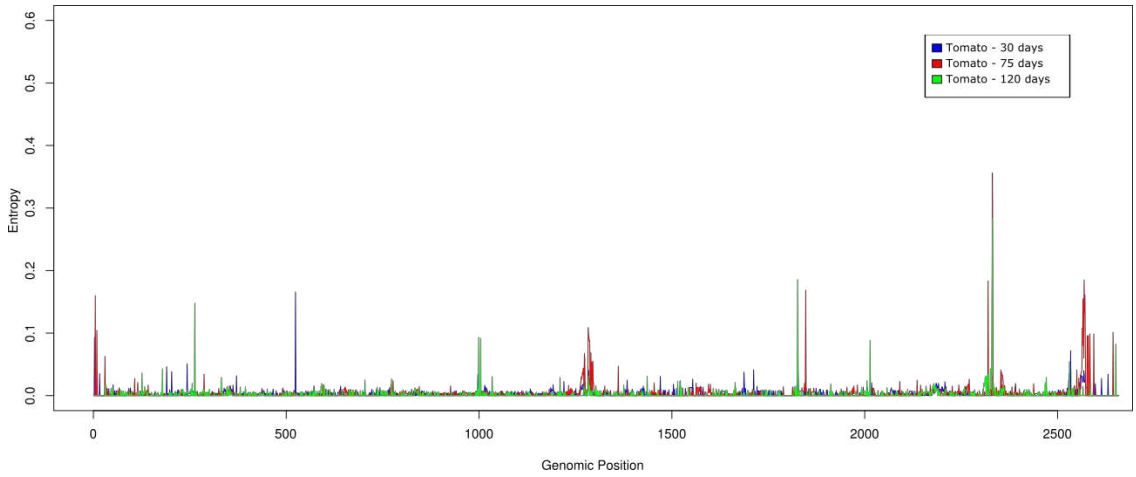
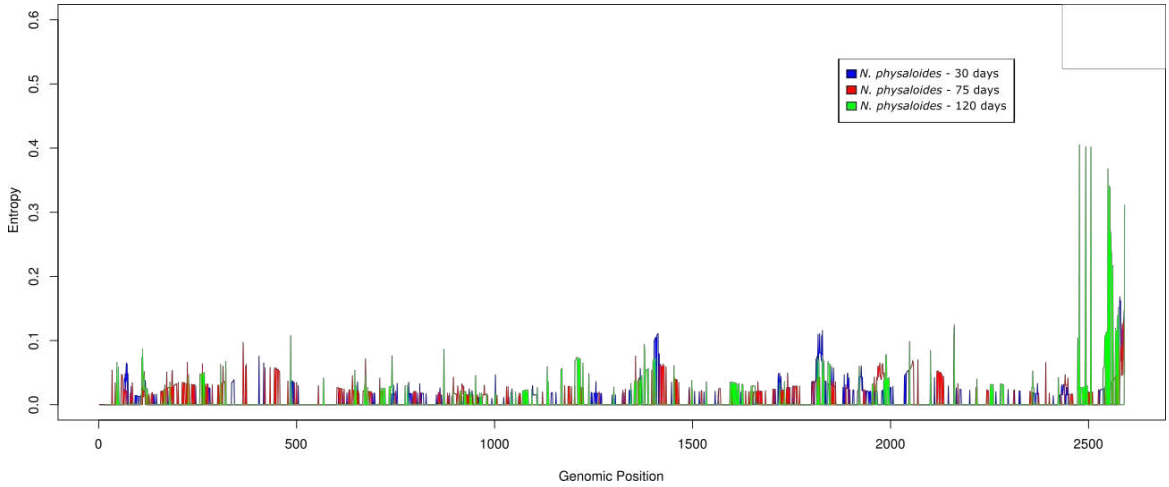
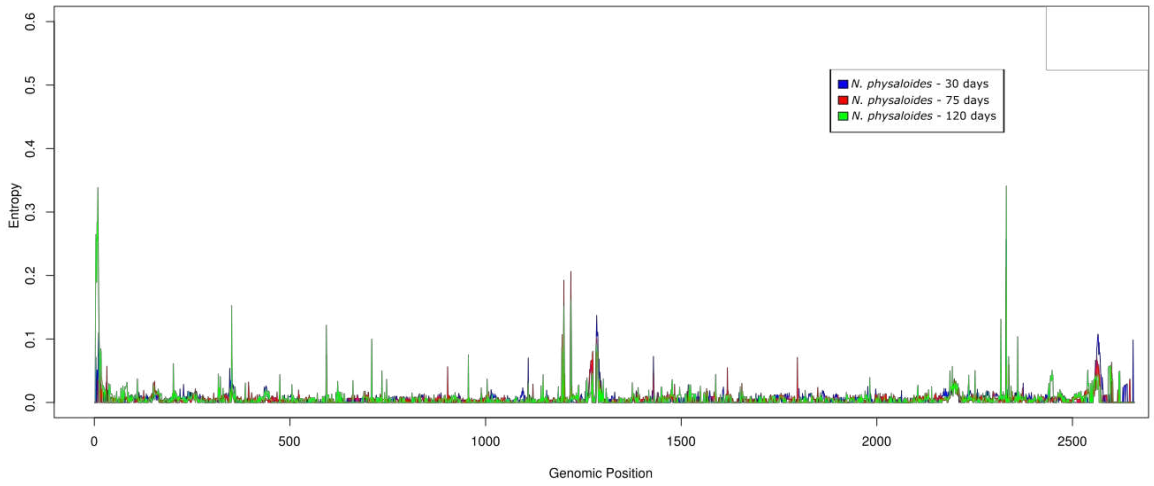

Tomato - 30 days
Tomato - 75 days
Tomato - 120 days
Entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
500
1000
1500
2000
2500
Genomic Position

Figure 6

A


N. physaloides - 30 days
N. physaloides - 75 days
N. physaloides - 120 days
Entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
500
1000
1500
2000
2500
Genomic Position

B


N. physaloides - 30 days
N. physaloides - 75 days
N. physaloides - 120 days
Entropy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
500
1000
1500
2000
2500
Genomic Position

Supplementary Table S1. Variability observed in the ToSRV CP gene after biolistic inoculation of tomato and Nicandra physaloides, at 30, 75 and 120 days after inoculation (dai).

	N. physaloides			**Tomato**		
	30 dai	**75 dai**	**120 dai**	**30 dai**	**75 dai**	**120 dai**
Total number of variations	69	126	33	57	42	26
Synonymous substitutions	12	20	8	19	10	5
Non-synonymous substitutions	57	106	25	38	32	21
Transitions	19	31	12	22	16	8
Transversions	20	34	13	20	17	13
Deletions	1	5	3	2	3	2
Insertions	2	2	0	1	0	0
Stop codons	2	6	2	1	0	2
% identical sites	87.6	76.8	94.8	91	93.4	96.2

Supplementary Table S2. Variability observed in the ToSRV Rep gene after biolistic inoculation of tomato and Nicandra physaloides, at 30, 75 and 120 days after inoculation (dai).

	N. physaloides			**Tomato**		
	30 dai	**75 dai**	**120 dai**	**30 dai**	**75 dai**	**120 dai**
Total number of variations	144	139	57	68	93	43
Synonymous substitutions	20	17	16	11	15	10
Non-synonymous substitutions	124	122	41	57	78	33
Transitions	30	28	21	18	24	15
Transversions	54	46	15	15	27	13
Deletions	6	2	2	1	7	2
Insertions	4	0	2	0	0	0
Stop codons	4	6	1	4	2	1
% identical sites	82.8	82	92.7	89.6	87.7	94.7

Supplementary Table S3. Variability observed in the ToSRV NSP gene after biolistic inoculation of tomato and Nicandra physaloides, at 30, 75 and 120 days after inoculation (dai).

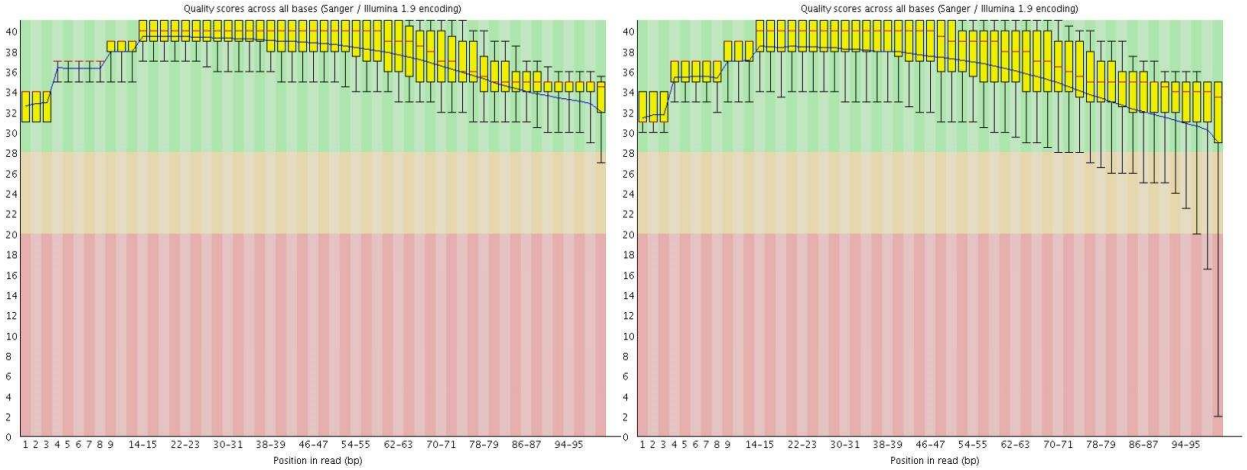
	N. physaloides			**Tomato**		
	30 dai	**75 dai**	**120 dai**	**30 dai**	**75 dai**	**120 dai**
Total number of variations	451	424	254	271	237	144
Synonymous substitutions	69	75	42	48	45	27
Non-synonymous substitutions	382	349	212	223	192	117
Transitions	125	132	84	101	91	58
Transversions	151	147	96	103	87	46
Deletions	14	18	10	3	14	7
Insertions	7	5	3	3	5	0
Stop codons	23	15	16	12	12	4
% identical sites	49.9	52.1	67.2	68.6	75.2	86.3

Supplementary Table S4. Variability observed in the ToSRV MP gene after biolistic inoculation of tomato and Nicandra physaloides, at 30, 75 and 120 days after inoculation (dai).

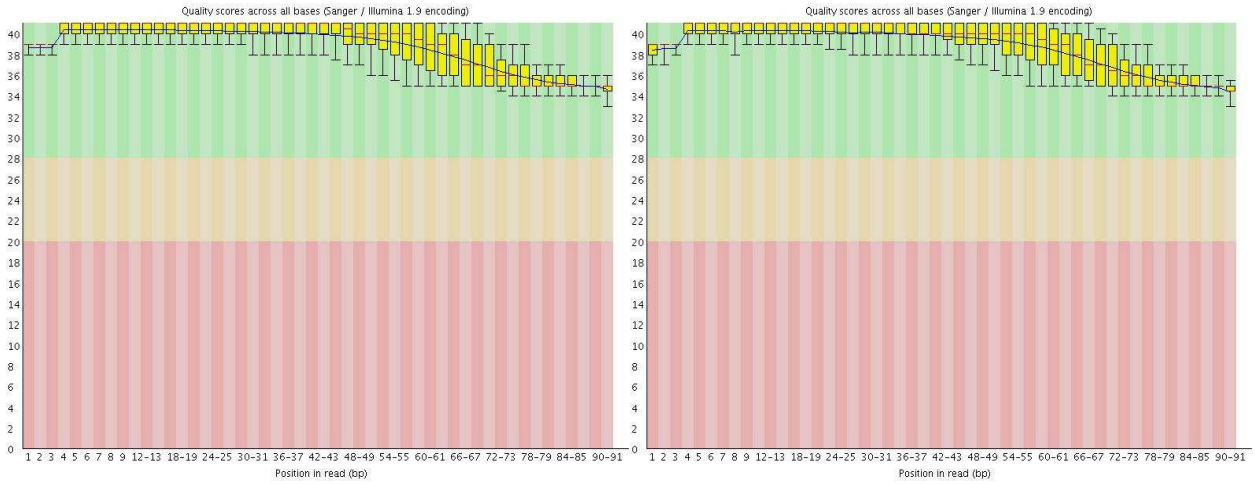
	N. physaloides			**Tomato**		
	30 dai	**75 dai**	**120 dai**	**30 dai**	**75 dai**	**120 dai**
Total number of variations	480	466	267	408	326	157
Synonymous substitutions	77	79	54	69	61	40
Non-synonymous substitutions	403	387	213	339	265	117
Transitions	149	131	82	143	121	63
Transversions	151	208	123	158	144	72
Deletions	23	14	8	13	15	8
Insertions	11	9	2	5	2	0
Stop codons	29	20	15	19	16	5
% identical sites	47.6	49.8	68.7	55.7	63.3	81.4

Supplementary Figure S1

A



B



Supplementary Figure S2