Virus genes: short communication

Geographically distant isolates of the persistent southern tomato virus (STV) show very low genetic diversity in the putative coat protein gene

Elvira-González L.[1,2], Rubio L.[1], Galipienso L.[1]

1. Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Valencia
2. Universitat Politècnica de València (UPV), Valencia

Corresponding author: galipienso_lui@gva.es

Instituto Valenciano de Investigaciones Agrarias (IVIA). CV-315, Km 10. Moncada, Valencia 46113

Abstract: Southern tomato virus (STV) from genus *Amalgavirus* (Family *Amalgaviridae*) is a persistent virus infecting tomato crops worldwide. Information on genetic diversity and evolutionary mechanisms for plant persistent viruses are very scarce in comparison with plant acute viruses. In this work, the putative coat protein gene of worldwide STV isolates was analyzed showing very low nucleotide diversity (< 0.0100). Phylogenetic analysis separated STV isolates into two clades, but no correlation was found between genetic and geographic distances. Also, no recombination events among STV isolates were detected. Comparison of synonymous and nonsynonymous substitutions indicated negative selection at the amino acid level.

Keywords: Southern tomato virus, *Amalgaviridae*, genetic diversity, phylogeny, evolution

Declarations

Funding. This research was supported by the INIA research project E-RTA2014-00010-C02 co-funded by FEDER 2014-2020 funds.

Conflict of interest. The authors declare that they have no conflict of interest.

Availability of data and material. Supplementary material Table S1 and Table S2 are available in virus genes.

Ethical approval. This article does not contain any studies with animals performed by any of the authors.

Introduction

Southern tomato virus (STV) is a persistent virus belonging to the genus *Amalgavirus* (family *Amalgaviridae)*. STV is widespread infecting tomato crops in different production areas in USA, Mexico, France, Italy, China, Bangladesh and Spain [1–6]. STV has been detected in both symptomatic and asymptomatic tomato plants with high incidence in peninsular Spain and Great Canary Island [7,8]. STV is transmitted vertically by seeds with rates of about 80% whereas no horizontal transmission has been reported [1,8]. Relationships between persistent plant viruses and their hosts have been poorly studied since most of these viruses did not induce any disease. Some persistent viruses establish mutualistic rather parasitic interactions, typical of acute viruses, with their hosts. For example, white clover cryptic virus 1 which modulates the nodulation process in clover plants and curvularia thermal tolerance virus enabling the fungus *Curvularia protuberata* to confer thermal tolerance to panic grass (*Panicum* sp.) so it can grow in soils with temperatures over 50°C [9,10]. For STV, beneficial effects in the tomato host have been described [11] although without statistical support. A recent report showed changes of microRNAs populations, involved in many developmental processes and stress responses, in STV-infected tomato plants even though no ultrastructural and macroscopy changes were observed [12].

STV has a small double-stranded RNA genome of 3.5 kb with two overlapping open reading frames. STV genome organization is similar to that of the family *Totiviridae*. The 5' encode for a putative coat protein (CP) named p42 and the 3' encode for a RNA-dependent RNA-polymerase (RdRp) as a fusion protein using a +1 ribosomal frameshift [13]. Phylogenetic analysis of the STV RdRp region showed that the family *Amalgaviriridae* is evolutionary closer to the family *Partiviridae.* [1]. However, the STV p42 did not show sequence similarities with the CP of viruses of the families *Totiviridae* or *Partitiviridae*, but it showed structural homologies with the CP of members of the genus *Tenuivirus* [1,14]. All this suggests that the *Amalgavirus* ancestor could have emerged from a recombination taking the *Partivirus* RdRp and a *Tenuivirus* CP [14].

Analysis of genetic diversity and evolution of plant viruses is crucial to understand their epidemiology [15] and develop accurate methods for virus detection and implement efficient measures of disease control [16]. Studies on genetic variability of persistent viruses are very scarce in contrast to the abundance of information regarding on acute viruses [17,18]. In this work, the nucleotide diversity of the putative CP of STV isolates from two regions of Spain was determined and compared with that from worldwide STV isolates. Phylogeny and evolution forces driving STV populations were studied.

Materials and methods

Eleven STV isolates from Valencian Community (Spain) and four isolates from Great Canary Island were collected (Supplementary Table S1). Total RNA was extracted from STV-infected plants by using a phenol/chloroform method [19]. The cDNA was synthesized from RNA extracts by reverse

transcription (RT) using SuperScript IV Reverse Transcriptase Kit (Thermofisher) with Random primers. The complete p42 gen (putative CP) was amplified by PCR using High fidelity polymerase Kit (Bio-rad) with the primers p42_F 5'GTC AGA TTT CTC GTC GTT GCT T'3, position 68-90 and p42_R 5'CGT GAC CGC GAG AAT GGA ATA G'3, position 1289-1311. RT-PCR products were purified by using the QIAquick® PCR Purification Kit (Quiagen). The consensus nucleotide sequences of the amplification products were determined in both senses using an ABI 377 DNA sequencer (PerkinElmer) and deposited in GenBank (MK026630 to MK026644). In addition, the nucleotide sequences of p42 from STV isolates from Mexico, China, South Korea, Dominican Republic, United Kingdom, Switzerland, Bangladesh, Spain and United States were retrieved from GenBank (Supplementary Table S1). The nucleotide sequences were aligned at the amino acid level using the program CLUSTAL W 2.0 [20]. The nucleotide substitution model that best fitted the sequence data was the Tamura-Nei model [21] which was used to assess nucleotide distances between STV isolates and to infer their phylogenetic relationships by the Maximum-likelihood (ML) method with 500 bootstrap replicates [22]. The role of natural selection at the molecular level was evaluated by comparing the rate of nonsynonymous substitutions per nonsynonymous site (dN) and the rate of synonymous substitutions per synonymous site (dS) according to the Pamilo-Bianchi-Li method [25]. All these analyses were performed with MEGA X [21]. Selection across the genomic coding regions was studied by estimation of the rates of dN and dS at each codon using the fixed effects likelihood method [28] implemented in the Datamonkey Server http://www.datamonkey.org/ [29]. Also, STV population evolution was studied using the Tajima's D [30], Fu and Li's D* and F* [31] neutrality test. Recombination among STV isolates was assessed by the program GARD [32] from the Datamonkey package and RDP v.4.97 [33]

Results and discussion

The CP coding sequence of STV showed low nucleotide variation, being the proportion of segregating sites 0.0212. Nucleotide diversity of the two STV populations analyzed in Spain was 0.0007 ± 0.0004 and 0.0006 ± 0.0003 for Great Canary Island and Valencia Community, which were similar to the nucleotide diversity between both populations, 0.0006 ± 0.0007 and about five times lower than that for the worldwide STV population, 0.0032 ± 0.0007. The genetic diversity of STV CP gene was compared with that of other persistent viruses phylogenetically related belonging to the families *Amalgaviridae*, *Partitiviridae* and *Totiviridae* [1], as well as with the most important acute viruses infecting tomato and members of the genus *Tenuivirus* given the structural homology with STV CP (Table 1 and Supplementary Table S2). The nucleotide sequences encoding for the CP from the different virus isolates were retrieved from the GenBank (Supplementary Table S2) and the nucleotide substitution model that best fitted the data for each virus was determined. In cases of viruses with too many sequences in GenBank, a manageable number of sequences from different geographical origins was randomly selected. Regarding the persistent viruses, the other member of the genus *Amalgavirus*, with enough available sequences, blueberry latent virus, showed also very low nucleotide diversity (0.0012 ± 0.0003). The six viruses of the family *Partiviridae* infecting plants, fungi and protozoa had low nucleotide

diversity ranging from 0.0058 ± 0.0017 to 0.0339 ± 0.0058. However, two members of the family *Totiviridae* showed high nucleotide diversity (> 0.100). Regarding the acute viruses, the nucleotide diversity was variable ranging from 0.0090 ± 0.0019 to 0.2846 ± 0.0402. The two members of the genus *Tenuivirus* infecting rice plants showed a nucleotide diversity of 0.0202 ± 0.0030 and 0.0338 ± 0.0038.

The phylogenetic tree of STV showed two clades or groups of STV isolates: Group A with isolates from Spain, Mexico, UK, Dominican Republic, USA, South Korea Bangladesh and China; and Group B with isolates from Switzerland and China (Figure 1 and Supplementary Table S3). The mean nucleotide distance between both groups was 0.0112 ± 0.0028 whereas within Group A and Group B were 0.0012 ± 0.0004 and 0.0018 ± 0.0010, respectively. No correlation between the geographic and nucleotide distance was found and some isolates from different countries had identical nucleotide sequences such as Florida, Mexico-1, DR (From Dominican Republic) or some Spanish isolates (Supplementary Table S3)

. ~~This absence of geographic structure has been observed also in populations of other plant viruses [23,24] and might be due to worldwide trade in tomato seeds and/or a negative selection .~~

~~The fact that some geographically distant STV isolates had identical or very similar nucleotide sequences might be consequence of a recent and rapid spread of STV-infected seeds. and/or a strong negative selection.~~

With regard on the role of natural selection at the STV evolution at molecular level, values of dN and dS were 0.0009 and 0.0072, respectively, and the value of the ratio dN/dS was 0.1241 indicating negative selection for amino acid change due to protein functional constrains (Supplementary Table S4). The value of dN/dS obtained for STV is in the range of most plant viruses. However, the value of dS (0.0009) of STV was much lower than most plant viruses, such as citrus leaf blotch virus (CLBV) and potato virus V (PVV), having dS values of 0.062 and 0.052, respectively) [17] (Vives et al., 2002). This result suggested a strong negative selection not only at amino acid level but also at nucleotide level which might due to secondary structure constrains [25] and/or codon usage bias [26]. This is supported by the fact that some isolates with identical sequence were collected in different years (Supplementary Table S1) and explains the genetic stability observed. Analysis of the natural selection pressure across the genomic coding regions showed that the positions 31, 36, 53, 138, 157 and 373 were under negative selection, whereas no codon was under positive selection. These sites could be involved in functional or structural domains. At the population level, analysis using the Tajima's D [27], Fu and Li's D* and F* neutrality test [28] gave negative values (D = -1.52041, D* = -0.41803 and F* = -0.90146) suggesting negative natural selection, although they were non-significant, which could be due to low statistical power because of the low genetic variation. Finally, no recombination events were found for STV by analysis with the program GARD [29] from the Datamonkey package and RDP v.4.97 [30]. The low genetic variation of STV isolates could preclude the detection of recombinants.

In conclusion, the absence of geographic structure of STV population determined in this work has been also observed in populations of other plant viruses [23,24] and might be due to worldwide trade in tomato seeds and/or the negative selection pressure.

Acknowledgements. We thank Ana Espino (Laboratory of Plant Health, Dirección General de Agricultura, Tenerife, Spain), Purificación Benito (Laboratory of Phytopathology, Cabildo de Gran Canaria, Great Canary Island, Spain), and Ana Alfaro Fernández and María Isabel Font San Ambrosio (Laboratory of Virology, Instituto Agroforestal Mediterráneo, Universistat Politècnica de València, Valencia, Spain) for providing some STV isolates.

Author's contributions. LEG carried out the experiments and LR performed the biocomputational analysis. LG designed the experimental procedures of this research work and wrote the manuscript as well. All authors read and approved the final manuscript.

References.

1. Sabanadzovic S, Valverde RA, Brown JK, Martin RR, Tzanetakis IE (2009) Southern tomato virus: the link between the families *Totiviridae* and *Partitiviridae*.Virus Res. **140**, 130.
2. Candresse T, Marais A, Faure C (2015) First report of southern tomato virus on tomatoes in southwest France.Genome **3**, e01226.
3. Iacono G, Hernandez-Llopis D, Alfaro-Fernandez A, Davino M, Font MI, Panno S, Galipenso L, Rubio L, Davino S (2015) First report of southern tomato virus in tomato crops in Italy. New Dis. Reports **32**.
4. Padmanabhan C, Zheng Y, Li R, Fei Z, Ling KS (2015) Complete genome sequence of southern tomato virus naturally infecting tomatoes in Bangladesh. Genome Announc. **3**, 10.1128/genomeA.01522.
5. Padmanabhan C, Zheng Y, Li R, Sun SE, Zhang D, Liu Y, Fei Z, Ling KS (2015) Complete genome sequence of southern tomato virus identified in China using next-generation sequencing. Genome Announc. **3**, 10.1128/genomeA.01226.
6. Verbeek M, Dullemans AM, Espino A, Botella M, Alfaro-Fernández A, Font MI (2015) First report of southern tomato virus in tomato in the Canary Islands. J. Plant Pathol. **97**.
7. Puchades AV, Carpino C, Alfaro-Fernandez A, Font-San-Ambrosio MI, Davino S, Guerri J, Rubio L, Galipienso L (2017) Detection of southern tomato virus by molecular hybridisation. Ann. Appl. Biol. **171**, 172.
8. Elvira-González L, Carpino C, Alfaro-Fernández A, Font-San Ambrosio MI, Peiró R, Rubio L, Galipienso L (2018) A sensitive real-time RT-PCR reveals a high incidence of southern tomato virus (STV) in Spanish tomato crops. Spanish J. Agric. Res. **16**, 1008.
9. Marquez LM, Redman RS, Rodriguez RJ, Roossinck MJ (2007) A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. Science **315**, 513.
10. Roossinck MJ (2011) The good viruses: viral mutualistic symbioses. Nat. Rev. Microbiol. **9**, 99.
11. Fukuhara T, Tabara M, Koiwa H, Takahashi H (2019) Effect of asymptomatic infection with southern tomato virus on tomato plants. Arch. Virol. 1.
12. Elvira-González L, Medina V, Rubio L, Galipienso L (2019) The persistent southern tomato virus

182 modifies miRNA expression without inducing symptoms and cell ultra-structural changes. Eur. J.
183 Plant Pathol.
184 13. Nibert ML, Pyle JD, Firth AE (2016) A 1 ribosomal frameshifting motif prevalent among plant
185 amalgaviruses. Virology **498**, 201.
186 14. Krupovic M, Dolja VV, Koonin EV (2015) Plant viruses of the *Amalgaviridae* family evolved via
187 recombination between viruses with double-stranded and negative-strand RNA genomes. Biol. Direct
188 **10**, 12.
189 15. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC (2004) Unifying the
190 epidemiological and evolutionary dynamics of pathogens. Science (80-.). **303**, 327.
191 16. Acosta-Leal R, Duffy S, Xiong Z, Hammond R, Elena SF (2011) Advances in plant virus evolution:
192 translating evolutionary insights into better disease management. Phytopathology **101**, 1136.
193 17. García-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus
194 populations. Annu. Rev. Phytopathol. **39**, 157.
195 18. Rubio L, Guerri J, Moreno P (2013) Genetic variability and evolutionary dynamics of viruses of the
196 family *Closteroviridae.* Front. Microbiol. **4**, 151.
197 19. Ferriol I, Ambrós S, da Silva DM, Falk BW, Rubio L (2016) Molecular and biological
198 characterization of highly infectious transcripts from full-length cDNA clones of broad bean wilt virus
199 1. Virus Res. **217**, 71.
200 20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F,
201 Wallace IM, Wilm A, Lopez R (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947.
202 21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics
203 analysis across computing platforms. Mol. Biol. Evol. **35**, 1547.
204 22. Efron B, Halloran E, Holmes S (1996) Bootstrap confidence levels for phylogenetic trees. Proc. Natl.
205 Acad. Sci. **93**, 7085.
206 23. Davino S, Panno S, Rangel EA, Davino M, Bellardi MG, Rubio L (2012) Population genetics of
207 cucumber mosaic virus infecting medicinal, aromatic and ornamental plants from northern Italy. Arch.
208 Virol. 1 (2012).
209 24. Walia JJ, Willemsen A, Elci E, Caglayan K, Falk BW, Rubio L (2014) Genetic variation and possible
210 mechanisms driving the evolution of worldwide fig mosaic virus isolates. Phytopathology **104**, 108.
211 25. Watters KE, Choudhary K, Aviran S, Lucks JB, Perry KL, Thompson JR (2017) Probing of RNA
212 structures in a positive sense RNA virus reveals selection pressures for structural elements. Nucleic
213 Acids Res. **46**, 2573.
214 26. Adams MJ, Antoniw JF (2004) Codon usage bias amongst plant viruses. Arch. Virol. **149**, 113.
215 27. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
216 polymorphism. Genetics **123**, 585.
217 28. Fu YX, Li WH (1993) Maximum likelihood estimation of population parameters. Genetics **134**, 1261.
218 29. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for
219 recombination detection. Bioinformatics **22**, 3096.
220 30. Martin 2nd RCG (2015) Irreversible electroporation of stage 3 locally advanced pancreatic cancer:
221 optimal technique and outcomes. J Vis Surg **1**, 4.

222

13 7
14