Acta virologica **64**: 88 – 92, 2020 doi: 10.4149/av_2020_107

Full genome sequence of a chrysanthemum-infecting tomato spotted wilt virus isolate from Zimbabwe obtained by next-generation sequencing

C. KARAVINA[1,2], J. D. IBABA[1], A. GUBBA[1]

[1]University of KwaZulu-Natal, School of Agriculture, Earth and Environmental Sciences, Department of Plant Pathology, Private Bag X01, Pietermaritzburg, South Africa; [2]Bindura University of Science Education, Department of Crop Science, Private Bag 1020, Bindura, Zimbabwe

Received April 12, 2019; accepted May 22, 2019

Summary. – Tomato spotted wilt virus (TSWV) is an economically important pathogen of many crops worldwide. However, prior to this study, only one complete genome sequence of an African TSWV isolate was available in public databases. This limits genetic diversity and evolutionary studies of the pathogen on the continent. TSWV was detected in symptomatic Zimbabwean chrysanthemum plants using lateral flow kits. The presence of the pathogen was subsequently confirmed by double antibody sandwich enzyme-linked immunosorbent assay and reverse transcription-polymerase chain reaction (RT-PCR). Total RNAs for RT-PCR and next-generation sequencing (NGS) were extracted using an RNA extraction kit. NGS performed on an Illumina HiSeq platform was used to recover the full TSWV genome and analyzed by different software packages. The tripartite genome of the Zimbabwe TSWV isolate consisted of L, M and S RNAs of 8914, 4824 and 2968 nucleotides, respectively. This isolate shared highest protein and nucleotide sequence identities with the isolate LK-1 from neighboring South Africa. The Zimbabwe TSWV isolate was found to be a non-recombinant and non-resistance-breaking. This study provides the first full genome of TSWV from Zimbabwe. It also adds useful information towards understanding the evolution of the pathogen.

Keywords: Africa; tospovirus; phylogenetic analysis; recombination; virus identification

Introduction

Tomato spotted wilt virus, TSWV (the genus *Orthotospovirus*, the family *Tospoviridae*) is an economically important pathogen of over 1300 plant species worldwide (Parrella *et al*., 2003). In Zimbabwe, TSWV was first reported infecting *Nicotiana tabacum* in 1939 (Hopkins, 1940). Its reported host range now includes at least 30 field and greenhouse-grown food and ornamental crops and weeds (Rothwell, 1982; Karavina and Gubba, 2017). TSWV is transmitted principally by *Frankliniella occidentalis* Pergande in a circulative-propagative manner (Rotenberg *et al*., 2015).

The TSWV virions are quasi-spherical enveloped particles that are 80–120 nm in diameter (German *et al*., 1992), while the genome is composed of three single-stranded linear RNA segments named according to their sizes as large (L), medium (M) and small (S) segments. The L segment encodes the RNA-dependent RNA polymerase (RdRp) in a negative sense (de Haan *et al*., 1991). The M and S segments each encode two proteins in the positive and negative-sense orientations. The M segment encodes the GN/GC glycoprotein in the antiviral sense and the non-structural movement (NSm) protein in the viral sense, while the S segment encodes the nucleocapsid (N) in the antiviral sense and non-structural (NSs) proteins in the viral sense (Heinze *et al*., 2001; de Haan *et al*., 1990).

E-mail: ckaravina@gmail.com; phone: +263-772-335-845.

Abbreviations: GRSV = groundnut ringspot virus; NGS = next-generation sequencing; TSWV = tomato spotted wilt virus

Traditionally, TSWV was detected by electron microscopy, serology and reverse transcription-polymerase chain reaction (RT-PCR) (Resende *et al*., 1991; Eiras *et al*., 2001). Recently, next-generation sequencing (NGS) technologies have proved to be useful and efficient tools for plant virus detection (Kreuze *et al*., 2009).

Several partial genomic sequences of African TSWV isolates are available in public databases like GenBank, EMBL and DNA Data Bank of Japan (Sivparsad and Gubba, 2009; Macharia *et al*., 2015; Kisten *et al*., 2016; Karavina and Gubba, 2017) compared to only one full genome sequence. This situation greatly compromises studies to understand TSWV evolutionary history in Africa and even worldwide, as few complete genome sequences are available (Lee *et al*., 2011; Kim *et al*., 2013; Margaria and Rosa, 2015; Margaria *et al*., 2015; Zhang *et al*., 2016). In this study, we describe the full genome of a TSWV isolate from Zimbabwe obtained by using the NGS technology.

Materials and Methods

Virus sampling and identification. The TSWV isolate was identified from chrysanthemum (*Dendranthema morifolium*) plants that were sampled in Zimbabwe in 2015. Plants targeted for sampling harbored high *F. occidentalis* populations and displayed symptoms of tospovirus etiology characterized by leaf necrotic spotting, flower petal browning and malformation, graying and stem collapse. Symptomatic leaves were tested on-site for tospoviruses using lateral flow kits (Loewe Biochemica GmbH, Germany) that could simultaneously detect TSWV, tomato chlorotic spot virus (TCSV) and groundnut ringspot virus (GRSV). Twelve tospovirus-positive leaf samples were collected, preserved in RNAlater® solution (Thermo Fisher Scientific, USA) and tested for TSWV, TCSV and GRSV by double antibody sandwich-enzyme linked immunosorbent assay (DAS-ELISA) with commercial diagnostic kits (Loewe Biochemica GmbH, Germany) at the University of KwaZulu-Natal, South Africa.

RNA extraction, RT-PCR and NGS. Total RNAs were extracted from the samples using the Quick RNA MiniPrep kit (Zymo Research, USA) and used as templates in RT-PCR and NGS experiments. For RT-PCR, first strand complementary DNA (cDNA) synthesis was undertaken using the RevertAid first strand cDNA synthesis kit (Thermo Fisher Scientific, USA) and PCR was performed using the TSWV-N-1F and TSWV-N-777c primer pair specific to the TSWV N gene (Navarre *et al*. 2009). RT reaction mixtures consisted of 10 µl (42 ng/µl) of total RNA pre-heated at 65°C for 10 min, 4 µl reaction buffer, 1 µl RevertAid M-MuLV reverse transcriptase enzyme, 1 µl RiboLock RNase inhibitor, 2 µl dNTPs, and 2 µl of the reverse primer TSWV-N-777c (TTAAGCAA GTTCTGTGAGTTTT). To produce cDNA, the reaction mixture was incubated for 60 min at 42°C followed by a termination step of 70°C for 5 min. Three microliters of the resultant cDNA were used to perform PCR in a 20 µl reaction consisting of 2 µl of forward primer TSWV-N-1F (ATGTCTAAGGTTAAGCTCACTA), 2 µl of reverse primer and 10 µl KAPA 2G Fast HotStart PCR kit (Kapa Biosystems, USA). Amplification was carried out under the following cycling conditions: initial denaturation at 95°C for 3 min; 35 cycles of denaturation at 95°C for 15 s, annealing at 51.3°C for 15 s, extension at 72°C for 15 s; and final extension at 72°C for 5 min. The PCR products were visualized on SYBR Safe (Life Technologies, USA) pre-stained 1.5% agarose gel electrophoresis in TAE buffer. NGS was performed at the Agricultural Research Council Biotechnology Platform (Pretoria, South Africa) on an Illumina HiSeq platform using the paired-end 2 x 125 bp. After sequencing, FastQC program was used to assess the quality of the reads. Adapter removal and quality trimming were conducted using Trimmomatic 0.36 program (Bolger *et al*., 2014). All reads were then used in *de novo* assembly on SPAdes 3.10.1 using the paired-end settings (Bankevich *et al*., 2012). The resultant contigs were subjected to MEGABLAST to identify sequences matching to the TSWV genome.

Data analysis. All contigs that matched to TSWV genome sequences were aligned using the Clustal W program embedded in MEGA7 (Kumar *et al*., 2016) to generate the consensus S, M and L genome sequences of the Zimbabwe TSWV isolate (Mum-A5). ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder) was used to identify the open reading frames (ORFs) on the Mum-A5 isolate genome. Protein molecular weights were determined using the online ExPASy bioinformatics tool (https://web.expasy.org/compute_pi/). Nucleotide and amino acid sequence identity were measured using the online SIAS program (http://imed.med.ucm.es/Tools/sias.html). Phylogenetic analyses were performed on all protein sequence products of the TSWV genome. Clustal W was used for multi-sequence alignments. Maximum Likelihood (ML) phylogenetic trees were inferred using the PhyML 3.0 online version (Guindon *et al*., 2010) based on the best-fit evolution model selected using the Smart Model Selection (Lefort *et al*., 2017) implemented on the PhyML 3.0. Bootstrap analysis was used to measure the reliability of the tree branches. The phylogenetic trees were drawn using the Interactive Tree of Life v4 (Letunic and Bork, 2016) with GRSV as an outgroup. Branches with bootstrap values below 70% were collapsed on the final trees. Recombination events of the complete S, M and L segments of the Mum-A5 isolate were investigated using Recombination Detection Program 4 (RDP4) software with the default settings (Martin *et al*., 2015). Recombination events detected by at least three programs at $p < 0.05$ were accepted.

Results and Discussion

All samples that were tested on-site with lateral flow kits were tospovirus-positive. The test line appeared in the viewing window within three minutes of testing, as an

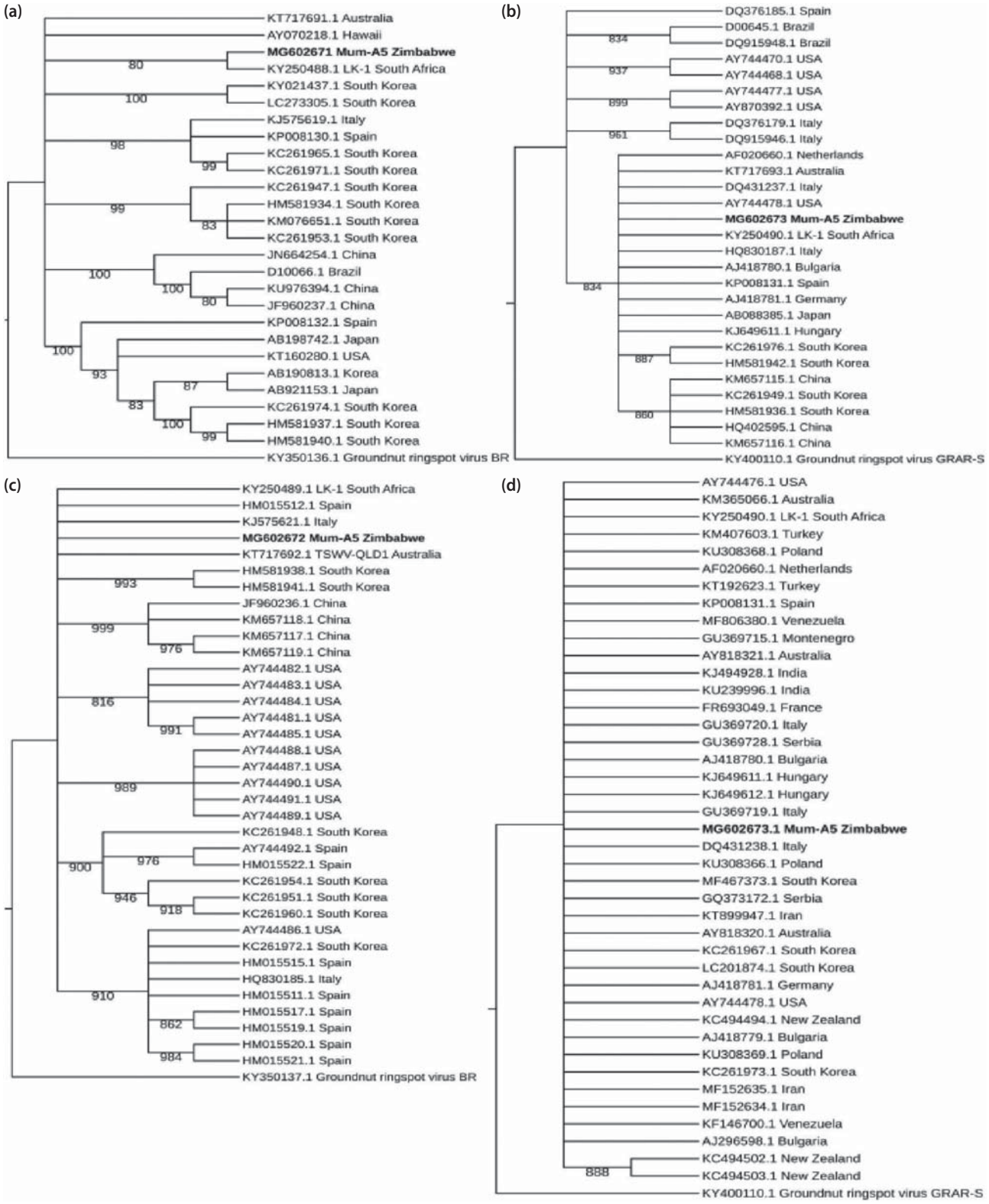

Fig. 1

TSWV rooted phylogenetic trees

TSWV rooted maximum likelihood phylogenetic tree using RdRp protein **(a)**, the NSs protein **(b)**, the G_N/G_C precursor protein **(c)**, and the N protein **(d)**. Bootstrap values are indicated on the branches.

Table 1. Characteristics of the Mum-A5 isolate genomic segments

Segment	Segment length	ORF polarity	ORF length	ORF position	Protein coded	Protein MW (kDa)
S	2968 nt	(+)	1404	89-1492	NSs	50.52
		(-)	777	2817-2041	N	27.53
M	4824 nt	(+)	909	101-1009	NSm	33.71
		(-)	3402	4740-1333	Gc-Gn	126.44
L	8913 nt	(-)	8640	8673-34	RdRp	331.72

indication of high virus titre in the samples. All samples were also positive for TSWV by DAS-ELISA. The 777- bp band indicative of TSWV presence was also produced from all samples tested by RT-PCR.

Sequences of the L, M and S segments of the Mum-A5 isolate were deposited in GenBank under Nos. MG602671, MG602672 and MG602673, respectively. The characteristics of the genomic segments are given in Table 1. The L segment enclosed an ORF of 8640 nt, located in position 241–8880. This ORF is predicted to encode the RdRp protein with an estimated molecular mass of 331.72 kDa. The Mum-A5 L segment shared the highest nucleotide sequence identity at 97.2% with the isolate LK-1 from South Africa. The M segment had two ORFs found at positions 101–1009 and 1333–4740, and encoded the NSm protein and the GN/GC precursor protein in the viral and antiviral senses, respectively. An intergenic region of 324 nt separated the two ORFs. The NSm protein has been identified as the avirulence determinant in Sw-5 gene-based resistance (Lopez *et al*., 2011). The Mum-A5 NSm amino acid sequence did not have the mutations responsible for resistance breakdown. The S segment had two ORFs separated by an intergenic AU-rich region. The ORF at the 3'end was 1404 nt long, encoding the NSs protein with a molecular mass of 50.52 kDa. The ORF at the 5'end, position 2041–2817, encodes the N protein in the complementary sense.

The phylogeny of the Mum-A5 was studied in order to understand its relationship to the other TSWV isolates previously described irrespective of their host and origin. Only complete and functional sequences of the studied proteins were analyzed. The TSWV N protein has been the preferred molecular marker in several studies, resulting in a plethora of sequences in public databases. The TSWV protein sequences in this study followed different evolutionary patterns. The Flu+G was the best-fit model describing the evolution of the NSm and the GN/GC precursor protein, while HIVw+G, JTT+G and JTT+G+I+F were the best-fit evolutionary models for the NSs, N and RdRp proteins, respectively. The outgroup-rooted ML trees using the NSm, GN/GC precursor and N proteins placed Mum-A5 into a distinct phylogenetic lineage (Fig. 1c and d). The NSm protein ML tree is not shown. Amino acid sequence identity values between Mum-A5 and the other TSWV isolates were 96.5–99.6% for the N protein; 95.7–99.3% for the NSm protein and 94.9–97.8% for the GN/GC precursor protein. Well-supported clades involving Mum-A5 were observed on the ML trees generated using the RdRp and NSs proteins. Mum-A5 forms a clade with the isolate LK-1 from neighboring South Africa on the RdRp protein ML tree (Fig. 1a and b). Both African isolates shared the highest RdRp amino acid sequence identity values of 99.1%. Incursion of TSWV could have occurred either through trade, migration or vector movement across the neighboring countries. On the NSs ML tree, the African isolates were grouped in a larger clade with isolates from Asia, Australia, Europe and the USA. No recombination events were detected in any of the Mum-A5 genomic segments. Our results suggest that the origin of Mum-A5 could have been the result of long-distance migration and reassortment.

This study provides the first full genome sequence of a TSWV isolate from Zimbabwe. It adds useful information towards understanding TSWV evolution. Since TSWV is an economically important pathogen, Zimbabwean farmers are advised to grow resistant cultivars to mitigate further virus spread and negative impact on agricultural production.

Acknowledgments. We are grateful to the Agricultural Research Council Biotechnology Platform (Pretoria, South Africa) for performing NGS on the test sample. This work was supported by The W.K. Kellogg Foundation Southern African Scholarship Program as part of CK's PhD studies at the University of KwaZulu-Natal, South Africa.

References

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov A et al. (2012): SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bolger AM, Lohse M, Usadel B (2014): Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. btu170. https://doi.org/10.1093/bioinformatics/btu170

de Haan P, Kormelink R, de Oliveira Resende R, van Poelwijk F, Peters D, Goldbach R (1991): Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. J. Gen. Virol. 72, 2207–2216. https://doi.org/10.1099/0022-1317-72-9-2207

de Haan P, Wagemakers L, Peters D, Goldbach R (1990): The S RNA segment of Tomato spotted wilt virus has an ambisense character. J. Gen. Virol. 71, 1000–1007. https://doi.org/10.1099/0022-1317-71-5-1001

Eiras M, Resende RO, Missiaggia AA, de Avila AC (2001): RT-PCR and dot blot hybridization for a universal detection of tospoviruses. Fitopatol. Bras. 26, 170–175. https://doi.org/10.1590/S0100-41582001000200009

German LT, Ulman DE, Moyer JW (1992): Tospoviruses: diagnosis, molecular biology, phylogeny and vector relationships. Ann. Rev. Phytopathol. 30, 315–348. https://doi.org/10.1146/annurev.py.30.090192.001531

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010): New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010

Heinze C, Letschert B, Hristova D, Yankulova M, Kauadjouor O, Willingman P et al. (2001): Variability of the N-protein and the intergenic region of the S RNA of Tomato spotted wilt tospovirus (TSWV). New Microbiol. 24, 175–187.

Hopkins JCF (1940): The tobacco ‚kromnek' virus in Rhodesia. Rhod. Agr. J. 37, 326–329.

Karavina C, Gubba A (2017): Detection and characterization of Tomato spotted wilt virus infecting field and greenhouse-grown crops in Zimbabwe. Eur. J. Plant. Pathol. 149, 933–944. https://doi.org/10.1007/s10658-017-1243-4

Kim J-H, Kim Y-S, Jang S-W, Joen Y-H (2013): Complete genome sequence of Tomato spotted wilt virus from paprika in Korea. E. Sci. J. Plant. Pathol. 2, 121–136.

Kisten L, Moodley V, Gubba A, Mafongoya P (2016): First detection of Tomato spotted wilt virus (TSWV) on Amaranthus thunbergii in South Africa. Plant. Dis. 100, 2176. https://doi.org/10.1094/PDIS-04-16-0470-PDN

Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I et al. (2009): Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a genetic method for diagnosis discovery and sequencing of viruses. Virology 388, 1-7. https://doi.org/10.1016/j.virol.2009.03.024

Kumar S, Stecher G, Tamura K (2016): MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Lee J-S, Cho WK, Kim M-K, Kwak H-R, Choi H-S, Kim K-H (2011): Complete genome sequences of three Tomato spotted wilt virus isolates from tomato and pepper plants in Korea and their phylogenetic relationship to other TSWV isolates. Arch. Virol. 156, 725–728. https://doi.org/10.1007/s00705-011-0935-x

Lefort V, Longueville J-E, Gascuel O (2017): SMS: Smart Model Selection in PhyML. Mol. Biol. Evol. 34, 2422–2424. https://doi.org/10.1093/molbev/msx149

Letunic I, Bork P (2016): Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acid Res. 44, W242–W245.

Lopez C, Aramburu J, Galipienso L, Soler S, Nuez F, Rubio L (2011): Evolutionary analysis of tomato Sw-5 resistance breaking isolates of Tomato spotted wilt virus. J. Gen. Virol. 92, 210–215. https://doi.org/10.1093/nar/gkw290

Macharia I, Backhouse D, Ateka EM, Wu S-B, Harvey J, Njahira M et al. (2015): Distribution and genetic diversity of Tomato spotted wilt virus following an incursion into Kenya. Ann. Appl. Biol. 166, 520–529. https://doi.org/10.1111/aab.12201

Margaria P, Miozzi L, Ciuffo M, Pappu HR, Turina M (2015): The first complete genome sequences of two distinct European Tomato spotted wilt virus isolates. Arch. Virol. 160, 519–595. https://doi.org/10.1007/s00705-014-2256-3

Margaria P, Rosa C (2015): First complete genome sequence of a Tomato spotted wilt virus isolate from the United States and its relationship to other TSWV isolates of different geographic origin. Arch. Virol. 160, 2915–2920. https://doi.org/10.1007/s00705-015-2589-6

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015): RDP4: detection and analysis of recombination patterns in virus genes. Virus Evol. 1, vev003. https://doi.org/10.1093/ve/vev003

Navarre DA, Shakya R, Holden J, Crosslin JM (2009): LC-MS analysis of phenolic compounds in tubers showing zebra chip symptoms. Am. J. Potato Res. 86, 88–95. https://doi.org/10.1007/s12230-008-9060-0

Parrella G, Gognalous P, Gebre-Selassie K, Vovlas C, Marchoux G (2003): An update on the host range of Tomato spotted wilt virus. J. Plant. Pathol. 85, 227–264.

Resende RO, de Haan P, de Avila AC, Kitajima EW, Kormelink R, Goldbach R et al. (1991): Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. J. Gen. Virol. 72, 2375–2383. https://doi.org/10.1099/0022-1317-72-10-2375

Rotenberg D, Jacobson AL, Schneweis DJ, Whitfield AE (2015): Thrips transmission of tospoviruses. Curr. Opin. Virol. 15, 80–89. https://doi.org/10.1016/j.coviro.2015.08.003

Rothwell A (1982): A revised list of plant diseases occurring in Zimbabwe. Kirkia 12, 233–352.

Sivparsad BJ, Gubba A (2009): Isolation and molecular characterization of Tomato spotted wilt virus (TSWV) isolates occurring in South Africa. Afr. J. Agr. Res. 3, 428–434.

Zhang Z, Wang D, Yu C, Wang Z, Dong J, Shi K, Yuan X (2016): Identification of three new isolates of Tomato spotted wilt virus from different hosts in China: molecular diversity, phylogenetic and recombination analysis. Virol. J. 13, 8. https://doi.org/10.1186/s12985-015-0457-3