Article Type : Original Article

Title Page

Serendipitous identification of a new *Iflavirus*-like virus infecting tomato, and its subsequent characterisation.

M. Saqib[*1,2], S. J. Wylie[2], M. G. K. Jones[2]

[1]Plant Gene Regulation Research Group, Bioproduction Research Institute.

National Institute of Advanced Industrial Science and Technology (AIST) Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8566, Japan.

[2]Plant Virus Section, Plant Biotechnology Research Group, Western Australian State Agricultural Biotechnology Centre, School of Veterinary and Life Sciences, Murdoch University, Perth, W.A. 6150, Australia.

*Corresponding author

 Fax: + 61 8 9360 2502

 Phone: + 61 8 9360 2424

E mail address: msaqib787@gmail.com

Other contact details:

s.wylie@murdoch.edu.au

m.jones@murdoch.edu.au

Abstract

The genomic sequence of a previously undescribed virus was identified from symptomless tomato plants (*Solanum lycopersicum*). The viral genome is a positive sense ssRNA molecule of 8,506 nucleotides. It is predicted to encode a single polyprotein of 314.5 kDa, which is subsequently processed into three coat protein components of 13.7, 17.9 and 13.5 kDa, and a viral replicase of approximately 207 kDa with conserved motifs for a helicase, a protease, and RNA-dependent RNA polymerase (RdRp). Pairwise analysis of the deduced amino acid sequence of the RdRp revealed that it shares closest identity with members of the family *Iflaviridae*, genus *Iflavirus* (19-47% identity). Evidence of replication in plants was detected by RT-PCR of the viral replicative strand, and short interfering RNAs (siRNAs) matching the virus. We propose the name Tomato matilda virus (TMaV), and further, propose that the genus Tomavirus (Tomato Matilda virus) be created within the family *Iflaviridae*. This is the first report of a plant-infecting virus resembling members of the *Iflaviridae*.

Introduction

New sequencing technologies known variously as high throughput nucleotide sequencing, next-generation sequencing (NGS), and deep sequencing, have been applied to the discovery of new plant viruses (Kreuze *et al*., 2009; Roossinck 2012; Wylie *et al*., 2012). The major advance over earlier technologies is that NGS-based approaches are generic (i.e. no previous knowledge of the virus structure or sequence is needed), and the depth of sequencing sometime enables the determination of complete genomes sequences.

In most previous studies on economically important virus infections of plants, visible symptoms of infection were evident, and this has lead to the view that most viruses induce symptoms in their hosts. Viruses that do not induce visible symptoms are designated as 'latent' or 'symptomless' viruses. Recent application of deep sequencing technologies to analyse RNA from apparently healthy, virus-free plants, has often revealed an abundance of viruses in such plants (e.g. Wylie *et al*., 2012; Roossinck, 2012; Wylie et al 2013[a,b]). Perhaps the incidence of asymptomatic virus infection in wild plants is not surprising, but even in well-studied plant species new viruses are being discovered from apparently healthy specimens (Coetzee *et al*., 2010).

Here we present the serendipitous discovery of a novel virus in tomato plants originally obtained from a seedling retailer for a transcriptome study on responses to phytoplasma infection. In this study, total mRNA and short-interfering RNAs (siRNAs) were analysed using the Illumina platform. Presence of a virus-like sequence was revealed after assembly of contigs. Despite very low levels of sequence identity with described viruses, the large open reading frame (ORF) encodes a polyprotein clearly recognizable as that of a virus. Aspects of the putative genome sequence of the virus, its

transmission to alternative hosts is described here. We discuss implications of its proposed classification within the family *Iflaviridae*.

Materials and methods

Plant materials and virus maintenance

Tomato seedlings were obtained in 2009 from a seedling supplier in Perth, Western Australia. They were maintained using a standard potting mix in an insect-proof glasshouse at 20-24°C under natural light. Virus cultures were transferred to tomato (*Solanum lycopersicum)*, capsicum and chilli plants (*Capsicum annuum*). For sap inoculation, infected leaves were ground in 0.05M potassium phosphate buffer pH 7.2 containing 0.05M sodium sulphite. The sap was mixed with diatomaceous earth (Aquacel) and rub-inoculated onto leaves of five plants each of tomato, capsicum*, Phaseolus vulgaris* (common bean)*, Vicia faba* (broad bean)*, Raphanus sativus* (radish)*, Brassica rapa* (canola)*, Pisum sativum* (pea)*, Solanum tuberosum* (potato)*, Solanum melongena* (egg plant)*, Nicotiana benthamiana, N. tabacum* (tobacco)*, Chenopdium quinoa, Daucus carota* (carrot) and *Choriandrum sativum* (coriander). Five control plants of each species were mock inoculated. Grafting to tomato seedlings was used to maintain virus inoculum.

RNA extraction

After grinding 100 mg tomato leaf material in liquid nitrogen, total RNA was extracted using an RNeasy Plant Miniprep kit (Qiagen) following the manufacturer's instructions.

Design of TMaV primers for PCR amplification

A series of overlapping primers (Table 1) were designed across the TMaV genome and used to resequence it using Sanger dideoxy sequencing, and to confirm diagnostic RT PCR assays. For quantitative real time PCR, the primer pair TMA forward and TMA reverse together with internal TMA probe, were designed using Integrated DNA Technologies (Coralville, Iowa) online software (Table 1). A primer was designed to synthesise cDNA from the viral replicative (negative) strand, and the cDNA was amplified by PCR (Table 1). Two full-length TMaV sequences were aligned with ClustalW Multiple Sequence Alignment Program version 1.82 for whole genome sequence analysis. The primer pairs were designed from consensus regions 500-800 nt apart.

Reverse Transcription and PCR

Specific reverse primers or random hexamer primers were used for first strand cDNA synthesis using a RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas). Reaction mixtures (25μl) contained 1/20 diluted cDNA as template, 2.5 mM $MgCl_2$, 1 X reaction buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.0) 150 μM dNTPs, and 10 μM of each primer and 0.5 unit *Taq* DNA polymerase. Incubation conditions were 94 °C for 5 min, 25 cycles of denaturation at 94 °C for 10 sec, annealing at 48°C- 56 °C for 30 sec (calculated for each primer pair), extension at 72°C for 1 min and a final extension of 72 °C for 10 min.

One step real time PCR was performed using Jumpstart Ready Master Mix (Sigma). Each reaction was performed in triplicate. Each reaction contained 50 mM $MgCl_2$, 300 nM each of forward and reverse primers and 250 nM of probe (Table 1). Cycling conditions were 42°C for 15 min followed by 95°C for 3 min and 40 cycles of 95°C for 15 sec and 60°C for 45 sec. Negative controls were RNase-free water and RNA extracted from uninfected plants.

Sanger sequencing

PCR products were purified from agarose gel using a QIAquick Gel Extraction Kit (Qiagen). Amplified products from each primer set were sequenced directly after purification using 5' and 3' end primers. Sanger sequencing was done using Big Dye Terminator 3.1 (Perkin-Elmer, Foster City, CA), with an Applied Biosystems 3730 sequencer. Sequencing was repeated twice for both DNA strands. After sequencing the PCR products were assembled using BioEdit (Hall 1999) and SeqEd 1.0.3. Sequences were confirmed at the 3' or 5' ends by comparing data from the complementary sequences amplified by adjacent primer sets. Sanger-generated sequences were mapped to the original Illumina-generated TMaV sequence.

Next Generation Sequencing

Total RNA (20 μg suspended in ethanol) from leaf tissue collected from three plants was sent to Macrogen Inc (Seoul, South Korea) for cDNA library synthesis of mRNA and single end sequencing of the cDNA using an Illumina GAIIx platform (Illumina Inc. San Diego, CA) over 78 cycles. For two RNA samples, mRNA was sequenced after cDNA synthesis using oligo-dT primers. For the other sample, siRNA was analysed after size fractionation for 18-30 nt fragments and randomly primed cDNA synthesis.

D*e novo* sequence assembly of the 78 nt reads was done using three short-read assemblers: Geneious Pro v 5.0.4 (Drummond *et al.,* 2010), CLC Genomics Workbench (CLC bio Finlandsgade Katrinebjerg, Denmark) and Velvet v 0.6.04 (Zerbino & Birney 2008) using default parameters. Contigs were produced and further assembled into longer contigs using the program ContigExpress, which is included in the Vector NTI package (Invitrogen, Carlsbad, CA). Assembled contigs were used to interrogate the GenBank database (NCBI) using nt (Blastn) and deduced amino acid (aa)

sequences translated from all six frames (Blastx). The entire assembly of 23,271contig sequences were analysed using MegaBlast (NCBI) against sequences at the International Tomato Annotation Group (ITAG) database, and the data was downloaded from Sol Genomic Networks for further analysis. Blast scores lower than 100 were recorded as no significant hit against the database. The remaining unmatched sequences were analysed further using Blastn to GenBank tomato mRNA, ITAG2.4 genomics and to ICTV virus sequences.

Genome annotation

Positions of ORFs, identities of deduced proteins, mature peptides, and domains encoded by them were predicted by Geneious Pro v6.0, the NCBI Conserved Domain Database (CDD), InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan), and by identity after alignment with characterised virus sequences.

Phylogenetic analysis

Alignment using Clustal W (Thompson *et al.,* 1994) and phylogenetic analysis using a Sankoff-Morel-Cedergren model implemented within MEGA5 (Tamura *et al.,* 2011) was done on deduced replicase aa sequences of TMaV together with those of 21 other related viruses. Publically available sequences of members of the *Iflaviridae* representing the genetic, host range, and geographical diversity of the viruses studied were selected for phylogenetic analysis. Picorna-like viruses of plants were also included (Table 2). Trees were drawn to scale, with branch lengths the same units as those of evolutionary distances used to infer the phylogenetic tree branch length in aa residue substitutions per site between sequences. The evolutionary history was inferred using neighbor-joining, minimum evolution and maximum parsimony methods. The aa identity scores were computed using the Poisson correction model (Zuckerkandi & Pauling, 1965).

Results

Genome assembly and sequence analysis

Illumina sequencing of total mRNAs isolated from two tomato plants tested generated 31,558,959 and 32,359,875 reads of 78 nt. Contigs were assembled *de novo.* The longest virus-like sequences were contigs of 8,506 nucleotides (nt) (isolate A from plant 1 assembled from 13,325 reads, 0.04% of total) and 8,506 nt (isolate B from plant 2 assembled from 3,570 reads, 0.01% of total) excluding 3' poly-adenylation signals. The virus genome sequences were polyadenylated, and the 5' untranslated region (UTR) of 104 nt and 3' UTR of 113 nt, suggested that the virus sequence was complete, although 5' rapid amplification of cDNA ends (RACE) was not done to confirm this. After sequence analyses, these sequences were designated as those of a previously undescribed virus, for which the name Tomato matilda virus (TMaV) is proposed. Isolate A of TMaV was granted GenBank accession code HQ260868.

The set of Illumina sequences from one sample was submitted to the Short Read Archive (NCBI) and was granted the accession no SRX506967. The entire assembled contig sequences 23,271 are provide in S1 supplemental data. A total of 96.6% of the contigs matched tomato sequences (Fig S2). Three contig sequences matched viruses. The complete raw data file and Genbank results are provided in supplemental data S3. Three contig sequence match viral sequences, which are *Varroa destructor virus-1* (VDV-1), *Grapevine leafroll-associated virus* and *Lone Star virus*. One contig (no 576) is the new virus sequence that closely matches with VDV-1. The complete Genbank results are provided in supplement data S3.

A third tomato plant was infected with TMaV and small RNA species generated were then sequenced. 24,010,322 sequences of 20-26 nt long were obtained. These were mapped to the TMaV genome sequence. The small RNA sequences mapped to only five regions of the TMaV genome (Fig. 1a). These were the 5' UTR at nt 61-83 where there was 2,197-fold coverage, at nt 2,534-2,556 where there was 40-fold coverage, at nt 4,228-4,250 where there was 14 fold coverage, at nt 6,256-6,278 where there was 8-fold coverage, and nt 7,396-7,416, which had 12 fold coverage. The presence of small RNAs that map to the virus genome provides evidence of a host response to TMaV infection.

Resequencing

To confirm accuracy of the genome sequence generated by Illumina sequencing, overlapping RT-PCR amplicons were generated to amplify the entire TMaV genome by the Sanger method using specific primer pairs (Table 1). Genome fragments of the expected sizes (667-836 nt) were obtained from all primer pairs, and their sequences determined. The re-sequenced genome shared 99.8 % sequence identity with the original sequence obtained. The primer pairs Matilda F2/R2 (Table 1) were used subsequently to detect TMaV in host range studies. The TMA forward and TMA reverse primer pair together with the TMA probe (Table 1) for RT qPCR was also used for TMaV diagnostics (data not shown).

Further evidence of virus replication in tomato plants was provided when the replicative (negative) RNA strand of the virus was detected from infected plants. Specific primers TMaV RCF, RCR1 (Table 1) detected the negative strand in five TMaV-infected tomato plants while RNA from the uninfected plant and negative control (water) did not amplify a product. The amplicon of 481 nt was sequenced, and it reverse compliment mapped with 100% identity to the virus genomic sequence.

Virus inoculation to alternate hosts

In order to understand more about symptom development and host range of TMaV, five plants each of *S. lycopersicum, C. annuum*, *S. melongena, P. vulgaris, V. faba, R. sativus, B. rapa, P. sativum, S. tuberosum, N. benthamiana, N. tabacum, C. quinoa, D. carota* and *C. sativum* were manually inoculated with sap from TMaV-infected tomato plants. Only plants of *S. lycopersicum, C. annuum*, and *S. melongena* were positive for TMaV in RT-PCR and qPCR analyses on new, uninoculated leaves two weeks post-inoculation. Only infected *C. annuum* plants developed a mild mosaic on the leaves of two plants, whereas the other systemically-infected plants remained asymptomatic. No insects were observed on the leaves when visualized by light microscope.

Molecular characterisation

The RNA genome of isolate A of TMaV was 8,506 nt in length, excluding the poly(A) tail: this is shorter than described iflaviruses, whose genomes range in size from 8.7 kb SBV to 10.9 kb. The genome is predicted to encode a single polyprotein from a single ORF. There was a 5' UTR of 104 nt. The first in-frame AUG occurred at nt 105-107 within the context CAAAAUGGA, which resembles translation initiation motifs of some plant viruses, including the flexiviruses, *Hardenbergia virus A* (GAAAAUGAG) and *Apple stem grooving virus* (AAAAAUGAG), and the iflavirus, VDV-1 (AAAAAUGGC), but not the iflaviruses SBV (isolate Korea) (UAUUAUGGA) and *Lygus lineolaris virus-1* (isolate LLV-1) (CACUAUGGC). The predicted polyprotein is 2,787 aa residues in length with a calculated mass of 314 kDa. It is terminated by an ochre (UAA) termination codon at nt 8,466-8,468, followed by a 3' UTR of 113 nt and a poly-A tail. Active domains within the polyprotein (Fig. 1a) were predicted, based on homology with other viruses that share similar domain architecture. Those with the most similar architecture were iflaviruses, including SBV and LLV. Three small rhinovirus-like (rhv-like) structural proteins with sequence identity to iflavirus capsid proteins were identified at the

N-terminus of the polyprotein. Rhv-like 1 protein was located at aa residues 307-427 and has a mass of 13.7 kDa, rhv-like 2 at residues 571-728 with a mass of 17.9 kDa, and rhv-like 3 at residues 858-977 has a mass of 13.4 kDa (Fig. 1b). Viruses that encode three similar capsid proteins at the N-terminus of their polyproteins are non-enveloped plus-strand ssRNA animal viruses with icosahedral capsids (Ehrenfeld *et al.,* 2009). An RNA helicase is predicted from residues 1,374-1,480. The TMaV helicase shares homology with the P-loop NTPase domain superfamily, which are characterised by the Walker A motif (GxxxxGK[S/T]) where x is any residue, and the Walker B motif (hhhh[D/E]), where h is a hydrophobic residue (Walker *et al.,* 1982). In TMaV the A motif exists as GASGIGKS at residues 1,376-1,383 and the B motif as VILVD at residues 1,425-1,429. A cysteine protease-like domain is located at residues 2,061-2,235. Picornaviral proteins are expressed as a single polyprotein cleaved by a cysteine protease, members of the protease 3C superfamily. The catalytic H located at aa residue 2,076, the E at 2,115, and the catalytic C within the motif GxCG (aa residues 2,209-2,212) were identified by homology with insect picornaviruses including SBV (NP_049374), LLV (AEL30247), SBPV (ADI46683), and DWV (ADK55525). Nucleotide binding sites were present at the C-terminal H and G, aa residues 2,228 and 2,231, respectively. A definitive VPg region was not found in TMaV but areas of sequence identity to VPg regions of *Cowpea mosaic virus* (CPMV) (Family *Secoviridae*, genus *Comovirus*) at 1,359-1,379 (DxxQxxYxxVP) and *Tomato ringspot virus* (Family *Secoviridae*, genus *Nepovirus*) at 2,024-2,048 (TxxxxxYxxxxxxxxxxxxFRxQxV) were identified that may represent a VPg. The RdRp was identified at aa residues 2,316-2,782. The catalytic centre of RNA virus RdRps is formed by several motifs containing a number of conserved amino acid residues. The conserved replicase core motif S/TGx3 Tx3 NS/Tx22 GDD (Koonin 1991) was present as SGx3 Tx3 NSx39 GDD at residues 2,607-2,659. Other conserved RdRp domains characterised were I (TLKDERR) at residues 2,463-2,469, II (VFCASPVDYTIALRQNLLHFCAATMKN) at residues 2,482-2,508, III (AVGINPLGPEWSKI) at residues 2,515-2,528, and IV (MDYSNFGPCFH) at residues 2,543-2,553.

Phylogenetic analysis

The entire polyprotein sequence of TMaV shared a maximum of 31-39% aa identity (85-95% coverage, e-values 0.0 - 2^{-108}) with those of SBV isolates from Korea, China and UK, and LLV-1 from Korea (Fig. 2). The individual structural proteins rhv1-3 also shared maximum identities (40-52%, coverage 97-100%, e-values $4^{-22} - 3^{-52}$) with isolates of SBV and LLV-1. Analysis of the RdRp domain confirmed TMaV's close association with iflaviruses. It shared maximum identities (34-41%, coverage 93-99%, e-values $3^{-108} - 2^{-154}$) with SBV, LLV-1, and an unclassified picorna-like virus that infects bats (*Eptesicus fuscus*) (Table 3). RdRP proteins of plant-infecting secoviruses *Apple latent spherical virus*, *Broad bean wilt virus*, CPMV, *Rice tungro spherical virus*, and *Tomato torrado virus* shared lower (20-28%) identities with that of TMaV.

Because of the relatively close sequence identity of the TMaV genome and gene products, and the similar genome architecture, we propose that Tomato matilda virus be classified within the family *Iflaviridae*. Due to the fact it replicates in plants, unlike other described iflaviruses that all replicate in arthropods, we propose that a second genus within the *Iflaviridae* be created. We propose that the name of this new monospecific genus be 'Tomavirus' (**To**mato **ma**tilda **virus**).

Discussion

Deep sequencing analysis of RNA extracted from asymptomatic tomato plants revealed the genome sequences of a virus for which we propose the name Tomato matilda virus. The name matilda was chosen from the iconic Australian ballad 'Waltzing Matilda' written by Banjo Paterson in 1895 that refers to an itinerant man travelling to foreign places. Thus, the name refers to the ancestral virus's improbable journey from replicating in an arthropod to replicating in a plant.

When the genome of TMaV was discovered, there was a possibility that the sequence represented an arthropod virus that had contaminated the sample. Blast analysis results showed that most contigs constructed from raw reads were from either the tomato plant or from TMaV. If the virus sequence were derived from an arthropod, its genetic signature would be clearly evident in the Blast output. There were no arthropod-derived sequences identified from over 23,000 contigs. This stands as strong evidence that TMaV is a plant virus. Furthermore, TMaV infected tomato and eggplant asymptomatically and induced mild symptoms of infection in some capsicum plants. It was transmitted between these solanaceous plants. Transmission between plants and subsequent systemic infection is strong evidence that TMaV replicates in plants. Detection of the replicative strand of the viral sequence by RT-PCR and transmission to other plants is also evidence of replication in plant hosts. This was supported by the detection of small RNAs that mapped to the viral genome sequence, indicating that the host was mounting a defensive response. Taken together, we feel the evidence presented is sufficient to ascribe TMaV as a plant virus.

Many questions about this unusual virus remain to be answered. These include establishing its mode of transmission, its natural host range, particle structure, replication cycle in different hosts, and pathology in other plants. There are also questions about its probable transition from an arthropod to a plant, notably the identity of molecular steps needed before the probable insect virus ancestor was able to replicate in a plant. Recent identification of a proposed plant-infecting member (Blackberry virus Z) of the picornavirus sister family *Dicistroviridae,* and the existence of the plant-infecting genus *Tospovirus* within the predominantly arthropod and vertebrate infecting family *Bunyaviridae* (Ullman *et al*., 2005), are evidence that viruses have made this transition before.

Analysis of the genome sequence of the virus found suggests that it belongs to the order *Picornavirales*. Like TMaV, members of the order exibit (i) a positive-sense RNA genome, usually with a 5′-bound VPg and 3′-polyadenylation signal, (ii) genome translation into auto-proteolytically processed polyprotein(s), (iii) capsid proteins organized in a module containing three related domains which form small icosahedral, non-enveloped particles, and (iv) a three-domain module containing a superfamily III helicase, a 3C-like cysteine proteinase, and an RNA-dependent RNA polymerase. The order includes the families *Picornaviridae*, *Dicistroviridae*, *Marnaviridae*, *Secoviridae* and *Iflaviridae,* and there are four unassigned members (Le Gall *et al*., 2008; Kapoor *et al.,* 2010). Plant infecting members of the *Picornavirales* are predominantly within the family *Secoviridae* (including the sub-family *Comovirinae*) (Sanfacon *et al.,* 2009), but the plant infecting blackberry virus Z is proposed as an unclassified member of the *Dicistroviridae*, which otherwise consists of arthropod infecting members. Two new insect infecting families have been reported in the *Picornavirales*: *Dicistroviridae* and *Iflaviridae.* Although both families have a monopartite genome, members of family *Dicistroviridae* are characterised by the presence of two cistrons each encoding a polyprotein, hence the name (Christian & Scotti 1998; Hunter *et al.,* 2006; Valles & Hashimoto 2009).

Members of the genus *Iflavirus* within the *Iflaviridae* have a single ORF and structural genes are located at the 5′ end of the genome. All seven iflavirus species described by the International Committee on the Taxonomy of Viruses (ICTV), and the other unclassified members, are non-enveloped, icosahedral particles that to date have only been reported to infect arthropods. The type species of *Iflavirus* is *Infectious flacherie virus* (IFV), and other species are *Deformed wing virus* (DWV), *Ectropis obliqua virus* (EoV), *Perina nuda virus* (PnV), *Sacbrood virus* (SBV), VDV-1 and *Slow bee paralysis virus* (SBPV) (van Oers 2010). To date, seven iflaviruses are accepted by the ICTV (van Oers, 2010). Current *Iflaviridae* family members infect insects from the orders Lepidoptera,

Hymenoptera, Heteroptera, Diptera, and Orthoptera, and have a linear ssRNA(+) genome of 8.8-9.7 kb. They have been identified as responsible for substantial economic losses in silk production (Aizawa & Kuruta, 1964) and apiculture (Bradbear, 1988).

Based on the genome organisation of TMaV and its ability to replicate in plants there is also a possibility that the plant host is in fact acting as a 'vector' for the transmission of TMaV to insects/animals. This concept needs to be studied further.

The rapid development of deep sequencing technologies will identify greater numbers of unusual viruses, and these will challenge current taxonomic groupings and force the creation of new ones. New virus taxa will be identified rapidly, and it is probable that the classical approach of methodical biological characterisation of new viruses will lag further behind. We envisage that new high throughput methods of biological characterisation will be developed to cope with the 'tsunami' of new virus discoveries from all domains of life that will inevitably challenge the science community in coming years.

References

Aizawa K, Kuruta K, 1964. Infection under aseptic conditions with the virus of infectious flacherie in the silkworm, *Bombyx mori*. *Journal of Insect Pathology* **6**, 130–132.

Bradbear N, 1988. World distribution of major honeybee diseases and pests. *Bee World* 15–39.

Christian PD, Scotti PD, 1998. *Picorna like virus of insects: The insect viruses*, In: Miller LK, Ball LA (eds). Plenum, New York. Pp 301–336.

Coetzee B, Freeborough MJ, Maree HJ, Celton JM, Rees DJG, Burger JT, 2010. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* **400**, 157–163.

de Miranda JR, Drebot M, Tyler S, Shen M, Cameron CE, Stoltz DB, Camazine SM, 2004. Complete nucleotide sequence of Kashmir bee virus and comparison with acute bee paralysis virus. *Journal of General Virology* **85**, 2263–2270.

de Miranda JR, Dainat B, Locke B, Cordoni G, Berthoud H, Gauthier L, Neumann P, Budge GE, Ball BV, Stoltz DB, 2010. Genetic characterization of slow bee paralysis virus of the honeybee (Apis mellifera L.). *Journal of General Virology* **91**, 2524–2530.

Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB, 2010. Metagenomic analysis of the viromes of three North American bat species: Viral diversity among different bat species that share a common habitat. *Journal of Virology* **84**, 13004–13018.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson, A, 2010a. *Geneious* v4.8. http://www.geneious.com

Ehrenfeld E, Modlin J, Chumakov K, 2009. Future of polio vaccines. *Expert Review in Vaccines* **8**, 899-905.

Fujiyuki T, Takeuchi H, Ono M, Ohka S, Sasaki T, Nomoto A, Kubo T, 2004. Novel insect picorna-like virus identified in the brains of aggressive worker honeybees. *Journal of Virology* **78**, 1093–1100.

Ghosh RC, Ball BV, Willcocks MM, Carter MJ, 1999. The nucleotide sequence of sacbrood virus of the honey bee: an insect picorna-like virus. *Journal of General Virology* **80**, 1541–1549.

Hall TA, 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium* **41**, 95–98.

Hunter WB, Katsar CS, Chaparro JX, 2006. Molecular analysis of capsid protein of *Homalodisca coagulata Virus-1*, a new leafhopper-infecting virus from the glassy-winged sharpshooter, *Homalodisca coagulate*. *Journal of Insect Science* **28**, 1536–2442.

Johnson KN, Christian PD, 1998. The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. *Journal of General Virology* **79**, 191–203.

Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, Simon R, 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* **25**, 1–7.

Kapoor A, Simmonds P, Lipkin WI, Zaidi S, Delwart E, 2010. Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *Journal of Virology* **84**, 10322–10328.

Kim MC, Kwon YK, Joh SJ, Kim SJ, Tolf C, Kim JH, Sung HW, Lindberg AM, Kwon JH, 2007. Recent Korean isolates of duck hepatitis virus reveal the presence of a new geno-and serotype when compared to duck hepatitis virus type 1 type strains. *Archives of Virology* **152**, 2059–2072.

Koonin EV, 1991. The phylogeny of RNA-dependent polymerase of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197–2206.

Lanzi G, de Miranda JR, Boniotti MB, Cameron CE, Lavazza A, Capucci L, Camazine SM, Rossi C, 2006. Molecular and biological characterization of deformed wing virus of honeybees (*Apis mellifera* L.). *Journal of Virology* **80**, 4998–5009.

Le Gall O, Christian P, Fauquet CM, King AMQ, Knowles NK, Nakashima N, Stanway G, Gorbalenya AE, 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T= 3 virion architecture. *Archives of Virology* **153**, 715–727.

Li C, Yoshikawa N, Takahashi T, Ito T, Yoshida K, Koganezawa H, 2000. Nucleotide sequence and genome organization of apple latent spherical virus: a new virus classified into the family *Comoviridae*. *Journal of General Virology* **81**, 541–547.

Lomonossoff GP, Shanks M, 1983. The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO Journal* **2**, 2253–2258.

Ma M, Ma C, Li M, Wang S, Yang S, Wang S, 2011. Loop-mediated isothermal amplification for rapid detection of Chinese sacbrood virus. *Journal of Virological Methods* **176**, 115–119.

Mason PW, Pacheco JM, Zhao QZ, Knowles NJ, 2003. Comparisons of the complete genomes of Asian, African and European isolates of a recent foot-and-mouth disease virus type O pandemic strain (PanAsia). *Journal of General Virology* **84**, 1583–1593.

Moore J, Jironkin A, Chandler D, Burroughs N, Evans DJ, Ryabov EV, 2011. Recombinants between *Deformed wing virus* and *Varroa destructor virus-1* may prevail in Varroa destructor-infested honeybee colonies. *Journal of General Virology* **92**, 156–161.

Ongus JR, Peters D, Bonmatin JM, Bengsch E, Vlak JM, van Oers MM, 2004. Complete sequence of a picorna-like virus of the genus Iflavirus replicating in the mite Varroa destructor. *Journal of General Virology* **85**, 3747–3755.

Reddick BB, Habera LF, Law MD, 1997. Nucleotide sequence and taxonomy of maize chlorotic dwarf virus within the family Sequiviridae. *Journal of General Virology* **78**, 1165–1174.

Roossinck M, 2012. Plant virus metagenomics: biodiversity and ecology. *Annual Review in Genetics* **46**, 357–67.

Rott ME, Gilchrist A, Lee L, Rochon D, 1995. Nucleotide sequence of tomato ringspot virus RNA1. *Journal of General Virology* **76**, 465–473.

Ryabov EV, 2007. A novel virus isolated from the aphid Brevicoryne brassicae with similarity to Hymenoptera picorna-like viruses. *Journal of General Virology* **88**, 2590–2595.

Sanfacon H, Wellink J, Le GO, Karasev A, van der Vlugt R, Wetzel T, 2009. Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Archives of Virology* **154**, 899–907.

Shen P, Kaniewska M, Smith C, Beachy RN, 1993. Nucleotide sequence and genomic organization of rice tungro spherical virus. *Virology* **193**, 621–630.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.

Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Turnbull-Ross AD, Mayo MA, Reavy B, Murant AF, 1993. Sequence analysis of the parsnip yellow fleck virus polyprotein: evidence of affinities with picornaviruses. *Journal of General Virology* **74**, 555–561.

Ullman DE, Whitfield AE, German TL, 2005. Thrips and tospoviruses come of age: mapping determinants of insect transmission. *Proceedings of the National Academy of Sciences (USA)* **102**, 4931–4932.

van Oers, M.M., 2010. Genomics and biology of Iflaviruses. In: Asgari, S. Johnson, K.N. (Eds.), Insect Virology Caister. Academic Press, Norfolk, pp. 231–250.

Valles SM, Hashimoto Y, 2009. Isolation and characterization of Solenopsis invicta virus 3, a new positive-strand RNA virus infecting the red imported fire ant, *Solenopsis invicta*. *Virology* **388**, 354–361.

Walker JE, Saraste M, Runswick MJ, Gay NJ, 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO Journal* **1**, 945–951.

Wilson JE, Powell MJ, Hoover SE, Sarnow P, 2000. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular Cell Biology* **20**, 4990–4999.

Wylie SJ, Luo H, Hua Li, Jones MGK, 2012. Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Archives of Virology* **157**, 271–284.

Wylie SJ, Li H, Jones MGK, 2013[a]. Donkey Orchid Symptomless Virus: A Viral 'Platypus' from Australian terrestrial orchids. *PLoS ONE* 8(11): e79587. DOI:10.1371/journal.pone.0079587

Wylie SJ, Li H, Dixon KW, Richards H, Jones MGK, 2013[b]. Exotic and indigenous viruses infect wild populations and captive collections of temperate terrestrial orchids (*Diuris* species) in Australia. *Virus Research* **171**:22-32.

Yamashita T, Sakae K, Tsuzuki H, Suzuki Y, Ishikawa N, Takeda N, Miyamura T, Yamazaki S, 1998. Complete nucleotide sequence and genetic organization of Aichi virus, a distinct member of the Picornaviridae associated with acute gastroenteritis in humans. *Journal of General Virology* **72**, 8408–8412.

Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Gene Research* **18**, 821–829.

Zuckerkandi E, Pauling L, 1965. Evolutionary divergence and convergence in proteins. Pp. 97-166. In: Evolving Genes and Protein. Edited by V. Bryson and H. J. Vogel. Academic press: New York and London.

Table legends

Table 1 Primer sequences used for detection, genome amplification and identification of the replicative negative strand of tomato matilda virus.

Table 2 Members of the order *Picornavirales* used for phylogenetic analysis.

Table 3 Pairwise comparisons of amino acid sequence identity of the RNA-dependent RNA polymerase of 24 confirmed members of the order *Picornavirales* and Tomato matilda virus. The genes with highest overall identity with that of Tomato matilda virus are indicated in bold.

Figure legends

Fig 1 (a) Predicted genome organisation of Tomato matilda virus showing the locations of proposed genes and domains. The red squares on the gray line indicate the regions where small RNAs mapped to the genome (drawn to scale). (For positions of additional motifs in the RdRp see text). (b) Schematic viral genome (drawn to scale) showing open reading frame (green), positions of genes and domains coding for structural and regulatory proteins (blue), and their positions in the polyprotein.

Fig 2 Neighbor-joining tree of the deduced amino acid sequences of twenty-four confirmed members of the order *Picornavirales* and tomato matilda virus. Amino acid sequences for the entire polymerase were aligned using ClustalX and the phylogenetic analysis was conducted using MEGA5 software with the default parameters. Numbers at nodes indicate bootstrap scores. The scale bar indicates an evolutionary distance in amino acid substitutions per site. For each virus, the order of the host is reported.

Table 1 Primer sequences used for detection, amplification of overlapping genome fragments, and identification of the negative (replicative) strand of tomato matilda virus.

Cover age	Primer name	Forward		Reverse	Amplicon Size
1-829	Matilda F	ATGATATTATATGTTATAAT ACAACG	R	TATTCCGCTAAAACTTTTC CAACTT	829
357-1038	Matilda F0	ATCACGACAGAGAGGACAA AGGCGC	R0	GACGACATAATGCATTTT CCACG	681
1004-1821	Matilda F1	ACGGCAGCCACGCTAAGAA A	R1	TCATTATGGCGCCTGTAT GG	817
1745-2277	Matilda F2	GTATAACTTTTGGTCCGGAA	R2	CTTTACGCACTGCTTCCTT T	532

2256-3006	Matilda F3	AAAAAGGAAGCAGTGCGTAA	R3	CGTCCCATTCTATAGCAGGA	750
2967-3684	Matilda F4	GTTGCTAATTTTTATGGCATTCCTG	R4	TCCGCTTCTTCTTTGGATCT	717
3655-4371	Matilda F5	GAGTGGTGTTAGATCCAAAG	R5	TGTATGCGGCATCAGTAACT	716
4345-5064	Matilda F6	TTAATCTAGTTACTGATGCCG	R6	CACCAAACACAAGACCAACT	719
5015-5788	Matilda F7	GTGGTTGCTGCTTAAGAATG	R7	GGAATTGAGTTTCGATGATA	773
5755-6511	Matilda F8	CAGCCCGACGTGATTATCAT	R8	CCAGTGTCTGCAACAGCACA	756
6485-7214	Matilda F9	TTTACCTTGTGCTGTTGCAG	R9	ACCTGCAGACGTTGTTAATT	729
7195-7809	Matilda F10	AATTAACAACGTCTGCAGGT	R10	TCTGTTGCTGGTAAACTGTA	614
7785-8506	Matilda F11	CAAGGTACAGTTTACCAGCA	R11	TTCTATCTACACTATCATACAAAT	720

Primers used for QPCR

7890-8062	TMA forward	TCTTTAGCCCAAGAACATTGCGCC	Rev	ACGATACCATAACGTGCCAGCGAT	172
7975	TMA probe	FAM5'ATGACTTGATCGCCGCTGTCTCAGAA3'BHQ1			

Replicative strand primer pair

7829-7349	TMaV RCF	GGCTCCTGATGGGCTGCCAC	RCR1	CCGGAGGAAGGGTGTTGTGCC	481

Table 2 Members of the order *Picornavirales* used for phylogenetic analysis.

Virus	GenBank accession	Geographical origin	Host name	Common name	Source
AV	NC_001918	Japan	*Homo sapiens*	Man	Yamashita *et al.*, 1998
ALSV	BAA90870	Japan	*Malus domestica*	Apple	Li *et al.*, 2000
BBPV	YP_001285409	UK	*Brevicoryne brassicae*	Cabbage aphid	Ryabov 2007
BBWV	AAD38152	South Korea	*Capsicum annuum*	Capsicum	Unpublished
CrPV	NP_647481	USA	*Teleogryllus commodus*	Cricket	Wilson *et al.*, 2000
CpMV	NP_613283.1	USA	*Vigna unguiculata*	Cowpea	Lomonossoff and Shanks 1983
DCV	NP_044945	Australia	*Drosophila melongaster*	Fruit fly	Johnson and Christian 1998
DWV	ADK55525	UK	*Apis mellifera*	Honey bee	Moore *et al.*, 2011
DWV	AAP49283	USA	*A. mellifera*	Honey bee	Lanzi *et al.*, 2006
DHV	DQ812092	South Korea	*Anas platyrhynchos*	Duck	Kim *et al.*, 2007
FMD	AJ539141	UK	*Bos taurus*	Cattle	Mason *et al.*, 2003
KBV	NP_851403	USA	*A. mellifera*	Honey bee	de Miranda *et al.*, 2004
KKV	YP 015696	Japan	*A. mellifera*	Honey bee	Fujiyuki *et al.*, 2004
LLV	AEL30247	USA	*Lygus lineolaris*	Tarnished plant bug	Unpublished
MCDV	NC003628	USA	*Zea mays*	Maize	Reddick *et al.*, 1997
PLV	ADR79389	USA	*Eptesicus fuscus*	Big brown bat	Donaldson *et al.*, 2010
PYFV	NC003628	UK	*Pastinaca sativa*	Parsnip	Turnbull-Ross *et al.*, 1993
RTSV	AAA66056.1	USA	*Oryza sativa*	Rice	Shen *et al.*, 1993
SBV	ADN38255	China	*A. mellifera*	Honey bee	Ma *et al.*, 2011
SBV	NP 049374	UK	*A. mellifera*	Honey bee	Ghosh *et al.*, 1999
SBV	ADZ98922	Korea	*A. cerana*	Asiatic honey bee	Unpublished
SBPV	YP_003622540	UK	*A. mellifera*	Honey bee	de Miranda *et al.*, 2010
TMaV	**HQ260868***	**Australia**	***Solanum lycopersicum***	**Tomato**	**This study**
ToRSV	NP_620765.1	Canada	*Rubus idaeus*	Raspberry	Rott *et al.*, 1995
VDV	YP 145791	Netherlands	*Varroa destructor*	Mite	Ongus *et al.*, 2004

Protein accession number not yet assigned.

Aichi virus (AV); *Apple latent spherical virus* (ALSV); *Brevicoryne brassicae picorna-like virus* (BBPV); *Broad bean wilt virus 1* (BBWV); *Cricket paralysis virus* (CrPV); *Cowpea mosaic virus* (CpMV); *Drosophila C virus* (DCV); *Deformed wing virus* (DWV); *Duck hepatitis virus* (DHV); *Foot-and-mouth disease virus* (FMD); *Kashmir bee virus* (KBV); *Kakugo virus* (KKV); *Lygus lineolaris virus* (LLV); *Maize chlorotic dwarf virus* (MCDV); *Parsnip yellow fleck virus* (PYFV); *Picorna-like virus* (PLV); *Rice tungro spherical virus* (RTSV); *Sac brood virus* (SBV); *Slow bee paralysis virus* (SBPV); *Tomato matilda virus* (TMaV); *Tomato ringspot virus* (ToRSV); *Varroa destructor virus* (VDV).

Table 3 Pairwise comparisons of deduced amino acid sequences of the RNA-dependent RNA polymerase of 24 members of the order *Picornavirales* with that of Tomato matilda virus. Identities with the Tomato matilda virus RdRp are indicated in bold. Accession codes provided in Table 2.

Virus	ToRSV (Canada)	BBWV (S. Korea)	CpMV (USA)	ALSV (Japan)	PYFV (UK)	MCDV (USA)	RTSV (USA)	KBV (USA)	CrPV (USA)	DCV (Australia)	LLV (USA)	TMaV (Australia)	PLV (USA)	SBV (Korea)	SBV (China)	SBV (UK)	SBPV (UK)	VDV (NL)	DWV (USA)	DWV (UK)	KKV (Japan)	DHV (Korea)	AV (Japan)	FMDV (UK)
Broad bean wilt virus (South Korea)	31.9																							
Cowpea mosaic virus (USA)	39.5	46.4																						
Apple latent spherical virus (Japan)		32.5	32.2																					
Parsnip yellow fleck virus (UK)	29.6	26.8	30.6	34																				
Maize chlorotic dwarf virus (USA)	26.5	26.3	30.4	32.2	34.5																			
Rice tungro spherical virus (USA)	28.8	26.1	30	31.7	38.5	61.3																		
Kashmir bee virus (USA)	23.4	23.5	23.5	24	27.1	24.9	27																	
Cricket paralysis virus (USA)	24.5	23.2	24.8	27.7	24.3	25.9	26.7	38.5																
Drosophila C virus (Australia)	24.8	22.7	24.8	25.8	26.1	27.3		39.2	66.2															
Lygus lineolaris virus (USA)	20.6	22.8	22.7	25	25.9	24.2	26.3	22.7	22.2	24.6														
Tomato matilda virus (Australia)	**24.6**	**25.7**	**24.2**	**29**	**28.5**	**27.9**	**28.9**	**29**	**31**	**31.9**	**40.4**													
Picorna-like virus (USA)	23.9	23.5	25.4	26.6	26.4	23.5	24.6	24.5	27.7	29.3	38	**39.4**												
Sacbrood virus (South Korea)	24.9	25.4	24.9	27.3	26.8	27.2	26.3	26.6	28.6	28.4	43.9	**47**	46.1											
Sacbrood virus (China)	24.6	25.2	24.9	27.3	26.8	27.2	26.1	26.1	28.6	28.4	42	**47**	46.9	97.9										
Sacbrood virus (UK)	24.9	25.7	25.2	27.1	26.4	27.5	26.3	26.1	28.6	28.1	49.3	**46.3**	46.1	97.4	97.4									
Slow bee paralysis virus (UK)	24.1	25.7	23.5	26.6	25.7	25.8	27	26.7	24.1	25.8	27.3	**28.7**	30.2	30.4	30.9	30.4								
Brevicoryne brassicae	23.	2	2	2	2	2	2	2	2	31	3	**32**	2	3	3	3	3							

Virus																								
picorna-like virus (UK)	6	4.7	4.9	7.3	7.8	8.3	5.5	5.4	6.8	.1	1.4	**.1**	9.2	3.7	4	4	5.9							
Varroa destructor virus-1 (Netherlands)	21.5	25	24.5	25	23.3	25.8	28.6	25.9	27.8	26.9	30	**30.7**	28.3	35	35.2	35.5	35.9	42.9						
Deformed wing virus (UK)	21	25.2	25.2	24.8	22.9	26	28.3	25.2	28.3	26.9	30	**30.1**	28.1	35.2	35.5	35.7	40.6	43.1	96.9					
Deformed wing virus (USA)	21	24.8	25	24.5	22.4	26	27.9	25.2	28.3	27.4	30	**30.5**	28.1	35.2	35.5	35.7	40.3	43.1	96.4	99				
Kakugo virus (Japan)	20.5	25	24.8	24.8	22.7	26	28.3	25.8	27.8	26.7	30	**30.2**	28.1	35.5	35.7	35	40	42.9	96.2	98.8	98.8			
Duck hepatitis virus (South Korea)	16.1	17.4	18.3	15.9	18.3	20.8	21.5	18.6	18.3	18.6	18	**19.6**	16.5	19.2	19.5	19	18.8	17.2	19	19	19	19.1		
Aichi virus (Japan)	18.4	17.5	17.5	18.4	20.7	18.7	20.6	19.6	20.5	19.3	21.4	**20.7**	21.4	21.1	21.1	21.1	18.2	20.3	19.1	19.1	19.4	19.4	21.5	
Foot-and-mouth disease virus (UK)	22.1	20.3	22.6	19.3	21.2	19.4	20.8	20.3	21.5	21.7	19	**21.5**	20.8	20.3	20.3	21	18.4	19.9	20.5	20.2	20.7	20.5	19.1	28.2

a

Structural protein — Non structural protein

Walker A motif / Walker B motif

5' UTR 1-104

RhV-like 1, 921-1281 nt
RhV-like 2, 1713-2184 nt
RhV-like 3, 2574-2931 nt
Possible Vpg, 4077-4017 nt?
RNA helicase, 4122-4442 nt
Possible Vpg, 6072-6144 nt?
Cystein protease, 6183-6705 nt
Nucleotide binding site, 6684-6691?
RNA dependent RNA polymerase
3' UTR 8393-8506 nt

b

ORF 2,787 aa residues 314KDa

RhV-like 1, 307-447 aa 13.7 KDa
RhV-like 2, 571-728 aa 17.9 KDa
RhV-like 3, 858-977 aa 13.7 KDa
RNA helicase 1,374-1,480 aa
Cystein protease 2,061-2,235 aa
RNA dependent RNA polymerase 2,316-2,782 aa