Virus Research 160 (2011) 256–263

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Detection of members of the *Tombusviridae* in the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA

Kay Scheets[a,*], Olga Blinkova[b,1], Ulrich Melcher[b], Michael W. Palmer[a], Graham B. Wiley[c,2], Tao Ding[b], Bruce A. Roe[c]

[a] Department of Botany, Oklahoma State University, Stillwater, OK 74078, USA
[b] Department of Biochemistry & Molecular Biology, Oklahoma State University, Stillwater, OK 74078, USA
[c] Department of Chemistry & Biochemistry, University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

Article history:
Received 26 April 2011
Received in revised form 27 June 2011
Accepted 29 June 2011
Available online 6 July 2011

Keywords:
Tombusviridae
Carmovirus
Tombusvirus
Ecogenomics
Genomic sequence
Fern host

ABSTRACT

Viruses are most frequently discovered because they cause disease in organisms of importance to humans. To expand knowledge of plant-associated viruses beyond these narrow constraints, non-cultivated plants of the Tallgrass Prairie Preserve, Osage County, Oklahoma, USA were systematically surveyed for evidence of the presence of viruses. This report discusses viruses of the family *Tombusviridae* putatively identified by the survey. Evidence of two carmoviruses, a tombusvirus, a panicovirus and an unclassifiable tombusvirid was found. The complete genome sequence was obtained for putative TGP carmovirus 1 from the legume *Lespedeza procumbens*, and the virus was detected in several other plant species including the fern *Pellaea atropurpurea*. Phylogenetic analysis of the sequence and partial sequence of a related virus supported strongly the placement of these viruses in the genus *Carmovirus*. Polymorphisms in the sequences suggested existence of two populations of TGP carmovirus 1 in the study area and year-to-year variations in infection by TGP carmovirus 3.

1. Introduction

Many species of plant viruses remain to be discovered due to relative lack of attention to viruses of plants from non-cultivated areas (Cooper and Jones, 2006; Wren et al., 2006). A project (Plant Virus Biodiversity and Ecology, PVBE) to uncover such viruses in plants of the Tallgrass Prairie Preserve (TPP, The Nature Conservancy, Oklahoma; Allen et al., 2009; Hamilton, 2007) was initiated (Melcher et al., 2008) and shown to result in the discovery of evidence of several novel viruses (Muthukumar et al., 2009; Roossinck et al., 2010). We report here the results of searches of the recovered sequences for evidence of members of the family *Tombusviridae*.

Known family *Tombusviridae* members are divided into eight genera (*Dianthovirus*, *Tombusvirus*, *Aureusvirus*, *Avenavirus*, *Carmovirus*, *Necrovirus*, *Panicovirus* and *Machlomovirus*) (Fauquet et al., 2005). Their particles are icosahedral and contain genomes consisting of one or two strands of positive-sense RNA. Their RNAs are uncapped (Kneller et al., 2006) and lack 3′ poly(A) tails (Fauquet et al., 2005). The viruses of the family *Tombusviridae* do not encode a viral helicase, and their RNA-dependent RNA polymerases (RdRPs) are expressed via suppression of a stop codon, or, for dianthoviruses, by a frameshift. Open reading frames (ORFs) initiating in the 3′ half of the genomes are expressed through production, depending on genus, of one (Castaño and Hernández, 2005; Castaño et al., 2009; Kinard and Jordan, 2002) or two subgenomic RNAs (sgRNAs). The number of sgRNAs produced has been proposed as a characteristic useful in taxonomic assignment within the genus *Carmovirus* (Castaño et al., 2009). Those producing only one sgRNA have been proposed to be placed in a new genus, designated "pelarspovirus". Although members of the family *Tombusviridae* are thought to infect plants exclusively, recently a complete viral sequence related to members of the family *Tombusviridae* and umbraviruses was assembled (Culley et al., 2007) from a coastal water metagenomic study of the Straits of Georgia. This virus lacked genes for a coat protein (CP) or movement protein (MP) suggesting it does not infect higher plants. The PVBE project uncovered evidence from plants of two carmoviruses (for discussion purposes named TGP carmoviruses 1 and 3), one tombusvirus (TGP tombusvirus 1), one unclassifiable tombusvirid (TGP tombusvirid 1) described here, and one panicovirus (Scheets, in preparation).

* Corresponding author. Department of Botany, 301 PS, Oklahoma State University, Stillwater, OK 74078, USA. Tel.: +1 405 744 5348; fax: +1 405 744 7074.
E-mail addresses: kay.scheets@okstate.edu (K. Scheets), blinkova.olga@gmail.com (O. Blinkova), u-melcher-4@alumni.uchicago.edu (U. Melcher), mike.palmer@okstate.edu (M.W. Palmer), wileyg@omrf.org (G.B. Wiley), tao.ding@okstate.edu (T. Ding), broe@ou.edu (B.A. Roe).
[1] Present address: Computercraft Corporation, 1360 Beverly Rd., McLean, VA 22101, USA.
[2] Present address: Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA.

0168-1702/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2011.06.023

2. Materials and methods

2.1. Sample collection and tissue processing

Collection, identification, and archiving of plant samples were described earlier (Melcher et al., 2008). Sample identification numbers specify collection year with the first two digits (ex. 05TGP00369 was collected in 2005). Processing of samples by preparation of viral particle fractions from which nucleic acid was isolated (VLP-VNA) and randomly amplified was also described therein. For the samples presented here, sequences from the randomly amplified cDNA were obtained by the 454 pyrosequencing method using a panel of 96 four-nucleotide tags as described by Roossinck et al. (2010). This method was also used for sequencing of DNA derived from dsRNA.

For analysis of bulked plant material, a geographically identified square of land 30 cm × 30 cm was denuded of above-ground plant material which was stored in bags for transport to the laboratory where it was frozen at −80 °C. For preparation of virus like particles, 10 g of mixed plant sample was transferred to a metallic flask and homogenized by using a homogenizer (Biospec, Bartlesville, OK). The homogenate was poured into 250 ml centrifuge bottles, and 75 ml of citrate buffer (Melcher et al., 2008) and 0.63 ml of 0.25 M iodoacetamide were added. After 10 min on ice, the bottles were centrifuged for 15 min at 12,000 × *g*. Of the supernatant, 40 ml were transferred to a 50 ml plastic tube, to which was added 3.8 ml of Triton X-100. Then, 43 ml of the mixture were transferred to a Beckman ultracentrifuge tube (94 ml tubes) and underlaid with 13 ml of 20% sucrose solution in citrate buffer. Centrifugation proceeded at 70,000 × *g* for 11 h at 21 °C in a Beckman Ti 45 rotor. The supernatant was decanted slowly and the pellet resuspended in 200 μL of 0.5× citrate buffer by pipetting. After centrifugation at 8000 × *g* for 10 min, 150 μl of the supernatant was transferred to an ultracentrifuge tube and underlaid with 50 μl of 20% sucrose solution in 0.5× citrate buffer. Final ultracentrifugation was at 150,000 × *g* for 65 min at 21 °C in the Beckman Ti 42.2 rotor. The supernatant was removed by a pipettor and the pellet was resuspended in 250 μl of viral resuspension buffer. The VLP fraction was treated for virus nucleic acid isolation as previously described (Melcher et al., 2008). The procedure was repeated for a total of three 10 g aliquots of the mixed plant sample.

2.2. Sequence processing

All contigs assembled by the 454 process pipeline were used as BLASTn and tBLASTx (Altschul et al., 1997) queries of the nr/nt database and as BLASTx queries of the nr protein database. All contigs assembled by the 454 process pipeline in the project were formatted for BLAST searching by formatdb. This database was queried by BLASTn using viral contigs identified from the initial screening as queries and by tBLASTn using protein sequences encoded by reference genomes of members of the family *Tombusviridae* as queries. The complete contigs from the TPP project metagenome are available at the MG-RAST (Meyer et al., 2008) server (http://mg-rast.mcs.anl.gov/mg-rast/FIG/linkin.cgi?metagenome=4444009.3). Assembled sequences of putative viruses inferred from the contig data were used to identify additional sequences of these viruses among unassembled sequences. The unassembled sequences were also screened with GenBank sequences for Maize necrotic streak virus (MNeSV, NC_007729) and *Cucumber Bulgarian latent virus* (CBLV, NC_004725). These singleton sequences were used in building the consensus sequences for the putative viruses. Because of the chance that an occasional read was assigned mistakenly to the wrong plant sample in a sequencing pool, only those plants belonging to species for which more than 0.5% of specimen reads and more than 8 reads belonged to the virus were considered as positive for the virus, unless it could be established that the sequences were unlikely to have arisen from contamination by a virus source in the pool. Assembled viral sequences were deposited in GenBank (IDs: NC_015227, JF437881, JF437885, HM640931, HM640936, and three pending submissions JN122349–51). Shorter sequences and a consensus sequence of TGP carmovirus 3 are provided in Supplementary files. An internal gap in the assembled sequence of TGP carmovirus 3 was filled by sequencing PCR amplicons made using primers from sequences flanking the gap with cDNA template prepared from VLP-VNA from specimen 08TGP00078 and primers 5′-AAA TGT TAT TCT ATG GCG ACT ACT-3′ and 5′-TAA TGC ACA CTA CCA CTA TTT CCT-3′.

2.3. RNA secondary structure analysis

RNA secondary structure predictions were performed with mfold version 2.3 at 28 °C (http://mfold.rna.albany.edu/?q=mfold) (Zuker, 2003) on TGP carmovirus 1 and the 17 carmoviruses with complete genomes in GenBank: Angelonia flower break virus (AnFBV, NC_007733), Calibrachoa mottle virus (CbMV, GQ244431), *Carnation mottle virus* (CarMV, NC_001265), *Cardamine chlorotic fleck virus* (CCFV, NC_001600), *Cowpea mottle virus* (CPMoV, NC_003535), *Hibiscus chlorotic ringspot virus* (HCRSV, NC_003608), Honeysuckle ringspot virus (HoRV, NC_014967), *Japanese iris necrotic ring virus* (JINRV, NC_002187), *Melon necrotic spot virus* (MNSV, NC_001504), Nootka lupine vein-clearing virus (NLVCV, NC_009017), *Pea stem necrosis virus* (PSNV, NC_004995), Pelargonium chlorotic ring pattern virus (PCRPV, NC_005985), *Pelargonium flower break virus* (PFBV, NC_005286), Pelargonium line pattern virus (PLPV, NC_007017), *Saguaro cactus virus* (SgCV, NC_001780), Soybean yellow mottle mosaic virus (SYMMV, NC_011643), *Turnip crinkle virus* (TCV, NC_003821). Three regions of the 5′ ends of viral RNA and the large sgRNA were folded for each genome: to the first AUG, the second AUG, and the first 120 nt. The complete 3′ UTRs were also folded. The recombinant necrovirus *Galinsoga mosaic virus* (GaMV, NC_001818) was also compared since it was misleadingly listed as a carmovirus (Fauquet et al., 2005).

2.4. Phylogenetic analysis of carmovirus coat proteins

Phylogenetic analysis of CP amino acid sequences of carmoviruses was performed on the CPs of carmoviruses with complete genomes, GenBank accessions of Elderberry latent virus (ELV, AAK74061.1), Pelargonium ringspot virus (PelRSV, AAK74063.1), and the non-carmoviral CP most similar to TGP carmovirus 1, Pelargonium necrotic spot virus (PeNSV, NP_945116.1). Aligned sequences were adjusted manually in a progressive fashion (Melcher, 1990) using parsimony trees generated by Protpars of the Phylip 3.66 package (Felsenstein, 1989) as guide. The adjusted alignment was used in Prottest (Abascal et al., 2005) to generate a maximum likelihood distance tree using PhyML and the LG scoring matrix (Le and Gascuel, 2008). The aligned sequences were also used to generate bootstrap confidence levels using Seqboot, Protdist and Consense of the Phylip package.

3. Results

3.1. Sequence identification and distribution

Contigs were identified with best similarity to sequences from three *Tombusviridae* genera: *Tombusvirus, Panicovirus* and *Carmovirus*. A single contig from the grass *Digitaria cognata* (fall witchgrass) and thus a single virus could be assigned to the *Tombusvirus* genus. For discussion purposes, we designate it TGP tombusvirus 1. The panicovirus Thin paspalum asymptomatic

Table 1
Members of the *Tombusviridae* identified in Tallgrass Prairie Preserve specimens.

Plant ID[a]	Plant family	Plant species	Virus nickname	% of total reads	Easting[b]	Northing[b]
05TGP00121	Pteridaceae	*Pellaea atropurpurea*	TGP carmovirus 1	1.4	734812	4076796
06TGP01062[c]	Fabaceae	*Melilotus officinalis*	TGP carmovirus 1	2.2	730589	4081049
06TGP01091	Fabaceae	*Lespedeza procumbens*	TGP carmovirus 1	59.4	734005	4078978
06TGP5030(2)	Mixed	Mixed[d]	TGP carmovirus 1	2.7	727250	4078545
06TGP5030(4)	Mixed	Mixed[d]	TGP carmovirus 1	8.3	727250	4078545
06TGP5030(6)	Mixed	Mixed[d]	TGP carmovirus 1	4.9	727250	4078545
05TGP00599	Juglandaceae	*Carya cordiformis*	TGP tombusivirid 1	1.5	729524	4080624
05TGP00009	Poaceae	*Sporobolus compositus*	TGP carmovirus 3	0.1	728417	4081476
05TGP00564	Rosaceae	*Amelanchier arborea*	TGP carmovirus 3	0.3	728559	4081444
06TGP01135	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.1	728997	4077015
07TGP00096	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.8	731021	4074000
07TGP00120	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.2	728000	4078000
07TGP00129	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.4	730000	4086000
07TGP00223	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.8	727000	4080000
08TGP00078	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	3.9	730030	4076001
08TGP00175	Asteraceae	*Ambrosia psilostachya*	TGP carmovirus 3	0.7	733995	4073990
05TGP00369	Poaceae	*Paspalum setaceum*	TPAV	79.6	730780	4082252
07TGP00199[c]	Poaceae	*Digitaria cognata*	TGP tombusvirus 1	0.2	732000	4075000

[a] First two digits indicate collection year.
[b] UTM, Sector 14, using NAD27.
[c] Plant 06TGP01062 exhibited yellowing of leaves; plant 07TGP00199 exhibited red spots on leaves; all others had no noticeable abnormalities.
[d] Mixed population consisting of *Baptisia australis*, *Ceanothus americanus*, *Croton monanthogynus*, *Elymus virginicus*, *Lespedeza procumbens*, *Monarda fistulosa*, *Physalis pumila*, and *Sporobolus compositus*. Three separate aliquots of the mixed population were tested.

virus (TPAV) is the subject of a separate communication detailing its complete sequencing and demonstration of infectivity of a cloned representative (Scheets, in preparation). The single sample contributing sequence information for this virus was the grass *Paspalum setaceum*, accounting for close to 80% of the total sequence reads for this plant specimen (Table 1).

The assumption that non-overlapping contigs obtained from the same plant specimen derive from the same viral species leads to the deduction of the presence of two carmoviruses in TPP plants. TGP carmovirus 1 accounted for 64% of sequence reads from a 2006 specimen of the perennial legume *Lespedeza procumbens* (trailing lespedeza) and 2.2% and 1.4% of the sequence reads of one of eight 2006 specimens of the annual legume *Melilotus officinalis* and a 2005 specimen of the perennial fern *Pellaea atropurpurea*, respectively. The latter species was not sampled in other years. TGP carmovirus 3 was found in 2005 in single samples of *Amelanchier arborea* and *Sporobolus compositus*. Recovery of further TGP carmovirus 3 sequences occurred from one 2006, four 2007, and two 2008 specimens of *Ambrosia psilostachya* (western ragweed), a frequently sampled species (22, 19, 20 and 19 samples, respectively in 2005, 2006, 2007 and 2008). Further analysis of a 186 nt putative RdRP sequence initially identified as TGP carmovirus 2 indicated that there was not enough sequence to identify the genus, and the putative virus was thus renamed TGP tombusvirid 1. This virus was detected solely from a single host, a 2005 *Carya cordiformis* (bitternut hickory) specimen. In the regions of nucleic acid sequence overlap (Fig. 1), TGP tombusvirid 1 differed from TGP carmovirus 3 in 70% of the 186 nucleotide residues, and from TGP carmovirus 1 at 72% of the positions, while TGP carmoviruses 1 and 3 shared 47% of residues in this same region. Of the 15 plant specimens positive for tombusvirus family sequences, only two (06TGP01062, *M. officinalis* and 07TGP00199, *D. cognata*) exhibited readily visible features that may have been symptoms of infection.

3.2. Sequence features

The complete 3929 nt sequence of TGP carmovirus 1 (GenBank ID: NC_015227) was contained within sample 06TGP01091 (Fig. 1). Characteristic carmovirus ORFs for a 28 kDa replicase associated protein (nt 27–773), an 88 kDa RdRP (nt 27–2339), and a 37 kDa CP (nt 2648–3688) were present. Synthesis of full-length RdRP would require readthrough of a UAG stop codon, a feature suggested by the presence of a consensus Group 3 (GGG) readthrough signal (Harrell et al., 2002). Two carmovirus MP ORFs (7 kDa and 8.8 kDa) were located at nt 2312–2515 and 2412–2651. Four additional ORFs potentially encoding polypeptides of about 100 residues were found. Their predicted amino acid sequences were not sufficiently similar to any protein sequence in the databases to produce a hit. New viral proteins may evolve by "overprinting" a known gene (Rancurel et al., 2009), and one of the two novel ORFs of HCRSV was shown to be expressed during replication and to be required for host-specific replication in kenaf (Liang et al., 2002). Thus, it is possible that one or more of these small ORFs encode functional proteins. The proposed MP2 ORF lacked an AUG for initiation of translation, as do the equivalent ORFs of the two proposed pelarspoviruses for which sequences were available in this region: PLPV and PCRPV. Translation of the equivalent p9.7 PLPV protein occurs from a GUG codon. Mutation of the GUG codon to a CUG codon preserved the translation initiation function (Castaño et al., 2009). A CUG codon in TGP carmovirus 1 RNA is the first UG-containing codon downstream of the RdRP UAG termination codon and is positioned correctly to give a polypeptide with sequence similarity to p9.7 of PLPV. PLPV, PCRPV and TGP carmovirus 1 have no AUG codons between the weak-context MP1 initiation codon and the CP-initiating AUG. This arrangement is similar to that in the panicoviruses *Panicum mosaic virus* (Turina et al., 2000), Cocksfoot mild mosaic virus (Ziegler et al., 2009), and TPAV (unpublished results) and the machlomovirus *Maize chlorotic mottle virus* (Scheets, 2000). These four viruses produce only one large sgRNA and also contain ORFs related to the second carmovirus MP using either CUG or GUG non-canonical start codons (Scheets, 2000; Turina et al., 2000; Scheets, unpublished results). In contrast, the 15 other carmoviruses with complete genome sequences encode one to eight additional AUGs in this region, which would require two sgRNAs for efficient expression of CP.

RNA secondary structure predictions were obtained for TGP carmovirus 1 and the 17 carmoviruses with complete genomes in GenBank to determine if any structures that could support 5′–3′ interactions and/or might support cap-independent translation or help regulate replication could be identified. Similar structures have been identified or predicted for many other viruses in the family *Tombusviridae* (Miller and White, 2006; Na and White, 2006).

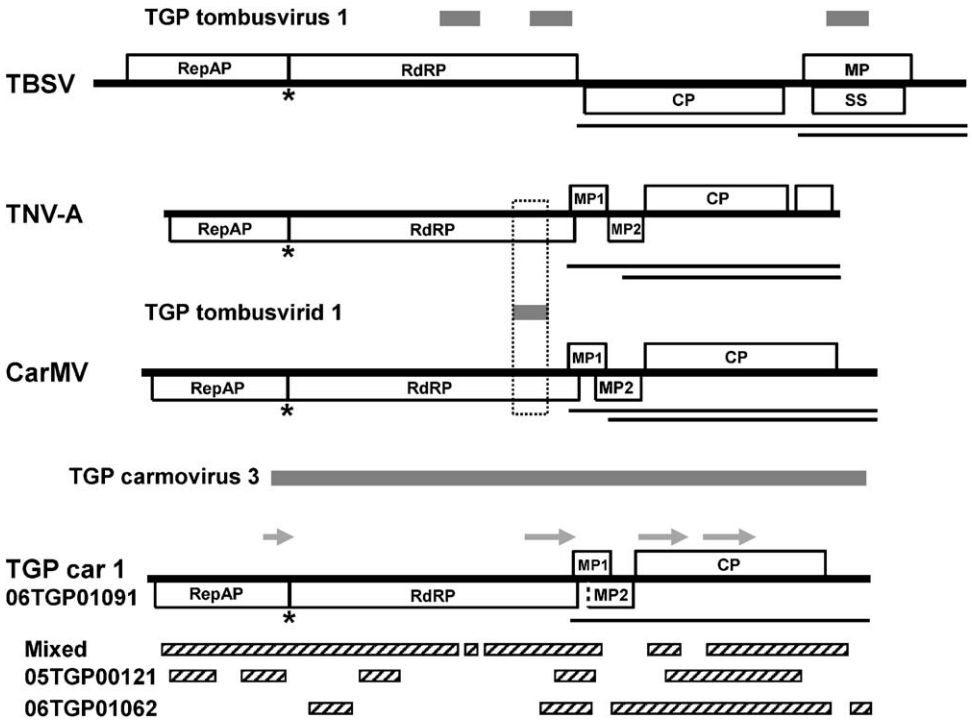

Fig. 1. Positions of TPP sequences from the family *Tombusviridae* relative to genomes of characterized family members. Diagram of the genome structure of TGP carmovirus 1 compared to the type viruses for the genera *Tombusvirus*, *Necrovirus*, and *Carmovirus*: TBSV (Tomato bushy stunt virus); TNV-A (Tobacco necrosis virus-A); CarMV (Carnation mottle virus). Thick black lines indicate the genomic RNAs and thin black lines are sgRNAs. Open boxes mark the ORFs for replicase associated protein (RepAP), RNA dependent RNA polymerase (RdRP), movement protein (MP), coat protein (CP), and suppressor of silencing (SS). Asterisks mark the suppressible stop codon in the RdRP ORFs. Wide gray lines mark the sequence fragments of TGP tombusvirus 1, TGP tombusvirid 1, and the consensus sequence of TGP carmovirus 3 aligned with the homologous regions of the type genomes. The box with dotted lines indicates that TGP tombusvirid 1 aligned with both necrovirus and carmovirus sequences. For TGP carmovirus 1 (TGP car 1) the complete 3929 nt genome assembled from 06TGP01091 is shown. Gray arrowed lines indicate ORFs for proteins with no similarity to known proteins. The dashed line beginning MP2 ORF marks the noncanonical start codon. Assembled contigs and singletons from 06TGP05030 (Mixed), 05TGP00121 and 05TGP01062 are aligned below 06TGP01091 (diagonally striped bars).

Analysis of the 241 nt 3′ UTR of TGP carmovirus 1 using mfold (Zuker, 2003) predicted several structures corresponding to structures known to be involved in control of replication or translation of TCV, which are found in the majority of carmoviruses (Yuan et al., 2010). Three stem-loop (SL) structures that correspond to the location and general structures of H4b, H5, and the (−) strand promoter Pr of TCV were predicted (Fig. 2). The 3′ terminal bases (GCCC) comprise the 3′ complementary silencer sequence which can pair with the replication silencer element in H5 to form a pseudoknot (Na and White, 2006). Carmoviruses fall into several different groups when examining RNA secondary structures for possible 5′–3′ base pairing interactions (Wang et al., 2011; Yuan et al., 2010). Like MNSV and SYMMV, the 5′ UTRs of TGP carmovirus 1 and its predicted sgRNA (26 and 23 nt, respectively) fold into short SLs beginning at their 5′ ends, and both contain identical terminal loop sequences capable of base-pairing with the terminal loop of a 60 nt SL near the beginning of the 3′ UTR (Fig. 2). These structures may be involved in cap-independent translation (Miller and White, 2006), but the 3′ SL of TGP carmovirus 1 does not have the consensus sequence identified as the eIF4F binding region for MNeSV 3′CITE and MNSV-264 (Nicholson et al., 2010).

A 3233 nt consensus sequence of TGP carmovirus 3, estimated to cover over 80% of the genome (Fig. 1 and Supplemental material), was assembled from reads from several plants and the sequence of a PCR product amplified from 08TGP00078 cDNA. Complete ORFs for both MPs, the CP and the readthrough portion of the RdRP were present. The translation readthrough signal UAGGGG present in TGP carmovirus 1 was also present in the TGP carmovirus 3 sequence. The MP1 (6.6 kDa) ORF had its initiation codon 28 nt 5′ of the RdRP termination codon. The other MP ORF lacked an AUG, but could be predicted to use a UUG or CUG 52 or 61 nt 3′ of the RdRP termination codon (encoding proteins of 8.9 or 8.7 kDa, respectively). Consequently, the two MP ORFs overlapped for 92 or 101 nt, consistent with arrangements in other carmoviruses. As seen in TGP carmovirus 1 and proposed pelarspovirus genomes, no AUGs occurred between the RdRP termination codon and the CP initiation codon.

To explore the probable taxonomic status of TGP carmoviruses 1 and 3, their encoded full length or partial RdRP sequences were used as queries of the protein RefSeq database using BLASTp. The species demarcation criterion (Fauquet et al., 2005) for the RdRP is less than 52% identity for this sequence. Comparison of the proteins for TGP carmovirus 1 and its closest relative AnFBV, determined by BLASTp search, resulted in 40.6% of the residues being identical. Similarly, for a comparison of the available RdRP sequence for TGP carmovirus 3 and the comparable region of PCRPV, the residue identity was 43.2%. In the more conserved readthrough region of the RdRP coding sequence, TGP carmovirus 1 differed from TGP carmovirus 3 at 51.3% of residues. The results suggest that both TGP carmoviruses 1 and 3 were distinct members of the genus *Carmovirus.*

The TGP tombusvirid 1 fragment encoded 62 aa of the RdRP readthrough region which showed 39% identity to GaMV as top BLASTp hit followed by two carmoviruses and the necrovirus Tobacco necrosis virus-A as next best hits. GaMV appears to be a recombinant virus with nonstructural ORFs most closely related to *Furcraea necrotic streak virus*, another recombinant virus, and necroviruses, while its CP is related to the aureusvirus *Pothos latent virus* and tombusviruses (Ciuffreda et al., 1998, and data not shown). Thus, the taxonomic placement of TGP tombusvirid 1 is unclear.

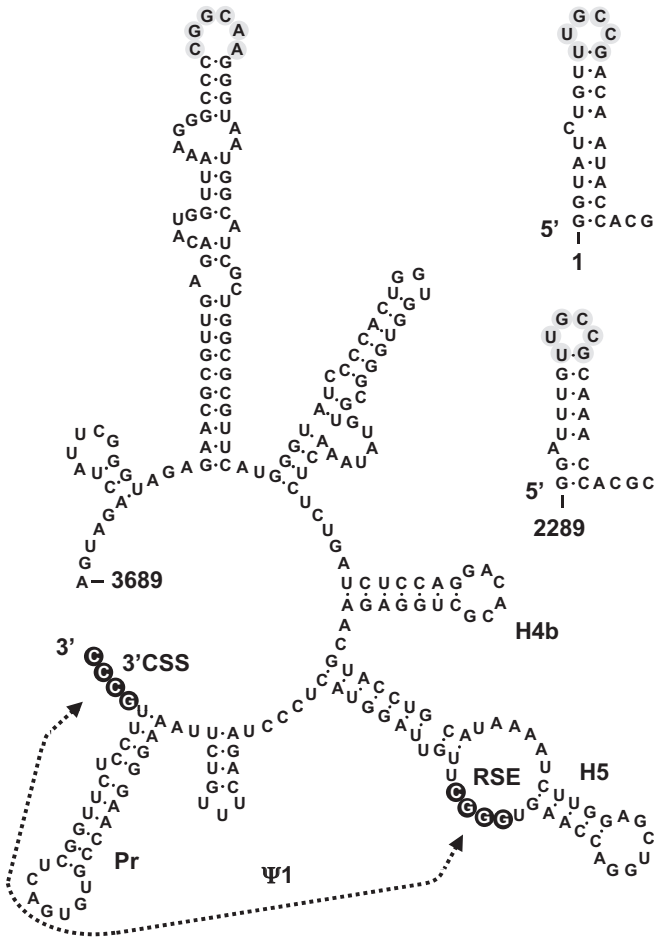

Fig. 2. Predicted RNA secondary structure of TGP carmovirus 1 UTRs. Structures were predicted by mfold, and numbering refers to genomic RNA sequence. Dotted line indicates base-pairing of the 3′-complementary silencer sequence (3′CSS) (white letters in black circles) with similarly marked bases in the replication silencer element (RSE) of SL H5 to form Pseudoknot1 (Ψ1). Structures similar to the TCV (−) strand promoter (Pr) and SL H4b are marked. Bases in the loop of the longest SL (black letters in gray circles) can potentially base pair with the similarly marked bases of the 5′ UTR and the predicted sgRNA 5′ UTR shown on the right.

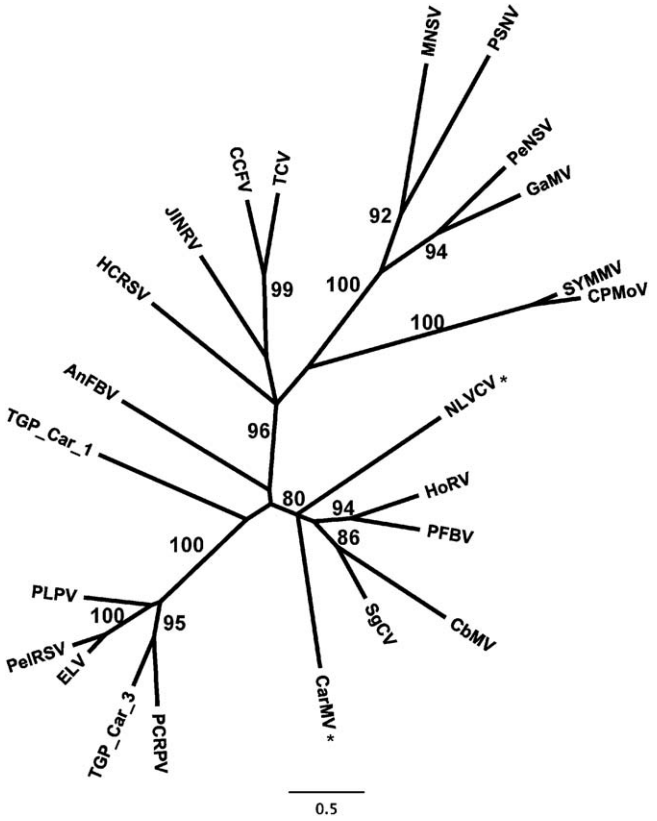

Fig. 3. Phylogenetic relationships for amino acid sequences of CPs of carmoviruses. Tree shown illustrates maximum likelihood distance relationships as determined by Prottest using the LG scoring matrix assuming some invariant sites and gamma distributed rates. Numbers on branches represent percent bootstrap support obtained by parsimony analysis using applications in the Phylip package. Maximum likelihood branches with less than 80% bootstrap parsimony support are unnumbered. Viruses designated with * were grouped with 87% bootstrap support. *Abbreviations used*: TGP_car_1, TGP carmovirus 1; TGP_car_3, TGP carmovirus 3.

The TGP tombusvirus 1 contig (HM640936) and two singletons (Supplemental material) were used for sequence comparisons. Two regions encoding 69 and 74 aa were derived from the RdRP readthrough ORF (Fig. 1), while a third fragment covered part of the overlapping MP/suppressor of silencing ORFs (73 and 74 aa). The most 5′ fragment encodes RdRP motifs B and C of the active site while the second fragment is within the thumb domain (O'Reilly and Kao, 1998). BLASTp searches indicate the active site region is most closely related (78% identity) to MNeSV, the only other grass-infecting tombusvirus, while the downstream RdRP fragment is most similar (81% identity) to CBLV, another smaller-than-typical tombusvirus. The suppressor of silencing fragment, representing about 44% of a 19 kDa protein, is most similar to that of MNeSV (75% identity) while the overlapping MP sequence is least similar to that of MNeSV and most similar to that of *Grapevine Algerian latent virus* (86% identity). This suggests that TGP tombusvirus 1 is a new grass-infecting tombusvirus.

Because of the availability of more carmovirus CP sequences than of those for other carmoviral proteins, the CP predicted amino acid sequences were used to explore the position of the two TGP carmoviruses in phylogeny (Fig. 3). The topology of the maximum likelihood tree differed from that derived by bootstrapped parsimony in one significant (greater than 75% bootstrap support) respect. Parsimony assigned CarMV and NLVCV to the same clade with 87% support while maximum likelihood assigned them to separate branches. As expected from the BLASTp search results, TGP carmovirus 3 CP branched with the PCRPV CP, and the pair clustered together with the CPs of other proposed pelarspoviruses PLPV, ELV, and PelRSV. TGP carmovirus 1 CP was located basally to this clade in the maximum likelihood tree, but the basal location was not strongly supported by other analyses. Six other CPs, those of CbMV, CarMV, SgCV, PFBV, HoRV and NLVCV, completed a clade well separated from the remaining carmovirus species. The tombusvirus with an aureusvirus-like CP, PeNSV, branched together with GaMV, which is consistent with the report that GaMV is a recombinant necrovirus with an aureusvirus or tombusvirus CP (Ciuffreda et al., 1998).

In 2006, an attempt was made to enlarge the scope of the metagenomic analysis by analyzing a mixture of plant parts representing all the above ground plant material in a defined area, referred to as the lawnmower approach. Three aliquots of plant material from this harvest were processed separately. All three revealed evidence of the presence of TGP carmovirus 1 (Table 1 and Fig. 1). Eight plant species comprised the plant mixture, one of which was *L. procumbens*.

3.3. Polymorphisms and populations

A scan of the aligned TGP carmovirus 1 sequences obtained from each virus positive plant sample and the mixture samples (Fig. 1) revealed that at least 295 of 3929 positions of TGP car-

Table 2
Differences between shared sequences of TGP carmovirus 1 isolates.

Plant ID	06TGP01091	06TGP01062	05TGP00121
Mixed	8.2% (3248)[a]	8.1% (1324)	8.1% (1463)
05TGP00121	0.7% (1630)	1.9% (925)	
06TGP01062	1.2% (1772)		

[a] % non-identity (total nt shared); determined by BLASTn comparisons.

movirus 1 exhibited nucleotide polymorphisms. That the bulk of these polymorphisms was not due to quasispecies variation or sequencing errors was supported by the observation that within-specimen comparisons at 5683 positions where comparisons were possible only uncovered 13 polymorphisms, 10 of which were associated with one specimen (05TGP00121). Table 2 compares the percentage of residues differing between pairs of specimens. The sequences from *P. atropurpurea* (05TGP00121), *M. officinalis* (06TGP01062) and *L. procumbens* (06TGP01091) differed from each other in less than 2% of the positions. All three of these sequences differed from the sequence of the mixed plant sequence in more than 8% of the positions. Interestingly, for specimen 05TGP00121, one of eleven positions whose nucleotide differed from that in the 06TGP01091 sequence resulted in alteration of the 06TGP01091 AUG initiation codon of MP1 to GUG, probably increasing the scanning by-pass ability of the ribosome at this position. The only noncoding nt from the partial sequences were in the 3′ UTR. The mixture showed three differences in the 91 nt shared with 06TGP01091, and two differences in the 15 nt shared with 06TGP01062, while there were no differences in the125 nt overlap between 06TGP01091 and 06TGP01062. Although the data do not lend themselves to careful population genetic analysis (due to unequal sampling of varying regions of sequence), they suggest strongly that the TGP carmovirus 1 population in the lawnmower plot was distinct from the other three. Similar to TGP carmovirus 1, evidence suggested that there were multiple strains of TGP carmovirus 3 in the TPP. In 3233 residues there were 2605 with sequences from more than one plant. Of these, 240 (or 9%) of the positions were polymorphic. Of the polymorphic positions, 61 were informatively so (Table 3). The sequences from the two 2008 *A. psilostachya* specimens shared the same residue at 86% of 35 common informative positions, while sequences from four 2007 *A. psilostachya* differed considerably from the 2008 samples but were quite similar to one another. A 2006 sample, 06TGP01135, from an *A. psilostachya* plant produced a sequence more resembling 2008 than 2007 samples. Sequences were available from two 2005 samples, 05TGP00009 and 05TGP00564, neither from an *A. psilostachya* plant. In a stretch of about 100 residues for which these sequences could be compared, there were 9 nt differences, suggesting that at least two strains were present in the TPP in 2005. At those positions that allow a distinction between 2007 and 2008 viruses, the 05TGP00564 sequence appeared more related to those from 2008 specimens, while the other sequence more resembled those from 2007 plants.

4. Discussion

With only five species, the paucity of species diversity in TPP members of the family *Tombusviridae* is paralleled by other families or genera of plant viruses in the TPP (*Sobemovirus, Tobamovirus*, for example, PVBE project, unpublished observations). Unlike the TPP plants, a diversity of viruses of the family *Tombusviridae* was present in the waters of Lake Needwood (Djikeng et al., 2009).

The level of polymorphisms noted for TGP carmoviruses 1 and 3 (Tables 2 and 3) both for the mixture specimen and the individual plant specimens, appeared lower than seen for the most predominant virus of the TPP, Asclepias asymptomatic virus (AsAV), a tymovirus (PVBE project, Min et al., submitted for publication, and unpublished observations). The statistical significance of this difference remains to be tested. Yet, the patterns of TGP carmovirus 1 nucleotide diversity suggest that their observed sequences result from two major bottleneck events, one giving rise to the strain of viruses in the mixed sample and the other giving rise to a strain consisting of all the other viruses. Since the location from which the mixed plant species sample was obtained was near the periphery of the sample locations, a biogeographical explanation for the apparent dual sources of this virus cannot be ruled out.

That the principal host for TGP carmovirus 3 was *A. psilostachya* made possible a year-by-year comparison of infection by strains of this virus because in each of four years about 20 specimens of this plant species were sampled. For the four other detected viruses, depth of sampling was too shallow to evaluate permanence. A temporal variation of strain appearance is suggested by the *A. psilostachya* results. In 2005, no *A. psilostachya* yielded evidence of the presence of TGP carmovirus 3, although single specimens of two other species had barely detectable amounts of a different strain (05TGP00564 and 05TGP00009) in each species. In 2006, strain 05TGP00564 appeared in one specimen of *A. psilostachya*. In 2007, four specimens of the plant species showed evidence of the 05TGP00009 strain. In 2008, the 05TGP00564 strain reappeared, being detected in two *A. psilostachya* plants. The significant sequence differences between 2007 and 2008 viruses could be consistent with numerous substitutions occurring from one year to the next. On the other hand, the observation that 2008-like viruses were detected in 2005 and 2006 is consistent with a model of reinitiation of infection by randomly selected strains in some years. That evolution may not be responsible for the difference calls into question the practice of using isolates from different years to calibrate the rate of evolution of RNA virus species (Harkins et al., 2009).

Castaño et al. (2009) suggested, based on the numbers of distinct sgRNAs, that carmoviruses with a single sgRNA be classified as a separate genus. TGP carmovirus 3, being closely related to PCRPV, is likely also a member of this clade. TGP carmovirus 1 shares sequence features with the single sgRNA cluster making it likely that also it produces a single sgRNA. The features include an absence of AUG codons in all frames between the MP1 and CP initiation codons, a non-canonical initiation codon for MP2 and

Table 3
Sharing of identities of informatively polymorphic positions among TGP carmovirus 3 sequences.

Plant ID	**05TGP00564**[a]	**06TGP01135**	07TGP00096	07TGP00120	07TGP00129	07TGP00223	**08TGP00078**	**08TGP00175**
05TGP00009	0/9[b]	1/2	0/0	0/0	0/0	9/11	3/11	4/11
05TGP00564		0/0	0/3	0/3	0/0	0/0	3/3	0/0
06TGP01135			0/0	5/18	0/0	0/1	14/19	1/1
07TGP00096				8/9	2/2	7/18	7/21	2/17
07TGP00120					0/0	9/12	3/21	2/11
07TGP00129						13/15	2/15	1/14
07TGP00223							5/36	4/35
08TGP00078								30/35

[a] The two sequence groupings discussed in the text are indicated by plant IDs in bold or non-bold font.
[b] Number of positions identical/number of polymorphic positions compared.

a phylogenetic position whose inclusion in the proposed pelarspovirus would not necessarily destroy the monophyletic nature of the clade. It is possible that the switch between double and single sgRNAs occurred only once during the early radiation of species of the family *Tombusviridae* resulting not only in the two TGP carmoviruses and the proposed pelarspovirus having a single sgRNA, but also the single sgRNA status of members of the genera *Panicovirus* and *Machlomovirus*. Without a reliable root to the tree, it cannot be decided whether a single sgRNA or a double sgRNA configuration is ancestral. It is also, of course, possible that this trait switch occurred multiple times during evolution of the *Tombusviridae*.

To the best of our knowledge, the TGP carmovirus 1 genome sequence is the first complete plant virus sequence assembled directly from an environmental survey sample. Other metagenomic studies have produced complete sequences of new viruses from initial data also; ten circular ssDNA viruses (1.1–3.2 kb) from an Antarctic lake (López-Bueno et al., 2009), and two unclassified *Circoviridae* (1.8 and 1.7 kb) from bat guano (Li et al., 2010). Two marine picorna-like virus genomes (8.8 and 9.2 kb) were assembled from a study of RNA viruses in coastal waters (Culley et al., 2006).

This is the first report of a carmovirus infecting a fern. There are only four other reports in the literature of viruses known to naturally infect ferns. Hart's tongue virus was tentatively identified as a tobravirus infecting *Phyllitis scolopendrium* (Hull, 1968), Fern potyvirus was found in *Polypodium vulgare* and *Dryopteris filix-mas* (Nienhaus et al., 1974), and *Adiantum pedatum* was found to be infected with *Cucumber mosaic virus* (Nameth and Steininger, 1997). More recently the phytovirus Japanese holly fern mottle virus was found in *Cyrtomium falcatum* (Valverde and Sabanadzovic, 2009). The *P. atropurpurea* (purple cliffbrake) sample infected with TGP carmovirus 1 was one of 21 samples of ferns and fern allies collected in the PVBE, representing 12 of the 18 species growing at the TGP. Initial analyses of the other 20 samples found AsAV in *Asplenium rhizophyllum* and five viruses with best tBLASTn similarities to various mycoviruses in two ferns and two fern allies. This shows that the PVBE survey approach potentially doubled the number of known viruses infecting ferns and potentially identified two viruses that infect fern allies.

That the number of viruses known to science can be substantially increased by sampling, without regard for symptoms, plants from non-cultivated environments for evidence of viruses was a central hypothesis of the PVBE project (Wren et al., 2006). The sampling of members of the family *Tombusviridae* reported here and other PVBE results (Melcher et al., 2008; Muthukumar et al., 2009; Roossinck et al., 2010) strongly support this view. The PVBE project was designed to be able to link an identified virus with a specific plant species (Roossinck et al., 2010). Additional metagenomic survey approaches sampling multiple plants simultaneously promise to provide a more global picture of viral diversity. In this study the lawnmower approach was tested as one way of assessing broader biodiversity questions. The discovery of TGP carmovirus 1 sequences in this preparation indicated that there is qualitative validity in the approach. However, the range of percentages of total reads attributable to the virus (2.7–8.3%) indicates that quantitative conclusions are not possible due to the difficulty of achieving homogeneity in sample aliquots. Other possible metagenomic approaches include surface water runoff from areas with plant litter (Djikeng et al., 2009); frass of herbivores (Zhang et al., 2006); and the contents of plant feeding insects.

Acknowledgements

This work was supported by the EPSCoR program of the National Science Foundation (EPS-0447262) and the Oklahoma Agricultural Research Station whose director has approved this manuscript. The authors thank Mukta Dutta for the PCR-derived TGP carmovirus 3 sequence, Marilyn Roossinck and colleagues for providing templates for sequencing dsRNA from individual plants, and Dan McGlinn, Vaskar Thapa, Fumiko Shirakura, Will Lowry, Matt Allen, Kelly Derennaux and Pete Earls who assisted OB, MWP, and GBW in plant collection.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2011.06.023.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.
Allen, M.S., Hamilton, R.G., Melcher, U., Palmer, M.W., 2009. Lessons from the Prairie: Research at The Nature Conservancy's Tallgrass Prairie Preserve. Oklahoma Academy of Sciences, Stillwater, OK.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Castaño, A., Hernández, C., 2005. Complete nucleotide sequence and genome organization of Pelargonium line pattern virus and its relationship with the family *Tombusviridae*. Arch. Virol. 150, 949–965.
Castaño, A., Ruiz, L., Hernández, C., 2009. Insights into the translational regulation of biologically active open reading frames of Pelargonium line pattern virus. Virology 386, 417–426.
Ciuffreda, P., Rubino, L., Russo, M., 1998. Molecular cloning and complete nucleotide sequence of Galinsoga mosaic virus genomic RNA. Arch. Virol. 143, 173–180.
Cooper, I., Jones, R.A.C., 2006. Wild plants and viruses: under-investigated ecosystems. Adv. Virus Res., vol. 67. Academic Press, pp. 1–47.
Culley, A.I., Lang, A.S., Suttle, C.A., 2006. Metagenomic analysis of coastal RNA virus communities. Science 312, 1795–1798.
Culley, A.I., Lang, A.S., Suttle, C.A., 2007. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. Virol. J. 4, 69.
Djikeng, A., Kuzmickas, R., Anderson, N.G., Spiro, D.J., 2009. Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One 4, e7264.
Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), 2005. Virus Taxonomy. Elsevier, London.
Felsenstein, J., 1989. PHYLIP – Phylogeny Inference Package. Cladistics 5, 164–166.
Hamilton, R.G., 2007. Restoring heterogeneity on the Tallgrass Prairie Preserve: applying the fire-grazing interaction model. In: Masters, R.E., Galley, K.E.M. (Eds.), Proceedings of the 23rd Tall Timbers Fire Ecology Conference: Fire in Grassland and Shrubland Ecosystems. Tall Timbers Research Station, Tallahassee, FL.
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.
Harrell, L., Melcher, U., Atkins, J.F., 2002. Predominance of six different hexanucleotide recoding signals 3′ of read-through stop codons. Nucleic Acids Res. 30, 2011–2017.
Hull, R., 1968. A virus disease of harts-tongue fern. Virology 35 (2), 333–335.
Kinard, G.R., Jordan, R., 2002. Genome organization of Pelargonium chlorotic ring pattern virus: further implications for *Tombusviridae* taxonomy. Acta Hortic. 568, 17–27.
Kneller, E.L.P., Rakotondrafara, A.M., Miller, W.A., 2006. Cap-independent translation of plant viral RNAs. Virus Res. 119, 63–75.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Li, L., Victoria, J.G., Wang, C., Jones, M., Fellers, G.M., Kunz, T.H., Delwart, E., 2010. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J. Virol. 84, 6955–6965.
Liang, X.-Z., Lucy, A.P., Ding, S.-W., Wong, S.-M., 2002. The p23 protein of *Hibiscus chlorotic ringspot virus* is indispensable for host-specific replication. J. Virol. 76, 12312–12319.
López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., Alcamí, A., 2009. High diversity of the viral community from an Antarctic lake. Science 326, 858–861.
Melcher, U., 1990. Similarities between putative transport proteins of plant viruses. J. Gen. Virol. 71, 1009–1018.
Melcher, U., Muthukumar, V., Wiley, G.B., Min, B.E., Palmer, M.W., Verchot-Lubicz, J., Nelson, R.S., Roe, B.A., Ali, A., Thapa, V., Pierce, M.L., 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. J. Virol. Methods 152, 49–55.
Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The

Metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386.
Miller, W.A., White, K.A., 2006. Long-distance RNA–RNA interactions in plant virus gene expression and replication. Annu. Rev. Phytopathol. 44, 447–467.
Min, B-E, Feldman, T., Ali, A.,Wiley, G., Muthukumar, V., Roe, B., Roossinck, M., Melcher, U., Palmer, M.W., Nelson, R.S. Physiological and ecological analysis of a novel tymovirus in *Asclepias viridis* (milkweed) from the Tallgrass Prairie Preserve of Oklahoma. Phytopathology, submitted for publication.
Muthukumar, V., Melcher, U., Pierce, M., Wiley, G.B., Roe, B.A., Palmer, M.W., Thapa, V., Ali, A., Ding, T., 2009. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. Virus Res. 141, 169–173.
Na, H., White, K.A., 2006. Structure and prevalence of replication silencer-3′ terminus RNA interactions in *Tombusviridae*. Virology 345, 305–316.
Nameth, S.T., Steininger, J., 1997. Cucumber mosaic virus associated with leaf malformation and stunting in *Adiantum pedatum*. Plant Dis. 81, 1333.
Nicholson, B.L., Wu, B., Chevtchenko, I., White, K.A., 2010. Tombusvirus recruitment of host translational machinery via the 3′ UTR. RNA 16, 1402–1419.
Nienhaus, F., Mack, C., Schinzer, U., 1974. The isolation of viruses from fern plants. Z. Pflkrankh. Pflsch. 81, 533–537.
O'Reilly, E.K., Kao, C.C., 1998. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structuresand computer predictions of secondary structure. Virology 252, 287–303.
Rancurel, C., Khosravi, M., Dunker, A.K., Romero, P.R., Karlin, D., 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into *de novo* protein creation. J. Virol. 83, 10719–10736.
Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., Chavarría, F., Shen, G., Roe, B.A., 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Mol. Ecol. 19, 81–88.
Scheets, K., 2000. Maize chlorotic mottle machlomovirus expresses its coat protein from a 1.47-kb subgenomic RNA and makes a 0.34-kb subgenomic RNA. Virology 267, 90–101.
Turina, M., Desvoyes, B., Scholthof, K.-B.G., 2000. A gene cluster encoded by *Panicum mosaic virus* is associated with virus movement. Virology 266, 120–128.
Valverde, R.A., Sabanadzovic, S., 2009. A novel plant virus with unique properties infecting Japanese holly fern. J. Gen. Virol. 90, 2542–2549.
Wang, Z., Parisien, M., Scheets, K., Miller, W.A., 2011. The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. Structure 19, 868–880.
Wren, J.D., Roossinck, M.J., Nelson, R.S., Scheets, K., Palmer, M.W., Melcher, U., 2006. Plant virus biodiversity and ecology. PLoS Biol. 4, e80.
Yuan, X., Shi, K., Young, M.Y., Simon, A.E., 2010. The terminal loop of a 3′ proximal hairpin plays a critical role in replication and the structure of the 3′ region of *Turnip crinkle virus*. Virology 402, 271–280.
Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. 4, e3.
Ziegler, A., Cowan, G., Torrance, L., 2009. Comparative sequence analysis and serological and infectivity studies indicate that Cocksfoot mild mosaic virus is a member of the genus *Panicovirus*. Arch. Virol. 154, 1545–1549.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.