Accepted Manuscript

Title: Identification of a nanovirus-alphasatellite complex in *Sophora alopecuroides*

Author: Jahangir Heydarnejad Mehdi Kamali Hossain Massumi Anders Kvarnheden Maketalena F. Male Simona Kraberger Daisy Stainton Darren P. Martin Arvind Varsani

PII: S0168-1702(17)30009-6
DOI: http://dx.doi.org/doi:10.1016/j.virusres.2017.03.023
Reference: VIRUS 97113

To appear in: Virus Research

Received date: 5-1-2017
Revised date: 15-3-2017
Accepted date: 18-3-2017

Please cite this article as: Heydarnejad, J., Kamali, M., Massumi, H., Kvarnheden, A., Male, M.F., Kraberger, S., Stainton, D., Martin, D.P., Varsani, A., Identification of a nanovirus-alphasatellite complex in *Sophora alopecuroides*, Virus Research (2017), http://dx.doi.org/10.1016/j.virusres.2017.03.023

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Identification of a nanovirus-alphasatellite complex in *Sophora alopecuroides*

Jahangir Heydarnejad\(^1\)*, Mehdi Kamali\(^1\), Hossain Massumi\(^1\), Anders Kvarnheden\(^2\), Maketalena F. Male\(^3\), Simona Kraberger\(^3,4\), Daisy Stainton\(^3,5\), Darren P. Martin\(^6\) and Arvind Varsani\(^3,7,8\)*

\(^1\)Department of Plant Protection, College of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

\(^2\)Department of Plant Biology, Uppsala BioCenter, Linnean Center of Plant Biology in Uppsala, Swedish University of Agricultural Sciences, Box 7080, SE-750 07 Uppsala, Sweden

\(^3\)School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, 8140, New Zealand

\(^4\)Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA

\(^5\)School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

\(^6\)Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, University of Cape Town Faculty of Health Sciences, Observatory 7925. South Africa

\(^7\)Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Rondebosch, 7701, Cape Town, South Africa

\(^8\)The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA

*Corresponding authors

Dr Jahangir Heydarnejad, Department of Plant Protection, College of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran, Fax: 0098 341 3222043, E-mail: jheydarnejad@uk.ac.ir

Assoc Prof. Arvind Varsani, The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA, Tel: +1 480 410 9366. E-mail: arvind.varsani@asu.edu
Key words: Nanovirus; *Sophora alopecuroides*; Iran; alphasatellite; high throughput sequencing
Abstract

Viruses in the genus *Nanovirus* of the family *Nanoviridae* generally have eight individually encapsidated circular genome components and have been predominantly found infecting Fabaceae plants in Europe, Australia, Africa and Asia. For over a decade *Sophora alopecuroides* L. (Fabaceae) plants have been observed across Iran displaying dwarfing, yellowing, stunted leaves and yellow vein banding. Using a high-throughput sequencing approach, sequences were identified within one such plant that had similarities to nanovirus genome components. From this plant, the nanovirus-like molecules DNA-R (n=4), DNA-C (n=2), DNA-S (n=1), DNA-M (n=1), DNA-N (n=1), DNA-U1 (n=1), DNA-U2 (n=1) and DNA-U4 (n=1) were amplified, cloned and sequenced. Other than for the DNA-R, these components share less than 71% identity with those of other known nanoviruses. The four DNA-R molecules were highly diverse, sharing only 65-71% identity with each other and 64-86% identity with those of other nanoviruses. In the *S. alopecuroides* plant 14 molecules sharing 57.7-84.6% identity with previously determined sequences of nanovirus-associated alphasatellites were also identified. Given the research activity in the nanovirus field during the last five years coupled with high-throughput sequence technologies, many more diverse nanoviruses and nanovirus-associated satellites are likely to be identified.
1. Introduction

The family *Nanoviridae* contains plant-infecting viruses with multi-component single-stranded DNA (ssDNA) genomes that are individually encapsidated within isometric 17-20 nm virions. The two genera within the family, *Nanovirus* and *Babuvirus*, are differentiated based on their biological and genomic properties. Viruses in both genera are transmitted by aphids with members of the genus *Nanovirus* infecting dicotyledonous host plants and those of the genus *Babuvirus* infecting monocotyledonous plants. Also, whereas members of the genus *Nanovirus* have eight genome components (named DNA-R, DNA-S, DNA-M, DNA-C, DNA-N, DNA-U1, DNA-U2 and DNA-U4) that are between 970 and 1021 nucleotides (nts) in length, members of the genus *Babuvirus* have six components (DNA-R, DNA-S, DNA-M, DNA-C, DNA-N, and DNA-U3) that are between 1013 and 1116 nts in length (Table 1) (Vetten et al., 2012).

There are presently eight recognized species in the genus *Nanovirus* (Table 1) (Abraham et al., 2012; Boevink et al., 1995; Chu and Helms, 1988; Grigoras et al., 2014; Grigoras et al., 2010; Grigoras et al., 2009; Katul et al., 1998; Sano et al., 1998), which have so far been found infecting various predominantly leguminous species (Table 1). The host symptoms include stunting, necrosis, leaf yellowing or reddening and leaf curling (Vetten et al., 2012).

The DNA-R, DNA-S, DNA-M, DNA-C and DNA-N components of nanoviruses and babuviruses are homologous and all encode a protein of known function. DNA-R encodes a replication-associated protein (Rep) (Burns et al., 1995; Hafner et al., 1997; Harding et al., 1993) which is involved in replicating all canonical components. DNA-S encodes the capsid protein (CP) (Wanitchakorn et al., 1997). DNA-C encodes a cell-cycle link protein (Clink) which is involved in switching the plant host into DNA replication or S-phase to increase replication of the other components (Aronson et al., 2000; Lageix et al., 2007; Wanitchakorn et al., 2000). DNA-M encodes a movement protein (MP) and based on cellular localisation studies DNA-N encodes a putative nuclear shuttle protein (NSP) (Wanitchakorn et al., 2000). The DNA-U1, DNA-U2, and DNA-U4 components have unknown functions.
The components of a nanovirus genome share two homologous regions: the common region stem-loop (CR-I) and the common region II (CR-II). In addition to the canonical genome components, nanoviruses have also been found associated with satellite molecules known as alphasatellites. Like the canonical DNA-R component, these molecules contain a rep-like gene, however, unlike DNA-R, they are unable to trans-replicate the canonical genome components with which they are associated (Horser et al., 2001; Timchenko et al., 1999; Timchenko et al., 2000). Alphasatellites that are related to those of nanoviruses are also found associated with viruses in the genera Begomovirus and Mastrevirus of the family Geminiviridae (Kumar et al., 2014; Zhou, 2013).

Sophora alopecuroides L. (Fabaceae) is a wild perennial herb that is widely distributed across the arid and semi-arid regions of Iran and other parts of Asia (Bisby et al., 1994). It is primarily used as livestock feed but in traditional Chinese medicine it is also used to treat fever and diarrhea (Song et al., 1999; Zhao et al., 2013). Over the past decade in Iran S. alopecuroides plants have been observed throughout most parts of the country with apparent disease symptoms including severe yellowing and stunting, shrunken leaves and yellow vein banding (Figure 1). In this report a putative nanovirus in the genus Nanovirus recovered from S. alopecuroides is described.

2. Materials and methods

2.1 DNA isolation, amplification of circular molecules and Illumina sequencing

A symptomatic sample of Sophora alopecuroides L. (Fabaceae; Figure 1) was collected at Shahid Bahonar University of Kerman (Kerman, Iran) in 2014 and total DNA was extracted according to Zhang et al. (1998). Circular DNA sequences were amplified using rolling circle amplification with Phi29 DNA polymerase as previously described (Shepherd et al., 2008). The
amplified circular DNA was sequenced on an Illumina HiSeq 2500 sequencer at Noveogene (Hong Kong).

2.2 De novo assembly of Illumina sequencing reads and analysis of resulting contigs

The Illumina sequenced paired-end reads were de novo assembled using ABySS v1.9 (Simpson et al., 2009) assembler. Contigs of >500 nts were analyzed using BLASTn and BLASTx (Altschul et al., 1990) against a viral database to identify viral-like contigs. In the 3366 contigs that were > 500 nts viral sequences were identified that shared similarities to nanovirus components and alphasatellite sequences.

2.3 Recovery and characterization of circular molecules with viral and alphasatellite sequences

Abutting primers were designed for all de novo assembled contigs that had similarities to nanovirus and alphasatellite sequences in order to recover, verify and archive the DNA molecules. In all but one case the overlapping primers were designed to contain a restriction enzyme site (Supplementary Table 1).

These abutting primer pairs were used to PCR amplify the circular molecules from the original DNA extracts with KAPA Hifi Hotstart DNA polymerase (Kapa Biosystems, USA) using the following protocol: initial denaturation at 95°C for 3 min followed by 25 cycles at 98°C for 20 s, 60°C for 15 s, 72°C for 45 s and a final extension at 72°C for 1 min. The amplicons of ~1.0 kb were resolved on a 0.7% agarose gel, gel purified and cloned into pJET1.2 plasmid (ThermoFisher, USA). The inserts of the resulting recombinant plasmids were Sanger sequenced by primer walking at Macrogen Inc. (South Korea) and the contigs assembled using DNAbaser v.4 (Heracle BioSoft S.R.L., Romania). The putative open reading frames were identified in the circular DNA molecule sequences using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder)
(Sayers *et al.*, 2012) and all pairwise nucleotide identities were determined using SDT v1.2 (Muhire *et al.*, 2014).

The recovered sequences determined by Sanger sequencing were aligned with representative nanovirus components and alphasatellite sequences from GenBank (downloaded on the 17th Oct 2016) using MUSCLE (Edgar, 2004). Phylogenetic trees were inferred using FastTree (*Price et al.*, 2010) with GTR+CAT substitution model. Branches with < 80% SH-like support were collapsed using TreeGraph2 (Stover and Muller, 2010). The phylogenetic trees were rooted with sequences of babuviruses for DNA-R, DNA-S, DNA-C, DNA-M and DNA-N datasets and midpoint rooted for DNA-U1, DNA-U2, DNA-U4 and the alphasatellite datasets. The CR-SL of all nanovirus components available in GenBank along with those from this study were aligned using MUSCLE (Edgar, 2004).

3. Results and discussion

3.1 Identification and recovery of nanovirus and associated alphasatellite molecules from *Sophora alopecuroides*.

In an effort to determine the etiological agent of a severe disease affecting *S. alopecuroides* in Iran (Figure 1), total DNA was purified from an affected plant, circular molecules were amplified and the resulting DNA was analyzed by high-throughput sequencing (HTS). Among *de novo* assembled contigs of the HTS reads nanovirus-like and nanovirus-associated alphasatellite-like molecules were identified. Using abutting primers designed from these contigs to amplify specific circular replicons, the amplicons for 14 alphasatellite molecules, four DNA-R molecules, two DNA-C molecules and one each of DNA-S, DNA-M, DNA-N, DNA-U1, DNA-U2 and DNA-U4 were cloned and Sanger sequenced (Figure 1). The sequences of these 26 circular molecules (Figure 1; Supplementary Table 1) have been deposited in GenBank as accessions KX534385 to KX534410. In addition, what appear to be six defective molecules (Supplementary Data 1) with an alphasatellite-like backbone and no detectable coding region were cloned and sequenced.
3.2 Sequence analysis of nanovirus-like genome components from S. alopecuroides

Analysis of the pairwise identities of the recovered nanovirus-like sequences from S. alopecuroides indicated that in general these share <72% pairwise identities with all other known nanovirus genome components, except DNA-R which shared up to 88% identity (Table 2, Figure 2). Nanoviruses whose CP amino acid sequence diversity is >15% and/or their overall genomes share <75% pairwise identity could be classified as new species (Vetten et al., 2012). The putative genome of the new nanovirus shares 66-69% pairwise identity with other nanovirus genomes and the predicted amino acid sequence of the CP of the virus share 44-54% identity with CP sequences of other nanoviruses. This implies that the newly determined components are likely derived from one or more novel nanovirus species. Based on the symptoms of the plant from which the virus was obtained (Figure 1), we refer to this virus as Sophora yellow stunt-associated virus (SYSaV) in the rest of the manuscript. It is noteworthy that the four recovered DNA-R molecules share less than 76.5% identity with each other and that they therefore likely each represents a different nanovirus species. DNA-R3 (KX534390) shares a higher degree of identity (78 - 88% pairwise identity) with the DNA-Rs of other known nanoviruses (Table 2; Figure 2) than with the other three DNA-R sequences isolated from the S. alopecuroides plant.

Similarly, the two DNA-C molecules from S. alopecuroides (KX534386 and KX534396) share only 71% pairwise identity with one another and therefore are also likely derived from different nanovirus species (Table 2; Figure 2). These molecules share 60.5-68.1 and 61.3-69.3% pairwise identities with DNA-C molecules of other nanoviruses.

The deduced amino acid sequence of the protein encoded by the two DNA-Cs of SYSaV contains the LXCXE motif that is conserved in nearly all other known nanoviruses. This protein is responsible for interactions with plant retinoblastoma-like proteins and is involved in cell cycle regulation (Aronson et al., 2000; Wanitchakorn et al., 2000).

The DNA-S component (KX534385) shares between 61.7 and 67.7% pairwise identity with the DNA-S components of other known nanoviruses (Table 2; Figure 3). The DNA-M (KX534387), DNA-N (KX534393) and DNA-U1 (KX534394) components, respectively, share 61.2-68.0%, 63.9-70.4% and 59.5-70.2% pairwise identities with those of other nanoviruses, and are most closely related to their homologous counterparts in subterranean clover stunt virus (SCSV).
SYSaV-U2 (KX534395) and SYSaV-U4 (KX534392), respectively, share 61.5-66.4% and 58.7-66.5% pairwise identities with their counterparts in other nanoviruses (Table 2; Figure 4).

3.3 Analysis of the CR-I, CR-II and identification of putative Rep recognition sequences

The CR-I which is conserved across all components of an individual nanovirus genome contains both the origin of virion strand replication and iterated direct and inverted repeat sequences that function as Rep recognition sites during the initiation and resolution of rolling circle replication (Hafner et al., 1997; Londono et al., 2010). All of the SYSaV components contain the “TAGTATTAC” nonanucleotide within the loop sequence of a likely hairpin structure that is highly conserved in all nanoviruses. Rep initiates rolling circle replication by nicking this nonanucleotide between the final T and A nucleotides. An alignment of the entire CR-I sequence of all nanoviruses indicated that they predominantly contain iterated sequences containing the trimer “TGA” (in the SYSaV components the full iterated sequence is TGACG) (Figure 5).

Pairwise identity analysis of the CR-I of all the nanovirus components reveals a diversity of 46.4% (Supplementary Data 2). Pairwise comparisons of the CR-I of DNA-R molecules of black medic leaf roll virus (BMLV), faba bean necrotic stunt virus (FBNSV), faba bean necrotic yellows virus (FBNYV), faba bean yellow leaf virus (FBYLV), milk vetch dwarf virus (MDV), pea necrotic yellow dwarf virus (PNYDV), pea yellow stunt virus (PYSV) and SCSV and that of their canonical DNA-C, -M, -N, -S, -U1, -U2 and U4 molecules indicate that there can be up to 31% diversity within a species (Supplementary Figure 1). It is noted that the CR-I of the four SYSaV DNA-Rs when compared with their canonical molecules shows diversity of ~15 - 40% (Supplementary Figure 1).

The CR-II of nanoviruses (~50 nts) is relatively smaller than that found in babuviruses. A highly conserved motif (CTCTGCGAAGCTATATG) was identified in the CR-II region (Figure 5). The
CR-II of three DNA-Rs (KX534388, KX534389, KX534391) shares >95% identity and the fourth one (KX534390) shares ~87% (Supplementary Data 3).

### 3.4 Nanovirus-associated alphasatellites and alphasatellites-like circular molecules

Alphasatellites are frequently associated with nanovirus infections (Vetten et al., 2012). Unlike begomovirus-alphasatellites which are ~1350 nts, the nanoviruses-associated alphasatellites are ~1000 nts and have an AT rich region domain downstream of the rep gene. The nanovirus-associated alphasatellites that have been so far identified all cluster together and their Reps are more closely related to Reps of begomovirus-associated alphasatellites. Fourteen of these molecules were identified in the examined S. alopecuroides plant (KX534397-KX534410) (Figure 1, Supplementary Table 1). All but two of these molecules had the consensus nonanucleotide that is generally found at alphasatellite virion-strand origins of replication (TAGTATTAC). The two exceptions, KX534397 and KX534398, had a CAGTATTAC sequence. The alphasatellites associated with SYSaV share between 57.7 and 99.7% pairwise identities with each other and 57.7-84.6% with other nanovirus-associated alphasatellites (Figure 6). Specifically, molecule KX534397 shares 84% pairwise identity with alphasatellites associated with FBNYV (AJ005964 and AJ132185). Analysis of the phylogeny of the 14 Sophora yellow stunt-associated alphasatellites revealed three well supported clades that accommodate all the currently identified nanovirus-associated alphasatellites except that of coconut foliar decay virus (CFDV; M29963; Figure 6). Both the pairwise identity and phylogenetic analysis of the available alphasatellite molecules indicated that there are no clear associations between specific groups of closely related molecules and particular nanovirus species.

Defective molecules which are similar to the canonical genome components but with insertions or deletions which likely render them as non-functional components have previously been found associated with nanoviruses (Stainton et al., 2016; Su et al., 2003) and geminiviruses (Al Rwahni et al., 2016; Bach and Jeske, 2014; Casado et al., 2004; Frischmuth and Stanley, 1992; Hadfield et al., 2012; Horn et al., 2011; Paprotka et al., 2010; Stanley and Townsend, 1985;
Stenger et al., 1992; van der Walt et al., 2009; Zaffalon et al., 2012). Six alphasatellite-like circular molecules sharing >85% nucleotide identity and with no detectable coding region (Supplementary Data 1) were recovered. A blast analysis of these shows that they share 78-90% identity with 24-33% coverage (mainly in the intergenic and the 3’ and 5’ termini of the rep gene with milk vetch dwarf alphasatellite molecules).

4. Conclusion

Here, using a high-throughput sequencing approach, at least one putative novel nanovirus is identified infecting S. alopecuroides, a legume that can be found growing wild throughout most parts of Iran. In a single diseased plant displaying yellowing and stunting symptoms 12 distinct molecules that appeared to be components of one or more nanovirus genomes and a further 14 distinct molecules which appear to be nanovirus-associated alphasatellites were identified. Given that such a diverse set of DNA-R and alphasatellite molecules have been recovered from the single S. alopecuroides plant, it is plausible that that this plant harboured a mixed infection. It would be very interesting to determine whether similarly symptomatic S. alopecuroides plants from elsewhere in Iran contain similar complements of canonical genome components and alphasatellite molecules to those described here. Based on all the analysis of the CR-I and CR-II of the DNA-Rs with other nanovirus-like components from S. alopecuroides plants, we are unable to provide a high confidence assemblage of what would be a novel nanovirus genome. It is highly likely, based on the four diverse DNA-Rs identified in this study and the two DNA-Cs, that these may represent four novel nanoviruses adding to the other nanovirus species that have so far been identified globally.

Acknowledgement

This study was supported by a grant from Shahid Bahonar University of Kerman, Kerman, Iran. AV and DPM are supported by the National Science Foundation of South Africa.
Figure legends and table text

**Figure 1:** a) Plant of *Sophora alopecuroides* (left) displaying symptoms associated with infection by Sophora yellow stunt-associated virus, including severe yellowing and stunting and small sized leaves, in comparison to a healthy plant (right). b) Early symptoms of yellow veins in a plant of the same species. c) Linearised illustration of the DNA molecules recovered from the infected *Sophora alopecuroides*.

**Figure 2:** Maximum likelihood phylogenetic tree and pairwise identity matrix of nanovirus DNA-R molecule nucleotide sequences. The phylogenetic tree is rooted with babuvirus DNA-R molecule sequences. DNA-R molecules from *Sophora alopecuroides* are in red bold font.

**Figure 3:** Maximum likelihood phylogenetic tree and pairwise identity matrix of nanovirus DNA-S molecule nucleotide sequences. The phylogenetic tree is rooted with babuvirus DNA-S molecule sequences. DNA-S molecules from *Sophora alopecuroides* are in red bold font.

**Figure 4:** Maximum likelihood phylogenetic trees of DNA-C, DNA-M, DNA-N, DNA-U1, DNA-U2 and DNA-U4 molecule nucleotide sequences of nanoviruses. DNA-C, DNA-M and DNA-N phylogenetic trees are rooted with corresponding babuvirus molecule sequences whereas DNA-U1, DNA-U2 and DNA-U4 phylogenetic trees are mid-point rooted. DNA molecules from *Sophora alopecuroides* are in red bold font.

**Figure 5:** Alignment of the CR-I (A) and CR-II (B) regions identified in the SYSaV sequences. The iterons and the nonanucleotide motif in CR-I and a highly conserved motif in CR-II are highlighted in grey boxes.

**Figure 6:** Maximum likelihood phylogenetic tree and pairwise identity matrix of nanovirus-associated alphasatellite molecule nucleotide sequences. Sophora yellow stunt-associated alphasatellites molecules from *Sophora alopecuroides* are in red bold font.

**Table 1:** Overview of all babuvirus and nanovirus species, including the size (nt) and presence/absence of components.
Table 2: Pairwise identities of SYSaV DNA-R, DNA-S, DNA-C, DNA-M, DNA-N, DNA-U1, DNA-U2 and DNA-U4 with those of other nanoviruses.

Supplementary Table 1: Details of primer pairs used to recover the DNA-R, DNA-S, DNA-C, DNA-M, DNA-N, DNA-U1, DNA-U2 and DNA-U4 molecules as well as alphasatellite molecules from Sophora alopecuroides. The underlined regions within the primer pairs are the regions which correspond to a restriction enzyme site. The GenBank accession numbers of the recovered components are included.

Supplementary Figure 1: Analysis of the percentage pairwise diversity of the CR-I of the DNA-R of BMLV, FBNSV, FBNYV, FBYLV, MDV, PNYDV, PYSV, SCSV and SYSaV with that of their canonical molecules.

Supplementary Data 1: Nucleotide sequence file (fasta format) of defective molecules recovered in this study.

Supplementary Data 2: Pairwise comparisons and alignment of the CR-I of all nanoviruses.

Supplementary Data 3: Pairwise comparisons and alignment of the CR-II of all nanoviruses.

References


Table 1:

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Acronym</th>
<th>Host species</th>
<th>Country</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Genus</td>
</tr>
<tr>
<td>Nanovirus</td>
<td>Subterranean clover stunt virus</td>
<td>SCSV</td>
<td>Trifolium subterranum</td>
<td>AU</td>
<td>1005</td>
</tr>
<tr>
<td>Fababean necrotic yellows</td>
<td>virus</td>
<td>FBNYV</td>
<td>Vicia faba</td>
<td>ES, AZ, ET, IR, SY, EG, MA</td>
<td>1003</td>
</tr>
<tr>
<td>Fababean necrotic stunt</td>
<td>virus</td>
<td>FBNSV</td>
<td>Phaseolus vulgaris Lens culinaris Vicia sativa</td>
<td>ET, IR, MA, AZ</td>
<td>1003</td>
</tr>
<tr>
<td>Pea necrotic yellow dwarf</td>
<td>virus</td>
<td>PNYDV</td>
<td>Pisum sativum</td>
<td>AT, GE</td>
<td>1002</td>
</tr>
<tr>
<td>Milk vetch dwarf virus</td>
<td>MDV</td>
<td></td>
<td></td>
<td>BD, CH, JP</td>
<td>1001</td>
</tr>
<tr>
<td>Fababean yellow leaf virus</td>
<td>FBYLV</td>
<td></td>
<td></td>
<td>ET</td>
<td>1002</td>
</tr>
<tr>
<td>Pea yellow stunt virus</td>
<td>PYSV</td>
<td></td>
<td>Pisum sativum</td>
<td>AT, AZ, SE</td>
<td>1002</td>
</tr>
<tr>
<td>Black medic leaf roll virus</td>
<td>BMLRV</td>
<td></td>
<td>Medicago lupulina Pisum sativum</td>
<td>AT, AZ, SE</td>
<td>1008</td>
</tr>
<tr>
<td>Sophora yellow stunt-</td>
<td>SYSaV</td>
<td></td>
<td>Sophora alopecuroides</td>
<td>IR</td>
<td>997</td>
</tr>
<tr>
<td>Babavirus</td>
<td>Banana bunchy top virus</td>
<td>BBTV</td>
<td>Musa spp</td>
<td>Banana worldwide</td>
<td>1110</td>
</tr>
<tr>
<td>Abaca bunchy top virus</td>
<td>ABTV</td>
<td></td>
<td>Musa spp</td>
<td>PIL, MY</td>
<td>1099</td>
</tr>
</tbody>
</table>
Table 2:

<table>
<thead>
<tr>
<th>Percentage pairwise identity</th>
<th>SYSaV R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>C1</th>
<th>C2</th>
<th>BMLRV</th>
<th>FBNSV</th>
<th>FBNYV</th>
<th>FBVLV</th>
<th>MVDV</th>
<th>PNYDV</th>
<th>PYSV</th>
<th>SCSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>74.6</td>
<td>76.1</td>
<td>74.0</td>
<td>75.6</td>
<td>74.8</td>
<td>75.4</td>
<td>75.7</td>
<td>76.7</td>
</tr>
<tr>
<td>R2</td>
<td>68.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>68.5</td>
<td>70.2</td>
<td>67.6</td>
<td>68.2</td>
<td>65.1</td>
<td>68.8</td>
<td>66.5</td>
<td>66.0</td>
</tr>
<tr>
<td>R3</td>
<td>76.5</td>
<td>67.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>85.8</td>
<td>88.2</td>
<td>83.7</td>
<td>84.4</td>
<td>84.2</td>
<td>85.0</td>
<td>83.8</td>
<td>81.8</td>
</tr>
<tr>
<td>R4</td>
<td>70.7</td>
<td>66.9</td>
<td>64.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>64.2</td>
<td>66.3</td>
<td>64.1</td>
<td>64.7</td>
<td>64.3</td>
<td>65.4</td>
<td>62.1</td>
<td>61.7</td>
</tr>
<tr>
<td>S</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>62.7</td>
<td>64.6</td>
<td>63.0</td>
<td>65.2</td>
<td>62.8</td>
<td>64.5</td>
<td>66.9</td>
<td>63.8</td>
</tr>
<tr>
<td>N</td>
<td>66.7</td>
<td>69.1</td>
<td>68.0</td>
<td>70.4</td>
<td>66.9</td>
<td>69.9</td>
<td>68.9</td>
<td>68.2</td>
<td>69.8</td>
<td>63.9</td>
<td>64.4</td>
<td>65.7</td>
<td>71.2</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>62.5</td>
<td>65.4</td>
<td>65.5</td>
<td>67.7</td>
<td>65.8</td>
<td>68.1</td>
<td>60.5</td>
<td>62.0</td>
</tr>
<tr>
<td>C2</td>
<td>71.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>64.4</td>
<td>65.4</td>
<td>67.3</td>
<td>69.3</td>
<td>65.3</td>
<td>67.1</td>
<td>63.5</td>
<td>64.5</td>
</tr>
<tr>
<td>M</td>
<td>62.3</td>
<td>65.6</td>
<td>63.0</td>
<td>66.3</td>
<td>60.9</td>
<td>65.0</td>
<td>66.7</td>
<td>61.6</td>
<td>66.7</td>
<td>65.8</td>
<td>66.1</td>
<td>64.4</td>
<td>64.0</td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>59.5</td>
<td>61.0</td>
<td>65.7</td>
<td>68.0</td>
<td>66.4</td>
<td>70.2</td>
<td>59.5</td>
<td>60.1</td>
</tr>
<tr>
<td>U2</td>
<td>63.5</td>
<td>65.3</td>
<td>63.3</td>
<td>66.4</td>
<td>61.5</td>
<td>66.1</td>
<td>61.8</td>
<td>63.8</td>
<td>65.1</td>
<td>63.3</td>
<td>65.5</td>
<td>62.3</td>
<td>62.7</td>
<td></td>
</tr>
<tr>
<td>U4</td>
<td>62.7</td>
<td>65.6</td>
<td>59.6</td>
<td>62.7</td>
<td>61.0</td>
<td>64.3</td>
<td>66.5</td>
<td>63.3</td>
<td>63.8</td>
<td>61.3</td>
<td>61.4</td>
<td>63.5</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Use of high throughput sequencing to identify novel virus and satellite molecules.

Detection of 12 nanovirus-like molecules associated with Sophora alopecuroides.

12 molecules shared less than 71% identity with those of other known nanoviruses.

14 alphasatellites molecules were recovered and characterized.