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

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

30 Key words: Nanovirus; *Sophora alopecuroides*; Iran; alphasatellite; high throughput sequencing

31

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)

(Table 2; Figure 4). 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 alphastaellites 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 Rwahnih *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

Abraham, A.D., Varrelmann, M., Vetten, H.J. (2012). Three distinct nanoviruses, one of which represents a new species, infect faba bean in Ethiopia. Plant Dis. 96, 1045–1053.

Al Rwahnih, M., Alabi, O.J., Westrick, N.M., Golino, D., Rowhani, A. (2017). Description of a novel monopartite geminivirus and its defective subviral genome in grapevine. Phytopathology 103, 240-251.

Alavinejad, E., Behjatnia, S.A.A., Izadpanah, K., Masoumi, M. (2011). Molecular detection of Faba bean necrotic yellows virus in legume fields of some North, North West and South provinces of Iran. In The 7th National Biotechnology Congress of I R Iran, p. 6.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. 215, 403-410.

Aronson, M.N., Meyer, A.D., Györgyey, J., Katul, L., Vetten, H.J., Gronenborn, B., Timchenko, T. (2000). Clink, a nanovirus-encoded protein, binds both pRB and SKP1. J. Virology 74, 2967-2972.

Bach, J., Jeske, H. (2014). Defective DNAs of beet curly top virus from long-term survivor sugar beet plants. Virus Res. 183, 89-94.

Bisby, F.A., Buckingham, J., Harborne, J.B. (1994). Phytochemical dictionary of the Leguminosae. Vol. 1: Plants and their constituents: Chapman and Hall, London.

Boevink, P., Chu, P.W.G., Kees, P. (1995). Sequence of subterranean clover stunt virus DNA: Affinities with the geminiviruses. Virology 207, 354-361.

Burns, T.M., Harding R.M., Dale, J.L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. J. Gen. Virol. 76, 1471-1482.

Casado, C.G., Javier Ortiz, G., Padron, E., Bean, S.J., McKenna, R., Agbandje-McKenna, M., Boulton, M.I. (2004). Isolation and characterization of subgenomic DNAs encapsidated in "single" T = 1 isometric particles of Maize streak virus. Virology 323, 164-171.

Chu, P.W.G., Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. Virology 167, 38-49.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids Res. 32, 1792-1797.

Frischmuth, T., Stanley, J. (1992). Characterization of beet curly top virus subgenomic DNA localizes sequences required for replication. Virology 189, 808-811.

Grigoras, I., Ginzo, A.I., Martin, D.P., Varsani, A., Romero, J., Mammadov, A., Huseynova, I.M., Aliyev, J.A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.J., Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. J. Gen. Virol. 95, 1178-1191.

Grigoras, I., Gronenborn, B., Vetten, H.J. (2010). First report of a nanovirus disease of pea in Germany. Plant Dis. 94, 642-642.

Grigoras, I., Timchenko, T., Katul, L., Grande-Perez, A., Vetten, H.J., Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. J. Virol. 83, 10778-10787.

Hadfield, J., Thomas, J.E., Schwinghamer, M.W., Kraberger, S., Stainton, D., Dayaram, A., Parry, J.N., Pande, D., Martin, D.P., Varsani, A. (2012). Molecular characterisation of dicot-infecting mastreviruses from Australia. Virus Res. 166, 13-22.

Hafner, G.J., Stafford, M.R., Wolter, L.C., Harding, R.M., Dale, J.L. (1997). Nicking and joining activity of banana bunchy top virus replication protein in vitro. J. Gen. Virol. 78, 1795-1799.

Harding, R.M., Burns, T.M., Hafner, G., Dietzgen, R.G., Dale, J.L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. J. Gen. Virol. 74, 323-328.

Horn, J., Lauster, S., Krenz, B., Kraus, J., Frischmuth, T., Jeske, H. (2011). Ambivalent effects of defective DNA in beet curly top virus-infected transgenic sugarbeet plants. Virus Res. 158, 169-178.

Horser, C.L., Harding, R.M., Dale, J.L. (2001). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. J. Gen. Virol. 82, 459-464.

Katul, L., Timchenko, T., Gronenborn, B., Vetten, H.J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. J. Gen. Virol. 79, 3101-3109.

Kumar, J., Kumar, J., Singh, S.P., Tuli, R. (2014). Association of satellites with a mastrevirus in natural infection: Complexity of Wheat dwarf India virus disease. J. Virol. 88, 7093-7104.

Lageix, S., Catrice, O., Deragon, J.M., Gronenborn, B., Pélissier, T., Ramirez, B.C. (2007). The nanovirus-encoded clink protein affects plant cell cycle regulation through interaction with the retinoblastoma-related protein. J. Virol. 81, 4177-4185.

Londono, A., Riego-Ruiz, L., Arguello-Astorga, G.R. (2010). DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. Arch. Virol. 155, 1033-1046.

Lotfipour, M., Izadpanah, K., Behjatnia, S.A.A. (2016). Identification and molecular characterization of Faba bean necrotic stunt virus, a new nanovirus in legume fields in Iran. in The 22nd Iranian Plant Prot Cong Karaj. Iran.

Makkouk, K.M., Fazlali, Y., Kumari, S.G., Farzadfar, S. (2002). First record of Beet western yellows virus, Chickpea chlorotic dwarf virus, Faba bean necrotic yellows virus and Soybean dwarf virus infecting chickpea and lentil crops in Iran. Plant Pathol. 51, 387.

Muhire, B.M., Varsani, A., Martin, D.P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PloS one 9, e108277.

Paprotka, T., Boiteux, L.S., Fonseca, M.E., Resende, R.O., Jeske, H., Faria, J.C., Ribeiro, S.G. (2010). Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. Virus Res. 149, 224-233.

Price, M.N., Dehal, P.S., Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one 5, e9490.

Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T., Kojima, M. (1998). Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. J. Gen. Virol. 79, 3111-3118.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Karsch-Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E., Ye, J. (2012). Database resources of the National Center for Biotechnology Information. Nucleic acids Res. 40, D13-25.

Shepherd, D.N., Martin, D.P., Lefeuvre, P., Monjane, A.L., Owor, B.E., Rybicki, E.P., Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. J. Virol. Methods 149, 97-102.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117-1123.

Song, J.Z., Xu, H.X., Tian, S.J., But, P.P. (1999). Determination of quinolizidine alkaloids in traditional Chinese herbal drugs by nonaqueous capillary electrophoresis. J. Chromatogr. A 857, 303-311.

Stainton, D., Martin, D.P., Collings, D.A., Thomas, J.E., Varsani, A. (2016). Identification and in silico characterisation of defective molecules associated with isolates of banana bunchy top virus. Arch. Virol. 161, 1019-1026.

Stanley, J., Townsend, R. (1985). Characterisation of DNA forms associated with cassava latent virus infection. Nucleic acids Res. 13, 2189-2206.

Stenger, D.C., Stevenson, M.C., Hormuzdi, S.G., Bisaro, D.M. (1992). A number of subgenomic DNAs are produced following agroinoculation of plants with beet curly top virus. J. Gen. Virol. 73, 237-242.

Stover, B.C., Muller, K.F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11, 7.

Su, H.J., Tsao, L.Y., Wu, M.L., Hung, T.H. (2003). Biological and molecular categorization of strains of Banana bunchy top virus. J. Phytopathol .151, 290-296.

Timchenko, T., De Kouchkovsky, F., Katul, L., David, C., Vetten, H.J., Gronenborn, B. (1999). A single Rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. J. Virol. 73, 10173-10182.

Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H.J., Gronenborn, B. (2000). The master Rep concept in nanovirus replication: Identification of missing genome components and potential for natural genetic reassortment. Virology 274, 189-195.

van der Walt, E., Rybicki, E.P., Varsani, A., Polston, J.E., Billharz, R., Donaldson, L., Monjane, A.L., Martin, D.P. (2009). Rapid host adaptation by extensive recombination. J. Gen. Virol. 90, 734-746.

Vetten, H.J., Dale, J.L., Grigoras, I., Gronenborn, B., Harding, R., Randles, J.W., Sano, Y., Thomas, J.E., Timchenko, T., Yeh, H.H. (2012). Nanoviridae. In Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruse, pp. 395-404. Edited by A.M.Q. King, M.J. Adams, E.B. Carstens, E.J. Lefkowitz. Amsterdam: Elsevier, Academic Press.

Wanitchakorn, R., Hafner, G.J., Harding, R.M., Dale, J.L. (2000). Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. J. Gen. Virol. 81, 299-306.

Wanitchakorn, R., Harding, R.M., Dale, J.L. (1997). Banana bunchy top virus DNA-3 encodes the viral coat protein. Arch. Virol. 142, 1673-1680.

Zaffalon, V., Mukherjee, S.K., Reddy, V.S., Thompson, J.R., Tepfer, M. (2012). A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of northwestern India. Arch. Virol. 157, 483-495.

Zhang, Y.P., Uyemoto, J.K, Kirkpatrick, B.C. (1998). A small-scale procedure for extracting nucleic acids from woody plants infected with various phytopathogens for PCR assay. J. Virol. Methods 71, 45-50.

Zhao, L.F., Xu, Y.J., Ma, Z.Q., Deng, Z.S. (2013). Colonization and plant growth promoting characterization of endophytic Pseudomonas chlororaphis strain Zong1 isolated from Sophora alopecuroides root nodules. Braz. J. Microbiol. 44, 623-631.

Zhou, X. (2013). Advances in understanding begomovirus satellites. Ann. Rev. Phytopathol. 51, 357-381.

Table 1:

											Components	
Genus	**Species**	**Acronym**	**Host species**	**Country**	**Rep**	**CP**	**MP**	**Clink**	**NSP**	**Unknown**	**Unknown**	**Unknown**
Nanovirus	*Subterranean clover stunt virus*	SCSV	*Trifolium subterraneum*	AU	1005	998	1001	991	1002	988	-	-
	Faba bean necrotic yellows virus	FBNYV	*Vicia faba*	ES, AZ, ET, IR, SY, EG, MA	1003	1006	992	999	986	996	1020	-
	Faba bean necrotic stunt virus	FBNSV	*Phaseolus vulgaris* *Lens culinaris* *Vicia sativa*	ET, IR, MA, AZ	1003	992	980	994	981	986	984	-
	Pea necrotic yellow dwarf virus	PNYDV	*Pisum sativum*	AT, GE	1002	981	988	988	993	978	985	-
	Milk vetch dwarf virus	MDV	*Astragalus sinicus* *Catharanthus roseus* *Dolichos lablab* *Glycine max* *Nicotina tabacum* *Pisum sativum* *Vicia faba*	BD, CH, JP	1001	997	985	990	977	989	981	-
	Faba bean yellow leaf virus	FBYLV	*Vicia faba*	ET	1002	1001	980	995	1000	990	995	-
	Pea yellow stunt virus	PYSV	*Pisum sativum*	AT, AZ, SE	1002	976	975	971	977	970	971	-
	Black medic leaf roll virus	BMLRV	*Medicago lupulina* *Pisum sativum*	AT, AZ, SE	1008	1017	1010	1017	1021	1011	1001	-
	Sophora yellow stunt-associatedvirus	SYSaV	*Sophora alopecuroides*	IR	997	990	982	985	983	974	982	
Babuvirus	*Banana bunchy top virus*	BBTV	*Musa* spp	Banana worldwide[†]	1110	1076	1047	1017	1090	-	-	1064
	Abaca bunchy top virus	ABTV	*Musa* spp	PH, MY	1099	1078	1076	1013	1073	-	-	1059

	Cardamom bushy dwarf virus	CBDV	*Amomum subulatum*	IN	1102	1086	1083	1027	1116	-	-	1088	

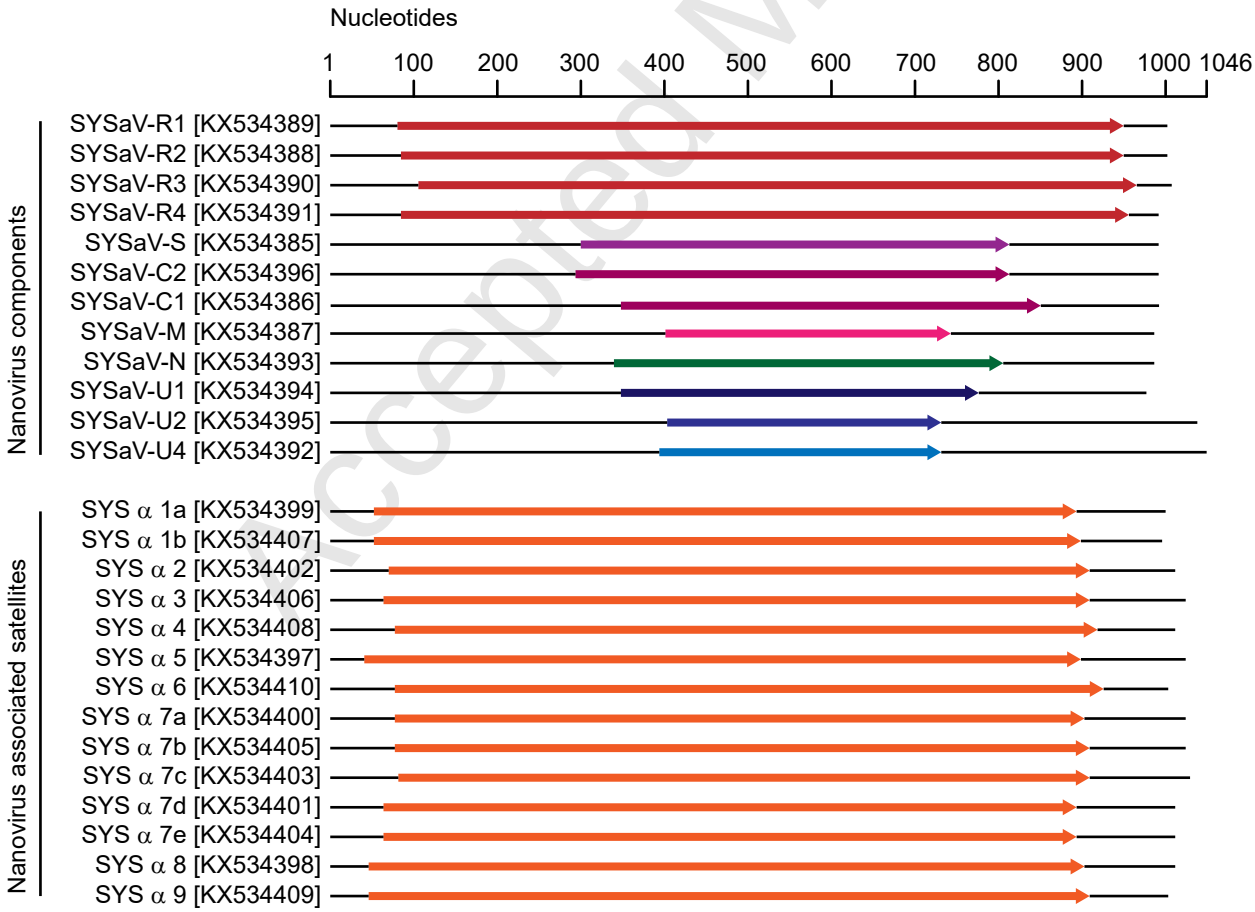
[†]Banana regions worldwide (except the Americas)

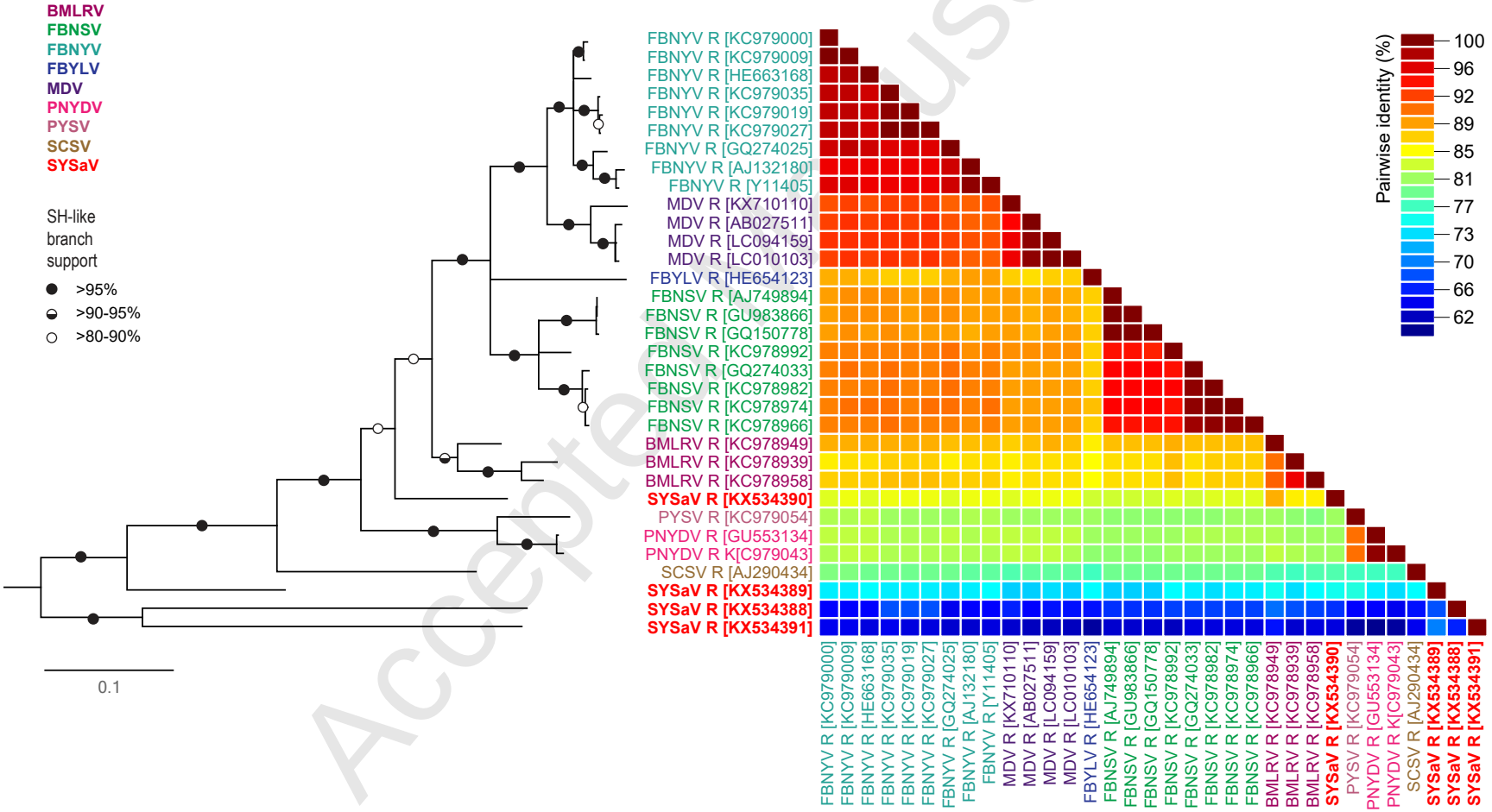
Table 2:

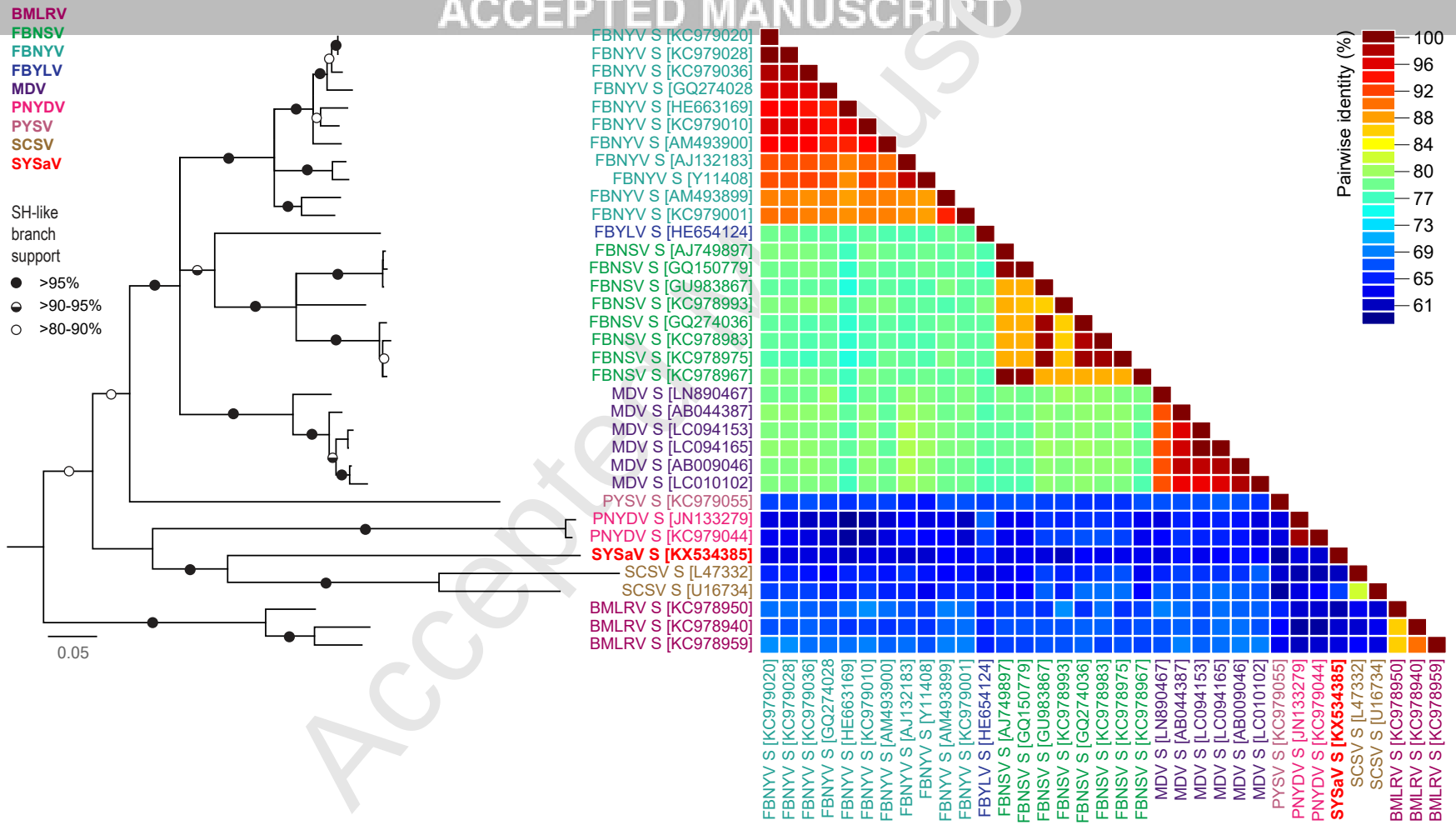
	Percentage pairwise identity												
	SYSaV						**BMLRV**	**FBNSV**	**FBNYV**	**FBYLV**	**MVDV**	**PNYDV**	**PYSV**	**SCSV**
DNA	R1	R2	R3	R4	C1	C2								
R1	-	-	-	-	-	-	74.6 - 76.1	74.0 - 75.6	74.8 - 75.4	75.7	73.9 - 74.4	74.7 - 74.9	76.7	75.7
R2	68.9	-	-	-	-	-	68.5 - 70.2	67.6 - 68.2	65.1 - 68.8	66.5	66.0 - 67.0	65.3 - 65.7	65.9	66.3
R3	76.5	67.7	-	-	-	-	85.8 - 88.2	83.7 - 84.4	84.2 - 85.0	83.8	84.4 - 84.9	81.8 - 81.9	81.9	78.4
R4	70.7	66.9	64.8	-	-	-	64.2 - 66.3	64.1 - 64.7	64.3 - 65.4	62.1	63.5 - 64.2	61.7 - 62.2	62.1	65.4
S	-	-	-	-	-	-	62.7 - 64.6	63.0 - 65.2	62.8 - 64.5	66.9	64.3 - 66.1	63.8 - 64.2	61.7	63.0 - 67.7
N	-	-	-	-	-	-	66.7 - 69.1	68.0 - 70.4	66.9 - 69.9	68.9	68.2 - 69.8	63.9 - 64.4	65.7	71.2
C1	-	-	-	-	-		62.5 - 65.4	65.5 - 67.7	65.8 - 68.1	60.5	65.6 - 67.5	62.0 - 62.9	62.7	64.0
C2	-	-	-	-	71.0	-	64.4 - 65.4	67.3 - 69.3	65.3 - 67.1	63.5	64.3 - 66.9	61.3 - 61.4	64.5	68.0
M	-	-	-	-	-	-	62.3 - 65.6	63.0 - 66.3	60.9 - 65.0	66.7	61.6 - 66.7	65.8 - 66.1	64.4	68.0
U1	-	-	-	-	-	-	59.5 - 61.0	65.7 - 68.0	66.4 - 70.2	59.5	68.0 - 68.9	60.0 - 61.6	64.7	63.1
U2	-	-	-	-	-	-	63.5 - 65.3	63.3 - 66.4	61.5 - 66.1	61.8	63.8 - 65.1	63.3 - 63.5	62.3 - 62.7	-
U4	-	-	-	-	-	-	62.7 - 65.6	59.6 - 62.7	61.0 - 64.3	66.5	63.3 - 63.8	61.3 - 61.4	63.5	-

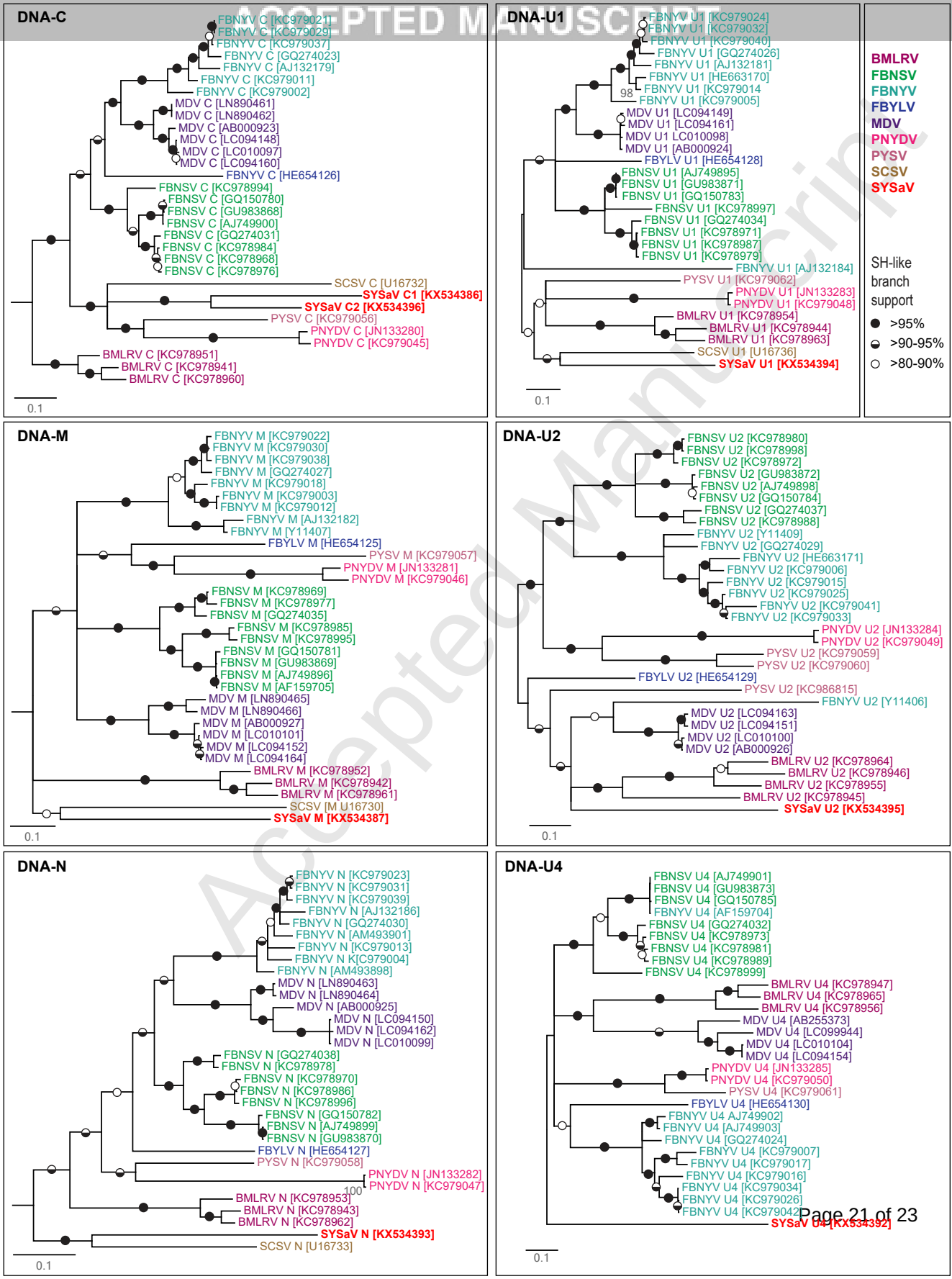
- Use of high throughput sequencing to identify novel virus and satellite molecules.
- Detection of 12 nanovirus-like molecules associated with Sophora alopecuroides.
- 12 molecules shared less than 71% identity with those of other known nanoviruses.
- 14 alphasatellites molecules were recovered and characterized.

c









A

B