

# SCIENTIFIC REPORTS



OPEN

## Discovery and small RNA profile of *Pecan mosaic-associated virus*, a novel potyvirus of pecan trees

Xiu Su<sup>1,2,\*</sup>, Shuai Fu<sup>1,\*</sup>, Yajuan Qian<sup>1</sup>, Liqin Zhang<sup>2</sup>, Yi Xu<sup>1</sup> & Xueping Zhou<sup>1,3</sup>

Received: 02 April 2016

Accepted: 06 May 2016

Published: 26 May 2016

A novel potyvirus was discovered in pecan (*Carya illinoensis*) showing leaf mosaic symptom through the use of deep sequencing of small RNAs. The complete genome of this virus was determined to comprise of 9,310 nucleotides (nt), and shared 24.0% to 58.9% nucleotide similarities with that of other *Potyviridae* viruses. The genome was deduced to encode a single open reading frame (polyprotein) on the plus strand. Phylogenetic analysis based on the whole genome sequence and coat protein amino acid sequence showed that this virus is most closely related to *Lettuce mosaic virus*. Using electron microscopy, the typical *Potyvirus* filamentous particles were identified in infected pecan leaves with mosaic symptoms. Our results clearly show that this virus is a new member of the genus *Potyvirus* in the family *Potyviridae*. The virus is tentatively named Pecan mosaic-associated virus (PMAV). Additionally, profiling of the PMAV-derived small RNA (PMAV-sRNA) showed that the most abundant PMAV-sRNAs were 21-nt in length. There are several hotspots for small RNA production along the PMAV genome; two 21-nt PMAV-sRNAs starting at 811 nt and 610 nt of the minus-strand genome were highly repeated.

RNA silencing is an antiviral immunity and fundamental cellular mechanism that regulates gene expression through the production of small RNAs (sRNAs). Based on the finding that the viral small silencing RNAs (vsRNAs) can be overlapping in sequence and assembling into long contiguous fragments of the invading viral genome<sup>1</sup>, deep sequencing of small RNAs combined with bioinformatics analysis has been widely used for detection and identification of novel viruses, especially for viruses with low titer or without any pathogenic symptoms<sup>2–5</sup>.

*Potyviridae* is the largest and most economically important family of plant viruses, accounting for ~30% of the currently known plant viruses and causing significant losses in agricultural, pastoral, horticultural and ornamental crops<sup>6,7</sup>. *Potyvirus* is a genus in the family *Potyviridae*. Most potyviruses are transmitted by aphids in a non-persistent manner. The virions of potyviruses are non-enveloped with a flexuous and filamentous nucleocapsid of 680 ~ 900 nm long and 11 ~ 20 nm in diameter. The typical genomic RNA is a linear positive sense single-strand RNA ranging in size from 9,000 ~ 12,000 nucleotides. It contains a single long ORF, which encodes a large polyprotein that is cleaved into 11 mature proteins including P3N-PIPO (P3 N-terminus-pretty interesting potyvirus ORF)<sup>8</sup>, which is produced by ribosomal frameshift during translation or transcriptional slippage during RNA synthesis in the P3 cistron<sup>9</sup>. And recently, a novel translational frameshift gene product named P1N-PISPO (pretty interesting sweetpotato potyvirus ORF) was discovered in sweet potato feathery mottle virus (SPFMV) infected plants<sup>10</sup>, and functions as an RNA silencing suppressor<sup>11,12</sup>.

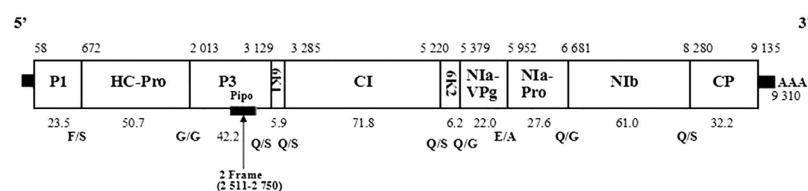
Pecan (*Carya illinoensis*) is native to central and southern North America. It is an excellent multipurpose tree, providing a source of nutritious edible nuts, furniture-grade wood, and aesthetic value for the landscape. Pecan nuts are ranked highest in antioxidant capacitance and are enriched with many nutrients, minerals, and vitamins that are essential for optimum health. Pecans have been extensively cultivated in many countries for their commercial value. In China, pecan is increasingly being grown in many regions, including Hebei, Henan, Jiangsu, Fujian and Zhejiang provinces. Diseases affecting pecans are mainly caused by bacteria, fungi, nematodes, and physiological problems. No viruses have been reported to occur in pecans until now. During our survey of plant

<sup>1</sup>State Key Laboratory of Rice Biology, Institute of Biotechnology, Zhejiang University, Hangzhou 310029, China.

<sup>2</sup>The Nurturing Station for the State Key Laboratory of Subtropical Silviculture, Zhejiang Agriculture and Forestry University, Lin'an 311300, China. <sup>3</sup>State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.X. (email: xuyiqdpd@zju.edu.cn) or X.Z. (email: zzhou@zju.edu.cn)



**Figure 1.** The mosaic symptom of pecan leaves.



**Figure 2.** Genome organization of PMaV. Dash areas represent 5' UTR and 3' UTR; the numbers above the figure represent the number of nucleotide of each proteins; the numbers below the figure represent the molecular mass of each proteins; the arrow represents protein PIPO in P3 protein zone, which is located at 2,511–2,750 nt of the PMaV genome; the 9 cleavage sites on this polypeptide also below the figure.

viruses of woody plants in southeast China, we found several pecan trees which showed typical symptoms of plant virus infection, with systemic mosaic on leaves. Further experiments were performed and finally a novel potyvirus was identified from these pecan trees. The virus is tentatively named Pecan mosaic-associated virus (PMaV). Furthermore, small RNA profiles of this virus suggest that the RNA silencing-mediated antiviral immunity mechanism is also applicable to pecan. To our knowledge, this the first viral disease found in pecan.

## Results

**Symptom of PMaV infected pecan trees.** Mosaic symptom was observed on the leaves of pecan trees during a field survey in pecan orchards. These plants show an overall lighter color along with mosaic patterns on some leaves, especially on the younger leaf (Fig. 1). Then, the leaves with symptom were used for RNA isolation.

**Deep sequencing and small RNA characterization.** A total of 17,832,459 raw reads was obtained, which contained 17,297,837 clean reads with length between 18 and 30 nt after filtering. All these small RNAs were assembled into contigs using Velvet software with k-mer value of 17<sup>13</sup>. After searching against the GenBank database, we found 35 contigs which shared some similarities with the genome sequences of other potyviruses, such as turnip mosaic virus (TuMV), lettuce mosaic virus (LMV), SPFMV and plum pox virus (PPV), suggesting that one or more potyviruses were present in our sample.

**Complete genome characterization.** Based on the contig sequences, the viral genome was obtained through several RT-PCR amplifications covering all gaps. After sequencing, we ensured that only one virus existed in our sample. RACE-PCR was performed to obtain the 5'- and 3'-termini sequences. As shown in Fig. 2, this virus was determined to comprise 9,310 nucleotides (nt) with a GC content of 39.15%. BLASTN search results showed that the virus was most similar to viruses of the genus *Potyvirus* (family *Potyviridae*). All thirty-six available complete genome sequences of *Potyviridae* viruses from GenBank were retrieved, and pairwise genome similarity comparison was performed. As a result (Table 1), the full-length nucleotide sequence similarities ranged from 24.0% (wheat yellow mosaic virus, WYMV) to 58.9% (LMV), and full-length amino acid (aa) sequence similarities from 19.7% (WYMV) to 53.6% (LMV). PMaV CP had the highest sequence identity with LMV on aa level (69.5%), followed by TuMV with 69% aa similarity. On CP nt level, PMaV CP had the highest sequence identity with TuMV (61.2%). All similarities are below the established threshold used to discriminate species of the genus *Potyvirus*<sup>14</sup>. Our results clearly demonstrated that this virus is a new species in the genus *Potyvirus*.

Virus <sup>a</sup>	ScMV	LMV	PRV	PPV	PMV	TuMV	PVY-O	PVY-N	SPFMV	WYMV
Genomic sequence (nt)	55.8	58.9	52.6	54.6	53.0	55.4	25.1	51.8	54.0	24.0
Whole polyprotein (aa)	50.4	53.6	46.2	52.0	47.0	50.7	47.5	47.3	49.9	19.7
CP (nt)	59.0	57.4	59.0	56.0	59.4	61.2	27.1	60.7	55.6	37.8
CP (aa)	66.2	69.5	63.8	63.4	66.7	69.0	65.7	66.2	60.1	9.9

**Table 1. Percent nucleotide (nt) and amino acid (aa) sequence identities of pecan mosaic-associated virus to ten other potyviruses.** <sup>a</sup>Sequences obtained from GenBank: scallion mosaic virus (ScMV, AJ316084.1), lettuce mosaic virus (LMV, KJ161194.1), papaya ringspot virus (PRV, NC\_001785.1), plum pox virus (PPV, NC\_001445), pepper mottle virus (PMV, NC\_001517), turnip mosaic virus (TuMV, NC\_002509), potato virus Y strain O isolate (PVY-O, JQ663997.1), potato virus Y strain N isolate (PVY-N, HM590405.1), sweet potato feathery mottle virus (SPFMV, NC\_001841.1) and wheat yellow mosaic virus (WYMV, NC\_002350).

**Genome structure analysis.** The whole genomic sequence of this virus is 9,310 nt long, excluding the 3' poly (A) tail, and has been submitted to the GenBank database under the accession number KT633868. The genomic organization of this virus is very similar with other potyviruses, consisting of a single open reading frame (ORF) (Fig. 2). The ORF starts at the first in-frame AUG (58–60 nt) and ends with a UAG (9136–9138 nt) termination codon; a polyprotein of 3,026 amino acids with a calculated molecular mass of 342.8 kDa was deduced to be translated. The nine putative protease cleavage sites were identified in the polyprotein (Fig. 2), that were predicted to generate eleven putative mature proteins, including a small P3N-PIPO protein was predicted to be expressed by polymerase slippage mechanism from the P3 ORF.

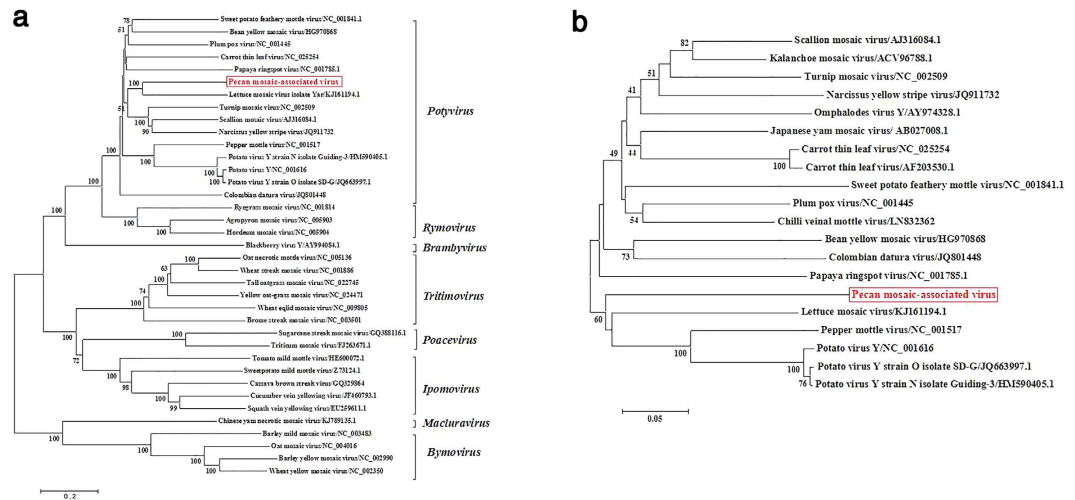
Proteinase P1 comprises 205 amino acids (23.5 kDa) encoded by nucleotides located at position 58–672 nt. The mature protein is predicted to be cleaved at the sequence WRGF/S, which is cut by P1 serine proteinase. P1 of this virus has a motif <sup>143</sup>GWSG<sup>146</sup> which is typical of potyviruses. The helper component proteinase (HC-Pro) comprises 447 residues (50.7 kDa) encoded by nucleotides located at position 673–2,013 nt, and is cleaved at the sequence YKVG/G, in agreement with the classical motif YXVG/G. The <sup>375</sup>FRNK<sup>378</sup> box, which is involved in RNA silencing suppression and associated with symptom development, is present<sup>15</sup>. Motifs of <sup>246</sup>RITC<sup>249</sup> and <sup>504</sup>PTK<sup>506</sup>, which are associated with aphid transmission<sup>16</sup> are also included. The P3 protein (2,014–3,129 nt) is composed of 372 residues (42.2 kDa), and is cleaved at the site ITLQ/S. An ORF for putative P3N-PIPO protein was identified that include the frame shift sequence <sup>2511</sup>GA-AAAAA<sup>2750</sup> and with a UAG stop codon at the position 2,751 nt. The cylindrical inclusion (CI) protein comprises 645 residues (71.8 kDa) encoded by nucleotides located at position 3,286–5,220 nt, and is cleaved at the site MTFQ/S. CI contains the <sup>1161</sup>GAVGSGKST<sup>1169</sup> and <sup>1250</sup>DECH<sup>1253</sup> motifs typical of the helicases of the superfamily 2<sup>17</sup>, and three other conserved motifs (<sup>1423</sup>GERIQLGRVGR<sup>1434</sup>, <sup>1380</sup>ATNIENGVTL<sup>1390</sup> and <sup>1277</sup>KVSATPD<sup>1283</sup>) whose functions are not clear. Nuclear inclusion body 'a' protein (NIa) includes the VPg (NIa-VPg) and NIa protein (NIa-Pro). Virus genome-linked protein (VPg) is a basic protein with 191 amino acids (22.2 kDa) encoded by nucleotides located at position 5,380–5,952 nt, and is cleaved at the sequence IQFE/A. Nuclear inclusion protein (NIa-Pro, 5,953–6,681 nt) comprises 243 residues (27.6 kDa), and is cleaved at the sequence MKFQ/G. NIa-Pro has cysteine protease active site which is responsible for cleavage at the P3/6K1, 6K1/CI, CI/6K2, 6K2/VPg, VPg/NIa-Pro, NIa-Pro/Nib, Nib/CP catalytic site with the consensus <sup>2011</sup>H-34X-D-67X-G-X-C-G-14X-H<sup>2132</sup>. Nuclear inclusion body 'b' protein (Nib) comprises 533 residues (61.0 kDa) encoded by nucleotides located at position 6,682–8,280, and was cleaved at the site ASAQ/S. It also contains the typical RdRp motif GDD (<sup>2515</sup>GNNSGQPSTVVDNTLMVILAIIVYSLIKLGHPPYRTHKDIKIFVNGDD<sup>2561</sup>). Finally, the CP (8,281–9,129 nt) comprises 285 residues (32.2 kDa), and also contains the highly conserved <sup>2735</sup>DAG<sup>2737</sup> motif involved in aphid transmission<sup>18</sup>.

**Phylogenetic and recombination analysis.** For phylogenetic analysis, multiple-alignments were first performed using the Muscle program<sup>19</sup>. Phylogenetic relationships between PMaV and other potyviruses were estimated based on their full genome sequences (Fig. 3a) and the amino-acid sequences of their coat proteins (Fig. 3b). The two phylogenetic trees had similar topologies, and both phylogenetic trees showed that this virus is most closely related to LMV. Based on the full genome sequence, this virus clusters with LMV, while on the amino acid sequence of CP, it forms a distinct branch adjacent to LMV.

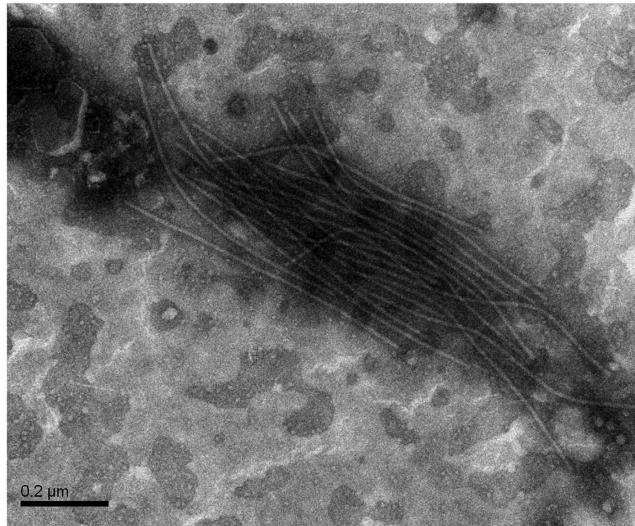
Recombination is a dominant feature of potyvirus evolution. To see whether or not the novel PMaV was generated by recombination, recombination analysis was performed using two different programs (TOPALi and RDP). No recombination events were detected, independently of the data set used for the analysis (data not shown).

**Detection of PMaV particle.** Using electron microscopy, the typical filamentous *Potyvirus* particles were found in infected pecan leaves with mosaic symptoms (Fig. 4), but were absent in healthy, control pecan leaves. The filamentous particle appeared to be non-enveloped and flexuous, measuring 680 ~ 800 nm long and 12 ~ 15 nm in diameter (Fig. 4). This result, together with sequence and structure composition, indicates that the virus identified in this study is a new member of the genus *Potyvirus* within the family *Potyviriidae*.

**Host range determination.** An experimental host range for PMaV was determined by mechanical inoculation of two plant species in *Solanaceae* family that are considered as common virus indicator hosts. No evidence of infection was detected in any of the tested *Nicotiana tabacum*, or *N. benthamiana* (data not shown). This result is in line with the fact that most potyviruses have narrow host ranges.



**Figure 3.** Neighbor-joining phylogenetic tree of *Potyviridae* viruses based on full genomic sequences (a) and CP amino acid sequences (b). Branch numbers represent the percent similarity in the bootstrap values out of 1000 replicates.

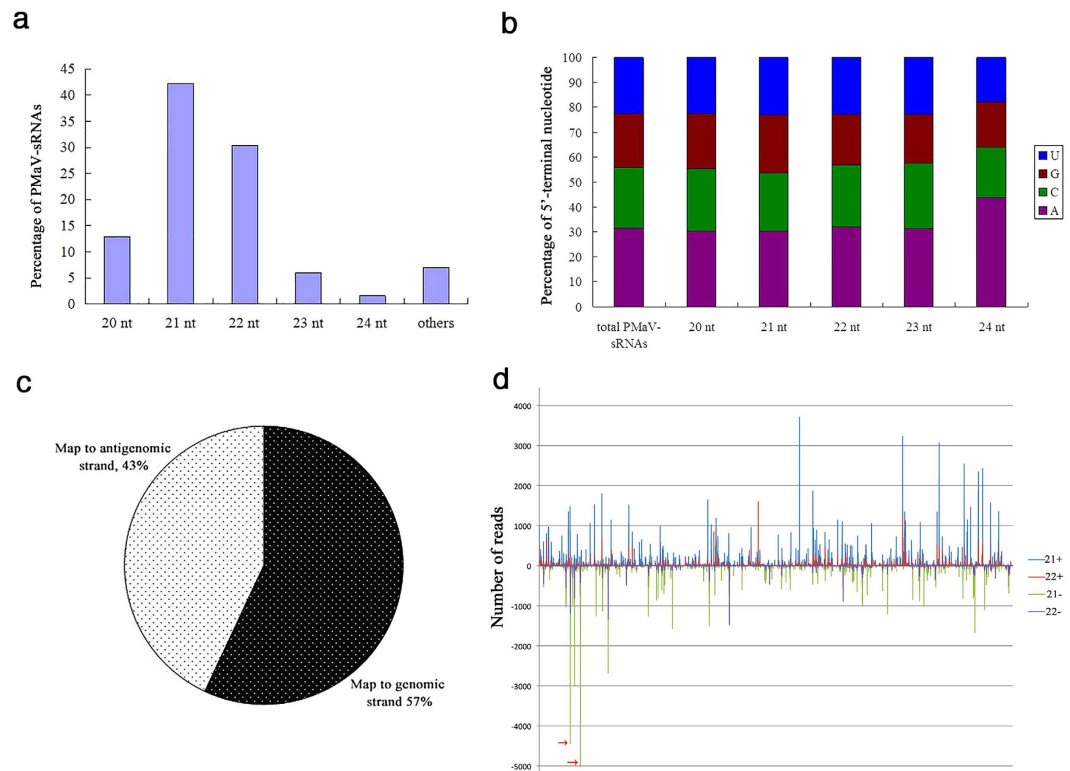


**Figure 4.** Filamentous virion in a leaf dip from infected pecan leaves with mosaic symptom. The bar represents 200 nm.

**Profiling of PMAV-derived small RNAs.** A total of 628,030 sRNAs can be mapped to the PMAV genomic sequence, which account for 3.77% of the total sRNAs. The size classes of PMAV-derived small RNAs (PMAV-sRNAs) were mostly 21 to 24-nt in length; among them the 21-nt class is most prominent, accounting for 42.22% of the total PMAV-sRNAs, followed by the 22-nt sRNAs (Fig. 5a). In *Arabidopsis*, the 5'-terminal nucleotide partially determines the preference of sRNAs for AGO proteins. Therefore, the distribution of 5'-terminal nucleotides was determined for the sequenced PMAV-sRNAs (Fig. 5b). For all four types of PMAV-sRNAs characterized (i.e., 20-, 21-, 22- and 23-nt), there is no significant difference for 5'-terminal nucleotide preference. In contrast, the 24-nt PMAV-sRNAs primarily contained adenine at the 5'-terminus (44%), indicating that they might be loaded into pecan AGO2 and AGO4 homologues, and may participate in transcriptional silencing for host transposons and repeated sequences.

Polarity analysis of redundant PMAV-sRNAs showed that they were derived slightly more often from the positive-sense genomic sequence (57%) than the antisense genomic sequence (43%) (Fig. 5c). Further analysis of those 21 and 22-nt PMAV-sRNAs showed that, within the PMAV genome, the distribution peaks representing the 21-nt PMAV-sRNAs almost overlapped perfectly with the peaks representing the 22-nt PMAV-sRNAs (Fig. 5d). The PMAV-sRNAs from both strands were discontinuous and unevenly distributed along the viral genome (Fig. 5d). However, three significant hotspots of small RNAs were identified along the PMAV genome. Among them, two are from the positive-sense genome, and are located in the CI-encoding region and CP-encoding region, respectively. The other hotspot is derived from the minus-strand genome, located in the regions encoding





**Figure 5. Small RNA profiling of PMAV from infected pecan leaves.** (a) Bar graph showing the mapped PMAV-sRNA size distributions; (b) Relative frequency of 5'-terminal nucleotide for 20 to 24-nt PMAV-sRNAs; (c) Statistical analysis of PMAV-sRNAs mapped to the genomic or antigenomic sequences; (d) Genome-wide mapping of PMAV-sRNAs. The horizontal axis represents the relative position along the PMAV genome. The vertical axis represents the number of PMAV-sRNAs reads mapped to the PMAV genomic (+) or antigenomic (-) sequences. Arrows represent 811 nt and 610 nt of the minus-genomic RNA sequence.

the P1 and HC-Pro proteins (bases from 590–811). In the region, two 21-nt PMAV-sRNAs, beginning at 811 nt and 610 nt, were highly repeated (Fig. 5d).

## Discussion

Pecan tree are increasingly being grown commercially all over the world because pecan nuts supply many valuable nutrients for a healthy diet. At present, there are no reported viral diseases found in pecans. In our research, we isolated and identified a new virus, tentatively named PMAV, from pecan in Lin'an, Zhejiang province, China using small RNA deep sequencing technology. The complete genomic sequence of PMAV was cloned and sequenced. The full-length nucleotide sequence only shared 58.9% similarity with its closest relative, LMV. Furthermore, using electron microscopy, we identified the typical *Potyvirus* filamentous particles in infected pecan leaves presenting mosaic symptoms. As a whole, these results support our finding that PMAV is a new species in the genus *Potyvirus*. During our investigation, we observed very severe mosaic symptoms in PMAV infected leaves; however, its pathogenicity in pecan is still being assessed. Also, this virus cannot be mechanically inoculated onto the two common hosts reported for other members of the family *Potyviridae*, indicating that PMAV may also have narrow host ranges. Other plants nearby the pecan trees are still under testing.

To date, several potyviruses have been reported as targets of the RNA silencing machinery of various herbaceous host plants. However, there is little information on this mechanism in woody plants. Here, profiling of small RNAs from PMAV-infected pecan leaves was performed. The presence of large quantities of PMAV-sRNAs (3.77% of total sRNAs) indicates that RNA silencing-mediated antiviral immunity is also utilized by pecan. In contrast to other potyvirus-derived small RNAs in herbaceous plants, which display a clear preference for adenine and uracil bases at the 5'-end<sup>20,21</sup>, there is no significant difference in the 5'-terminal nucleotide preference in PMAV-sRNAs from pecan (Fig. 5b). Although polarity analysis of redundant PMAV-sRNAs showed that they were derived slightly more often from the positive-sense genomic sequence (57%) than the antisense genomic sequence (43%) (Fig. 5c), we still found that two 21 nt sRNAs, starting at 811 nt and 610 nt of the minus-genomic sequence, were highly repeated (Fig. 5d). This is unexpected, because for positive-strand RNA viruses, the genomic-strand viral RNAs (plus-strand) are typically more abundant than the complementary-strand RNAs (minus-strand genome). The secondary structure of RNA should not be responsible for this difference, because there is no difference in the secondary structure between the two strands of RNA. To better understand this phenomenon, polarity analysis of non-redundant PMAV-sRNAs was performed; we found an almost perfect match of hotspot distribution between the plus- and minus-genome (Supplementary Fig. S1). Therefore, our results give some clue as

to the PMAV-sRNAs' biogenesis: PMAV-derived nonredundant sRNAs originate from viral dsRNAs and subsequent biogenesis of redundant vsRNAs might be a combination of primary vsRNAs from dsRNAs and secondary vsRNAs from host RDRs. RDRs responsible for generating secondary vsRNAs has been shown to preferentially amplify vsRNAs from distinct hot spots within the viral genome in plants<sup>22</sup>. Further research is needed to unveil the function of this kind of minus-genome derived PMAV-sRNA.

There are still many questions regarding PMAV epidemiology that need to be answered. First, pecan trees are native to central and southern North America, but we still do not know the geographic origin of the virus. Second, although most potyviruses in nature are transmitted by aphids, we do not yet know the mode of PMAV transmission. There are several species of aphids that inhabit pecan trees, such as the black margined aphid (*Monellia caryella*) and the yellow pecan aphid (*Monelliopsis pecanis*), but we do not know whether these aphids can transmit PMAV. In contrast to herbaceous plants, woody plants are usually bred by grafting; whether or not this virus can be transmitted to pecan tree through grafting is currently under investigation. Our study, identifying a novel potyvirus that infects pecan trees, will assist in the development of molecular diagnosis of this virus in pecan trees all over the world.

## Methods

**RNA isolation and deep sequencing.** Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, USA) following the manufacturer's instructions. The small RNA library was constructed as described<sup>23</sup>. Briefly, low molecular mass (LMW) RNA in the total RNA samples was enriched using PEG (molecular mass 8000) and NaCl (5M). LMW RNAs (sRNAs) were sequentially ligated to a 3' and a 5' adapter. After ligation, sRNAs were purified by electrophoresis using a 15% denaturing PAGE. The final purified ligation products were reverse transcribed into cDNAs using the Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA). The first strand cDNAs were PCR amplified using the Taq polymerase (Roche Applied Science, Basel, Switzerland) and DNA fragments were purified for high-throughput sequencing using the Solexa platform (BGI, Shenzhen, China).

**Genome assembly and sequence analysis.** Assembled contigs were scanned against the GenBank database using BLASTN (<http://www.ncbi.nlm.nih.gov/>). Then the whole viral genome was obtained through several RT-PCR amplifications covering all the gaps, using a battery of sense and antisense-specific primers designed based on the contig sequences. The 5'- and 3'-terminal sequences were determined using 5'/3' RACE PCR (Clontech, Palo Alto, CA, USA). The primer sequences were shown in Supplementary Table S1. PCR was performed in a total volume of 50  $\mu$ L with Bio-rad thermal cycler S1000<sup>TM</sup> (Life Science, Foster City, CA, USA) with Phusion<sup>®</sup> high-fidelity DNA polymerase (New England BioLabs, Ipswich, MA, USA). The amplified fragment was recovered with the AxyPrep<sup>TM</sup> DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) and cloned using the ZeroBack Fast Ligation Kit (Tiangen, Beijing, China). The positive colonies were further confirmed by DNA sequencing.

The obtained sequences were assembled into a contiguous sequence at a standard of  $\geq 99.9\%$  similarity at each overlapping region using DNASTar 7.01 package (DNASTAR Inc., Madison, USA). The putative cleavage sites and conserved domains of the putative functional proteins in the PMAV polyprotein were determined by comparison of its own sequence with other potyviruses<sup>14</sup>.

**Phylogenetic and recombination analysis.** Thirty-six whole genome sequences and twenty CP amino-acid sequences of representative recognized members of each of the eight genera comprising the family *Potyviridae* were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/>). Phylogenetic trees were constructed by the Neighbor-joining method using MEGA 6.06 software<sup>24</sup>. Recombination events were searched within the PMAV genome or in the CP genes, using two distinct data sets: (i) complete genome sequences representing the major lineages of potyviruses; (ii) all CP sequences deposited in GenBank corresponding to potyviruses; the analysis was performed with TOPALi v2.5 and RDP v4.43. Recombination analyses were performed using default settings for each detection program and a Bonferroni corrected *P*-value cut-off of 0.05.

**Electron microscopy.** Samples from pecan leaves with mosaic symptoms and non-mosaic leaves were prepared for transmission electron microscopy (TEM). The samples were grinded with 0.5 ml ddH<sub>2</sub>O in a mortar, dipped the grid in the serosity for about 10 seconds, then dry with filter paper. The grid was covered with a drop of 2% tungsten phosphate staining solution (pH 7.0) for about 20 seconds, and the copper grid was dried before EM observation.

**Bioinformatic analyses of sRNA sequences.** After removal of the adapter sequences and low quality reads by in-house Perl script, the resulting clean reads were then assembled into larger contigs using Velvet 0.7.31 with the *k*-mer value of 17<sup>13</sup>. For hotspot analysis, PMAV-derived small RNAs (PMAV-sRNAs) were extracted by alignment with the PMAV genome via bowtie1 allowing zero mismatch (version 0.12.8, Langmead, 2009), and then PMAV-sRNA hotspots along the PMAV genome were generated by executing in-house Perl scripts.

## References

1. Wu, Q. *et al.* Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. USA* **107**(4), 1606–1611 (2010).
2. Kreuze, J. F. *et al.* Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* **388**(1), 1–7 (2009).
3. Hagen, C. *et al.* Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses. *Arch. Virol.* **156**(7), 1209–1216 (2011).
4. Xu, Y. *et al.* Identification of Himetobi P virus in the small brown planthopper by deep sequencing and assembly of virus-derived small interfering RNAs. *Virus Res* **179**(3), 235–240 (2014).

5. Su, X., Fu, S., Qian, Y. J., Xu, Y. & Zhou, X. P. Identification of Hop stunt viroid infecting *Citrus limon* in China using small RNAs deep sequencing approach. *Virology* **12**(1), 1–5 (2015).
6. Riechmann, J. L., Lain, S. & Garcia, J. A. Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol* **73**(1), 1–16 (1992).
7. Whitworth, J. L., Nolte, P., McIntosh, C. & Davidson, R. Effect of Potato virus Y on yield of three potato cultivars grown under different nitrogen levels. *Plant Dis* **90**(1), 73–76 (2006).
8. King, A. M. Q., Lefkowitz, E., Adams, M. J. & Carstens, E. B. *Virus taxonomy: Ninth report of the international committee on taxonomy of viruses*. Elsevier Academic Press: San Diego, USA 1069–1089 (2011).
9. Chung, B. Y. W., Miller, W. A., Atkins, J. F. & Firth, A. E. An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. USA* **105**, 5897–5902 (2008).
10. Clark, C. A. *et al.* Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Dis* **96**, 168–185 (2012).
11. Mingot, A. *et al.* The P1N-PISPO trans-frame gene of sweet potato feathery mottle potyvirus is produced during virus infection and functions as RNA silencing suppressor. *J. Virol* **90**(7), 3543–3557 (2016).
12. Untiveros, M. *et al.* A novel sweet potato potyvirus ORF is expressed via polymerase slippage and suppresses RNA silencing. *Mol. Plant Pathol.* doi: 10.1111/mpp.12366 (2016).
13. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**(5), 821–829 (2008).
14. Adams, M. J., Antoniw, J. F. & Beaudoin, F. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol. Plant Pathol* **6**(4), 471–487 (2005).
15. Shibolet, Y. M. *et al.* The conserved FRNK box in HC-Pro, a plant viral suppressor of gene silencing, is required for small RNA binding and mediates symptom development. *J. Virol* **81**(23), 13135–13148 (2007).
16. Carrington, J. C., Cary, S. M., Parks, T. D. & Dougherty, W. G. A second proteinase encoded by a plant potyvirus genome. *J. Euro. Mol. Biol. Org* **8**(2), 365–70 (1989).
17. Fernández, A. *et al.* The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic Acids Res* **25**(22), 4474–4480 (1997).
18. López-Moya, J. J., Wang, R. Y. & Pirone, T. P. Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *J. Gen. Virol* **80**(12), 3281–3288 (1999).
19. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(3), 1792–1797 (2004).
20. Naveed, K., Mitter, N., Harper, A., Dhingra, A. & Pappu, H. R. Comparative analysis of virus-specific small RNA profiles of three biologically distinct strains of Potato virus Y in infected potato (*Solanum tuberosum*) cv. Russet Burbank. *Virus Res* **191**, 153–160 (2014).
21. Ogwok, E., Ilyas, M., Alicai, T., Rey, M. E. & Taylor, N. J. Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* crantz) infected with cassava brown streak viruses. *Virus Res* **215**, 1–11 (2016).
22. Wang, X. B. *et al.* RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**(1), 484–489 (2010).
23. Xu, Y., Huang, L. Z., Fu, S., Wu, J. & Zhou, X. P. Population diversity of Rice Stripe Virus-derived siRNAs in three different hosts and RNAi-based antiviral immunity in *Laodelphax striatellus*. *Plos One* **7**(9), 1602–1603 (2012).
24. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol* **30**(4), 2725–2729 (2013).

## Acknowledgements

This research was supported by grants from the National Key Basic Research and Development Program of China (2012CB114004) and the Open Fund for the Top-Priority-Discipline Forestry of Zhejiang province (KF201330).

## Author Contributions

X.Z. and Y.X. conceived and designed the experiments; X.S. and S.F. performed the experiments; X.S., Y.X., S.F., Y.Q. and L.Z. analyzed the data; X.S., Y.X. and X.Z. wrote the paper.

## Additional Information

**Accession codes:** Sequencing data were deposited in NCBI GenBank database with the accession number KT633868.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Su, X. *et al.* Discovery and small RNA profile of *Pecan mosaic-associated virus*, a novel potyvirus of pecan trees. *Sci. Rep.* **6**, 26741; doi: 10.1038/srep26741 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>