

RESEARCH

Open Access



New isolate of sweet potato virus 2 from *Ipomoea nil*: molecular characterization, codon usage bias, and phylogenetic analysis based on complete genome

Kun-Jiang Wei^{1†}, Ai-Ming Jiang^{1†}, Shuo Jiang^{1†}, Yang-Jian Huang¹, Song-Yu Jiang¹, Xiao-Ling Su¹, Carlos Kwesi Tettey², Xiao-Qiang Wang^{3*}, Wei Tang^{4*} and De-Jie Cheng^{1*}

Abstract

Background Viral diseases of sweet potatoes are causing severe crop losses worldwide. More than 30 viruses have been identified to infect sweet potatoes among which the sweet potato latent virus (SPLV), sweet potato mild speckling virus (SPMSV), sweet potato virus G (SPVG) and sweet potato virus 2 (SPV2) have been recognized as distinct species of the genus *Potyvirus* in the family *Potyviridae*. The sweet potato virus 2 (SPV2) is a primary pathogen affecting sweet potato crops.

Methods In this study, we detected an SPV2 isolate (named SPV2-LN) in *Ipomoea nil* in China. The complete genomic sequence of SPV2-LN was obtained using sequencing of small RNAs, RT-PCR, and RACE amplification. The codon usage, phylogeny, recombination analysis and selective pressure analysis were assessed on the SPV2-LN genome.

Results The complete genome of SPV2-LN consisted of 10,606 nt (GenBank No. OR842902), encoding 3425 amino acids. There were 28 codons in the SPV2-LN genome with a relative synonymous codon usage (RSCU) value greater than 1, of which 21 end in A/U. Among the 12 proteins of SPV2, P3 and P3N-PIPO exhibited the highest variability in their amino acid sequences, while P1 was the most conserved, with an amino acid sequence identity of 87-95.3%. The phylogenetic analysis showed that 21 SPV2 isolates were clustered into four groups, and SPV2-LN was clustered together with isolate yu-17-47 (MK778808) in group IV. Recombination analysis indicated no major recombination sites in SPV2-LN. Selective pressure analysis showed d_N/d_S of the 12 proteins of SPV2 were less than 1, indicating that all were undergoing negative selection, except for P1N-PISPO.

[†]Kun-Jiang Wei, Ai-Ming Jiang and Shuo Jiang contributed equally to this work.

*Correspondence:
Xiao-Qiang Wang
wangxiaoqiang@caas.cn
Wei Tang
tangv0001@163.com
De-Jie Cheng
chengdejie@gxu.edu.cn

Full list of author information is available at the end of the article



Conclusion This study identified a sweet potato virus, SPV2-LN, in *Ipomoea nil*. Sequence identities and genome analysis showed high similarity between our isolate and a Chinese isolate, yu-17-47, isolated from sweet potato. These results will provide a theoretical basis for understanding the genetic evolution and viral spread of SPV2.

Keywords Sweet potato virus 2, Complete genomic sequence, Codon usage bias, Selective pressure, Phylogenetic analysis

Introduction

Sweet potato (*Ipomoea batatas*) plants are grown extensively throughout tropical and temperate regions worldwide and are the third most crucial root crop, following Irish potato and cassava. Viral diseases of sweet potatoes have become widespread, causing severe crop losses worldwide. More than 30 viruses have been identified to infect sweet potatoes [1]. Among the different viruses infecting sweet potato, sweet potato latent virus (SPLV), sweet potato mild speckling virus (SPMSV), sweet potato virus G (SPVG) and sweet potato virus 2 (SPV2) have been recognized as distinct species of the genus *Potyvirus* in the family *Potyviridae*. SPV2 was first isolated from the Taiwan province of China and Nigeria [2] and has since then, the virus has been found in sweet potato crops in mainland China, South Korea, South Africa, Spain, the United States, Australia and other countries [3–7]. The sweet potato SPV2 has also been detected in plants belonging to the genus *Chenopodium* and *Nicotiana*. However, there has been no report of SPV2 infecting *Ipomoea nil* so far.

SPV2 is a positive single-stranded RNA virus that forms scroll-like inclusions in the cytoplasm [8]. SPV2 encodes a polyprotein characteristic of a Potyvirus, which is cleaved into ten functional proteins: the first protein (P1), helper-component proteinase (HC-Pro), the third protein (P3), the first 6 K protein (6K1), cytoplasmic inclusion protein (CI), the second 6 K protein (6K2), viral genome-linked protein (VPg), nuclear inclusion protein a-proteinase (NIa), nuclear inclusion protein b (Nib) and coat protein (CP) [9]. In addition, there is a small protein, pretty interesting sweet potato potyvirus ORF (PISPO) and pretty interesting potyvirus ORF (PIPO) encoded by a translational frameshift in P1 and P3 proteins [10, 11].

SPV2 has not been studied extensively, but its complete genome sequence analysis was recently reported [4, 12–14]. The P1 protein was highly conserved in potyviruses that infect sweet potatoes [12]. During turnip mosaic virus (TuMV) infection, the P1 protein influenced virus proliferation. The host factor NOD19 interacts with the P1 and is necessary for strong infection [15]. The P1 of TuMV also interacts with the chloroplast protein cpSRP54 leading to its degradation through the 26 S proteasome and autophagy pathways. This ultimately inhibits the synthesis of Jasmonic acid (JA) by suppressing cpSRP54 thereby enhancing viral infection [16]. Additionally, the P1 protein of some viruses are involved in

host selection. The nucleotide and amino acid sequence of P1N-PISPO has been identified among SPVG, SPV2, sweet potato feathery mottle virus (SPFMV) and sweet potato virus C (SPVC).

Recombination is an important force in the evolution of plant viruses, including several *Potyviridae* members [17, 18]. Potential recombination events occurred in the P1, HC-Pro, and Nla-Nib regions within the SPFMV and SPLV species. The P1 is under a looser evolutionary constraint and is vulnerable to recombination in the *Potyviridae* family [18–20]. The GmEF1A and GmVAT have a strong interaction with SMV-P3 and play a critical role in virus replication. Blocking GmEF1A and GmVAT can inhibit the infection of soybeans by SMV [21]. The interaction between the membrane protein SMV-P3 and GmRHP may also be involved in the infection of potato virus Y. GmRHP may be an important host factor for P3 to participate in the replication of potato virus Y [22]. The analysis of protein interaction networks revealed that some proteins interacting with SMV-P3N-PIPO are involved in the host signal transduction and chloroplast photosynthetic system pathways, which are related to the formation of soybean symptoms caused by SMV. This indicates that P3N-PIPO may also play a role in the virus's pathogenicity besides mediating the virus's intercellular transport [23]. The nucleotides and amino acid sequences of CP are highly conserved among SPFMV, SPVC, SPVG, SPV2 and Sweet potato virus-Zimbabwe [12]. The variations in CP is mainly at the N-terminal region among the different species or different strains of these viruses [12]. The high level of homology at the C-terminal regions of the isolates indicates that this region may play a critical function in these viruses. Studying the genetic structure and diversity of viruses is important to understand the molecular evolutionary histories of their virulence, dispersion and emergence of epidemics [24].

In this study, we reported for the first time the occurrence of SPV2 in *I. nil* in China and analyzed the complete genomic sequence of SPV2-LN. Our results will provide a theoretical basis for understanding the genetic evolution and viral spread of SPV2, and could provide a theoretical basis for comprehensive control of the virus.

Materials and methods

Sample collection

Symptomatic leaf samples of *I. nil* showing yellow vein symptoms were collected in November 2021 at Guangxi

University, Nanning City, Guangxi Province, China (N22°50'28.41", E108°17'9.00") and photographed. Inoculum was prepared from the symptomatic leaf (1 g/10 mL 1×PBS) and used to mechanically inoculate *N. benthamiana* plants. Symptom expression was observed and photographed at 4 days post inoculation.

RNA extraction, RT-PCR

Total plant RNA was extracted from the leaves of the symptomatic *I. nil* using the Plant RNA Extraction Kit (Vazyme Biotech, Nanjing, China) according to the manufacturer's instructions. The first strand of cDNA was synthesized by mixing 5 µL of total RNA and 1 µL oligo(dT)₁₈ in a PCR tube and incubating at 65 °C for 5 min and quickly transferring into ice for 2 min. Residual DNA was removed by adding 2 µL of 4×gDNA wiper mix (Vazyme Biotech, Nanjing, China) to the above mixture and incubating at 42 °C for 3 min. 2 µL of 5×Hiscrip-tII qRT SupermixII (Vazyme Biotech, Nanjing, China) was added to the above mixture and incubated at 50 °C for 30 min. Finally, the mixture was maintained at 85 °C for 5 s, and stored in a -20 °C freezer till further use.

Small RNAs sequencing

An RNA library was constructed using the Small RNA Sample Pre Kit (Vazyme Biotech, Nanjing, China). Total RNA was used as the starter sample, and adapters were directly added to both ends of the Small RNA before reverse transcription to synthesize cDNA. Subsequently, the target DNA fragments were amplified by PCR using primers SPV2-1 F / SPV2-1390R, SPV2-1 F / SPV2-1350R, SPV2-7103 F / R, SPV2-10147 F / oligo(dT)₁₈ (Supplementary table A1). A total of 50 µL PCR reaction system was made consisting of 2 µL forward primer (10 mM), 2 µL reverse primer (10 mM), 2 µL cDNA, 25 µL 2×taq master mix (Vazyme Biotech, Nanjing, China), and 19 µL ddH₂O were added into the PCR tube. The PCR reaction condition was pre-denaturation at 95 °C for 5 min. The mixture was denatured at 95 °C for 15 s, annealed at 55 °C for 15 s, and extended at 72 °C for 90 s (elongation time according to 1 kb/min) with 32 cycles. Finally, the mixture was maintained at 72 °C for 10 min and stored in a 4 °C. The amplicons were separated by PAGE gel electrophoresis and sequenced using the Illumina Novase sequencing platform.

Rapid amplification of cDNA ends (5' RACE)

The 5'-UTR sequence was amplified using the Rapid Amplification Kit for cDNA Ends (RACE) (Sangon Biotech, Shanghai, China) according to the manufacturer's protocol. The first strand of cDNA was synthesized as a DNA-RNA hybrid using total RNA as a template and SPV2-243R as a primer (Supplementary table A1). The mixture with 15 µL DNA-RNA mix, 15 µL 5×Hybrid RNA

Degeneration Buffer, 1 µL RNase H, and 44 µL DEPC water was kept at 30°C for 1 h. Single-stranded DNA was purified with a rapid gel recovery kit (Sangon Biotech, Shanghai, China). The mixture with 26 µL cDNA, 10 µL 5×TdT Buffer, 12.5 µL 0.1% BSA, 1 µL 100 M dATP, and 0.5 µL TdT polymerase was kept at 37°C for 10 min. SPV2-243R, 5' RACE-Q_T, and 5' RACE-Q₀ were used as primers for PCR amplification, and SPV2-136R/5' RACE-Q₁ were used as primers for the second round of nested PCR [25, 26].

Genomic sequence analysis

The original data files obtained after sequencing were converted into sequenced reads (Raw data) by base calling analysis. To ensure the quality of sequence reads, we obtained clean reads by removing the low-quality ends. 18 to 26 nt small RNAs were acquired from the above clean reads for subsequent analysis. We compared the above small RNAs with GenBank virus RefSeq using Bowtie to identify the specific viruses infecting the plants. Longer contigs were generated by mapping small RNA assemblies and splicing with SPAdes and Velvet. We annotated the contigs using the databases of NCBI Nr, NCBI Nt and GenBank Virus RefSeqs. We have submitted the raw data related to this study to NCBI. The accession to cite for these SRA data is PRJNA1040476. For detailed information, please refer to this website <https://www.ncbi.nlm.nih.gov/sra/PRJNA1040476>.

The complete genome of SPV2-LN was obtained by fragment splicing using DNAMAN (version 10) and SEQMAN [27]. The complete genome of SPV2-LN and 20 SPV2 isolates obtained from the NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) database (Supplementary Table A3) were used for sequence alignment using CLUSTAL W of MEGA11. Polyprotein cleavage site analysis was performed on the website <https://www.dpvweb.net/potycleavage/index.html> to determine the protease recognition sites for SPV2-LN and 20 SPV2 isolates.

Genome base composition and codon usage bias

To study the evolution and environmental adaptation of different virus isolates, the base composition of SPV2-LN was analyzed by using DNAMAN. The RSCU (Relative synonymous codon usage) value was calculated by codonW1.4.2, and the codon bias was analyzed using the GGLOT2 and APLOT packages of the R language [28]. If the RSCU value is <1, the codon is used less frequently than other synonymous codons; If RSCU value >1, the codon is used more frequently than other synonymous codons; If RSCU value =1, the codon has no preference.

Sequence consistency rate analysis

Based on the coding region of SPV2-LN and information from SPV2 isolates in GenBank, the start and end positions of each protein were determined. The CLUSTAL W program of MEGA11 was used for sequence alignment of the SPV2 genomes, and the nucleotide and amino acid sequence identities were determined using BIOEDIT [29].

Recombination analysis

All the SPV2 nucleotide sequences were aligned with CLUSTAL W in MEGA11. Recombinant analysis was performed using the RDP4 package [30, 31]. The package contains RDP, CENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN and 3SEQ. Each software parameter was set to the default value, and the *P*-value was 0.05. When more than six algorithms support recombination, the isolate is considered to be recombinant, otherwise, it is considered to be non-recombinant [30, 32].

Phylogenetic and genetic distance analysis

To explore the phylogenetic relationship between the SPV2 isolates and other isolates, a phylogenetic tree was constructed using the neighbor-joining method, with the bootstrap value set to 1000 [33]. The Kimura two-parameter model (K2P) method in MEGA11 was used to calculate the genetic distance within and between groups, and SPFMV (GenBank number NC_001841.1) was used as an out-group.

Selection pressure analysis

To reveal the evolutionary pressure on the SPV2 species [34], the online analysis software Datamonkey (<http://www.datamonkey.org/>) was used to calculate the d_N/d_S (non-synonymous/synonymous) values between different genes. The values of d_N/d_S were used to analyze and predict the selection pressure on each SPV2 protein. The codon selection pressure of 12 proteins at different positions of 21 SPV2 isolates were determined. $d_N/d_S=1$, $d_N/d_S<1$ and $d_N/d_S>1$ indicate neutral, negative and positive selection, respectively.

Results

Symptoms of SPV2-LN

The *I. nil* infected with SPV2 showed mottling and yellow vein symptoms (Fig. 1A). The *Nicotiana benthamiana* plants inoculated with this sample slurry by mechanical inoculation also showed leaf curling and yellowing symptoms at 4 dpi (Fig. 1B). The SPV2 infecting the *I. nil* plant was detected by small RNA sequencing, and the isolate was named SPV2-LN. All spliced sequences yielded a complete genomic sequence of SPV2-LN with a characteristic *Potyvirus* genome structure (Fig. 1C).

Base composition and codon preference of SPV2-LN genome

Based on genome sequencing and splicing results, the complete genome size of SPV2-LN was predicted to be 10,606 nt. This was deposited in the GenBank with accession number OR842902. The length of the coding region was 10,278 nt, which was composed of 3,266 A (31.8%), 2,558 T (24.9%), 2,497 G (24.3%) and 1,957 C (19.0%). The CDS encoded a polyprotein containing 3,425 amino acids of which leucine (Leu) was the most abundant amino acid, accounting for 8.82% and Tryptophan (Trp) was the least abundant amino acid, accounting for only 1.20%. The relative synonymous codon usage (RSCU) of the degenerate codon of SPV2-LN (Fig. 2), showed a preference for all amino acids, except for Cys, Glu, Lys and Gln: GCC (Ala), GAU (Asp), UUU (Phe), GGA (Gly), CAU (His), AUA (Ile), UUG (Leu), AAU (Asn), CCA (Pro), AGA (Arg), AGU and UCA (Ser), ACA (Thr), GUU (Val), UAU (Tyr). The third codon GC content (GC3s) was 39.4%, and the frequency of A/U was higher than that of G/C. The RSCU value was >1 and with 28 codons, among them 21 codons ending in A/U, indicating that A and U were the preferred codons in the genome (Supplementary table A2). It was revealed that the codon terminal bases of the SPV2-LN genome were mainly A/U.

Protease recognition site analysis

The genome structure of SPV2-LN contains a conserved gene sequence of *Potyvirus*. The SPV2-LN polyprotein is cleaved into 12 mature proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa, NIB, CP, P1N-PISPO and P3N-PIPO proteins.) by protease hydrolysis and transcriptional sliding. The conserved motifs of the protease cleavage sites of each coding region of the SPV2-LN polyprotein were determined (Table 1). The protease recognition sites of 20 SPV2 isolates downloaded from NCBI were shown in Supplementary Table A5.

Consistency rate of SPV2-LN with other SPV2 isolates

The amino acids and nucleotides sequence identities of SPV2-LN with other 20 SPV2 isolates were analyzed (Supplementary table A4). SPV2-LN had the highest nucleotide and amino acid sequence identity with isolate yu-17-47 (GenBank No. MK778808), 92.7% and 97.0% respectively. Among the 12 proteins, the amino acid sequence identities between SPV2-LN and all other SPV2 isolates were 87.0–95.3% for P1, 55.8–98.0% for HC-Pro, 28.1–94.6% for P3, 59.6–96.1% for 6K1, 64.6–98.7% for CI, 47.1–90.5% for 6K2, 59.2–98.9% for VPg, 61.7–98.7% for NIa, 65.8–97.6% for NIB, 52.1–97.7% for CP, 27.0–96.9% for P3N-PIPO and 59.3–93.6% for P1N-PISPO. The nucleotides sequence identities between SPV2-LN and all other SPV2 isolates were 89.2–96.8% for P1, 58.3–87.5% for HC-Pro, 44.0–87.6% for P3, 55.1–92.9% for

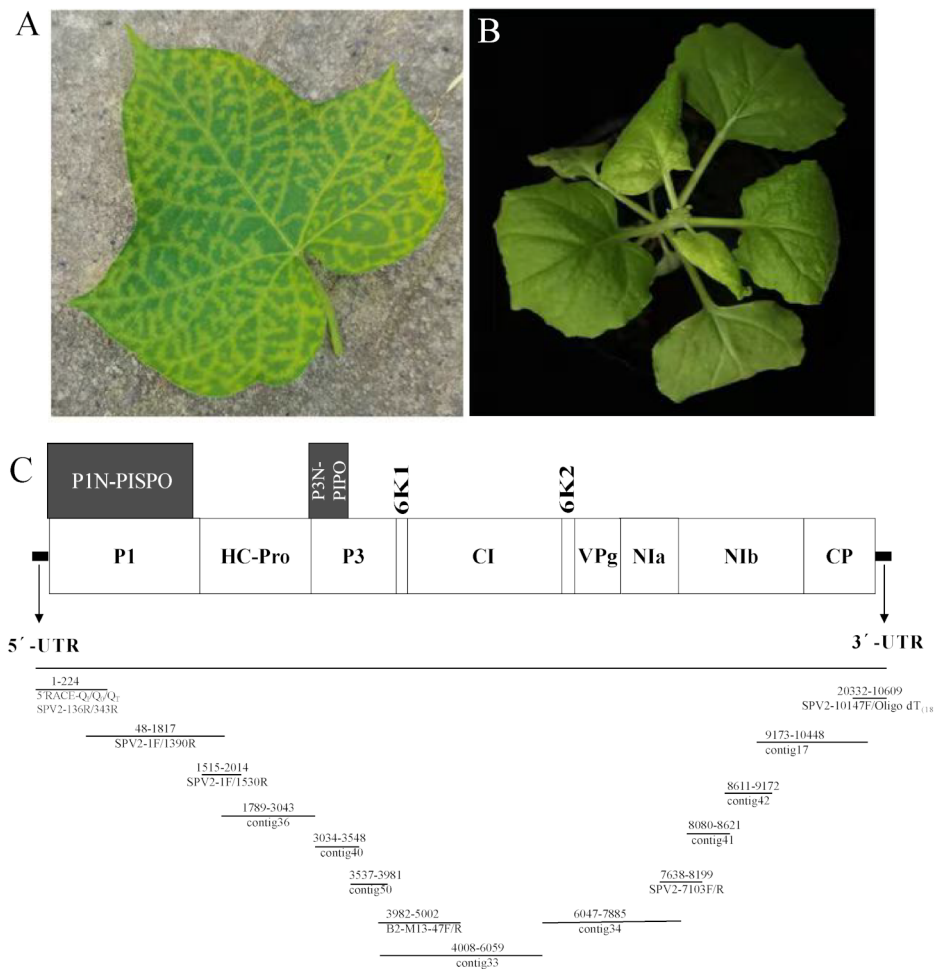


Fig. 1 The symptoms on different hosts and amplification strategy of SPV2-LN. **(A)** Field symptoms on *I. nil*. **(B)** Symptoms on *N. benthamiana* plants at 4 dpi. **(C)** Genome structure of SPV2-LN and genome sequencing primer design

6K1, 63.0–92.0% for CI, 54.0–89.3% for 6K2, 59.4–93.7% for VPg, 59.3–95.4% for NIa, 62.2–94.4% for Nib, 52.8–95.5% for CP, 33.8–85.4% for P3N-PIPO and 61.0–96.7% for P1N-PISPO. The nucleotide sequence identities of the P1N-PISPO and P1 proteins were higher than those found for the other proteins.

P1N-PISPO, HC-Pro, CP and Nib contained recombination sites

The nucleotide sequence analysis of 21 SPV2 isolates using RDP4 showed that four SPV2 isolates had similar recombination sites (Table 2). The recombination isolates yu-17-47 were from China, SSBles-74_ZA and SPV2 were from South Africa, and SPV2-SP1 was from Greece. Meanwhile, the recombination analysis of the nucleotide alignment of each of the genes showed no obvious recombination sites in the sequences of P3, P3N-PIPO, 6K1, CI, 6K2, VPg and NIa. Recombination sites were only found in P1N-PISPO of SSBles-74_ZA, HC-Pro of SPV2, CP of SPV2-SP1 and Nib of SPV2-SP1 (Table 3).

Phylogenetic and genetic distance analysis

The phylogenetic tree was constructed using the Neighbor-Joining method based on the complete genome sequences of 21 SPV2 isolates (Fig. 3). The results showed that the 21 SPV2 isolates grouped into four clusters (Fig. 3). The SPV2-LN isolate was most closely related to isolate yu-17-47 in group IV (Fig. 3). SSBles-74_ZA and SPV2 from South Africa were in the group III and II. The rest of the SPV2 isolates were clustered in group I.

Genetic distances within and between groups were analyzed (Table 4). The results showed that the genetic distance within group 1 was the lowest. The genetic distance between group IV and other groups (group IV vs. III=0.447, group IV vs. II=0.443, and group IV vs. I=0.475), was more than the genetic distance between the other three groups. The majority of SPV2 isolates were isolated from sweet potato, and only SPV2-LN was isolated from *I. nil* for the first time, indicating that SPV2 may have a broad host range.

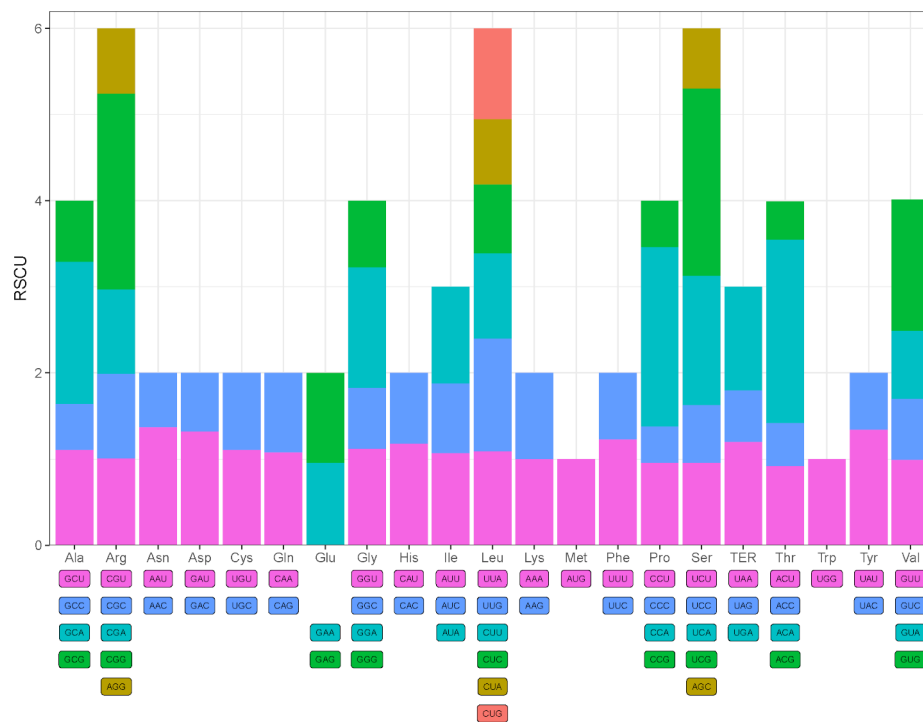


Fig. 2 The RSCU value of codons in SPV2-LN. RSCU value > 1, indicating relative positive codon bias

Codon selection pressures

We examined the codon selection pressures at different positions of the 12 proteins in SPV2 isolates. The analysis showed that all of the d_N/d_S of SPV2 isolate proteins were less than 1. There was only one positive selection site in the P1N-PISPO protein. However, the other proteins were under strong negative selection pressure, the selection pressure of HC-Pro was the lowest among all the proteins (Table 5).

Discussion

This is the first report of SPV2 isolated from *I. nil*. This study obtained the complete nucleotide sequences of SPV2-LN by miRNA sequencing and RT-PCR. The structural characteristics and phylogenetic relationships of the SPV2 genome were analyzed by sequence consistency comparison and phylogenetic analysis [35, 36]. This research could provide a basis for revealing the evolution and pathogenicity of SPV2-LN at the molecular level and help with sweet potatoes viral control.

The complete genomes of 21 SPV2 isolates ranged between 10,562 and 10,747 nt, excluding the poly-A tail which contained a polyprotein. Isolate SPV2-LN had genome size of 10,606 nt excluding the poly-A tail which falls within the range of previous isolates. Genome size may differ significantly between closely related species due to various factors including host cell type, and environmental stress, among other factors [37]. SPV2-LN

was closely related to yu-17-47 and it is closely related to the common ancestor of the four groups. The position of SPV2-LN in Group IV after phylogenetics analysis is of great significance, as it not only displays the phylogenetic relationships and evolutionary processes between species but also provides an important scientific basis for our understanding of the origin and evolution of virus biodiversity

Twelve mature proteins are formed through polyprotein processing and frameshift translation strategies including the P1 protein. Previous studies have shown that the P1 protein exhibited high variation and is a recombination hotspot in potyviruses infecting sweet potato [38]. However, in SPV2, the P1 protein was more conserved, and the sequence of P3 was more variable than the other 11 proteins. The P1 protein is conserved in potyviruses that infect sweet potato, and some viral P1 protein play a role in host selection [39]. This suggests that the conservation of P1 in SPV2 may play a role in the virus's ability to infect and replicate in specific host plants. Previous studies have shown that the genomes of three potyviruses, including SPFMV, sweet potato virus C (SPVC) and SPVG, encoded P1N-PISPO protein. In contrast, the P1N-PISPO protein was not encoded in sweet potato latent virus (SPLV), sweet potato mild mottle virus (SPMMV) and other 78 potyviruses [12, 38, 40]. On the other hand, the high variability observed in P3 and P3N-PIPO may influence the virus's ability to evade

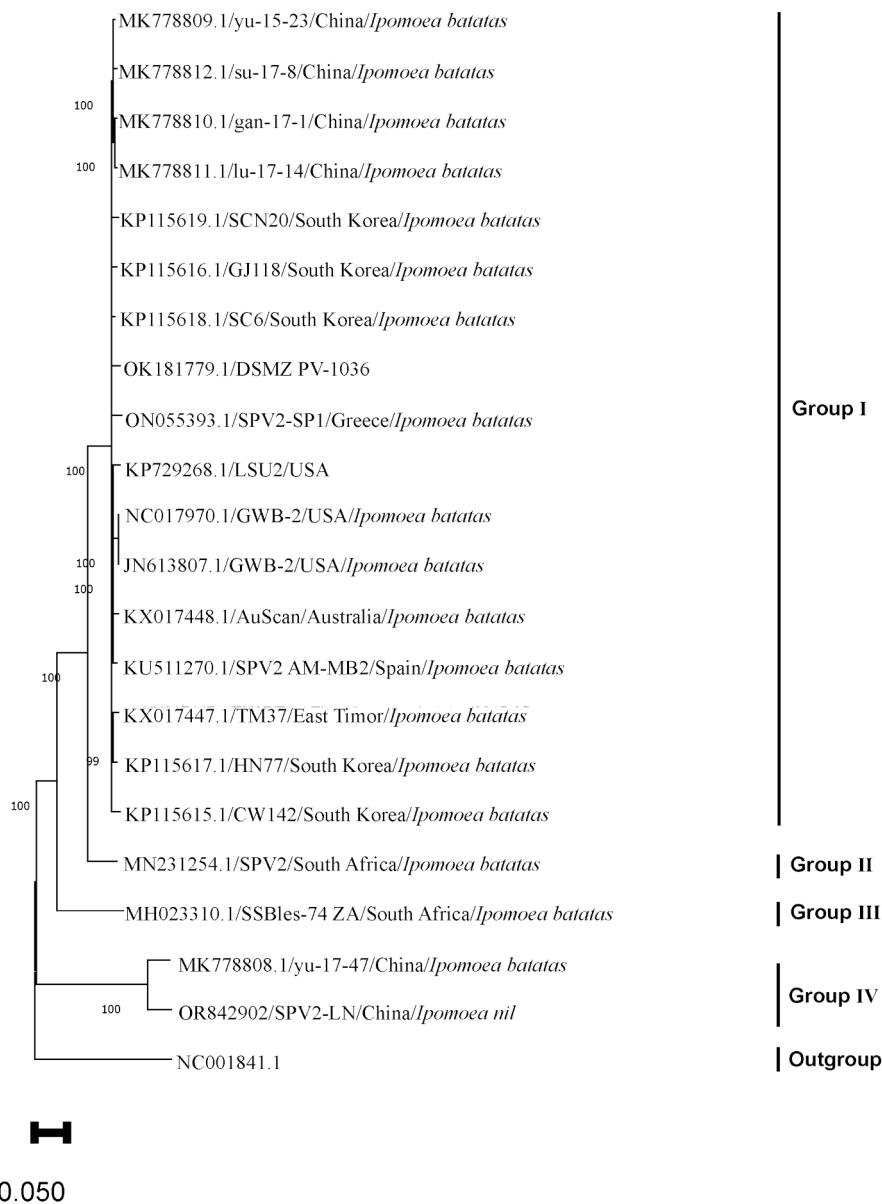


Fig. 3 Phylogenetic relationship between SPV2 isolates. Neighbor-Joining tree generated from the genomic sequences of SPV2-LN and the other 20 isolates of SPV2. ▲ indicates isolate obtained in this study. The scale bars showed a genetic distance of 0.050

the host's immune system allowing the virus to infect a wider host range. P3N-PIPO produced from the P3 gene via transcriptional slippage P3N-PIPO is critical in virus cell-to-cell movement and contributes to the suppression of host resistance [41, 42]. Here we showed that this variability in P3 and P3N-PIPO may be due to recombination which is an important force in the evolution of plant viruses, including several *Potyviridae* members [43]. Further research about the functional implications of these variations would be beneficial to better understand the mechanisms by which SPV2 interacts with its host and causes disease. Future studies can also focus on

investigating the specific roles of these proteins in the virus's life cycle and their potential as targets for antiviral strategies.

Codon use bias is affected by a variety of biological factors, such as GC composition, genome size, tRNA abundance, gene expression level, and mutation pressure [44]. This correlates with tRNA abundance in highly expressed genes which determines the translation efficiency and quality of the protein synthesized [45]. Natural selection and mutational stress are the main factors influencing codon preference [46]. The effect of mutation on a genome is not random but has a directionality toward

Table 1 The genomic structure and protease cleavage sites of SPV2-LN by comparison with other SPV2 isolates

Gene name	Start-end site	Size (nt/aa)	Cleavage site (C-terminus)
5'-UTR	1-134	134/-	
P1	135-1988	1854/618	PYMEQ F/S
HC-Pro	1989-3368	1380/460	KYYV VG/G
P3	3369-4424	1056/352	ENV V HQ/A
6K1	4425-4580	156/52	QDV H HQ/S
CI	4581-6509	1929/643	SCV Q HQ/N
6K2	6510-6668	159/53	TKV S HQ/G
VPg	6669-7247	579/193	DTV I HE/S
N1a	7248-7976	729/243	M Q V Y AQ/S
N1b	7977-9530	1554/518	AN V R H Q/A
CP	9531-10,412	882/293	
P1N-PISPO	135-1938	1803/600	
P3N-PIPO	3369-3857	488/162	
3'-UTR	10,413-10,606	194/-	

Note: conserved amino acids in the cleavage sites are bolded

Table 2 Recombinant analysis of SPV2 coding sequence

Recombinant	Major parent	Minor parent	Region	Detection methods						
				R	G	B	M	C	S	T
yu-17-47	Unknown	lu-17-14	1318-1995	+	+	+	+	+	+	+
SSBles-74_ZA	SPV2	Unknown	1430-2000	+	+	+	+	+	+	+
SPV2	Unknown	HN77	1-2552, 9624-10,795	+	+	+	+	+	+	+
SSBles-74_ZA	Unknown	GWB-2	1-1413, 10,739-10,795	+	+	+	+	+	+	+
SPV2-SP1	SC6	Unknown	10,126-10,582	+	+	+	+	+	+	+
yu-17-47	Unknown	gan-17-1	92-1036	+	+	+	+	+	+	+
SPV2-SP1	lu-17-14	Unknown	7538-7792	+	+	+	+	+	+	+

Table 3 Recombinant analysis of P1, P1N-PISPO, HC-Pro, N1b and CP in SPV2

Recombinant	Parent(major×minor)	Region	Detection methods						
			R	G	B	M	C	S	T
SSBles-74_ZA	Unknown×GWB-2	P1N-PISPO	+	+	+	+	+	+	+
SPV2	Unknown×GJ118	HC-Pro	+	+	+	+	+	+	+
SPV2	SSBles-74_ZA×Unknown	N1b	+	+	+	+	+	+	+
SPV2-SP1	Unknown×DSMZ PV-1036	CP	+	+	+	+	-	+	+

Table 4 Genetic distances within and between groups

Group	I	II	III	IV
I	0.020			
II	0.102	/		
III	0.233	0.225	/	
IV	0.447	0.443	0.475	0.076

higher or lower guanine-plus-cytosine content [47, 48]. In general, the higher the level of gene expression, the stronger the codon bias [49]. The codons used by SPV2-LN are mainly A/U, indicating that A and U are the preferred codons of this genome [50].

Recombination is prevalent in the evolution of potyviruses. In this study, recombination was absent in SPV2-LN, but present in SSBles-74_ZA, SPV2, yu-17-47 and SPV2-SP1. The lack of recombination in SPV2-LN indicates a low recombination frequency, suggesting that recombination is not the primary driving force of SPV2

molecular evolution. However, phylogenetic analysis showed a close relationship between SPV2-LN and yu-17-47 but relatively far apart from other isolates. This shows the existence of some variations among the SPV2 isolates, which may be the result of recombination events and mutations.

The frequency of non-synonymous mutations (d_N) and homologous mutations (d_S) are important for assessing selection pressure. A strong negative selection pressure reduces variation, especially in functionally conserved proteins [48]. For all genes of SPV2, the d_N/d_S ratio was lower than 1, indicating negative selection. In general, strong negative selection pressure tends to suppress variation, which generally occurs in functionally conserved proteins. Through recombination, genetic variation can be generated, mutation load can be reduced and new strains that change the biological characteristics of the virus can be generated contributing to the evolution of

Table 5 Selection pressure analysis of SPV2

Gene	d_n/d_s	Sites under positive selection	Sites under negative selection
P1	0.3570	0	19
HC-Pro	0.0596	0	37
P3	0.1350	0	68
6K1	0.0848	0	15
CI	0.0649	0	166
6K2	0.1250	0	13
VPg	0.1140	0	33
NIa	0.0865	0	44
NIb	0.1010	0	89
CP	0.2020	0	10
P1N-PISPO	0.7470	1	13
P3N-PIPO	0.1040	0	40

the virus [51–53]. Only one end of the P1N-PISPO protein was under positive selection while the rest was under negative selection. This indicates that recombination may not be the main driving force of SPV2 molecular evolution, but negative selection may be the main force driving SPV2 evolution.

Conclusion

This is the first report on the occurrence of SPV2 in *I. nil* in China. SPV2-LN presented clear symptoms in *I. nil* and using small RNA sequencing, RT-PCR and RACE, we characterized the genome of SPV2-LN. This discovery provides evidence of viruses' ability to replicate, spread, and adapt to multiple host environments, thereby accelerating the speed and scope of virus transmission. Though single-potyvirus infections may cause mild or no symptoms in sweet potato, synergistic infection may arise which may cause severe symptoms. We advocate strengthening the monitoring and control of SPV2, especially in areas where SPCSV is severe to prevent co-infection and yield loss of sweet potatoes. This study will provide a basis for the in-depth study of the host range and the development of management or preventive measures against SPV2.

Abbreviations

NIa	Pro-nuclear inclusion protein a-proteinase
SPVG	Sweet potato virus G
SPMSV	Sweet potato mild speckle virus
SPLV	Sweet potato latent virus
SPMMV	Sweet potato mild mottle virus
RT-PCR	Reverse transcription polymerase chain reaction
CP	Coat protein
PISPO	Pretty interesting sweet potato potyvirus ORF
NIb	Nuclear inclusion protein b
PIPO	Pretty interesting potyvirus ORF
RACE	Rapid amplification of cDNA ends
PAGE	Polyacrylamide gel electrophoresis
cDNA	Complementary DNA
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
μ L	Microlitre
dATP	Deoxyadenosine triphosphate
TdT	Terminal deoxynucleotidyl transferase

JA	Jasmonic acid
RefSeq	Reference sequence
RSCU	Relative synonymous codon usage

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12985-024-02500-0>.

Supplementary Material 1: Table A1: The name, sequence and amplified fragment of primers used in this study.

Supplementary Material 2: Table A2: The amino acid compositions and codon usage frequency of SPV2-LN.

Supplementary Material 3: Table A3: The accession number, name, geographical and host origins of SPV2 isolates used in this study.

Supplementary Material 4: Table A4: Nucleotide and amino acid sequence identities (%) between the isolates SPV2-LN and other SPV2 isolates infecting sweet potato.

Supplementary Material 5: Table A5: Protein cleavage sites of 21 SPV2 isolates.

Author contributions

Conceptualization, D.J.C. and X.Q.W.; methodology, D.J.C., A.M.J. and K.J.W.; data curation, K.J.W., A.M.J., S.J., Y.J.H., S.Y.J. and X.L.S.; data analysis, K.J.W., A.M.J., S.J. and X.L.S.; chart plotting, K.J.W.; writing—original draft preparation, K.J.W., D.J.C., A.M.J., C.K.T., and S.J.; writing—review and editing, D.J.C., K.J.W., C.K.T., X.Q.W., V.T., S.J., and A.M.J.; All authors have read and agreed to the published version of the manuscript; All authors have read and agreed to the published version of the manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the Natural Science Foundation of Guangxi Province (No. 2024GXNSFBA010322) and the Research Foundation Ability Enhancement Project for Young and Middle-Aged Teachers in Guangxi Universities (No. 2022KY0006).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Guangxi Key Laboratory of Agro-Environment and Agric-Products Safety, Agricultural College, Guangxi University, Nanning 520004, China

²Department of Molecular Biology and Biotechnology, University of Cape Coast, PMB UCC, Cape Coast, Ghana

³Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China

⁴Xuzhou Institute of Agricultural Sciences in Jiangsu Xuhuai Area, Xuzhou 221131, China

Received: 2 July 2024 / Accepted: 11 September 2024

Published online: 19 September 2024

References

- Clark CA, Davis JA, Abad JA, Cuellar WJ, Fuentes S, Kreuze JF, Gibson RW, Mukasa SB, Tugume AK, Tairo F, Valkonen JPT. Sweet potato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Dis*. 2012;96:168–85.
- Rossel HW, Thottappilly G. (1988) Complex virus diseases of sweet potato. In: Exploration, maintenance, and utilization of sweet potato genetic resources. Report of 1st Sweet Potato Planning Conference 1987. International Potato Centre, Lima, Per.
- Trenado HP, Lozano G, Valverde RA, Navas-Castillo J. First report of *Sweet potato virus G* and *Sweet potato virus 2* infecting sweet potato in Spain. *Plant Dis*. 2007;91:1687.
- Perez-Egusquiza Z, Ward LI, Clover GRG, Fletcher JD. Detection of *Sweet potato virus 2* in sweet potato in New Zealand. *Plant Dis*. 2009;93:427.
- Tairo F, Jones RAC, Valkonen JPT. Potyvirus complexes in sweet potato: occurrence in Australia, serological and molecular resolution, and analysis of the sweet potato virus 2 (SPV2) component. *Plant Dis*. 2006;90:1120–8.
- Ateka EM, Barg E, Njeru RW, Lesemann DE, Vetten HJ. Further characterization of 'sweet potato virus 2': a distinct species of the genus *Potyvirus*. *Arc Virol*. 2004;149:225–39.
- Almeyda CV, Abad JA, Pesic-VanEsbroeck Z. First report of *Sweet potato virus G* and *Sweet potato virus 2* infecting sweet potato in North Carolina. *Plant Dis*. 2013;97:1516.
- Ateka EM, Barg E, Njeru RW, Thompson G, Vetten HJ. Biological and molecular variability among geographically diverse isolates of sweet potato virus 2. *Arch Virol*. 2007;152:479–88.
- Urcuqui-Inchima S, Haenni AL, Bernardi F. Potyvirus proteins: a wealth of functions. *Virus Res*. 2001;74:157–75.
- Rodamilans B, Valli A, San León D, López-Moya JJ, García JA. An atypical RNA silencing suppression strategy provides a snapshot of the evolution of sweet potato-infecting potyviruses. *Sci Rep*. 2018;8:15937.
- Miao R, Ye Z, MacFarlane S, Li Y, Mao Q, Tian Y, Deng Z, Sun Z, Yang J, Li J, Yan F, Chen J, Yu C. The P3N-PIPO protein encoded by wheat yellow mosaic virus is a pathogenicity determinant and promotes its pathogenicity through interaction with NbRLK6 in *Nicotiana Benthamiana*. *Viruses*. 2022;14:2171.
- Li F, Xu D, Abad J, Li R. Phylogenetic relationships of closely related potyviruses infecting sweet potato determined by genomic characterization of *Sweet potato virus G* and *Sweet potato virus 2*. *Virus Genes*. 2012;45:118–25.
- Rodríguez Pardina PE, Beijerman N, Luque AV, Di Feo L. Complete nucleotide sequence of an Argentinean isolate of *sweet potato virus G*. *Virus Genes*. 2012;45:593–5.
- Wang M, Abad J, Fuentes S, Li R. Complete genome sequence of the original Taiwanese isolate of sweet potato latent virus and its relationship to other potyviruses infecting sweet potato. *Arch Virol*. 2013;158:2189–92.
- Bello EO, Yang Y, Fang Y, Chai M, Jiang X, Luan Y, et al. P1 of turnip mosaic virus interacts with NOD19 for vigorous infection. *Front Microbiol*. 2023;14:1216950.
- Ji M, Zhao J, Han K, Cui W, Wu X, Chen B, et al. Turnip mosaic virus P1 suppresses JA biosynthesis by degrading cSPRP54 that delivers AOCs onto the thylakoid membrane to facilitate viral infection. *PLoS Pathog*. 2021;17:e1010108.
- Jonson MG, Seo JK, Choi HS, Kim JS, Kim KH. Effects of recombination on the pathogenicity and evolution of pepper mottle virus. *Plant Pathol*. 2009;25:417–21.
- Seo JK, Ohshima K, Lee HG, Son M, Choi HS. Molecular variability and genetic structure of the population of *soybean mosaic virus* based on the analysis of complete genome sequences. *Virology*. 2009;393:91–103.
- Nguyen HD, Tran HTN, Ohshima K. Genetic variation of the *turnip mosaic virus* population of Vietnam: a case study of founder, regional and local influences. *Virus Res*. 2013;171:138–49.
- Valli A, López-Moya JJ, García JA. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *Gen Virol*. 2007;88:1016–28.
- Luan H, Shine MB, Cui X, Chen X, Ma N, Kachroo P, Zhi H, Kachroo A. The potyviral P3 protein targets eukaryotic elongation factor 1A to promote the unfolded protein response and viral pathogenesis. *Plant Physiol*. 2016;172:221–34.
- Cui X, Lu L, Wang Y, Yuan X, Chen X. The interaction of soybean reticulon homology domain protein (GmRHP) with soybean mosaic virus-encoded P3 contributes to the viral infection. *Biochem Biophys Res Commun*. 2018;495:2105–10.
- Song PW. Identification and analysis of soybean gene interacting with P3N-PIPO gene of soybean mosaic virus. Nanjing Agricultural University; 2016.
- Retel C, Märkle H, Becks L, Feulner PGD. Ecological and evolutionary processes shaping viral genetic diversity. *Viruses*. 2019;11:220.
- Kokkinos CD, Clark CA. Interactions among *Sweet potato chlorotic stunt virus* and different potyviruses and potyvirus strains infecting sweet potato in the United States. *Plant Dis*. 2006;90:1347–52.
- Takumi T. Rapid cDNA cloning by PCR screening (RC-PCR). *Methods Mol Biol*. 2002;192:385–9.
- Park DJ. 3' RACE LaNe: a simple and rapid fully nested PCR method to determine 3'-terminal cDNA sequence. *Biotechniques*. 2004;36:586–8.
- Green MR, Sambrook J. Rapid amplification of sequences from the 3' ends of mRNAs: 3'-RACE. *CSH Protoc*. 2019;5:426–30.
- Zhao Y, Su H, Zhou J, Feng H, et al. The APSES family proteins in fungi: characterizations, evolution and functions. *Fungal Genet Biol*. 2015;81:271–80.
- Maag JLV. Gganatogram: an R package for modular visualisation of anatomograms and tissues based on ggplot2. *F1000Res*. 2018;7:1576.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–80.
- Martin DP, Lemey P, Posada D. Analysing recombination in nucleotide sequences. *Mol Ecol Resour*. 2011;11:943–55.
- Martin DP, Khoosal A, Murrell B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015;1:003.
- Martin DP, Murrell B, Khoosal A, Muhire B. Detecting and analyzing genetic recombination using RDP4. *Methods Mol Biol*. 2017;1525:433–60.
- Mingot A, Valli A, Rodamilans B, et al. The P1N-PIPO *trans*-frame gene of sweet potato feathery mottle potyvirus is produced during virus infection and functions as an RNA silencing suppressor. *Virology*. 2016;90:3543–57.
- Jones RAC, Dwyer GI. Detection of sweet potato chlorotic fleck virus and *sweet potato feathery mottle virus*-strain O in Australia. *Australasian Plant Pathol*. 2007;36:591–4.
- Boulesteix M, Weiss M, Biémont C. Differences in genome size between closely related species: the drosophila melanogaster species subgroup. *Mol Biol Evol*. 2005;23:162–7.
- Kwak HR, Kim J, Kim MK, Seo JK, et al. Molecular characterization of five potyviruses infecting Korean sweet potatoes based on analyses of complete genome sequences. *Plant Pathol J*. 2015;31:388–401.
- Ruiz-Bedoya T, McTavish KJ, Av-Shalom TV, Desveaux D, Guttman DS. Towards integrative plant pathology. *Curr Opin Plant Biol*. 2023;75:102430.
- Qin Y, Zhang Z, Qiao Q, Zhang D, Tian Y, Wang Y. Molecular variability of sweet potato chlorotic stunt virus (SPCSV) and five potyviruses infecting sweet potato in China. *Arch Virol*. 2013;158:491–5.
- Choi SH, Hagiwara-Komoda Y, Nakahara KS, et al. Quantitative and qualitative involvement of P3N-PIPO in overcoming recessive resistance against *Clover yellow vein virus* in pea carrying the *cyv1* gene. *J Virol*. 2013;87:7326–37.
- Wen RH, Hajimorad MR. Mutational analysis of the putative *pipo* of soybean mosaic virus suggests disruption of PIPO protein impedes movement. *Virology*. 2010;400:1–7.
- Jonson MG, Seo JK, Cho HS, et al. Effects of recombination on the pathogenicity and evolution of pepper mottle virus. *Plant Pathol J*. 2009;25:417–21.
- Wan XF, Xu D, Kleinhofs A, Zhou J. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol*. 2004;4:19.

45. Kimchi-Sarfaty C, Schiller T, Hamasaki-Katagiri N, et al. Building better drugs: developing and regulating engineered therapeutic proteins. *Trends Pharmacol Sci.* 2013;10:534–48.
46. Li L, Hu Y, He M, Zhang B, Wu W, et al. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics.* 2021;22:38.
47. Sueoka N. Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci.* 1988;85:2653–7.
48. Sueoka N. Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. *Mol Evol.* 1993;37:137–53.
49. Zhou Z, Dang Y, Zhou M, Li L, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci.* 2016;113:E6117–25.
50. Zhang J. On the evolution of codon volatility. *Genetics.* 2005;169:495–501.
51. Nagy PD, Pogany J. Multiple roles of viral replication proteins in plant RNA virus replication. *Methods Mol Biol.* 2008;451:55–68.
52. Padhi A, Ramu K. Genomic evidence of intraspecific recombination in sugarcane mosaic virus. *Virus Genes.* 2011;42:282–5.
53. Gell G, Sebestyén E, Balázs E. Recombination analysis of Maize dwarf mosaic virus (MDMV) in the sugarcane mosaic virus (SCMV) subgroup of potyviruses. *Virus Genes.* 2015;50:79–86.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.