

Uncovering Partial Genomes of Sweet Potato Feathery Mottle Virus, Sweet Potato Virus G, and Sweet Potato Virus C Using Public Next-Generation Sequencing Data

Aakansha Manav¹, Jitender Singh^{2*}, Rekha Dixit¹, Pankaj Kumar¹, Pushpendra Kumar¹, Satya Prakash¹ and Ramesh Singh¹

¹Sardar Vallabhbhai Patel University of Agriculture & Technology, Meerut 250 110, U.P., India

²Department of Microbiology, Chaudhary Charan Singh University, Meerut 250 004, U.P., India

(Received 3 July, 2025; Accepted 18 September, 2025)

ABSTRACT

Sweet potato (*Ipomoea batatas* L.) production is heavily constrained by viral infections, particularly those caused by members of the *Potyviridae* family, such as Sweet potato feathery mottle virus (SPFMV), Sweet potato virus G (SPVG), and Sweet potato virus C (SPVC). Although extensive high-throughput sequencing (HTS) data are available in public repositories, Indian sweet potato datasets remain underexplored for viral diversity. In this study, a publicly available RNA-Seq dataset (SRR20083564) from Indian sweet potato samples was mined to identify viral genomes using a bioinformatic pipeline built in Galaxy. After quality control, trimming, and assembly with SPAdes, BLASTx analyses revealed three viral contigs exhibiting high similarity to SPFMV (9521 nts, 98% identity), SPVG (8133 nts, 99.1% identity), and SPVC (10,217 nts, 98.7% identity). The sequences were annotated, verified through multiple sequence alignments, and deposited in GenBank (BK071748–BK071750). Pairwise identity matrices indicated high conservation in SPVC and SPVG isolates, while SPFMV displayed greater nucleotide divergence but strong amino acid conservation. Phylogenetic analyses of coat protein sequences confirmed that the Indian isolates clustered with global representatives, including isolates from Africa, South America, and Asia, reflecting their worldwide distribution and possible movement through germplasm exchange. A sequence demarcation tool (SDT)-based heatmap further illustrated genetic clustering and highlighted differential diversity among the three viruses. This work demonstrates the utility of *in silico* viral mining to extend genomic resources, provides the partial genomic sequences of SPFMV, SPVG, and SPVC from India, and emphasizes the importance of leveraging public datasets for plant virus surveillance, phylogeography, and crop health management.

Key words: Viral, GenBank, High-throughput sequencing, Divergence

Introduction

Sweet potato (*Ipomoea batatas* L.) stands as a remarkably sustainable crop that addresses critical global challenges including food security, climate change,

and environmental conservation. Sweet potato has low water requirements, prevents soil erosion by creating ground cover, performs weed control, is drought-tolerant and pest-resistant, and can grow under high CO₂ concentrations. Both potatoes and

sweet potatoes produce more calories per hectare than rice or wheat, and with less water, making them exceptionally resource-efficient for feeding growing populations. All parts of the sweetpotato plant, including leaves, stems, and tuberous roots, can be used for human and animal consumption, maximizing agricultural productivity while minimizing waste. As one of the most economically important crops for addressing global food security and climate change issues, sweet potato offers small-holder farmers a resilient, low-input crop that supports multiple Sustainable Development Goals, from poverty reduction to zero hunger, while providing essential nutrition through its high vitamin A content and complete protein profile (Kumar *et al.*, 2020).

But due to vegetative propagation it is highly susceptible to viral infections, particularly those caused by potyviruses, significantly hamper sweet potato production through yield loss, quality deterioration and, severely impacting crop productivity and economic returns (Gibson *et al.*, 2020). Among the numerous viruses infecting sweet potato, Sweet potato feathery mottle virus (SPFMV), Sweet potato virus G (SPVG), and Sweet potato virus C (SPVC) are predominant and economically important pathogens with a global distribution (Adams and Kreuze, 2020).

Sweet potato feathery mottle virus (SPFMV) is the most common and genetically diverse potyvirus infecting sweet potato worldwide. It causes mild to severe symptoms depending on strain and host conditions, and synergizes with Sweet potato chlorotic stunt virus (SPCSV) to induce the severe sweet potato virus disease (SPVD), which leads to significant crop losses (Mbanzibwa *et al.*, 2011). SPFMV exhibits high nucleotide variability but conserved protein sequences, reflecting adaptation under functional constraints (Aritua *et al.*, 2021). It is transmitted by aphids in a non-persistent manner, facilitating rapid spread among crops (Untiveros *et al.*, 2020).

Sweet potato virus G (SPVG) is less extensively characterized but has been increasingly detected in sweet potato-growing regions including South America, Asia, and Africa (Kwak *et al.*, 2023). SPVG infections often produce mild or latent symptoms, complicating accurate diagnosis and limiting epidemiological data (Stewart *et al.*, 2021). Phylogenetic analyses show SPVG isolates form well-defined clades with limited genetic diversity compared to SPFMV (Tugume *et al.*, 2016). Transmission prima-

rily involves aphid vectors, and knowledge of its genetic diversity is critical for developing sensitive detection tools (Adams and Kreuze, 2020).

Sweet potato virus C (SPVC) is an emerging potyvirus identified in multiple continents with distinct genetic lineages (Mbanzibwa *et al.*, 2011). Its symptomology ranges from mild leaf mottling to more severe chlorotic patterns depending on host and strain (Tugume *et al.*, 2016). SPVC has shown lower genetic diversity than SPFMV, but recombination events suggest ongoing viral evolution (Kumar *et al.*, 2020). Like other potyviruses, SPVC is aphid-transmitted and its incidence is increasing in global sweet potato production systems (Wainaina *et al.*, 2022).

In India, Sweet Potato Virus C (SPVC) has limited reports however; there is evidence of SPVC reported in nearby regions such as East Java and Bali in Indonesia with distinct molecular identification and symptom descriptions. Within Indian literature, direct reporting or detection of SPVC in sweet potato is scarce or not well documented in recent studies (Tibiri *et al.*, 2020). A study reported SPVG in combination with other viruses in sweet potato samples in different regions, and molecular analyses suggest SPVG isolates have some genetic differentiation. Though SPVG is primarily confirmed in other countries such as Korea and Taiwan, the virus is recognized as potentially present in the Indian sweet potato virome but not widespread or extensively reported in the current Indian literature. Some global virus studies including India mention SPVG occasionally as a component of mixed infections, but detailed prevalence and incidence data in India remain limited (Ding *et al.*, 2024). Sweet Potato Feathery Mottle Virus (SPFMV) is reported and naturally occurring in India in sweet potato. A study conducted in West Bengal during 2004-2005 involving 26 Indian sweet potato cultivars detected SPFMV among other viruses using nitrocellulose membrane enzyme-linked immunosorbent assay (NCM-ELISA) and protein coat characterization. SPFMV was found in many leaf samples as a single or mixed infection with a frequency of 26.92%. The disease incidence varied across cultivars, with some like Pol-4-9 showing incidence up to 25.19% in 2005. The virus causes symptoms including chlorotic feathery discoloration and vein clearing on leaves, and partial virus purification revealed viral coat proteins at molecular weights consistent with SPFMV (Sinha *et al.*, 2007).

Hence, sweet potato viruses, especially when occurring as mixed infections such as Sweet Potato Virus Disease (SPVD) caused by Sweet Potato Feathery Mottle Virus (SPFMV) and Sweet Potato Chlorotic Stunt Virus (SPCSV); lead to dramatic yield reductions often exceeding 70% and in extreme cases reaching total crop failure. Symptoms include leaf mottling, chlorosis, vein clearing, leaf deformation, and plant stunting, which affect photosynthesis and overall plant vigor. Infected plants produce fewer, smaller, and lower-quality tubers often exhibiting defects like cracked or rough skin, which diminishes marketability (Zou *et al.*, 2025). The spread of these viruses through insect vectors such as aphids and whiteflies further exacerbates the problem, demanding integrated management strategies including use of virus-free planting stock, removal of infected plants, vector control, and breeding for resistance to safeguard sweet potato production and food security. Economically, virus diseases in sweet potato impose substantial losses globally and regionally, undermining both smallholder and commercial cultivation (Tibiri *et al.*, 2020).

Therefore, the advent of next-generation sequencing (NGS) and the accumulation of vast public datasets deposited in the Sequence Read Archive (SRA) have revolutionized virus discovery and genomic characterization approaches (Wainaina *et al.*, 2022). Mining SRA datasets enables the identification of viral sequences from host transcriptomes and metagenomes without additional sampling, providing a cost-effective and powerful method to explore viral diversity and distribution on a global scale (Lauber *et al.*, 2024; Edgar *et al.*, 2022).

Despite increasing metagenomic studies of sweet potato viruses globally, Indian genomic resources in the SRA remain underexploited, representing a significant knowledge gap given the country's major role in sweet potato production and virus disease emergence (Kumar *et al.*, 2020). Mining public Indian sweet potato SRA dataset can uncover partial viral genomic sequences, deepen understanding of SPFMV, SPVG, and SPVC diversity, and clarify their evolutionary relationships within a global context.

This study reports the identification of partial genomic sequences of SPFMV, SPVG, and SPVC from Indian sweet potato SRA dataset. Through sequence similarity analyses, phylogenetic reconstruction, and diversity assessment, the study expands the genomic landscape of these viruses and demonstrates the value of public data mining for plant vi-

rus genomics and epidemiology.

Methodology

SRA data retrieval and pre-processing

A previously reported RNA Sequencing dataset of Sweet potato from India having accession number SRR20083564 was retrieved on 15 June, 2024. This HTS RNA-Seq data generated using Illumina HiSeq 2500 having 46.3% GC content had 23.1M spots and around 6.8G bases. It was submitted by Central Tuber Crops Research Institute under a study titled 'Whole Transcriptome Analysis in Tuber samples of different colours using NextSeq 500 technology'.

This data was then uploaded on freely accessible online platform, Galaxy Australia (Abueg *et al.*, 2024) for further processing using the tool 'Download and Extract reads in FASTQ format from NCBI SRA (Galaxy Version 3.1.1+galaxy1)'. The uploaded datafiles of paired end data were assessed using 'FastQC Read Quality reports (Galaxy Version 0.74+galaxy1)' tool. This tool act as a quality control checks that detect major issues in the data. It shows basic statistics; per base-sequence quality, sequence content and N content; per sequence-GC content and quality score; overrepresented sequences; adapter content; and sequence length distribution. The trimming of reads in paired end dataset files was performed with the help of 'Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)'. The files generated thereafter were again checked by 'FastQC Read Quality reports (Galaxy Version 0.74+galaxy1)'. The reads in the data were then assembled into contigs and scaffolds using 'SPAdes genome assembler for genomes of regular and single cell projects (Galaxy Version 4.1.0+galaxy0)'.

Similarity searching and alignment checks

The nucleotide sequences obtained in the contig file were then searched as translated query sequence for similarity against protein databases using Blastx (Altschul *et al.*, 1990). Mapped contig showing high similarity to viral proteins was then analyzed for the presence of various protein coding regions, by aligning them with the various CDS of the putative reference virus genome available at NCBI using multiple sequence alignment tool Clustal W (Thompson *et al.*, 1994) in BioEdit (Version 7.7.1) (Hall, 1999). The protein coding sequences, their length and their posi-

tions were predicted using ORF Finder (Rombel *et al.*, 2002) of NCBI.

Phylogenetic analysis

Phylogenetic analysis was conducted using the MEGA11 (Tamura *et al.*, 2021) software. The neighbor-joining method with 1000 bootstrap replicates was employed to construct phylogenetic trees from multiple sequence alignments. Clustal W (Thompson *et al.*, 1994) was utilized for performing multiple alignments of nucleotide sequences. The tree construction dataset comprised the coat protein sequence of the viral contigs and multiple global isolated of Sweet potato virus C, Sweet potato feathery mottle virus, Sweet potato virus G, Sweet potato virus 2, and other potyviruses including Leek yellow stripe virus, Yam bean mosaic virus, Onion yellow dwarf virus, Potato virus A, Potato yellow blotch virus, Yam mild mosaic virus, Lily mottle virus along with an outgroup Wheat streak mosaic virus belongs to genus Tritimovirus.

The BioEdit (Hall, 1999) software was utilized to determine the nucleotide and amino acid percent identities of the CP across all sequences in the dataset, including comparisons with obtained viral contig. A nucleotide and amino acid identity matrix was created using BioEdit (Hall, 1999) and MS Excel. SDT heat maps were generated using Sequence Demarcation Tool version 1.3 (SDT v1.3) (Muhire *et al.*, 2014).

Results

The transcriptome analysis of SRA data having accession number SRR20083564 has been performed in this study using Galaxy Australia (Abueg *et al.*, 2024) through a customized pipeline with no major issue detected in the FastQC reports.

A contigs file of 73,886 sequences (59.5 MB size) was obtained using SPAdes program. These sequences were then analyzed using Blastx (Altschul *et al.*, 1990) and a contig of 10217 nts showed the

presence of reads corresponding to proteins of a plant virus i.e. Sweet potato virus C with maximum percent identity of 98.7% (AOW69224) having query coverage of 98%; contig of 9521 nts showed the presence of reads corresponding to proteins of a plant virus i.e. Sweet potato feathery mottle virus with maximum percent identity of 98% (QEV81531) having query coverage of 98%; whereas contig of 8133 nts showed the presence of reads corresponding to proteins of a plant virus i.e. Sweet potato virus G with maximum percent identity of 99.16% (AUD0057) having query coverage of 97% were found and then these viral contigs has been characterized and analyzed using BioEdit software (Hall, 1999) are submitted to GenBank, where it has been assigned the accession numbers BK071748, BK071749 and BK071750.

Sequence Similarity and Phylogenetic Analysis

The current species demarcation criteria for the genus Potyvirus (family *Potyviridae*), suggested by ICTV based on sequence similarity in coat protein gene is <76–77% nucleotide and <80% amino acid identity. The combined sequence identity matrix (Table 2) indicated that BK071748_Sweet potato virus C_C2_India: Meerut showed (95.1 to 98.8) and (94.8 to 99.3) pairwise percentage identity at nucleotide and amino acid levels respectively including (95.1 nts; 94.8% aa) with NC_014742_Sweet potato virus C_C1_Peru; (96.15 nts; 96.8% aa) with OQ102062_Sweet potato virus C_DSMZ PV-0553_Nigeria; (95.9% nts; 97.4% aa) with OR233836_Sweet potato virus C_SPVC_30_Uganda: Mbale; (96.4% nts; 97.1% aa) with KU511269_Sweet potato virus C_PVC AM-MB2_Spain; (98.8% nts; 99.3% aa) with KU877879_Sweet potato virus C_SPVC-China1_China. This indicates that BK071748_Sweet potato virus C_C2_India: Meerut shared sequence similarity with Sweet potato virus C.

BK071749_Sweet potato feathery mottle virus_FMIND_India: Meerut showed (91.7 to 93.5)

Table 1. Statistics of FastQC reports of SRR20083564

S. No.	Basic Statistics	Raw Data		Trimmed Reads	
		Forward Reads	Reverse Reads	Forward Reads	Reverse Reads
1.	Total sequences	23138799	23138799	23128719	23128719
2.	Total bases	3.3 Gbp	3.3 Gbp	3.2 Gbp	3.2 Gbp
3.	GC%	46%	46%	46%	46%
4.	Sequence length	100-151	100-150	2-151	2-151

Table 2. Comparisons of coat protein nucleotide sequence (red) and amino acid (black) identity of pairwise combinations of partial genome sequences of Sweet potato virus C (BK071748), Sweet potato feathery mottle virus (BK071749) and Sweet potato virus G (BK071750) with other coat protein sequences of *Potyviridae* viruses

Seq->nts/aa	BK071748_S_PVC_C2_India:Meerut	KU877879_S_PVC_China1_China	KU511269_S_PVC_AM-MB2_Spain	OR233836_S_PVC_S_PVC_30_Uganda	OQ102062_S_PVC_DSMTZ_PV-0553_Nigeria	NC_014742_S_PVC_C1_Peru	BK071749_S_PFMV_FMIND_India:Meerut	MH763686_S_PFMV_Rw-Nyamagabel_Rwanda	MH763687_S_PFMV_Ug-Iganga3_Uganda	MH763682_S_PFMV_Ke-Nakuru1_Kenya	FJ155666_S_PFMV_Piu3_Peru	MN852852_S_PFMV_sw12353xxx_China	BK071750_S_PVC_G7_India:Meerut	NC_018093_S_PVC_Jesus Maria_Argentina	MF185716_S_PVC_S_PVC_UNB-01_Brazil	NC_017970_S_PV2_GWB-2_USA	KP168261_LYSV_INDIA_INDIA	NC_016441_YBMV_SR_Peru	NC_005029_OYDV_Yuhang_China	CDNAclone_Hungary	JX294310_PYBV_United Kingdom	NC_019412_YMMV_Brazil	NC_005288_LMoV_Sb_China	AF057533_WSMV-Sidney 81_USA
	ID	99.3	97.1	97.4	96.8	94.8	83.8	83.1	83.8	82.8	82.8	82.5	63.1	64.2	63.4	67.9	51.1	48	53.6	53	55.9	47.4	50.1	18.3
	98.8	ID	97.7	98	97.4	95.5	84.1	83.4	84.1	83.1	83.1	82.8	63.1	64.2	63.4	67.6	51.1	48	53.6	53	55.9	47.4	50.4	18.3
	96.4	96.5	ID	98.4	99	97.1	83.4	83.1	83.4	82.5	82.5	82.2	62.2	63.4	62.5	67	51.1	47.7	53.6	52.7	55.5	47.4	50.4	18
	95.9	96	96.9	ID	97.4	95.5	83.8	83.1	83.4	82.5	82.5	82.2	62.8	63.9	63.1	67.6	51.1	48	53.6	53	55.9	47.7	50.4	18
	96.1	96.4	98.8	96.1	ID	97.4	83.1	82.8	83.1	82.2	82.2	81.9	62	63.1	62.2	66.7	51.1	47.7	53.6	52.7	55.9	47.4	50.4	18
	95.1	95.4	97.5	95.3	97.4	ID	81.9	81.5	81.9	80.9	80.9	80.9	60.8	62	61.1	66.1	50.7	47.4	53.3	52.3	55.5	47.4	50.1	17.7
	77.3	77.5	76.7	77.2	76.6	76.5	ID	97.1	97.4	96.1	96.8	96.8	63.5	64	63.7	67.4	50.3	47.4	51.4	51.1	54.2	46.3	50.1	17.5
	77.5	77.4	77.4	77.9	77.3	77.2	93.5	ID	97.4	97.7	98	96.1	62.9	63.7	63.5	67.4	50	47.4	51.4	51.4	53.9	46.6	50.1	17.5
	77.6	77.4	77.4	77.9	77.3	77	92.9	95.9	ID	97.1	97.4	96.1	63.2	64	63.7	67.1	50.9	47.4	51.4	51.4	53.9	46.6	50.1	17.5
	77	76.8	77.1	77.6	77	76.8	91.8	95.9	95.7	ID	98.4	96.5	63.2	64	63.7	67.4	50	48.4	51.4	51.4	53.9	46.9	49.8	17.2
	77.3	77.1	76.9	77.8	76.8	76.5	93.3	97.5	96	96.1	ID	95.5	62.9	63.7	63.5	67.1	49.6	47.7	51.4	51.4	53.9	46.3	50.1	17.7
	76.8	76.6	76.6	77.2	76.6	76.6	91.7	92.3	92.9	93.6	92.3	ID	62.9	63.5	63.5	67.1	50.3	47.4	51.4	51.1	54.2	46.9	50.1	17.5
	60.2	60.4	59.4	60	59.5	59.1	61.3	61.1	61.1	60.9	60.7	60	60.8	60	97.1	69.7	44.5	41.6	44.2	46.7	45.9	43.5	44.5	19
	60	60.2	59.2	59.6	59.3	58.7	60.9	60.4	60.4	60.2	60	59.7	60	98.4	ID	70.5	44.8	41.6	44.5	46.4	45.6	43.8	44.2	19
	63.6	63.5	63.2	63.8	62.8	63.2	65.8	65.3	65.3	65.6	64.9	64.7	68.5	68.5	ID	46.5	43.7	43.7	47.2	47.2	47.8	42.7	45.7	18.6
	51.1	50.5	50.4	50.1	50.7	50.6	50.9	51.8	52.6	52	51.8	51.1	45.5	45.5	45.6	48.9	ID	52	53.4	56.2	54.5	56.7	54.3	20.9
	49.6	49.7	49.7	50.3	49.5	49.5	51.6	52.1	51.3	51.9	51.5	51.1	46.6	46.4	46.6	47.6	59.3	ID	55.5	57	57	51.4	56	20.4
	51.4	51.3	51.4	50.8	51.6	51.2	51.3	51.1	50.9	50.7	50.4	50.5	45.7	45.9	45.9	48.1	54	57.5	ID	53.1	55.3	51.7	57.2	16.3
	53.5	53.8	54.2	54	54.2	54.6	55.1	54.6	54.6	54.7	54.4	53.8	47.8	47.6	47.3	49.6	55.2	58.7	58.6	ID	82.5	53.5	54.3	20.6
	52.6	52.7	53	53.4	53.3	52.9	54.5	54.4	54.3	53.5	53.5	53.2	46.1	46.6	46.6	49.8	53.9	58	59.9	76.7	ID	53.5	55	19.5
	50.6	50.6	50.9	50.3	50.9	50.7	49.8	49.8	49.8	49.8	49.7	48.8	44.6	45.2	45.2	48.3	53.3	56.7	61.2	58.6	58.1	ID	53.5	20.9
	50	49.6	50.5	50.5	50.1	50.3	50.6	51.6	51.4	51.6	51.6	50.7	45.5	46	45.3	47.4	56.2	57.5	55.6	56.5	54.9	57.2	ID	20.9
	34.8	34.8	34.9	34.7	35.1	34.8	35.1	35	35.8	35.1	35.4	35.6	35	34.7	34.4	34.4	33	34.5	31.1	32.9	31.8	31.8	32.4	ID

and (96.8 to 97.1) pairwise percentage identity at nucleotide and amino acid levels respectively including (91.7 nts; 96.8% aa) with MN852852_Sweet potato feathery mottle virus_swt12353xxx_China; (93.3 nts; 96.8% aa) with FJ155666_Sweet potato feathery mottle virus_Piu3_Peru; (91.8% nts; 96.1% aa) with MH763682_Sweet potato feathery mottle virus_Ke-Nakuru1_Kenya; (92.9% nts; 97.4% aa) with MH763687_Sweet potato feathery mottle virus_Ug-Iganga3_Uganda; (93.5% nts; 97.1% aa) with MH763686_Sweet potato feathery mottle virus_Rw-Nyamagabe1_Rwanda. This indicates that BK071749_Sweet potato feathery mottle virus_FMIND_India: Meerut shared sequence similarity with Sweet potato feathery mottle virus.

BK071750_Sweet potato virus G_G7_India: Meerut showed (97.7 to 98.4) and (97.1) pairwise percentage identity at nucleotide and amino acid levels respectively including (97.7 nts; 97.1% aa) with MF185716_Sweet potato virus G_SPVG-UNB-01_Brazil: Pernambuco; (98.4 nts; 97.1% aa) with NC_018093_Sweet potato virus G_JesusMaria_Argentina. This established the ge-

netic relatedness of BK071748_Sweet potato virus C_C2_India: Meerut to Sweet potato virus G.

To investigate and confirm the evolutionary relationships between sequences retrieved after processing of SRA data (SRR20083564) with other *Potyviridae* family members, a phylogenetic tree was constructed (Figure 1). This tree incorporated Coat protein sequences of BK071748_Sweet potato virus C_C2_India:Meerut, BK071749_Sweet potato feathery mottle virus_FMIND_India:Meerut and BK071750_Sweet potato virus G_G7_India:Meerut along with 21 coat protein sequences sourced from NCBI, which included multiple global isolates of Sweet potato virus C, Sweet potato feathery mottle virus, Sweet potato virus G, Sweet potato virus 2, and other potyviruses including Leek yellow stripe virus, Yam bean mosaic virus, Onion yellow dwarf virus, Potato virus A, Potato yellow blotch virus, Yam mild mosaic virus, Lily mottle virus along with an outgroup Wheat streak mosaic virus belongs to genus Tritimovirus.

The analysis showed that BK071748_Sweet potato virus C_C2_India:Meerut was clustered with isolates KU511269 from Spain; OQ102062 from Nigeria; NC_014742 from Peru; KU877879 from China and OR233836 from Uganda: Mbale as they were clustered together in same clade, demonstrating that sequence BK071748 showed close phylogenetic relationship with Sweet potato virus C.

BK071749_Sweet potato feathery mottle virus_FMIND_India:Meerut was clustered with isolates MN852852 from China; MH763687 from Uganda; MH763682 from Kenya; MH763686 from Rwanda and FJ155666 from Peru under the same clade indicating close phylogenetic relationship between them whereas BK071750_Sweet potato virus G_G7_India:Meerut was clustered with isolates NC_018093 from Argentina and MF185716 from Brazil: Pernambuco demonstrating that sequence BK071750 showed close phylogenetic relationship with sweet potato virus G. Therefore, it is demonstrating the worldwide distribution of these economically important sweet potato pathogen.

Genetic diversity among SPVC, SPFMV and SPVG isolates infecting sweet potato was assessed using the Sequence Demarcation Tool (SDT v1.2) (Muhire *et al.*, 2014). This tool calculates pairwise sequence identity and visualizes the results as a color-coded heat map. In the matrix, darker shades represent higher sequence similarity, while lighter tones indicate greater genetic divergence; the result-

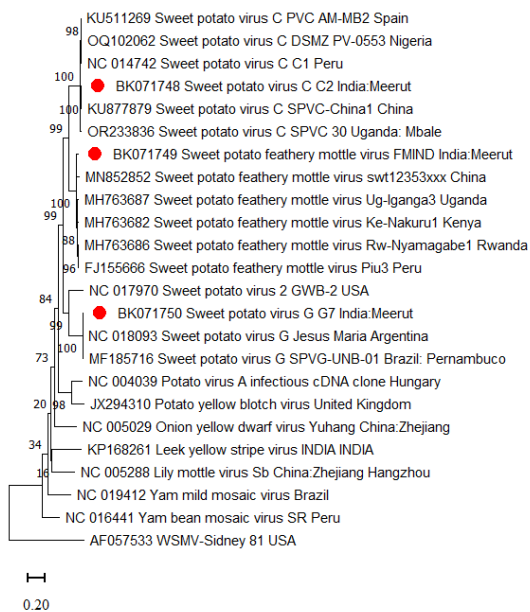


Fig. 1. Phylogenetic tree depicting the evolutionary relationship of BK071748_Sweet potato virus C_C2_India:Meerut, BK071749_Sweet potato feathery mottle virus_FMIND_India: Meerut and BK071750_Sweet potato virus G_G7_India:Meerut with other coat protein sequences of 21 global isolates using Neighbor joining algorithm. The evolutionary distances were computed using p-distance method with 1000 bootstrap replicates.

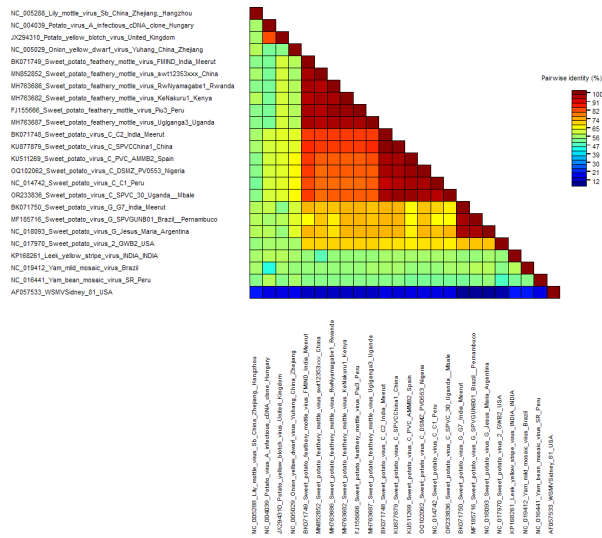


Fig. 2. Heat map showing pairwise percentage identities of coat protein amino acid sequences of *Potyvirus* species

Heatmap (Figure 2) revealed clusters of closely related isolates, suggesting the presence of genetically similar strains, while more divergent sequences may represent novel variants. This finding highlights the evolutionary relationships among the viral isolates and may inform understanding of virus-host interactions within the sweet potato virome.

Discussion

The transcriptome analysis of the SRA dataset SRR20083564 using the Galaxy Australia platform (Abueg *et al.*, 2024) resulted in high-quality sequencing data, evidenced by consistent FastQC metrics including total sequences, bases, and stable GC content (~46%), both before and after trimming (Table 1). These results confirm the reliability of the dataset for downstream viral genome assembly and analysis, aligning with previous studies that emphasize the importance of rigorous quality control in transcriptomic investigations of plant virus infections (Barba *et al.*, 2014; Abueg *et al.*, 2024).

De novo assembly with SPAdes yielded contigs file of 73,886 sequences, among which three significant viral contigs were identified via Blastx analysis (Altschul *et al.*, 1990) as belonging to Sweet potato virus C (SPVC), Sweet potato feathery mottle virus (SPFMV), and Sweet potato virus G (SPVG). The high identity scores (98.7% for SPVC, 98% for

SPFMV, and 99.16% for SPVG) with known viral proteins and substantial query coverage underscores the successful detection of these viruses in the Indian sweet potato sample. These findings are consistent with earlier reports using transcriptomic and metagenomic approaches that have effectively identified mixed viral infections in sweet potato crops worldwide (Kreuze *et al.*, 2009; Cuellar *et al.*, 2015).

Sequence analyses following ICTV species demarcation criteria for *Potyvirus*, which recommend <76–77% nucleotide and <80% amino acid identity thresholds for species differentiation, confirmed that the identified isolates (BK071748, BK071749, BK071750) belonged to their respective species. The nucleotide and amino acid pairwise identities with global reference isolates from Peru, Nigeria, Uganda, Kenya, Rwanda, Spain, Argentina, and Brazil ranged from 91.7% to 99.3%, indicating these Indian isolates are genetically closely related to widely distributed strains. This aligns with global surveys that document the widespread distribution and genetic conservation of these viruses across continents (Mukasa *et al.*, 2006; Lebot, 2019).

Phylogenetic analysis further corroborated these relationships, clustering the Indian isolates with other geographically diverse SPVC, SPFMV, and SPVG sequences (Figure 1). This phylogenetic clustering reflects the global dissemination and evolutionary stability of these viruses, likely facilitated by vegetative propagation and international germplasm exchange (Kreuze *et al.*, 2009; Cuellar *et al.*, 2015). Similar tree topologies have been reported in recent studies, emphasizing the co-circulation of these viruses and the need for integrated management strategies across production regions (Villamor *et al.*, 2019; Mukasa *et al.*, 2021).

The genetic diversity assessment using the Sequence Demarcation Tool (SDT v1.2) (Muhire *et al.*, 2014) revealed clusters of highly similar isolates interspersed with divergent sequences, visualized by the heatmap in Figure 2. This observation is consistent with previous studies demonstrating the utility of SDT in delineating viral strain diversity and detecting emerging variants within complex viromes (Muhire *et al.*, 2014; Mbanzibwa *et al.*, 2011). The presence of genetically distinct variants may have important implications for virus epidemiology, host adaptation, and resistance breeding programs (Lebot, 2019).

Therefore, the mining of public SRA datasets yielded partial genomic sequences corresponding to

Sweet potato feathery mottle virus (SPFMV), Sweet potato virus G (SPVG), and Sweet potato virus C (SPVC), three major potyviruses infecting sweet potato worldwide. The recovered viral sequences from India contribute to the existing body of genomic information on these pathogens and highlight the potential of *in silico* metagenomic approaches for virus discovery and diversity assessment (Adams *et al.*, 2011; Wainaina *et al.*, 2022).

Collectively, this study demonstrates that public SRA mining is an effective and rapid approach for uncovering hidden viral genomes from existing datasets. The recovered partial sequences enhance the genomic representation of Indian sweet potato potyviruses and underscore the country's role as a hotspot for viral diversity given its large sweet potato cultivation zones. Importantly, insights into genetic diversity and global relationships of SPFMV, SPVG, and SPVC provide the framework for designing diagnostic markers, developing virus-free propagation protocols, and implementing biosecurity measures to control virus spread.

Future whole-genome sequencing of these Indian isolates, coupled with recombination detection and population dynamics studies, would further clarify the evolutionary forces shaping their genetic architecture. This work also contributes to the growing need for integrated sweet potato virome mapping projects in Asia, which remain underrepresented compared to African and South American efforts (Wainaina *et al.*, 2022).

Conclusion

This study effectively utilized high-quality transcriptome data (SRA accession SRR20083564) processed through the Galaxy Australia platform to identify and characterize three significant sweet potato viruses: Sweet potato virus C (SPVC), Sweet potato feathery mottle virus (SPFMV), and Sweet potato virus G (SPVG). The viral contigs demonstrated high sequence identity and coverage against known global isolates, confirming their taxonomic classifications in accordance with ICTV species demarcation criteria for Potyvirus. Phylogenetic analyses revealed that the Indian isolates clustered closely with geographically diverse isolates from multiple continents, underscoring the widespread distribution and evolutionary relatedness of these pathogens. Furthermore, genetic diversity assessment using the Sequence Demarcation Tool highlighted both

highly conserved clusters and more divergent variants, providing insights into viral evolution and potential emergence of novel strains. Collectively, these findings enhance the understanding of sweet potato virus diversity and evolution, and the submitted sequences contribute valuable genomic resources for future surveillance, comparative genomics, and integrated disease management strategies for this economically important crop.

Funding

This research received no external funding.

Data Availability Statement

The partial nucleotide sequences of Sweet potato virus C (SPVC), Sweet potato feathery mottle virus (SPFMV) and Sweet potato virus G (SPVG) have been deposited in NCBI under accession number BK071748, BK071749 and BK071750.

Acknowledgments

The authors acknowledge the K.K Singh, Vice-Chancellor, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut 250110, Uttar Pradesh, Bioinformatics facility, Department of Biotechnology, India for providing financial support and the facilities to carry out this research work.

Conflicts of Interest

All authors declare that they have no conflict of interest.

References

- Abueg, L. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*. 52(W1): W83-W94.
- Adams, I. P. and Kreuze, J.F. 2020. Genomic biology of sweetpotato-infecting viruses. *Advances in Virus Research*. 106: 89-140.
- Adams, I. P. 2011. Recent advances in plant virus discovery and diagnostics through metagenomics. *Virus Research*. 159(2): 228-235.
- Altschul, S. F. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215(3): 403-410.
- Barba, M. 2014. High-throughput sequencing technologies for plant virus detection and discovery. *Virus Research*. 188: 138-148.
- Cuellar, W. J. 2015. Metagenomic analysis of sweet potato viral communities. *Plos One*. 10(3): e0118540.
- Ding, Z., Zeng, F., Li, H., Zheng, J., Chen, J., Chen, B. and

- Yue, X. 2024. Identification of sweetpotato virus disease-infected leaves from field images using deep learning. *Frontiers in Plant Science*. 15: 1456713.
- Edgar, R. 2022. Serratus: ultra-high throughput search of the Sequence Read Archive. *Bioinformatics*.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symposium Series*. 41(41): 95-98.
- ICTV, 2025. Virus Taxonomy Profile: Potyviridae. International Committee on Taxonomy of Viruses.
- Kreuze, J. F. 2009. Deep sequencing of plant viral genomes. *Virus Research*. 145(1): 74-83.
- Kumar, P. L. 2020. Seed degeneration and viruses affecting sweetpotato. *Frontiers in Plant Science*. 11: 857.
- Lauber, C. 2024. Deep mining of the Sequence Read Archive reveals viral genomic innovations. *bioRxiv/preprint*.
- Lebot, V. 2019. Sweet potato: Crop improvement and virus disease management. *Agricultural Science*. 10(1): 45-56.
- Mbanzibwa, D.R. 2011. Genetic diversity and evolution of Sweet potato feathery mottle virus. *Virus Genes*. 42: 173-185.
- Mbanzibwa, D.R. 2011. Sequence Demarcation Tool (SDT): A useful tool for the classification of viral isolates. *Virology Journal*. 8(1): 172.
- Muhire, B. M., Varsani, A. and Martin, D. P. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One*. 9(9): e108277.
- Mukasa, S. B. 2006. Genetic variability and population structure of sweet potato feathery mottle virus isolates from East Africa. *Phytopathology*. 96(9): 974-983.
- Mukasa, S. B. 2021. Molecular epidemiology of sweet potato viruses: Insights from recent high-throughput sequencing studies. *Virus Research*. 292: 198209.
- Rombel, I.T., Sykes, K. F., Rayner, S. and Johnston, S.A. 2002. ORF-FINDER: a vector for high-throughput gene identification. *Gene*. 282(1-2): 33-41.
- Sinha, B. and Tarafdar, J. 2007. Occurrence and detection of sweet potato virus disease (SPVD) in West Bengal. *Journal of Applied Horticulture*. 9(2): 123-126.
- Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*. 38(7): 3022-3027.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22(22): 4673-4680.
- Tibiri, E. B., Pita, J. S., Tiendrébéogo, F., Bangratz, M., Néya, J. B., Brugidou, C. and Barro, N. 2020. Characterization of virus species associated with sweetpotato virus diseases in Burkina Faso. *Plant Pathology*. 69(6): 1003-1017.
- Tugume, A.K. 2016. Genetic variability of Sweet potato virus C. *Molecular Plant Pathology*. 17: 431-447.
- Villamor, D.E.V. 2019. Sweet potato virus epidemiology and management. *Annual Review of Phytopathology*. 57: 371-398.
- Wainaina, J. M. 2022. Detecting and characterizing viruses in Africa's sweetpotatovirome using metagenomics. *Nature Communications*. 13: 4989.
- Zou, W., Chen, S. P., Yang, Z. and Chen, X. 2025. Key Sweet Potato Viruses in Fujian Province and Their Distribution, Harmfulness, and Implications in China. *Current Issues in Molecular Biology*. 47(4): 242.
-