



RESEARCH ARTICLE

RNA-dependent RNA polymerase domain-based molecular characterisation and evolutionary analysis of potato virus S

Swati Bhuria^{1,2}, Chinnakaruppan Marimuthu^{1,2}, Niranjana Prem Minipreman^{1,2}, Abhilasha Singh^{1,2}, Nikeshun Vivekananthan¹, Susheel Kumar^{1,2}, Yogita Maheshwari^{1,2*} & Vijayanandraj Selvaraj^{1,2*}

¹Plant Molecular Virology Laboratory, CSIR-National Botanical Research Institute, Lucknow 226 001, Uttar Pradesh, India

²Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201 002, Uttar Pradesh, India

*Correspondence email - maheshwari.yogita@gmail.com; vijayanandraj.s.nbri@csir.res.in

Received: 05 February 2026; Accepted: 22 June 2026; Available online: Version 1.0: 14 September 2026

Cite this article: Swati B, Chinnakaruppan M, Niranjana PM, Abhilasha S, Nikeshun V, Susheel K, Yogita M, Vijayanandraj S. RNA-dependent RNA polymerase domain-based molecular characterisation and evolutionary analysis of potato virus S. Plant Science Today (Early Access). <https://doi.org/10.14719/pst.13996>

Abstract

Potato virus S (PVS) is a globally distributed *Carlavirus* that infects potato (*Solanum tuberosum* L.) and related *Solanum* species. While it often causes mild symptoms, it can have a notable impact on crop productivity. This study examined the genetic diversity and phylogenetic structure of PVS isolates from India using complete genome sequences and the RNA-dependent RNA polymerase (RdRP) domain of the replicase gene, with the aim of assessing the suitability of RdRP as a molecular marker for evolutionary inference. A global dataset of 151 PVS sequences was analysed, along with RdRP domain sequences from 11 Indian isolates and complete genome sequences from two representative isolates. Sequence demarcation tool (SDT) analysis showed pairwise nucleotide identities ranging from 73.8–100 % for complete genomes and 82.2–100 % for RdRP domains. Using a 92 % nucleotide identity threshold, the PVS isolates were grouped into eight intraspecific phylogenetic clades (G1–G8). Phylogenetic trees inferred from RdRP domain sequences showed high topological congruence with complete genome phylogenies, as confirmed by tanglegram analysis. In addition, these phylogenetic groups also broadly matched previously reported PVS phylogroups, including the predominant PVS-O lineage. Population genetic analyses indicated stronger host-associated structuring than geographic differentiation, while the RdRP domain exhibited low nucleotide diversity and strong purifying selection. Collectively, these findings indicate that RdRP domain-based analysis provides a reliable and efficient approach for PVS molecular characterisation, comparative evolutionary studies and the identification of divergent or putative novel PVS variants.

Keywords: genetic diversity; phylogenetic analysis; potato virus S; RNA-dependent RNA polymerase domain; reverse transcription quantitative polymerase chain reaction; virus detection

Introduction

Potato (*Solanum tuberosum* L.) is one of the most important food crops worldwide and ranks third in food production after wheat and rice. Despite its economic importance, potato crops are highly susceptible to various viral diseases, which reduce overall yield and compromise tuber quality. Among the numerous viruses infecting potato, potato virus S (PVS) is a frequently detected and widely distributed pathogen and is considered endemic in nature (1, 2). It is readily spread through tuber cutting, via infected seed tubers and plant-to-plant contact (3). It is also transmitted by aphids such as *Myzus persicae* and *Aphis nasturtii* in a non-persistent manner (4–6). Potato virus S is not transmitted through pollen or seeds (7). Potato virus S has been reported from major potato-growing regions worldwide (8) and can infect hosts belonging to several plant families, including Solanaceae, Amaranthaceae and Apiaceae. The virus often exhibits mild symptoms that make it difficult to identify visually, especially in asymptomatic potato cultivars, although it may cause undulation of leaf margins and rugose leaf symptoms in susceptible cultivars (9–11). The yield loss from PVS infection can range from 5–20 % when it is associated with potato virus X (PVX) and potato virus Y (PVY) (12).

Potato virus S belongs to the *Carlavirus* genus within the family Betaflexiviridae, characterised as a positive-sense, single-stranded RNA virus with a genome of approximately 8.5 kb (13). Its structure consists of a 5' untranslated region (UTR), six open reading frames (ORFs), encodes RNA-dependent RNA polymerase (RdRP), the triple gene block proteins (TGB1, TGB2 and TGB3), a coat protein (CP) and a nucleic acid binding protein, along with a 3' UTR containing a poly (A) tail. Genetic studies have identified distinct groups of PVS isolates, suggesting that genetic diversity is influenced more by host interactions than geographical factors. Based on biological and molecular characteristics, PVS isolates have been grouped into distinct categories. Early classification, based on symptom expression in *Chenopodium quinoa* Willd., categorised PVS into PVS-O and PVS-A (14). Potato virus S-O causes local lesions, while PVS-A leads to systemic infection (15). Subsequent molecular phylogenetic analyses based on complete genome sequences revealed additional lineages within PVS, identifying distinct phylogroups such as PVS-I, PVS-II and PVS-III (16–18), indicating a broader evolutionary divergence within the species. More recently, comprehensive genome-based phylogenetic analyses identified four distinct phylogroups within PVS, further expanding the current understanding of its genetic diversity and evolutionary relationships (8).

The Carlavirus genus is classified based on CP or RdRP gene sequences, with less than 72 % nucleotide identity or 80 % amino acid identity among members. Among these genomic regions, the RdRP gene and its encoded protein are particularly important for evolutionary investigations because they are universally conserved among RNA viruses (19). The RdRP catalytic domain plays a central role in the viral replication cycle by facilitating synthesis of the viral RNA genome (20) and has therefore been widely utilised in molecular evolutionary studies (21, 22). Structurally, RdRPs belong to a large family of polymerases characterised by conserved palm, finger and thumb subdomains (23). Six conserved motifs (A–F) have been identified within these domains, with motifs A–E located primarily in the palm subdomain and motif F in the finger subdomain. These motifs are essential for maintaining polymerase fidelity and ensuring accurate nucleotide incorporation during RNA synthesis (23, 24). Owing to its high degree of conservation, the RdRP domain has become a valuable molecular marker for investigating viral evolution, phylogenetic relationships and virus–host co-evolutionary patterns (19, 24, 25). Although the International Committee on Taxonomy of Viruses (ICTV) does not recognise taxonomic ranks below the species level, increasing availability of nucleotide sequence data has facilitated the identification of genetically distinct viral lineages and phylogroups. However, the absence of supporting biological and serological data for many reported isolates, together with the occurrence of recombinant variants, continues to complicate classification below the species level (26). Recent advances in next-generation sequencing (NGS) have substantially increased the availability of complete viral genome sequences, providing new opportunities for molecular evolutionary and phylogenetic analyses. In this context, the present study focuses on RdRP domain-based molecular characterisation of PVS isolates from different regions of India to assess their genetic diversity and evaluate the suitability of the RdRP domain as a molecular marker for evolutionary and phylogenetic investigations.

Materials and Methods

Source of plant and viral materials

Asymptomatic potato leaf samples were collected from fields and used to obtain the complete genome of PVS. Potato tubers were collected from vegetable markets for obtaining the RdRP domain sequences and virus characterisation. The samples were frozen in -80°C until RNA extraction.

Virus detection

The samples were macerated and tested for the presence of the virus using direct antigen coating-enzyme-linked immunosorbent assay (DAC-ELISA) using in-house generated polyclonal antibody. Briefly, plates were coated with diluted sap (1 : 20) in carbonate buffer (pH 9.6), blocked with 0.5–1 % bovine serum albumin (BSA) and sequentially incubated with primary antisera (1 : 10000) and alkaline phosphatase-conjugated goat anti-rabbit IgG (1 : 20000) at 37°C . Colour development was achieved using p-nitrophenyl phosphate substrate and absorbance was read at 405 nm. Samples with OD $\geq 2\times$ the negative control was considered positive. Healthy potato leaf/tuber samples were used as negative control for ELISA. Samples that tested positive in the ELISA were subsequently subjected to a one-step reverse transcription quantitative polymerase chain reaction (RT-qPCR) assay. Total RNA

was extracted from ELISA-positive potato leaf samples using the cetyltrimethylammonium bromide (CTAB) method. Detection of PVS was performed by one-step RT-qPCR using the SuperScript™ III Platinum™ SYBR™ Green One-Step qRT-PCR kit according to the manufacturer's instructions. The primers were designed from a conserved region near the 3' end of the PVS genome (Forward: 5'-GGATCCTGYAATTCTAAGTGTGATAAT-3'; Reverse: 5'-TGGTATCACCTCAGTTACTCCAA-3'). The RT-qPCR reaction was carried out in a final volume of 50 μL containing 25 μL of 2 $\times$ SYBR® Green Reaction Mix, 1 μL of SuperScript® III RT/Platinum® Taq Mix (including RNaseOUT™), 1 μL each of forward and reverse primers (10 μM), 1 μg of template RNA and nuclease-free water to make up the final reaction volume. Amplification was performed using a QuantStudio™ 5 Real-Time PCR System. The thermal cycling conditions consisted of reverse transcription at 50°C for 5 min, initial denaturation at 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15 sec and annealing/extension at 63°C for 40 sec. RNA extracted from virus-free potato plants and PCR-grade water (non-template control (NTC)) were included as negative controls in each RT-qPCR assay.

Cloning of genome

The first strand of cDNA was synthesised using Revertaid First Strand cDNA synthesis kit as per manufacturer's instruction. The complete genome of the PVS isolates was obtained as five overlapping amplicons using PrimeSTAR GXL Taq polymerase. The primer sets were designed based on full length PVS sequences obtained from National Centre for Biotechnology Information (NCBI) (Supplementary Table 1). The RdRP domain was amplified using generalised primers and resulted in amplification of ~1200 bp amplicon. Polymerase chain reaction amplicons were cloned into the pGEM-T Easy vector (Promega) and transformed into *Escherichia coli* DH5 α competent cells. The clones were sequenced using Sanger sequencing strategy and sequence analysis was performed using BioEdit sequence alignment editor version 7.2.5. The sequences were confirmed by National Center for Biotechnology Information (NCBI) basic local alignment search tool (BLAST) analysis.

Phylogenetic and co-phylogenetic analysis

A comprehensive dataset of 151 complete PVS genome sequences was obtained from the NCBI database and used for evolutionary analysis along with two complete genome sequences and 11 RdRP domain sequences from Indian isolates developed in this study. This dataset was used to conduct an in-depth analysis of the genetic diversity and evolutionary relationships among PVS isolates. The sequence data were subjected to multiple sequence alignment using ClustalW implemented in BioEdit software version 7.2.5. To infer the evolution of the PVS isolates, pairwise distance matrices of the complete genome and RdRP domain sequences were calculated at the nucleotide (nt) level using sequence demarcation tool (SDT) v1.2. The Neighbor-Joining method was employed using MEGA software version 12. To evaluate the evolutionary relationships between complete genome and RdRP domain sequences, a co-phylogenetic analysis was performed using the phytools package in RStudio (version 2025.09.1+401). The aligned, rooted phylogenetic trees were visualised as a Tanglegram, enabling a direct comparison of their topologies. This analysis highlighted the evolutionary congruence between the datasets and provided insights into potential co-evolutionary relationships between the complete genome and the RdRP domain.

Genetic diversity and selection pressure analysis

DnaSP v. 6.0 was used to calculate various parameters to investigate nucleotide polymorphism. Nucleotide diversity (π) of PVS was estimated using a sliding window length of 100 bp with a step size of 25 bp. Genetic diversity was assessed by estimating the total number of mutations (Eta), nucleotide diversity (π), the average number of nucleotide differences (k), theta per site from Eta, Watterson's theta (Θ_w), the number of haplotypes (h) and haplotype diversity (Hd). Neutrality was assessed using Tajima's D, Fu and Li's D* and Fu and Li's F* tests available in DnaSP v. 6.0. The ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated for each dataset using standard parameters in MEGA12. A dN/dS ratio greater than 1 was considered indicative of positive (diversifying) selection, a ratio less than 1 of purifying (negative) selection and a ratio equal to 1 of neutral evolution. Sites subject to positive and negative selection were identified using the single-likelihood ancestor counting (SLAC) method implemented on Datamonkey web server (www.datamonkey.org). The analysis of molecular variance (AMOVA) was performed using Arlequin v3.5 to measure genetic differences among PVS populations. Populations were grouped by geographic region and host species to examine the effects of country and host on PVS genetic diversity. Recombinant sequences were removed before carrying out the AMOVA analysis.

Recombination detection

Comprehensive analysis of recombination was conducted for the complete genome and RdRP domain nucleotide sequences using RDP4 v4.101, employing various methods such as RDP, BootScan, SiScan, Chimaera, GENECONV, MaxChi and 3Seq. The objective was to obtain a consensus of regions potentially resulting from recombination. To identify recombination events among PVS isolates, only those events supported by five or more methods with *p*-values less than 0.01 were considered. Events for which the parental sequences could not be identified were excluded from the analysis.

Results

Survey and detection of potato virus S

Asymptomatic leaves and healthy commercial tuber samples collected from three different states viz., Himachal Pradesh, Uttar Pradesh and Madhya Pradesh covering 12 different locations were used for detection of PVS (Table 1). The samples were tested by one-step RT-qPCR using 3' region specific primers and one-step

RT-qPCR using CP gene specific primers for further confirmation. The qPCR method specifically detected PVS without cross-reacting with other viruses such as PVX or PVY positive samples and no amplification was observed with the NTC.

Genome amplification

Complete genome characterisation of PVS was conducted using isolates from Palampur, Himachal Pradesh and Lucknow, Uttar Pradesh. The full genome was amplified as six overlapping PCR fragments. The PCR fragments were cloned and sequenced using Sanger sequencing. The Palampur isolate, with a genome length of 8460 bp and the Lucknow isolate, with a genome length of 8483 bp, were both sequenced and submitted to NCBI under accession numbers MW331321.1 and OQ473813.1 respectively. The RdRP domain sequence was identified in the InterPro website using PVS replicase gene amino acid sequence. Primers adjoining to the RdRP domain were designed and used for amplification from potato tubers and suspected leaf samples, which resulted in an amplification of 1214 bp. Eleven RdRP domain sequences of PVS were cloned, sequenced and analysed using InterPro. Later, the RdRP domain sequences (993 bp) obtained through InterPro were submitted to NCBI and used for further analysis (Table 1).

Sequence demarcation tool analysis

Sequence demarcation tool analysis was performed using both complete genome and RdRP domain sequences. Pairwise nucleotide identities ranged from 73.8 to 100 % for complete genomes and from 82.2 to 100 % for RdRP sequences. Frequency distribution plots revealed two major peaks for both datasets. The first peak corresponded to species-level demarcation, while the second peak represented variation among isolates within the species (Fig. 1). A 92 % nucleotide identity threshold was selected for isolate discrimination below the species level. Applying this threshold consistently resolved eight groups (G1–G8) in both datasets (Supplementary Table 2 and 3).

Phylogenetic analysis

Phylogenetic analyses based on complete genome and RdRP domain sequences resolved eight major groups (G1–G8) (Fig. 2 and 3). Group G1 predominantly contained *S. tuberosum* isolates with broad geographic distribution and included a few *Solanum lycopersicum* L. isolates. Group G2 consisted mainly of *S. tuberosum* isolates together with single representatives of *Solanum phureja* Juz. et Buk. and *S. lycopersicum*. Group G3 included isolates from Bolivia, the Netherlands and India infecting *S. tuberosum* and *Solanum muricatum* Aiton. Group G4 contained

Table 1. List of potato virus S isolates used in this study, including their source, geographic origin and sequence information

State	Location	Isolate name	Genome region sequenced	Accession number
Himachal Pradesh	Palampur	Palampur	Complete genome	MW331321
	Lucknow	Lucknow	Complete genome	OQ473813.1
	Lucknow	LKO-4	RdRP domain	PP350766
	Bareilly	Brly-36	RdRP domain	PP350767
Uttar Pradesh	Gorakhpur	Grk-2	RdRP domain	PP350772
	Faizabad	Fzd-13	RdRP domain	PP350773
	Hardoi	Hrd-23	RdRP domain	PP350774
	Kushinagar	KN-1	RdRP domain	PP350775
	Raebareli	Rb-47	RdRP domain	PP350776
	Datia	Dt-10	RdRP domain	PP350768
Madhya Pradesh	Morena	Mn-4	RdRP domain	PP350769
	Guna	Gh-25	RdRP domain	PP350770
	Tekanpur	GWT-49	RdRP domain	PP350771

RdRP: RNA-dependent RNA polymerase.

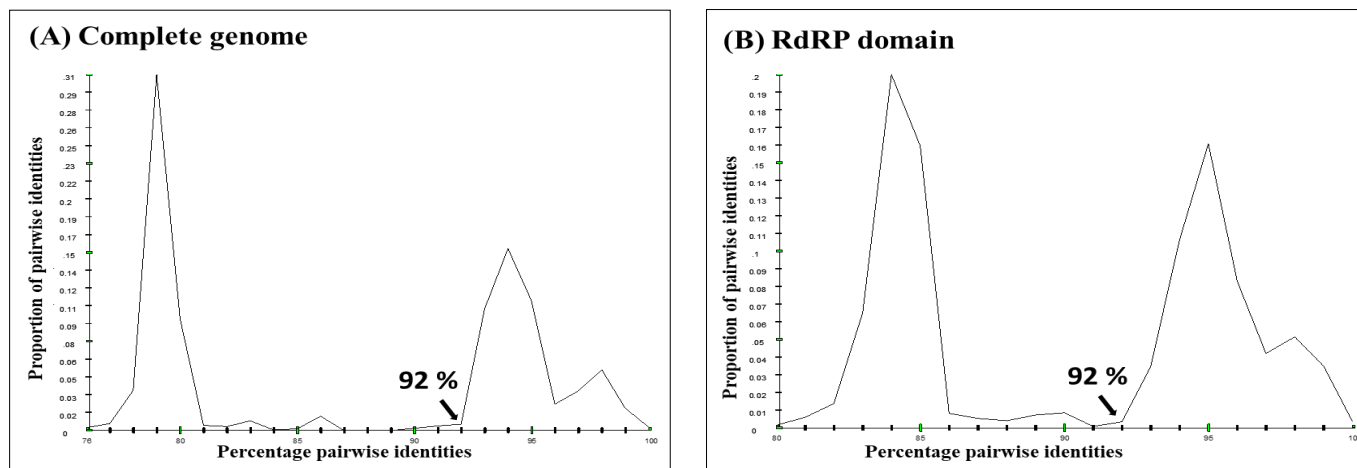


Fig. 1. Potato virus S frequency distribution plot of pairwise nucleotide identity. A: Complete genome; B: RdRP domain. The 92 % cut off value for complete genome and RdRP domain is marked with arrows. RdRP: RNA-dependent RNA polymerase.

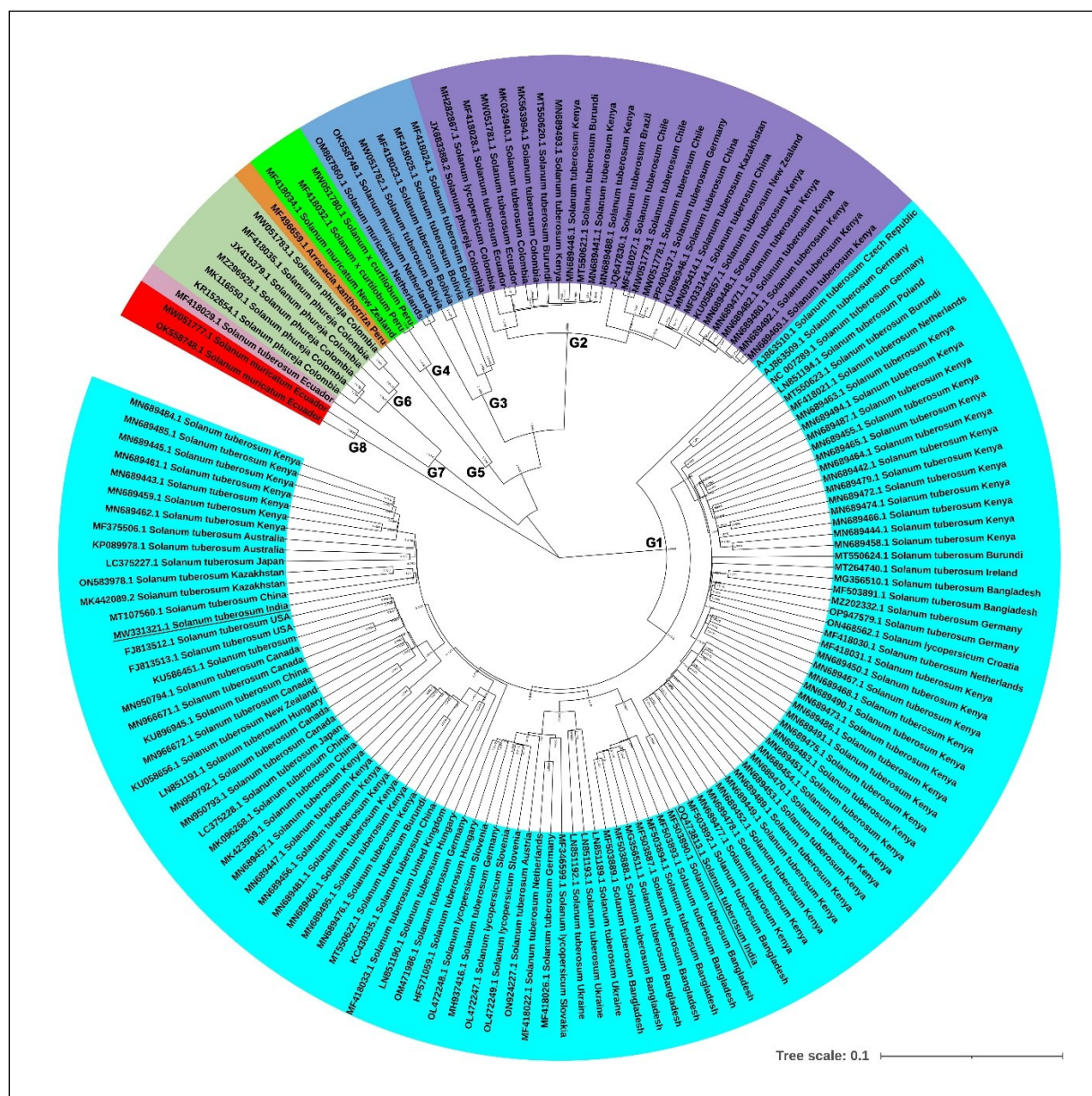


Fig. 2. Phylogenetic analysis of potato virus S (PVS) complete genome sequences.

Circular maximum-likelihood phylogenetic tree constructed using 153 complete genome nucleotide sequences of PVS. The analysis resolved eight major phylogenetic groups (G1–G8), reflecting the genetic relationships among isolates obtained from different host species and geographic regions. Distinct colour coding was used to visualise the clustering of isolates within each phylogenetic group. The observed grouping pattern provides an overview of the genetic diversity and population structure of PVS at the whole-genome level. Potato virus S isolates characterised in the present study are indicated by underlined labels. The scale bar represents the number of nucleotide substitutions per site.

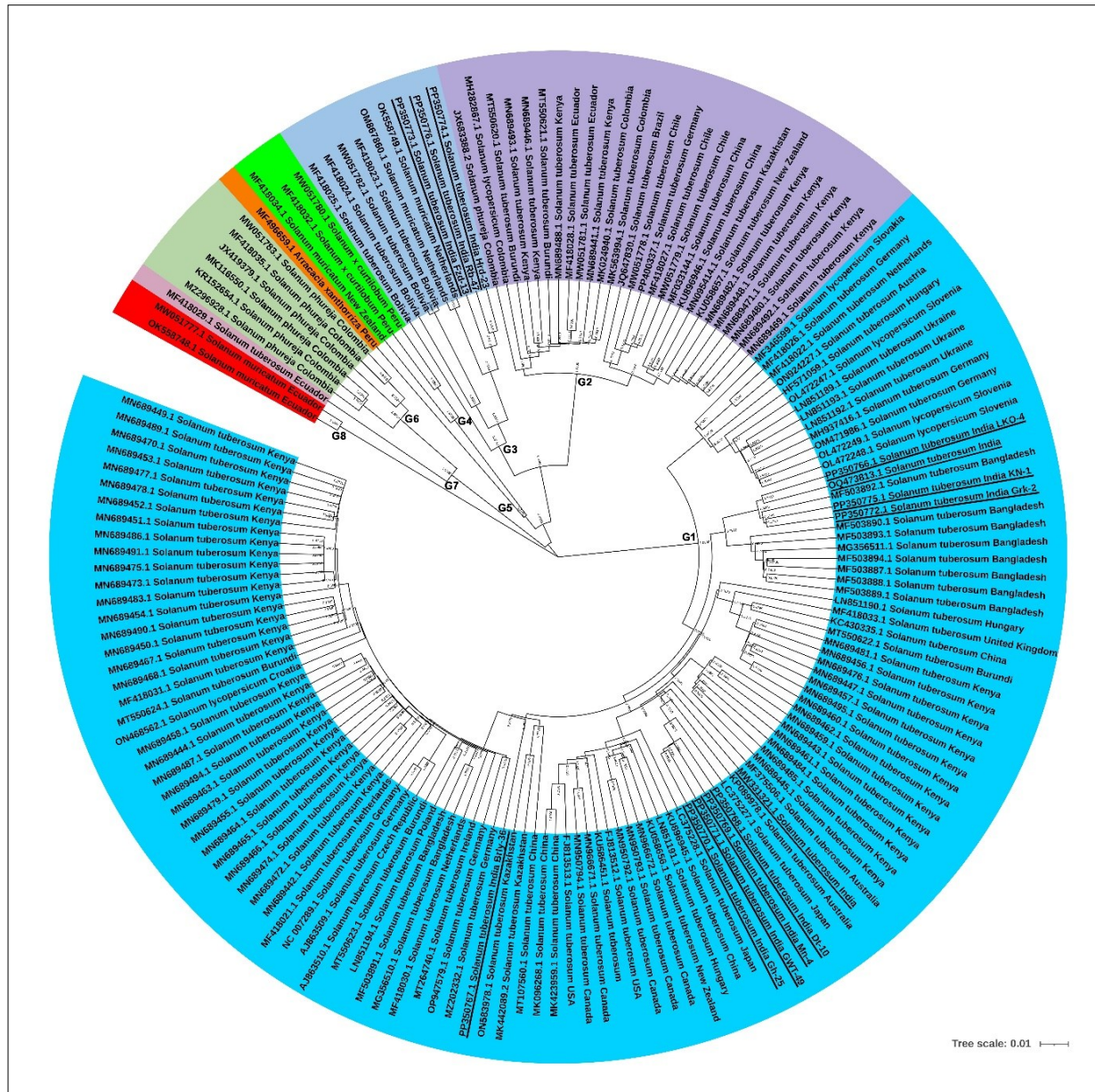


Fig. 3. Phylogenetic analysis of the RNA-dependent RNA polymerase (RdRP) domain sequences of potato virus S (PVS).

Circular maximum-likelihood phylogenetic tree constructed using 164 PVS RdRP domain nucleotide sequences (993 bp). The analysis resolved eight major phylogenetic groups (G1–G8), represented by different colours. Isolates were obtained from diverse host species and geographic regions, illustrating the genetic diversity and population structure of PVS. The clustering pattern observed in the RdRP domain dataset was consistent with that obtained from complete genome sequence analysis. Potato virus S isolates generated in the present study are indicated by underlined labels. The scale bar represents the number of nucleotide substitutions per site.

isolates from Peru associated with *Solanum curtilobum* Juz. & Bukasov. and one isolate from New Zealand infecting *S. muricatum*. Group G5 comprised a single isolate from Peru obtained from *Arracacia xanthorrhiza*. Group G6 contained *S. phureja* isolates from Colombia, while Groups G7 and G8 included isolates from Ecuador associated with *S. tuberosum* and *S. muricatum* respectively. The overall grouping pattern was similar for both complete genome and RdRP datasets. Tanglegram analysis further demonstrated the congruence between the two phylogenies, with cross-linking lines occurring primarily within the same clades and no links between different groups (Fig. 4).

Genetic diversity and selection pressure

The mean nucleotide identity among PVS isolates was 85 % for the complete genome and 89 % for the RdRP domain. Complete genome analysis identified 6087 mutations, with nucleotide diversity (π) of 0.13, theta from Eta (θ) of 0.14 and Watterson's theta (θ_w) of

0.09. The average number of nucleotide differences (k) was 961.21 and haplotype diversity (H_d) was 1.00. Among genomic regions, the nucleotide-binding protein (NBP) exhibited the highest nucleotide diversity ($\pi = 0.16$), whereas TGB1 and the RdRP domain showed the lowest diversity ($\pi = 0.09$). Tajima's D values were negative across all regions, ranging from -0.76 in TGB1 to -0.22 in NBP. Fu and Li's D and F statistics were positive or near zero for most regions, except TGB1, where both values were negative. The overall genome-wide dN/dS ratio was 1.80, while individual genomic regions showed values ranging from 0.23 in TGB3 to 0.03 in the RdRP domain. Site-by-site selection analysis identified 661 positively selected sites and 274 negatively selected sites across the genome, with the replicase region containing the highest number of negatively selected sites (1435). The RdRP domain exhibited the lowest nucleotide diversity and dN/dS ratio among all analysed regions (Table 2).

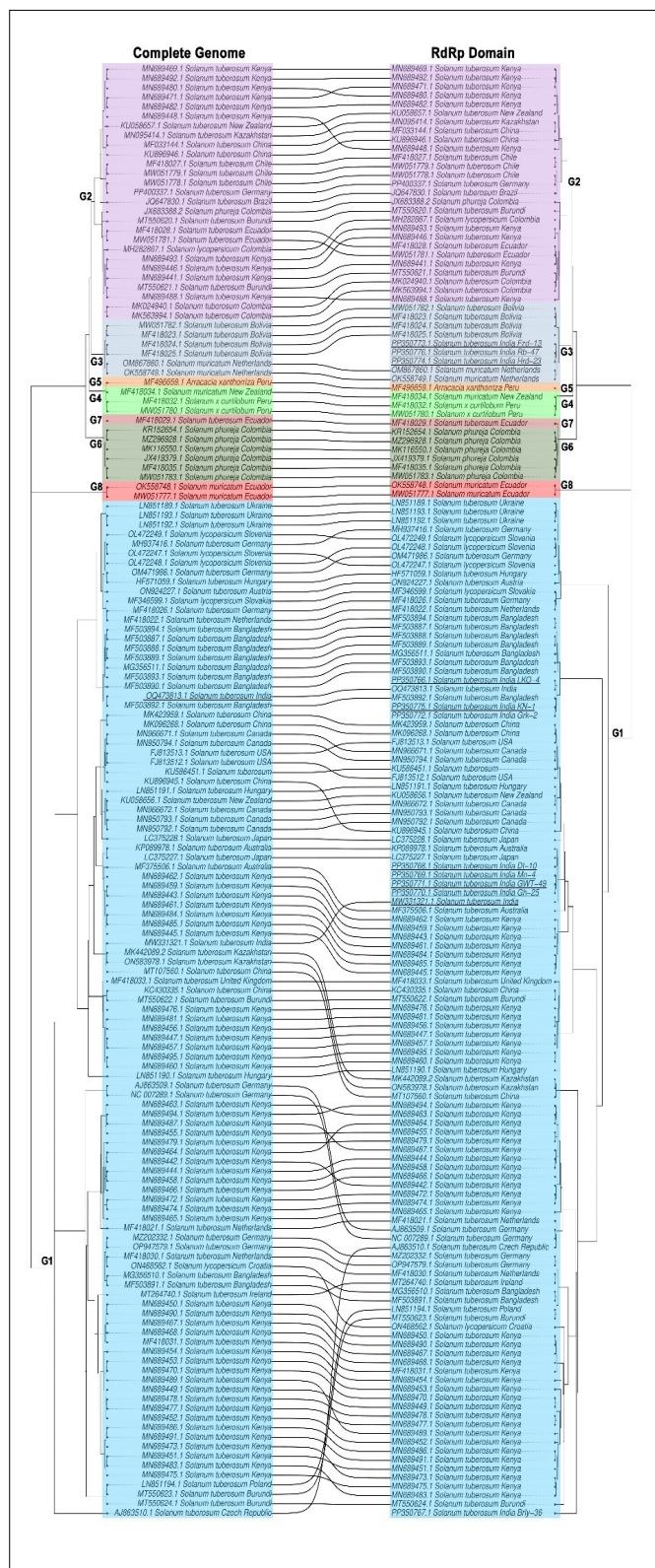


Fig. 4. Tanglegram comparing phylogenetic relationships inferred from complete genome and RNA-dependent RNA polymerase (RdRP) domain sequences of potato virus S (PVS).

The tanglegram illustrates the correspondence between phylogenetic trees constructed from complete genome and RdRp domain nucleotide sequence datasets. Lines connecting identical isolates between the two trees demonstrate the degree of topological congruence. Most connections occur within the same phylogenetic groups (G1–G8), indicating that the evolutionary relationships inferred from the RdRp domain closely reflect those obtained from complete genome analysis. The limited crossing of connecting lines highlights the overall consistency between the two datasets and supports the suitability of the RdRp domain for phylogenetic and molecular evolutionary studies of PVS.

Table 2. Genetic diversity and selection pressure parameters of potato virus S complete genome and individual genomic regions

Genomic region	No. of mutations (Eta, η)	Nucleotide diversity (π)	Average number of nucleotide differences (k)	Theta (per site) from Eta (θ)	Theta-w haplotypes, diversity, Tajima's D Fu and Li's D Fu and Li's F	Number of haplotypes, diversity, Tajima's D Fu and Li's D Fu and Li's F	dN/dS	Total no. of amino acid sites under SLAC		
								AA under positive diversifying selection	AA under negative purifying selection	AA under neutral
PVS complete genome	6087	0.13	961.21	0.14	0.09	136	0.10	661	274	1435
Replicase	4646	0.13	736.44	0.15	0.10	136	0.15	9	0	148
TGB1	455	0.09	63.61	0.11	0.08	126	-0.66	0	0	74
TGB2	209	0.10	33.04	0.12	0.08	115	0.01	0	0	29
TGB3	125	0.10	20.52	0.11	0.09	94	0.35	0	16	225
CP	751	0.11	122.22	0.12	0.08	125	0.22	0	0	45
NBP	58	0.16	9.79	0.17	0.11	61	0.28	0	0	268
RdRp domain	556	0.09	93.50	0.10	0.07	125	0.26	0	0	268

TGB1: triple gene block 1; TGB2: triple gene block 2; TGB3: triple gene block 3; CP: coat protein; NBP: nucleic acid-binding protein; RdRp domain: RNA-dependent RNA polymerase catalytic region; dN/dS: ratio of non-synonymous (dN) to synonymous (dS) substitution rates indicating selection pressure; AA: amino acid; SLAC: single-likelihood ancestor counting.

Genetic structure of potato virus S

The AMOVA for PVS indicates that most of the genetic variation occurred within populations rather than among them. When isolates were grouped according to host species, 42.4 % of the total variation for the complete genome was among populations and 57.6 % within populations ($F_{ST} = 0.42398$). When grouped by country, 38.5 % of variation was among populations and 61.5 % within populations ($F_{ST} = 0.38477$). The RdRP domain showed a similar pattern: host-based grouping accounted for 41.8 % of variation among populations and 58.2 % within populations ($F_{ST} = 0.41796$), while country-based grouping explained 37.3 % among populations and 62.7 % within populations ($F_{ST} = 0.37297$) (Table 3).

Recombination

Recombination analysis was performed independently on both the complete genome and RdRP domain datasets. Analysis of the complete genome sequences identified six significant recombinant events among isolates originating from the Czech Republic, Ukraine, Kenya, Poland and China. The recombinant sequences involved both minor and major parental sequences from related *S. tuberosum* isolates as well as closely related species such as *S. lycopersicum*. Recombination breakpoints were supported by significant *p*-values. In contrast, no evidence of recombination was detected in the RdRP domain dataset, indicating that this region is highly conserved and less prone to recombination events (Supplementary Table 4).

Discussion

RNA viruses are generally characterised by high mutation rates and rapid evolution, often existing as quasispecies within their hosts, thereby contributing to genetic diversity and adaptation (27, 28). Consistent with this pattern, PVS exhibited moderate genome-wide diversity despite its relatively narrow host range. Sequence demarcation tool analysis using complete genome and RdRP domain sequences identified a 92 % nucleotide identity threshold for distinguishing isolates below the species level. This threshold resolved eight distinct groups (G1–G8), which were fully supported by phylogenetic analyses. Group G1 (PVS-O) comprised the largest number of isolates, predominantly associated with *S. tuberosum* and displaying a broad geographic distribution, consistent with the widespread occurrence of the ordinary strain. In contrast, Groups G2–G4 corresponded to previously reported PVS-A-like variants and included isolates associated with Andean hosts such as *S. phureja*, *S. curtilobum* and *S. muricatum*, as well as isolates originating from Bolivia and Peru, recognised centres of diversity for these lineages

(16–18). The additional groups (G5–G8) provided further phylogenetic resolution and were associated with more restricted host ranges and geographic distributions. For example, G5 included isolates from *Arracacia xanthorrhiza* Bancr. in Peru, G6 was primarily composed of *S. phureja* isolates from Colombia and G8 contained isolates associated with *S. muricatum* in Ecuador. The placement of the Czech Republic isolates (PVS-Vltava) within Group G1 (PVS-O), despite previous reports suggesting affiliation with PVS-A or a recombinant origin (16, 17), further demonstrates the utility of combining SDT-based classification with phylogenetic analysis for accurate isolate assignment.

The strong concordance observed between complete genome- and RdRP domain-based phylogenies, as confirmed by tanglegram analysis, demonstrates that the evolutionary signal contained within the RdRP domain closely reflects that of the complete viral genome. This finding highlights the suitability of the RdRP domain as a robust molecular marker for phylogenetic reconstruction and molecular characterisation. The high degree of sequence conservation observed in this region is further supported by its low nucleotide diversity, low dN/dS ratio and absence of detectable recombination. These characteristics are consistent with the essential role of the RdRP domain in viral replication, where functional constraints are expected to limit the accumulation of non-synonymous substitutions. The strong purifying selection detected in this region suggests that most amino acid changes are selectively removed, thereby maintaining the structural and functional integrity of the viral polymerase. Such evolutionary conservation supports the use of the RdRP domain for comparative evolutionary studies, particularly when complete genome sequencing is not feasible.

Genetic diversity analyses indicated substantial haplotypic variation across PVS isolates despite moderate nucleotide diversity. The higher nucleotide variability observed within the NBP region suggests greater evolutionary flexibility relative to other genomic regions, whereas the RdRP domain remained highly conserved. This contrasting pattern indicates differential evolutionary pressures acting across the viral genome, with essential replication-associated regions remaining constrained while other genomic regions may tolerate greater sequence variation. Population genetic analyses revealed that most genetic variation was distributed within populations rather than among populations. Although both host species and geographic origin contributed to population structure, differentiation was consistently greater when isolates were grouped according to host species. This pattern suggests that host-associated factors play an important role in shaping the genetic structure of PVS

Table 3. Analysis of molecular variance for the effects of potato virus S geography and host species

Factor	Sum of squares			Variance components			Percentage variation		Fixation index
	Among populations	Within populations	Total	Among populations	Within populations	Total	Among populations	Within populations	
Complete genome - Host	27226.2	132847.4	160073.5	326.3	443.4	769.7	42.4	57.6	0.424
Complete genome - Country	64501.6	94293.4	158795.0	213.1	340.7	553.8	38.5	61.5	0.385
RdRP domain - Host	2471.6	13162.0	15633.6	29.4	40.9	70.2	41.8	58.2	0.418
RdRP domain - Country	6165.8	9467.8	15633.6	18.8	31.7	50.5	37.3	62.7	0.373

RdRP: RNA-dependent RNA polymerase.

populations. The consistency of this observation across both complete genome and RdRP datasets further support the influence of host adaptation on viral diversification. Overall, the integration of SDT-based sequence demarcation, complete genome and RdRP domain phylogenetic analyses, tanglegram validation and population genetic assessments resolved eight distinct PVS groups and provided insight into the evolutionary structure of the virus. The predominance of the PVS-O lineage within the analysed dataset is consistent with its widespread distribution, while the high sequence stability of the RdRP domain reinforces its value as a reliable marker for molecular characterisation, phylogenetic inference and evolutionary investigations of PVS.

Conclusion

This study provides a comprehensive assessment of the genetic diversity and phylogenetic structure of potato virus S (PVS) based on complete genome sequences and the RNA-dependent RNA polymerase (RdRP) domain. Sequence demarcation tool (SDT) analysis using a 92 % nucleotide identity threshold effectively distinguished PVS isolates below the species level and was consistent with phylogenetic clustering. The strong topological congruence between RdRP domain- and complete genome-based phylogenies, supported by tanglegram analysis, demonstrates the suitability of the RdRP domain as a reliable marker for molecular characterisation and evolutionary studies. The RdRP domain exhibited low genetic diversity, strong purifying selection and no detectable recombination, highlighting its evolutionary conservation and functional importance. Population genetic analyses further indicated that host species contribute more strongly to PVS genetic structuring than geographic origin. Collectively, these findings demonstrate that RdRP domain-based analysis provides an efficient and cost-effective approach for PVS characterisation and large-scale evolutionary investigations when complete genome sequencing is not feasible. Future research should focus on expanding the dataset with additional global isolates, integrating biological and epidemiological data and exploring the evolutionary dynamics of other genomic regions to further improve our understanding of PVS diversity, adaptation and host-virus interactions.

Acknowledgements

The authors thank the Director, National Botanical Research Institute, Lucknow, Uttar Pradesh, India, for providing the necessary facilities to carry out the work. YM acknowledges the DST-INSPIRE faculty award. This research work was supported by funding (DST/INSPIRE/04/2018/001057) from the Department of Science and Technology, Ministry of Science and Technology, Government of India. The CSIR-NBRI manuscript number is CSIR-NBRI_MS/2026/04/06.

Authors' contributions

SB, CM and VS contributed to the conceptualisation of the study. SB, CM, NPM, AS and NV were responsible for the experimental design and methodology. Data analysis, review and editing were carried out by SB, CM, NPM, AS, NV, SK, YM and VS. The original draft of the manuscript was prepared by SB, CM, NV and VS. Funding acquisition and overall supervision were provided by YM and VS. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

- Kreuze JF, Souza-Dias JA, Jeevalatha A, Figueira AR, Valkonen JPT, Jones RAC. Viral diseases in potato. In: Campos H, Ortiz O, editors. *The Potato Crop*. Cham: Springer; 2020. p. 389–430. https://doi.org/10.1007/978-3-030-28683-5_11
- Kumar R, Tiwari RK, Sundaresha S, Kaundal P, Raigond B. Potato viruses and their management. In: Singh BP, Raigond B, Dutt S, editors. *Sustainable Management of Potato Pests and Diseases*. Singapore: Springer; 2022. p. 309–35. https://doi.org/10.1007/978-981-16-7695-6_12
- Kumar R, Jeevalatha A, Baswaraj R, Sharma S, Nagesh M. A multiplex RT-PCR assay for simultaneous detection of five viruses in potato. *J Plant Pathol*. 2017;99:37–45.
- Wardrop EA, Gray AB, Singh RP, Peterson JF. Aphid transmission of potato virus S. *Am Potato J*. 1989;66:449–59. <https://doi.org/10.1007/BF02855437>
- Kostiwi M. Transmission of potato virus S by *Aphis nasturtii*. *Potato Res*. 1975;18(4):641–3. <https://doi.org/10.1007/BF02365690>
- LP A, HN V. Current status of viral diseases of potato and their ecofriendly management—a critical review. *Virol Res Rev*. 2017;1(4). <https://doi.org/10.15761/VRR.1000122>
- Goth RW, Webb RE. Lack of potato virus S transmission via true seed in *Solanum tuberosum*. *Phytopathology*. 1975;65(12):1347–9. <https://doi.org/10.1094/Phyto-65-1347>
- Lei J, Liang B, Yang H, Zhang B. Origin, evolution, dispersal and global population genetic structure of *Carlavirus sigmasolani*. *Front Plant Sci*. 2025;16:1667771. <https://doi.org/10.3389/fpls.2025.1667771>
- Slack SA. Identification of an isolate of the Andean strain of potato virus S in North America. *Plant Dis*. 1983;67(7):786–9. <https://doi.org/10.1094/PD-67-786>
- Foster GD, Mills PR. The 3' nucleotide sequence of an ordinary strain of potato virus S. *Virus Genes*. 1992;6:213–20. <https://doi.org/10.1007/BF01702560>
- Čeřovská N, Filigarova M. Specific detection of the Andean strain of potato virus S using monoclonal antibodies. *Ann Appl Biol*. 1995;127(1):87–93. <https://doi.org/10.1111/j.1744-7348.1995.tb06653.x>
- Khurana SP. Potato viruses and their management. In: Naqvi SAMH, editor. *Diseases of Fruits and Vegetables*. Dordrecht: Springer; 2004. p. 389–440.
- De Bokx JA, Waterreus HA. Electron microscopy of plant tissues infected with potato viruses A and S. *Neth J Plant Pathol*. 1971;77(4):106–12. <https://doi.org/10.1007/BF02000550>
- Li X, Hataya T. Construction and characterization of an infectious cDNA clone of potato virus S. *Virol J*. 2019;16:17. <https://doi.org/10.1186/s12985-019-1124-x>
- Mackenzie DJ, Tremaine JH, Stace-Smith R. Organization and interval homologies of the 3' terminal portion of potato virus S RNA. *J Gen Virol*. 1989;70(5):1053–63. <https://doi.org/10.1099/0022-1317-70-5-1053>
- Duan G, Zhan F, Du Z, Ho SYW, Gao F. Europe was a hub for the global spread of potato virus S in the 19th century. *Virology*. 2018;525:200–4. <https://doi.org/10.1016/j.virol.2018.09.022>
- Santillan FW, Fribourg CE, Adams IP, Gibbs AJ, Boonham N, Kehoe MA, et al. Biology and phylogenetics of potato virus S isolates from the Andean region. *Plant Dis*. 2018;102(5):869–85. <https://doi.org/10.1094/PDIS-09-17-1414-RE>
- Topkaya Ş, Çelik A, Santosa AI, Jones RAC. Molecular analysis of the global population of potato virus S redefines its phylogeny. *Viruses*. 2023;15(5):1104. <https://doi.org/10.3390/v15051104>

19. Bartholomäus A, Wibberg D, Winkler A, Pühler A, Schlüter A, Varrelmann M. Deep sequencing analysis reveals the mycoviral diversity of the virome of an avirulent isolate of *Rhizoctonia solani*. PLoS One. 2016;11(11):e0165965. <https://doi.org/10.1371/journal.pone.0165965>
20. Ferrer-Orta C, Ferrero DS, Verdaguer N. RNA-dependent RNA polymerases of picomaviruses: From structure to regulatory mechanisms. Viruses. 2015;7(8):4438–60. <https://doi.org/10.3390/v7082829>
21. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions. Nature. 2006;439(7074):344–8. <https://doi.org/10.1038/nature04388>
22. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, et al. Origins and evolution of the global RNA virome. mBio. 2018;9(6):e02329-18. <https://doi.org/10.1128/mBio.02329-18>
23. Farias ST, Santos Junior AP, Rêgo TG, José MV. Origin and evolution of RNA-dependent RNA polymerase. Front Genet. 2017;8:125. <https://doi.org/10.3389/fgene.2017.00125>
24. Jácome R, Becerra A, Ponce de León S, Lazcano A. Structural analysis of monomeric RNA-dependent RNA polymerases: Evolutionary and therapeutic implications. PLoS One. 2015;10(9):e0139001. <https://doi.org/10.1371/journal.pone.0139001>
25. Černý J, Černá Bolfíková B, Valdes JJ, Grubhoffer L, Růžek D. Evolution of tertiary structure of viral RNA-dependent RNA polymerases. PLoS One. 2014;9(5):e96070. <https://doi.org/10.1371/journal.pone.0096070>
26. Colson P, Bader W, Fantini J, Dudouet P, Levasseur A, Pontarotti P. From viral democratic genomes to viral wild bunch of quasispecies. J Med Virol. 2023;95(11):e29209. <https://doi.org/10.1002/jmv.29209>
27. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012;76(2):159–216. <https://doi.org/10.1128/MMBR.05023-11>
28. Llaure AS, Andino R. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog. 2010;6(7):e1001005. <https://doi.org/10.1371/journal.ppat.1001005>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.