Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complex diversityin papaya ringspot virus (PRSV) populations is governed by the host species and the geo-climatic locations

Shiksha Bhandari
Indian Agricultural Research Institute
Y. B. BASAVARAJ
rajayb@gmail.com

Indian Agricultural Research Institute
Kumar Ashwini
Indian Agricultural Research Institute
Aruna Kumar
Amity University
Susheel Kumar Sharma
Indian Agricultural Research Institute
Rakesh Kumar Jain
Indian Agricultural Research Institute

Research Article

Keywords: Evolutionary dynamics, Genetic diversity, Gene flow, Haplotype, Phylogeny

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11470720/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Complex diversityin papaya ringspot virus (PRSV) populations is governed by the host species and the geo-climatic locations

Shiksha Bhandari[1,2], Ashwini Kumar[1,3], Aruna Kumar[2], Susheel Kumar Sharma[1], Basavaraj Y.B[1,4]*

[1]Advanced Centre for Plant Virology, ICAR- Indian Agricultural Research Institute, New Delhi 110012, India
[2]Amity Institute of Biotechnology, Amity University Noida, Uttar Pradesh 201313, India
[3]Acharya Narendra Dev University of Agriculture and Technology, Ayodhya, Uttar Pradesh – 224229, India
[4] ICAR-Indian Agricultural Research Institute, Regional Station Pune, Maharashtra 411067, India

*Correspondence to: Email: rajayb@gmail.com
ORCID ID: https://orcid.org/0000-0002-9216-877X

ABSTRACT

Papaya ringspot virus (PRSV) is one of the most important viral pathogens limiting papaya and cucurbit production globally. However, the evolutionary processes governing its population structure across diverse agro-ecological environments remain insufficiently understood. This study examined the genetic diversity, recombination patterns, and population structure of PRSV-P and PRSV-W isolates collected from different hosts and geo-climatic regions of India. Twenty-seven PRSV isolates originated from papaya and multiple cucurbit species were analyzed based onthe sequences of three protein coding regions: helper component proteinase (HC-Pro), nuclear inclusion proteinase-a (NIa-Pro), and coat protein (CP). The CP gene displayed substantial variation, ranging from 837 to 876 nucleotides, with the amino acid (aa)sequence diversities reaching up to 20%. Higher diversity was observed among cucurbit-derived PRSV-P isolates from warm semi-arid (WSA) regions, primarily through frequent aainsertion and deletion events. Phylogenetic analyses revealed varying evolutionary patterns among genomic regions. CP sequences grouped according to geo-climatic origin, whereas NIa-Pro differentiated based on host-& pathotype-associated lineages (PRSV-P and PRSV-W), while HC-Pro showed no distinct clustering patterns. Recombination analysis identified gene-specific evolutionary dynamics, with recombination detected in CP and NIa-Pro but not in HC-Pro. Population genetic analyses indicated extremely high haplotype diversity (*Hd* = 0.996), strong population subdivision (*Snn* = 0.88889–1.00000; $p \leq 0.004$), and moderate to high genetic differentiation (*Fst* = 0.1899- 0.7713), despite restricted gene flow (*Nm* = 0.20). Negative Tajima's D values suggested recent population expansion and/or strong purifying selection. Overall, environmental variation, host association, and gene-specific constraints jointly influence PRSV evolution, while genetic isolation and limited gene flow promote diversification and ecological adaptation across distinct geo-climatic regions.

Keywords: Evolutionary dynamics, Genetic diversity, Gene flow, Haplotype, Phylogeny

INTRODUCTION

Papaya (*Carica papaya* L.) is a widely grown fruit crop in tropical and subtropical areas valued for its nutritional, digestive, and medicinal properties (Manshardt, 1992; Saeed et al. 2014).To date, 22important virus diseases affecting papaya cultivation have been reported worldwide, belonging to different viral groups (Nivetha et al. 2025). Of these, the aphid-transmitted papaya ringspot virus (PRSV) (Family *Potyviridae*, Genus *Potyvirus*, Species*Potyvirus papayanuli*) is the leading cause of widespread, destructive disease, posing serious economic and agronomic challenges to papaya production globally (Tripathi et al. 2008; Lu et al. 2008; Mishra et al. 2019; Singh et al. 2017; Gogoi et al. 2019).PRSV isolates are typically divided into two groups, PRSV-W that infects plants from the *Chenopodiaceae* and *Cucurbitaceae* plant families, while PRSV-P extends its infection to include members of the family *Caricaceae* (Purcifulet al.1984; Yeh et al. 1984; Chen et al. 2008; Basavaraj et al. 2019).

For managing PRSV, transgenic methods have gained prominence in recent years, outpacing conventional approaches that have shown limited success rates (Yeh et al.1994; Fermin et al. 2010). Nevertheless, the effectiveness of transgenic resistance is limited by the geographical variations of PRSV and the transgene sequence homology(Bau et al. 2003). Substantial genetic diversity within PRSV populations may impose limitations on the global management of the disease(Jain et al. 2004). The genetic variability among the IndianPRSVpopulation is up to 20%.The genetic variability observed in potyviruses, including PRSV, arises predominantly from mutation and recombination (Basavaraj et al. 2018, 2019). Exploring the function of each gene is essential to unveil the evolutionary dynamics of this virus and gain insights into the variations among conserved domains of the genome (Mangrauthia et al. 2008). Consequently, a thorough understanding of the sequence divergence and evolutionary dynamics of genes is crucial to grasp the potential breakdown of transgenic resistance (Kunget al. 2015). Factors determining the genetic diversity among different plant virus populations have been reviewed by several researchers (García-Arenal et al. 2001; Bateson et al. 2002;García-Arenal et al.2013).

Insights into the sequence diversity among PRSV isolates and their distribution can enhance our knowledge of viral origins, disease etiology, and epidemiology. This knowledge, in turn, can prove valuable for formulating effective strategies in disease management (Gonsalves 1994; Tripathi et al. 2008; Basavaraj et al. 2019). Assessment of the genetic diversity and phylogenetic analyses on diverse PRSV isolates unveiled a strong genetic correlation among isolates originating from the same geographic location (Bateson et al. 1994, 2002). PRSV isolates originating from various geographic regions such as America, Africa, Europe, and Oceania, exhibited a shared strain (Quiot-Douine et al. 1990). Nonetheless, the Indian isolates showed significant heterogeneity among PRSV isolates from different locations within India (Jain et al.1998, 2004; Akhter et al. 2013).

It has been speculated that geo-ecological regions play a crucial role in shaping the genetic diversity of plant viruses, and agriculture stands out as a pivotal factor influencing their spread (Bernardo et al. 2018).The usage pattern of agricultural land significantly influences virus distribution and prevalence, with cultivated areas exhibiting greater virus diversity and prevalence compared to uncultivated regions (Bernardo et al. 2018).Moreover, climate change is predicted to impact the spread of plant viruses by changing host plant and insect vector populations, their geographical distributions, and the physiological processes vital to virus-plant interactions (Canto et al. 2009).These factors may shift over time with changing climatic conditions (Bellard et

al. 2012; Scheffers et al. 2016).However, very limited attempts have been made to systematically examine those hypotheses that genetic diversity in PRSV populations is whether governed by geographical conditions (Basavaraj et al. 2019). Understanding the genetic diversity and organization of plant virus populations is vital for revealing their evolutionary biology and epidemiology. This information is key to devising strategies to lessen the impact of viral diseases on agricultural and natural ecosystems (Paradell 2017).

Most of the available literature predominantly addresses genetic diversity studies concentrated on specific gene fragments of PRSV, especially highlighting the coat protein (CP) gene (Yeh et al. 1992; Batesonet al. 1994; Silva et al. 2000; Akhter et al. 2013). PRSV isolates from different locations in India exhibited high heterogeneity in CP length and amino acid sequences (Bag et al. 2007). In recent studies, the evolutionary basis of PRSV has been examined through a comprehensive comparative analyses of the entire genome or some vital protein-coding regions of the virus (Tripathiet al. 2008; Basavaraj et al. 2019). Very recently, it has also been shown that the different hosts from where isolates originate play a significant role in generating genetically diverse populations of the virus through genetic recombination and mutations (Basavaraj et al. 2019). In that study, although very few cucurbit-originated P and W isolates were examined, it was shown that the diversity was largely through the recombination between PRSV isolates originated from papaya(PRSV-P-P), cucurbits(PRSV-C-P), and PRSV-W pathotypes.

This research envisages the work plan to re-examine the impact of different host plant species and/or pathotypes on the genetic variability in the PRSV genome using more cucurbits-originated P and PRSV-W isolates over PRSV-P-P. Besides, it had been shown that geo-ecological locations from where the PRSV isolates originate also have a significant effect in shaping the genome sequences to enable the isolates to the existing conditions. Thus, in the present investigation, we have characterized and analyzed PRSV-P and -W isolates originated from varied geo-climatic locations and hosts. The primary focus was on three crucial protein-encoding regions, namely, HC-Pro (helper component proteinase), NIa-Pro (Nuclear inclusion protein-a), and CP to assess the diversity within and between PRSV populations.

MATERIALS AND METHOD

Sample collection, virus detection and diagnosis

A total of 110 (23 papaya and 87 cucurbit) leaf samples suspected of PRSV infection were collected from seven different states of India (Table 1). Samples exhibiting typical virus disease-like symptoms, including mosaic, blistering, leaf distortion, curling, puckering, filiformy, shoe-stringing, and plant stunting (Fig. 1), were subjected to the following diagnostic assays to confirm PRSV infection.

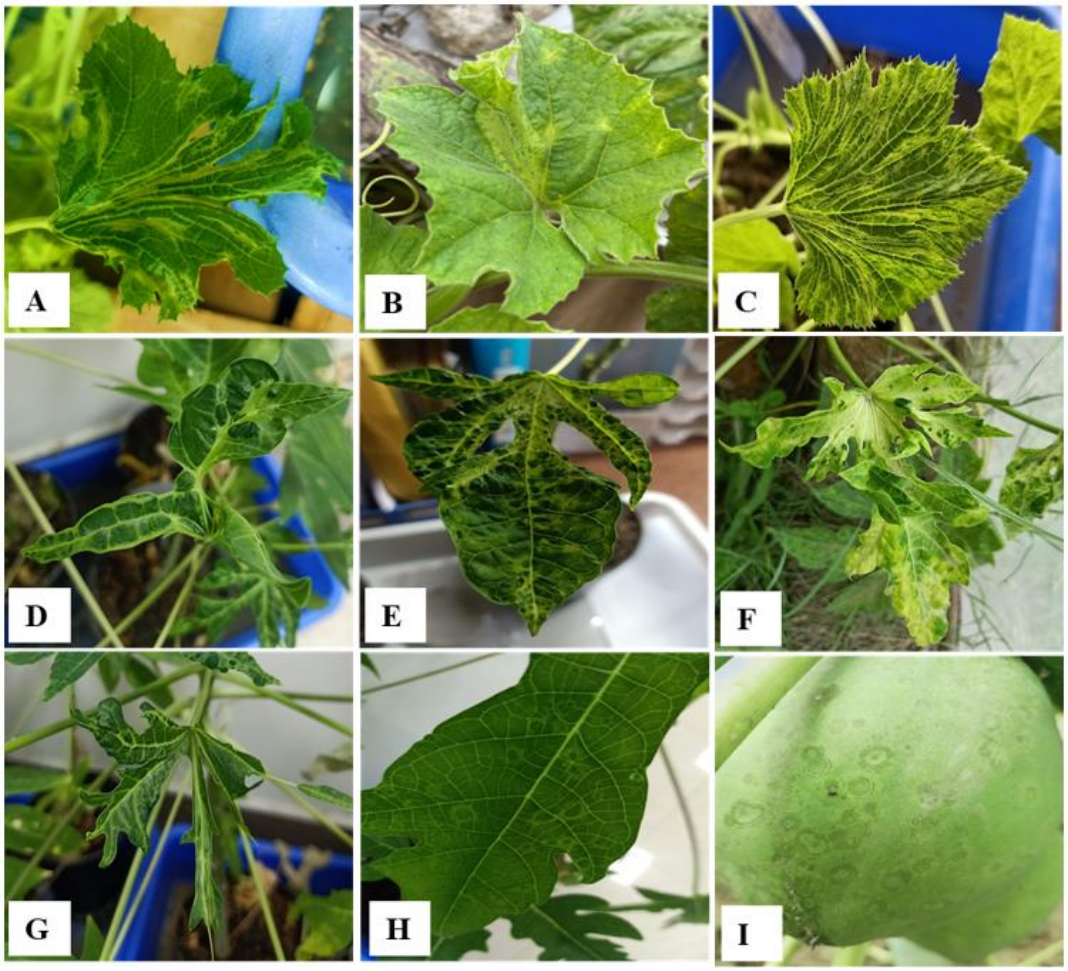


Figure 1. Various symptoms of Papaya ring spot virus (PRSV) on cucurbits and papaya were observed under field conditions. A. Leaf distortion and Vein banding; B. Vein clearing; C. Vein banding and Inter-veinal chlorosis; D. Leaf distortion and blistering; E. Leaf distortion and vein clearing; F. Leaf distortion, blistering and chlorosis; G. Filiformy and shoe stringing; H. Ring spots on leaf; I. Ring spots on fruits.

Table 1: Details of different region sequences of PRSV cloned and sequenced in the present study along with the genome sequences used for comparative analysis.

Geographical Origin	Sl. No.	Acronym [$]	Host	Pathotype	Ecological origin	Gene bank accession no.			Length of CP [@]	References
						HC-pro	NIa-pro	CP		
Sequence information generated in this study										
Delhi	1	***IND-DL3 (W)***	Muskmelon	W	WSA	ON015705	-	ON015693	855	This study
	2	***IND-DL4 (P-P)***	Papaya	P		ON015712	ON015753	ON015697	848	
	3	***IND-DL5 (W)***	Pumpkin	W		-	ON015732	ON015692	855	
	4	***IND-DL6 (W)***	Pumpkin	W		ON015726	ON015731	ON015663	861	
	5	***IND-DL7 (C-P)***	Pumpkin	C-P		-	ON015748	ON015665	876	
	6	***IND-DL8 (W)***	Pumpkin	W		ON015728	ON015729	ON015667	861	
Gujarat	7	***IND-GJ1 (P-P)***	Papaya	P	HST	-	-	ON015670	846	
	8	***IND-GJ2 (P-P)***	Papaya	P		-	ON015749	ON015671	846	
Karnataka	9	***IND-KA4 (C-P)***	Cucumber	C-P	WSA	ON015711	ON015743	ON015676	861	
	10	***IND-KA5 (W)***	Cucumber	W		-	ON015742	ON015680	858	
	11	***IND-KA6 (P-P)***	Papaya	P		ON015724	ON015735	ON015660	858	
	12	***IND-KA7 (P-P)***	Papaya	P		ON015725	ON015752	-	-	
	13	***IND-KA8 (W)***	Pumpkin	w		-	ON015745	-	-	
	14	***IND-KA9 (P-P)***	Papaya	P		ON015727	ON015739	ON015682	855	
	15	***IND-KA10 (P-P)***	Papaya	P		ON015709	ON015738	ON015674	858	
	16	***IND-KA11 (P-P)***	Papaya	P		ON015710	ON015747	ON015677	858	
	17	***IND-KA12 (W)***	Pumpkin	W		-	ON015736	ON015672	861	
	18	***IND-KA13 (P)***	Papaya	P		-	-	ON015661	837	
	19	***IND-KA14 (C-P)***	Watermelon	C-P		-	-	ON015688	855	
	20	***IND-KA15 (C-P)***	Watermelon	C-P		-	ON015746	ON015675	858	
Odisha	21	***IND-OD (P-P)***	Papaya	P	HST	-	-	ON015681	852	
Uttar Pradesh	22	***IND-UP6 (C-P)***	Bottle gourd	C-P	WSA	ON015701	-	-	876	
	23	***IND-UP7 (P-P)***	Papaya	P		ON015702	ON015734	ON015684	858	
Uttarakhand	24	***IND-UK1 (P-P)***	Papaya	P	WSA	ON015719	-	ON015700	858	
	25	***IND-UK2 (P-P)***	Papaya	P		ON015716	-	ON015679	861	
	26	***IND-UK3 (P-P)***	Papaya	P		ON015722	-	ON015659	861	
West Bengal	27	***IND-WB2 (P-P)***	Papaya	P	HST	-	-	ON015662	852	
Previously published partial genome sequences were used in the current study to compare respective coding regions										
Assam	28	IND-AS (P-P[*])	Papaya	P	HST	-	KC155247	KC149500	852	Basavaraj et al., 2019
Goa	29	IND-GA1 (P-P)	Papaya	P	Mo	KY448322	KY448316	KY448328	855	
	30	IND-GA2 (P-P)	Papaya	P		-	KY448313	KY448329	855	
Haryana	31	IND-HR (C-P)	Bottle gourd	C-P	WSA	KY448323	KY448318	KY448332	876	
Karnataka	32	IND-KA1 (W)	Watermelon	W	WSA	JQ394692	-	JN979405	855	
	33	IND-KA2 (P-P)	Papaya	P		KU196788	KU196797	JN979404	861	
	34	IND-KA3 (P-P)	Papaya	P		KC149506	KC155246	KU196792	855	

	35	IND-KA (W)	Watermelon	W		-	KC155252	JN979403	861	
Maharashtra	36	IND-MH (P-P)	Papaya	P	WSA	KC149505	KC155251	KC149502	858	
Punjab	37	IND-PB1 (W)	Summer squash	W	WSA	-	KY448312	KY448326	855	
	38	IND-PB2 (C-P)	Summer squash	C-P		-	KY448317	KY448327	861	
Jharkhand	39	IND-JK (P-P)	Papaya	P	HST	-	KC155249	JQ394693	852	
Uttar Pradesh	40	IND-UP1 (C-P)	Bottle gourd	C-P	WSA	KY448319	KY448310	KY448324	876	
	41	IND-UP2 (C-P)	Pumpkin	C-P		KY448320	KY448311	KY448325	876	
	42	IND-UP3 (P-P)	Papaya	P		KU196789	KU196798	KU196793	858	
	43	IND-UP4 (W)	Ash gourd	W		KU196790	-	KU196795	861	
	44	IND-UP5 (C-P)	Pumpkin	C-P		KU196791	KU196801	KU196796	855	
	45	IND-UP8 (P-P*)	Papaya	P		-	KU196799	-	-	
West Bengal	46	IND-WB1 (P-P)	Papaya	P	HST	KC149504	KC155250	JN979406	852	
Previously published complete genome sequences used in the current study to compare respective protein coding regions										
Brazil	47	BRA-1 (C-P)	Cucurbit	P	OC	KP462721			861	Unpublished
	48	BRA-2 (W)	Cucurbit	W		DQ374152			861	Inoue-Nagata et al. (2007)
	49	BRA-3 (W)	Cucurbit	W		DQ374153			867	Inoue-Nagata et al. (2007)
China	50	CHI-1 (P-P)	Papaya	P	MO	EF183499			858	Lu et al. (2008)
	51	CHI-2 (P-P)	Papaya	P		KT895257			867	Unpublished
	52	CHI-3 (P-P**)	Papaya	P		KT633943			858	Unpublished
Columbia	53	COL-1(P-P)	Papaya	P	MO	KT275937			861	Ortiz-Rojas and Chaves-Bedoya (2017)
	54	COL-2 (P-P)	Papaya	P		KT275938			861	Ortiz-Rojas and Chaves Bedoya (2017)
France	55	FRA (C-U)	Cucurbits	U	OC	KC345609			861	Romay et al. (2014)
Korea	56	KOR (C-P)	Cucurbits	P	MO	AB369277			861	Unpublished
Mexico	57	MEX (P-P)	Papaya	P	OC	AY231130			855	Noa-Carrazana et al. (2006)
Taiwan	58	TAI-1(P-P**)	Papaya	P	MO	DQ340770			861	Unpublished
	59	TAI-2 (P-P*)	Papaya	P		X97251			861	Wang and Yeh (1997)
	60	TAI -3 (W)	Cucurbit	W		AY027810			867	Unpublished
Hawaii	61	HAW-1 (P-P*)	Papaya	P	OC	EU126128			855	Unpublished
	62	HAW-2(P-P)	Papaya	P		S46722			861	Unpublished
Thailand	63	THAI 1(P-P)	Papaya	P	MO	AY162218			858	Unpublished
	64	THAI 2-W	Cucurbit	W		AY010722			858	Attasart et al. (2002)
USA	65	USA-W	Gourd	W	OC	KY039583			861	Ali (2017)
Hyderabad	66	IND-HD (P-P)	Papaya	P	WSA	KP743981			858	Unpublished
Delhi	67	IND-DL1 (P-P)	Papaya	P	WSA	EF17707			855	Parameswari et al. (2007)
	68	IND-DL2 (W)	Cucurbit	W		EU475877			861	Mangrauthia et al. (2009)
Rajasthan	69	IND-RJ (P-P**)	Papaya	P	MO	KJ755852			861	Unpublished

P–P: PRSV-P pathotype from papaya host; **C-P :**PRSV-P from cucurbits; **W :** PRSV-W (Exclusively from cucurbit host); **C-U or U :** Originated from cucurbit host, but pathotype unknown; **P–P***: PRSV-P from papaya & passed through cucurbit host before sequence characterization; **P–P**:**PRSV-P from papaya & probably passed through cucurbit host before sequence characterization; **@** : The length of HC-pro (1320 bp) & NIa-pro (725 bp) are constant in all the isolates, hence not mentioned in table; **–** : The sequence information could not be generated in current study; $: The PRSV isolates characterized in the current investigation are indicated in bold and Italics.

Leaf dip electron microscopy: The association of virus particles with symptomatic samples was examined by leaf-dip electron microscopy (Gibbs and Harrison 1976) using a transmission electron microscope (TEM) (JEOL TEM-10111).

Serological and bio-assay: The TEM-positive papaya and cucurbit samples were analyzed further by direct antigen-coated enzyme-linked immunosorbent assay (DAC-ELISA) (Clark and Bar-Joseph 1984) to confirm the association of PRSV usingin-house polyclonal antibodies (PAbs) targeting PRSV, capsicum chlorosis orthotospovirus (CACV,*Orthotospoviruscapsiciflavi*), cucumber mosaic virus (CMV, *Cucumovirus CMV*), groundnut bud necrosis orthotospovirus (GBNV, *Orthotospovirusarachinecrosis*), watermelon bud necrosis orthotospovirus (WBNV,*Orthotospovirus citrullonecrosis*),and zucchini yellow mosaic virus (ZYMV, *Potyvirus cucurbitaflavitesselati*)available at Advanced Centre for Plant Virology (ACPV), ICAR-Indian Agricultural Research Institute, New Delhi.

ELISA-positive samples were subjected to bio-assay with mechanical sap inoculation on squash and papaya seedlings to identify the pathotypes (P & W) of associated PRSV isolates. Following inoculation, both symptomatic and asymptomatic plants were tested through ELISA again to confirm the presence of the virus andto determine the contagious nature of PRSV isolates.

Molecular detection through RT-PCR: The total RNA was extracted from ELISA positive samples, using PureLink^{TM}RNA Mini kit (Invitrogen, Life Technologies). 500 ng of total RNA and 20 nmol of specific reverse primersdesigned specifically to amplify HC-Pro, NIa-Pro, and CP coding regions of the PRSV genome (Table S1) were used for cDNA synthesis. Each cDNA (2.0 µL) was used as a template for the amplification of respective coding regions by PCR. The PCR assay was carried out utilizing the Bio-Rad^{TM} Thermal Cycler (T100 Thermal Cycler, Bio-Rad Laboratories, Inc) employing specified temperature conditions. The PCR amplification conditions included an initial denaturation cycle of 5 min at 94 °C followed by 32 cycles of denaturation for 40 s at 94 °C, respective annealing temperature (Table S1), and extension for 1 min at 72 °C, with a final extension for 10 min at 72 °C.

Table S1: Details of primers designed for PCR amplification of HC-pro, NIa-pro, & CP coding regions of PRSV.

Sl. No	Gene/ Target region	Primer Sequence (5′-3′)	Amplicon Size	Annealing Temperature (Tm)
1.	HC-Pro	RKJ 116- AATGACGTGGCTGAAAAATTCTG RKJ 117- AATGAAGCACTACATTGTCGG	1.3 Kb	54°C
2.	Partial VPg+ NIa-Pro	RKJ 158- GTGGAACATTTTGGTGA RKJ 159- CGAACAAAGCGGTAG	999 bp	52°C
3.	Partial NIb+ CP	HRP 50- ATGATAGAGTCATGGGGTTA RKJ 3- CTCTCCTGGGTATGCGCAAC	1.1 Kb	52°C

Molecular cloning,sequencing and sequence analyses

PCR-amplified DNA fragments weregel-purifiedusing the NucleoSpin® Gel and PCR clean-up kit (MACHEREY-NAGEL, GmbH & Co.) andcloned into pGEM-T easy T/A cloning vector (Promega, Madison, WI, USA). The resultant recombinant clones were sequenced using Sp6 and T7 primers by outsourcing through Sanger sequencing. The consensus sequences of all three genomic regions from each isolate were deposited in the GenBank NCBI database to obtain their specific Accession numbers.

Multiple sequence alignment, phylogeny, and sequence identity studies

The PRSV isolates originating from different geoclimatic locations and hosts, was examined across three important protein-coding regions (HC-Pro, NIa-Pro, and CP). These coding region sequences were analyzed to determine their phylogenetic grouping patterns, and the underlying sequence identitiesor diversities among PRSV isolates originating from different geo-climatic locations and hosts. The PRSV isolates analyzed in this investigation (highlighted with bold and italic tests), in relation to those characterized in this study and those that were previously characterized globally, are described in Table 1. These isolates predominantly originated from four distinct climatic zones as described by Basavaraj et al (2019): Humid Sub-Tropical (HST), Monsoon (Mo), Oceanic (Oc), and Warm Semi-Arid (WSA). The genomic sequences of PRSV isolates originated from corresponding zoneswere retrievedfrom the NCBI GenBank database (http://www.ncbi.nlm.nih.gov) for comparative sequence analysis.Sequence alignments, at nt and aa levels, andaa sequence identity analyseswereperformed using the Clustal W multiple alignment toolfollowed by Sequence identity matrix tool, respectively (BioEdit 7.2 version). Phylogenetic trees, using aa sequences, were constructed throughNeighbour-joining (NJ) method based on the Tamura-Nei model, with 1000 bootstrap replicates on MEGA11 (Molecular Evolutionary Genetics Analysis) software with default parameters (Tamura et al. 2021).

Genetic differentiation, gene flow and haplotype analysis

Three statistical parameters were utilized to assess genetic differentiation between the PRSV isolates analyzed in this investigation and from various other geographical regions based on their geo-climatic zones of origins focusing on their CP coding region sequences.Other genomic regions (HC-Pro and NIa-Pro) were not included in these analyses due to their inconsistent clustering patterns as per the geo-climatic perspectives. In the alignment of CP sequences, an analysis was conducted using a permutation-based statistical test, where*Kst**(a sequence-based statistical test) value close to zero indicates no differentiation, and *Snn*(a measure of the frequency of nearest neighbor sequences within the same locality) value close to one indicates differentiation (Hudson 2000). The evaluation of gene flow among PRSV populationswas performed through the assessment of *Fst* value or fixation index. The analysis was carried out with 1,000 replicates. The *Fst* value signifies the degree of inter-population diversity that ranges from zero (indicating frequent gene flow) to one (indicating infrequent gene flow) between populations (Hudson et al. 1992). The *Fst*represents the standardized variance in allele frequencies across populations or the inter-population component of genetic variation. The *Fst* coefficient estimates the degree of genetic differentiation among populations. An *Fst* value greater than 0.33 suggests there is a negligible gene flow, while less than 0.33 indicates frequent gene flow (Rozas et al. 2003).

PRSV CP coding sequences of 65 isolates,divided into different geo-ecological zones as per the phylogenetic clustering patterns, i.e.,HST I (n=7), HST II (n= 4), and MO (n = 10), OC (n =10), WSA I (n =6),

WSA II (n = 4),WSA III (n= 5), WSA IV (n=13), WSA V (n=2), and WSA VI (n=4)were divided into a net number of 10 populations. The PRSV population from each geo-climatic zone was compared with the rest of the 9 populations individually and was analysed for gene flow and genetic differentiation analysis. In the analysis, the sites with alignment gaps were excluded. After defining populations, the genetic differentiation and gene flow analyseswere performed using DnaSP program.Nucleotide diversity indices, such as π (nucleotide diversity per site), *k* (average number of nucleotide differences), and *Hd* (Haplotype diversity), were also computed using DnaSP program ver. 6.12.03 (Rozas et al. 2017). Additionally, neutrality tests, including Tajima's D and Fu Li's F and D testswere performed.

Recombination analysis

The aligned sequences of PRSV isolates of this study were compared with sequences from corresponding PRSV isolates retrieved from the NCBI GenBank database to detect potential events of recombination. The potential influence of recombination on sequence diversity was investigated through seven distinct automated algorithms: RDP, GeneConv, BootScan, MaxChi, Chimaera, SiScan, and 3Seq as implemented in Recombination Detection Program, ver. 4.0 (RDP4). A multiple comparison corrected *p*-value cutoff of 0.05 was applied (Martin et al. 2015). Evidence for recombination was considered valid only when it was recognized by a minimum of three distinct algorithms.

RESULTS AND DISCUSSION

Out of the 110 suspected samples tested (23 papaya and 87 cucurbits) by DAC- ELISA, 62 (23 papaya and 39 cucurbits) turned out positive for the association of PRSV either singly (n=37) or as a mixed infection with the other viruses (n=73) tested in this study, such as GBNV, WBNV, CACV & ZYMV (data not shown). Out of the selected representative samples that possess the association of PRSV alone (15 papaya + 22 cucurbits = 37), only 26 samples (papaya-14 and cucurbits-12) induced symptoms suggestive of PRSV infection on papaya as well as squash plants during bioassay, thus indicated the association of PRSV pathotype-P. While 9 out of 22 cucurbit samples induced the symptoms exclusively on squash plants and suggested the association of PRSV pathotype-W. The DAC-ELISA results (data not shown)suggested that the symptoms observed on suspected samples were not solely due to PRSV infection alone, as other viruses,such as,CMV, papaya leaf distortion mosaic virus (PLDMV, *Potyvirus caricae*), ZYMV, and zucchini tigremosaic virus (ZTMV, *Potyvirus pepotigris*) are also capable of inducing comparable symptoms (Maoka and Hataya 2005; Al-Ani et al. 2011; Chen et al. 2023), etc. This explains why a smaller numberof samples out of 110 were tested positive for PRSV infection alone, emphasizing the need for exploring comprehensive diagnostic approaches to identify viral infections and their overlapping symptoms accurately.Furthermore, the identification of only 9 out of 22 cucurbit samples as the W-pathotype suggested the predominant occurrence of PRSV pathotype-P over W in the studied crops and locations.

Molecular characterization of PRSV: The RT-PCR assay was performed using HC-Pro, NIa-Pro, and CP-specific primers (Table S1), which resulted in the expected PCR amplicon of ~1.3 kb, ~1 kb, and ~1.1 kb, respectively. Among them, the RT-PCR amplified products of HC-Pro (n=15), NIa-Pro (n=17), and CP (n=24) coding regions yielded sequences of 1371 nt, 723 nt, and 837-876 nt, respectively, upon trimming the flanking

region sequences. The resultant consensus sequences were submitted to the NCBI GenBank database and obtained their respective accession numbers (Table 1).

Although37 samples were tested positive for PRSV infection in DAC-ELISA and bio-assays, but only 27 wereamplified successfully in the RT-PCR assay. This may be due to possible sequence variations in the primer binding sites. Previous studies have shown that mismatches in the primer binding sequences specifically at the 3' end of the primers can significantly destabilize the primer-template duplex, potentially reducing PCR specificity (Kwok et al. 1990; Christophersonet al. 1997; Klein et al. 2001; Bru et al. 2008). So,genetic diversity studies based on PCR amplification using target-specific primers will not be so successful and suggestive in exploring all the available populations (strains/pathotypes) of any virus. This highlights the limitations of relying solely on PCR-based methods and Sanger sequencing for viral detection.A more comprehensive approach, such as Next-Generation Sequencing (NGS), could provide greater sensitivity, accuracy, and high-throughput, holistic detection of all the viruses or their strains associated with the diseased samples. Further, NGS help in generating complete/near-complete genome sequences, facilitating more precise and reliable identification of viral infections.

Sequence topology: The HC-Pro coding sequences measured 1371 nucleotides (encoding 457 amino acids) demonstrated no variation in length and showed the aa sequence variability only up to9%.Similar previous observations have highlighted the remarkable sequence conservation(0-9% at aa level) in the HC-Pro coding sequences of PRSV in other studies as well (Mishra et al. 2019; Yeh et al. 1992). This emphasizes its potential role in the pathogenicity of the virus and host interaction mechanisms. Although the 5' terminal protein-coding sequences of South Indian isolates of PRSV displayed a significant variability, with the P1 region being the most variable, HC-Pro remained the most conserved region (Mishra et al. 2019).

Similarly, the NIa-Pro coding sequences measured 723 nt(encoding 241 aa) also exhibited a negligible aa sequence diversity (0-10%). Notably, as usual, at the 27thposition of NIa-Pro coding sequences of this study featured a distinct amino acid unique to the pathotypes of PRSV-P (Lysine or Arginine) and -W (Aspartic acid or Glutamic acid), known for conferring host specificity (Chen et al. 2008; Basavaraj et al. 2019). Remarkably, similar results were observed in the present study as well.

Unlike HC-Pro& NIa-Pro, the lengths of CP coding sequences varied among papaya and cucurbit isolates. Interestingly, in the present study, the Indian PRSV isolates showed considerably higher diversity, from 279 aa to 292 aa, with a variation of 39 nt (13 aa). The variability in length of CP sequences was more significant among PRSV-P and PRSV-W isolates. This observation suggests that the CP sequence lengths of Indian PRSV isolates were highly variable, i.e., from 837 to 876 nt. Out of the isolates of the present study, apapaya-originated PRSV-P isolate from Karnataka [IND-KA13 (P-P) ON015661] had the shortest length of 837 nt, whereas two cucurbit-originated PRSV-P isolates one from Uttar Pradesh [IND-UP6 (C-P) ON015690], and another Delhi [IND-DL7 (C-P) ON015665], had the longest length of 876 nt.

In the previous studies, papaya-PRSV-P isolates comprised 846-861 nt (encoding 282-287 aa), while the cucurbit-PRSV-P isolates, comprised 855-876 nt (encoding 285-292 aa). Subsequent study showed that, the lengths of CP coding regions ofPRSV have ranged from 840-873 nt (encoding 280-291 aa) (Jain et al.1998, 2004; Wei et al. 2007). This variability was observed in isolates from different locations, including India and China, and was not correlated with geographical origin (Jain et al.1998, 2004; Wei et al.2007). In a previous

study, the CP gene sequence of the papaya malformed leaf strain of PRSV was the shortest (861 nucleotides) in length, with a three-nucleotide deletion as compared to other strains (Guo et al. 2001).

Several studies have emphasized considerable diversity in the length of CP sequences of Indian PRSV isolates as compared to those from other countries (Jain et al.1998, 2004; Akhter et al. 2013; Verma et al. 2016; Singh et al. 2017; Basavaraj et al. 2019). While the lengths of other Asian and American isolates were less variable, which ranged from only 858-867 nt (encoding 286-289 aa) and 855-867 nt (encoding 285-289 aa), respectively. Variations in the CP region have been shown to affect symptom severity in PRSV (Medina-Salguero et al. 2021). Additionally, greater sequence divergence was observed in the CP regions of PRSV isolates from Mexico and India as compared to those from the other isolates of the USA and Australia (Tripathi et al. 2008).

Elevated variability can be linked to the EK repeat region lying at the 5' end of CP, which varied from 16aa to 23 aa. In this study, the shortest EK region (16 aa) was observed in the papaya PRSV-P isolates such as IND-GJ1(P-P) (ON015670) and IND-GJ2 (P-P) (ON015671), followed by 17 aa in IND-DL4 (P-P) (ON015697), 18 aa in IND-AS (P-P*) (KC149500), IND-WB2 (P-P) (ON015662), JQ394693 IND-JK (P-P*), IND-WB1 (P-P) (JN979406), and IND-OD (P-P) (ON015681).Whereas, the longest EK repeat region (22 aa) was found in the cucurbit-originatedisolates such as IND-HR (C-P) (KY448332), IND-UP1 (C-P) (KY448324), IND-UP2 (C-P) (KY448325), and IND-DL7(C-P)(ON015665) among Indian isolates. Similarly, a PRSV-W isolate from Taiwan, TAI 3-W (AY027810), also possessed the longest EK repeat region of 23aa (Fig. S1).In the previous studies also, all the differences in CP length were observed within the first 50 amino acids at the amino-terminal end of the CP. This region was known for its variability, which has been attributed to changes in the number of EK repeats (Bateson et al. 2002; Jain et al.1998, 2004).The sequence variations in this region of the PRSV genome mainly cause conservative amino acid substitutions, which are likely to preserve CP function (Tripathi et al. 2008).

Interestingly, the CP length of different pathotypes of PRSV exhibited limited but distinct variation across geo-climatic regions also, ranging from 837 to 876 bp. The majority of isolates from HST, Mo, and Oc clustered within a narrow range of 855-861 bp, indicating a high degree of conservation. However, notable deviations were observed among the isolates of WSA regions, where P-P isolates showed shorter lengths (837-861 bp), and C-P isolates possessed an extended CP length of up to 876 bp. This pattern suggests that although the CP gene is generally conserved, subtle length polymorphisms occur and may be associated with host-driven selection or adaptive evolution rather than strictly geo-climatic influences. The predominance of similar CP lengths across diverse regions supports functional constraints on the CP gene, while the observed variations, particularly in the WSA isolates may reflect localized evolutionary pressures, possibly linked to host specificity or viral fitness.

In terms of variability with respect to the geo-climatic origin of the isolates, the widest range of differences in EK length was observed between PRSV isolates originating from HST (13 aa) and WSA (22 aa) regions. Among all PRSV isolates, those from the HST region had the shortest EK region (13 aa) in P-P isolates, while those from the WSA region had the longest EK region (up to 22 aa) in C-P isolates. Similarly, nearly all the W isolates displayed comparatively longer EK regions (21-22 aa), regardless of their climatic location. This suggests that the length of EK repeat regions in the CP is influenced by the host of origin rather than by geo-climatic conditions.

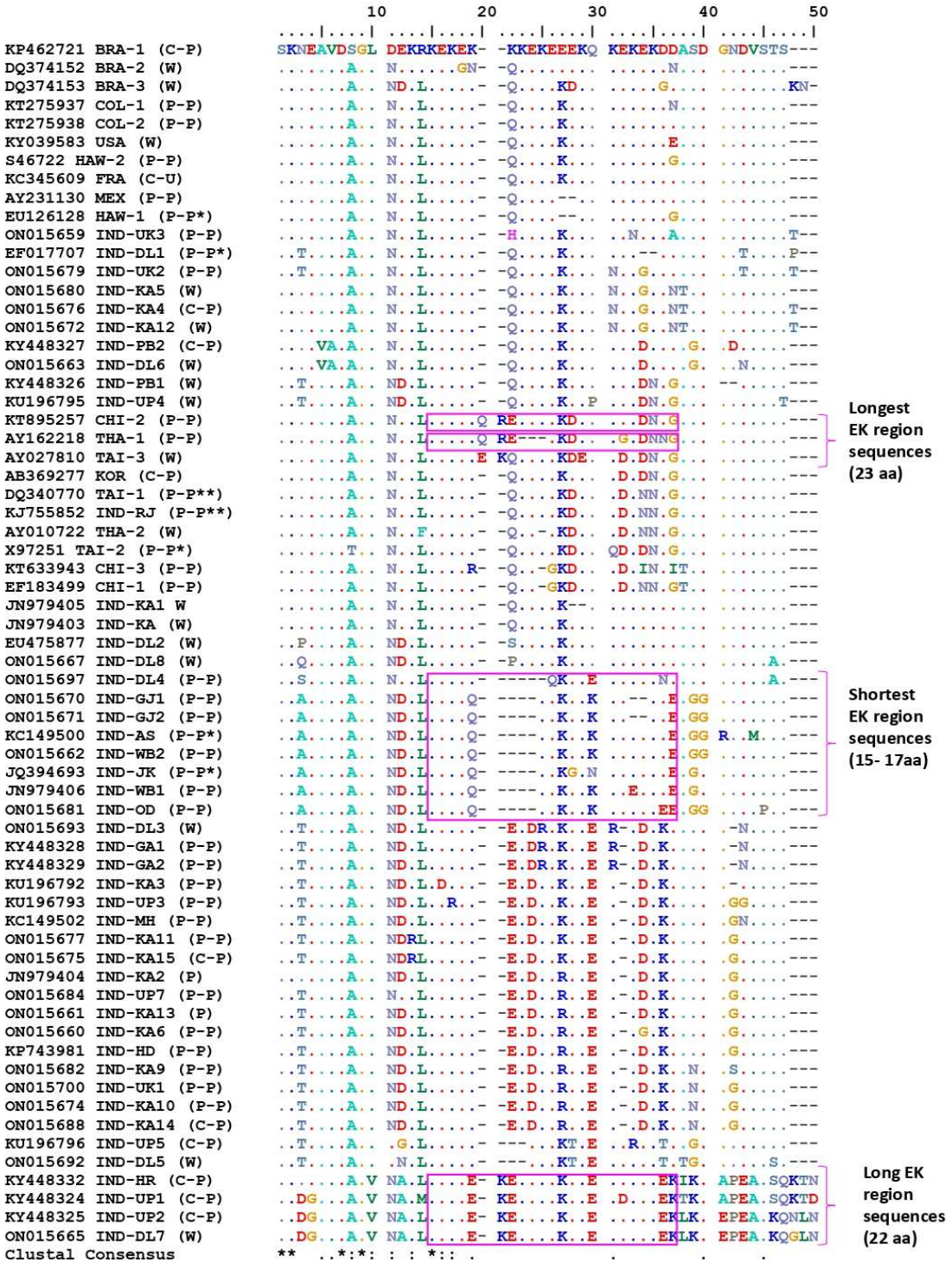


Figure S1. Sequence alignment of PRSV coat protein (CP) coding region showing the variation in EK repeat region.

These EK repeat regions have been associated with the genetic diversity and occurrence of recombination events within the virus (Mishra et al. 2019; Saleem et al. 2021).This diversity is noticeable especially within the CP coding region, where Indian PRSV isolates exhibited notable heterogeneity (Jain et al.1998, 2004; Singh et al. 2017).Previously, the increased frequency of EK repeats in the CP gene might be a result of the virus's specific traits, such as its geographic distribution (Hema and Prasad 2004). However, this statement did not hold as per the observation of this study. This could also be influenced by the virus's capability

to form S-S bonds, as seen in the CP of raspberry ringspot virus (RpRSV, *Nepovirus ruby*) (Acostaand Mayo 1990). These observations suggest that the development of effective resistance strategies may require a subtle understanding of the genetic makeup and evolutionary patterns of PRSV isolates, especially in locations where high genetic diversity is observed like India.

Phylogenetic Evolutionary Patterns in relation to pathotypes, hosts, and Geo-Climatic Locations: Phylogenetic analysis of HC-Pro coding amino acid sequences (Fig.2) revealed the formation of fivemajor clusters.The clusters represented different agro-climatic locations from where they have originated, namely, Mo, Oc,HST & WSA, WSA I, and WSA II.The isolates in these clusters included those originated from different hosts and belonged to both the pathotype and did not show any correlation with either their pathotypes or the host of their origin while forming their clusters. Notably, there found some distinct aa residues (encircled with solid line in Fig. 2) that are conserved among the clusters of WSA I and WSA II isolates and found some inconsistent aa residues (encircled with dotted line in Fig. 2) among the other clusters (Fig. 2). These conserved aa residues must have the direct relationship with the existing geo-climatic conditions to shape the adaptive populations to those ecological niches. This weak correlation aligns with the findings of Basavaraj et al. (2019), who also observed no significant relationship between genetic diversity and geographical distribution in the HC-Procoding sequences.

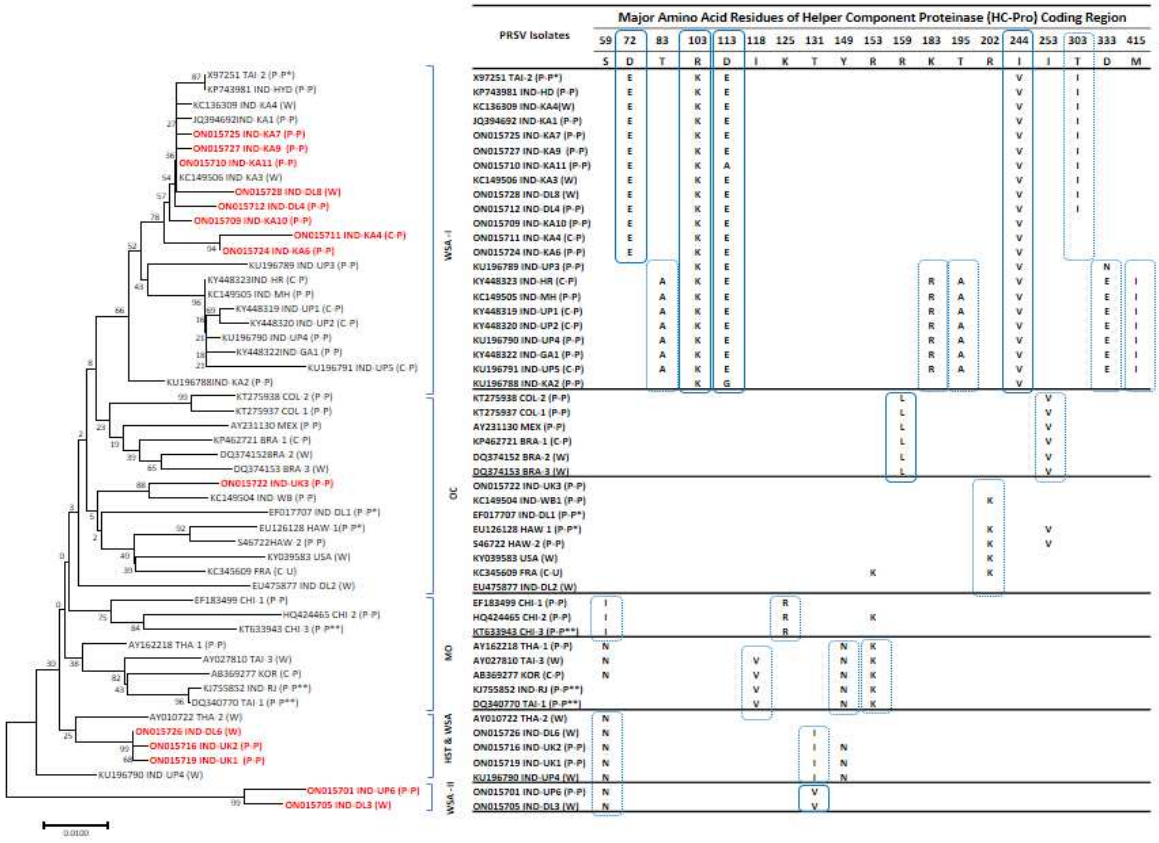
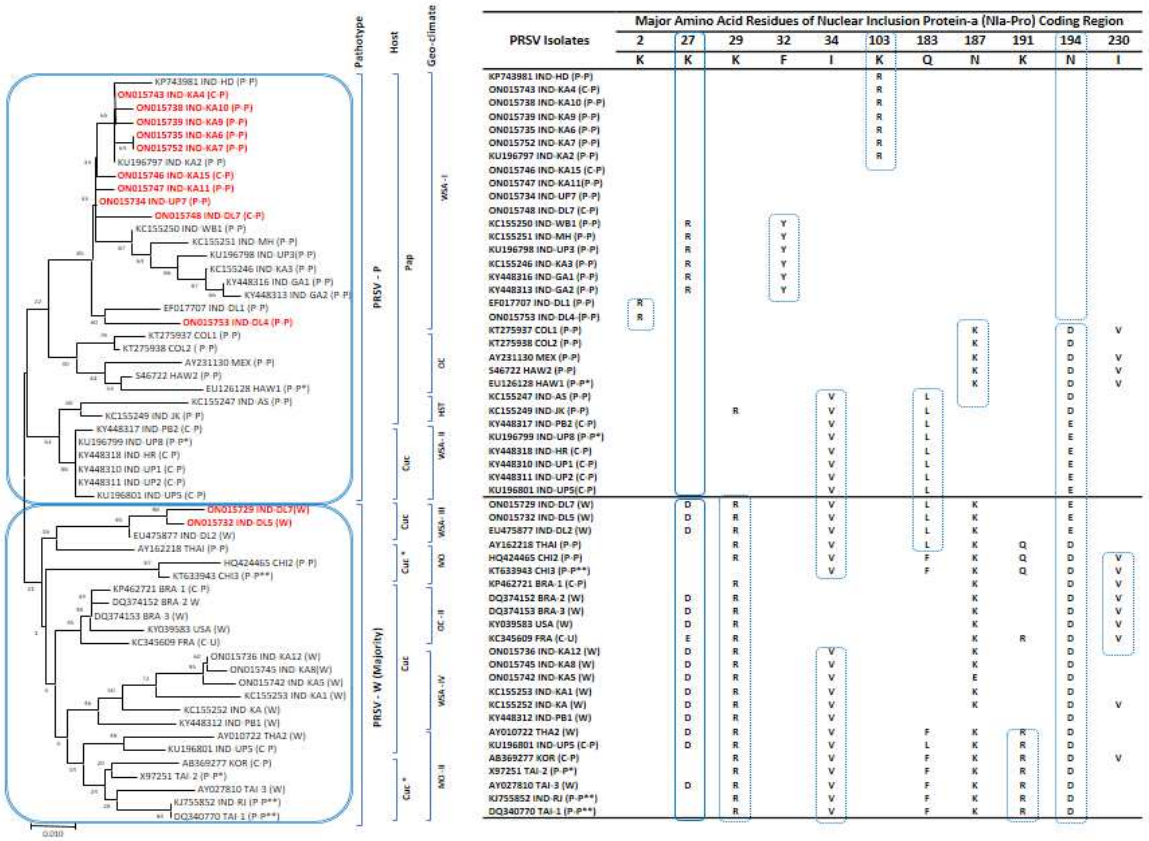


Figure 2: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the Hc-Pro coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

The phylogenetic tree based on the amino acid sequences of NIa-Pro coding region (Fig.3) indicated the formation of two unique clusters, PRSV P and PRSV W clusters. Notably, these clusters are primarily organized based on the pathotypes and hosts of origin of the isolates, showing a negligible or no correlation with their geo-climatic origin. Here, the isolates of pathotype-P were largely grouped in the major cluster I. Almost all the isolates of pathotype-W, along with some PRSV-P isolateswhichhave either originated from cucurbit hosts or were passed experimentally through any cucurbit hosts before generating their sequence information, were grouped under another major cluster II(encircled with double solid lines in Fig. 3). The isolates of PRSV-P cluster were uniquely displayed Lys^{27,29}(Fig. 3). Whereas, the isolates of PRSV-W cluster displayed Asp^{27} and Arg^{29}, as known previously (Chen et al. 2008; Basavaraj et al. 2019).



PRSV Isolates	Major Amino Acid Residues of Nuclear Inclusion Protein-a (NIa-Pro) Coding Region										
	2	27	29	32	34	103	183	187	191	194	230
	K	K	K	F	I	K	Q	N	K	N	I
KP743981 IND-HD (P-P)						R					
ON015743 IND-KA4 (C-P)						R					
ON015738 IND-KA10 (P-P)						R					
ON015739 IND-KA9 (P-P)						R					
ON015735 IND-KA6 (P-P)						R					
ON015752 IND-KA7 (P-P)						R					
KU196797 IND-KA2 (P-P)						R					
ON015746 IND-KA15 (C-P)											
ON015747 IND-KA11(P-P)											
ON015734 IND-UP7 (P-P)											
ON015748 IND-DL7 (C-P)											
KC155250 IND-WB1 (P-P)		R		Y							
KC155251 IND-MH (P-P)		R		Y							
KU196798 IND-UP3 (P-P)		R		Y							
KC155246 IND-KA3 (P-P)		R		Y							
KY448316 IND-GA1 (P-P)		R		Y							
KY448313 IND-GA2 (P-P)		R		Y							
EF017707 IND-DL1 (P-P)	R										
ON015753 IND-DL4 (P-P)	R										
KT275937 COL1 (P-P)								K		D	V
KT275938 COL2 (P-P)								K		D	
AY231130 MEX (P-P)								K		D	V
S46722 HAW2 (P-P)								K		D	V
EU126128 HAW1 (P-P*)								K		D	V
KC155247 IND-AS (P-P)					V		L			D	
KC155249 IND-JK (P-P)			R		V		L			D	
KY448317 IND-PB2 (C-P)					V		L			E	
KU196799 IND-UP8 (P-P*)					V		L			E	
KY448318 IND-HR (C-P)					V		L			E	
KY448310 IND-UP1 (C-P)					V		L			E	
KY448311 IND-UP2 (C-P)					V		L			E	
KU196801 IND-UP5(C-P)					V		L			E	
ON015729 IND-DL7 (W)		D	R		V		L	K		E	
ON015732 IND-DL5 (W)		D	R		V		L	K		E	
EU475877 IND-DL2 (W)		D	R		V		L	K		E	
AY162218 THAI (P-P)			R		V		L	K	Q	D	
HQ424465 CHI2 (P-P)			R		V		F	K	Q	D	V
KT633943 CHI3 (P-P**)					V		F	K	Q	D	V
KP462721 BRA-1 (C-P)			R					K		D	V
DQ374152 BRA-2 (W)		D	R					K		D	V
DQ374153 BRA-3 (W)		D	R					K		D	V
KY039583 USA (W)		D	R					K		D	V
KC345609 FRA (C-U)		E	R					K	R	D	V
ON015736 IND-KA12 (W)		D	R		V			K		D	
ON015745 IND-KA8 (W)		D	R		V			K		D	
ON015742 IND-KA5 (W)		D	R		V			E		D	
KC155253 IND-KA1 (W)		D	R		V			K		D	
KC155252 IND-KA (W)		D	R		V			K		D	V
KY448312 IND-PB1 (W)		D	R		V					D	
AY010722 THA2 (W)		D	R		V		F	K	R	D	
KU196801 IND-UP5 (C-P)		D	R		V		L	K	R	D	
AB369277 KOR (C-P)			R		V		F	K	R	D	V
X97251 TAI-2 (P-P*)			R		V		F	K	R	D	
AY027810 TAI-3 (W)		D	R		V		F	K	R	D	
KJ755852 IND-RJ (P-P**)			R		V		F	K	R	D	
DQ340770 TAI-1 (P-P**)			R		V		F	K	R	D	

Figure 3: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the NIa-Pro coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

Interestingly, the aa Arg^{29} was found in the cucurbit-originated PRSV-P [i.e., PRSV (C-P)] isolates as well as in some PRSV-P isolates that were passed experimentally through some cucurbit hosts [i.e., PRSV (P-P*)]. This observation indicates that the Arg^{29} may be correlated to the host from where the sequence information was generated rather than the pathotype. Besides the aa residues at 27th and 29th position, there are several other aa residues that are conserved among different sub-clusters, but are inconsistent (isolates are encircled with double dotted lines and aa residues with single dotted lines in Fig. 3).

ThisNIa-Pro sequence-based phylogenetic grouping pattern revealed a distinct clustering according to the host species of their origin, which may be due to specific interaction with some internal factors existing within the hosts from which the PRSV isolates originated. This observation suggests that the protease may play a significant role in the adaptation of PRSV to specific hosts, with variations in their sequence that may potentially influence host-specific interactions. Such host-driven clustering aligns with previous findings that emphasize the importance of NIa-Pro in viral host range determination and adaptation (Rodamilans et al. 2018; Gong et al. 2020; Kim et al. 2021). The NIa-Pro plays a key role in regulating cleavage efficiency at polyprotein junctions, which subsequently influences the host range and viability of potyviruses. Variations in protease activity can determine the virus's ability to adapt to different hosts and its capacity to persist and propagate within those hosts.

In the Previous studies on PRSV isolates, it was evident that a single aa change in NIa-Pro imparted a host shift from cucurbits to papaya or vice-versa, although the exact role of the protease's catalytic activity in this shift remained unclear(Chen et al. 2008). Similarly, studies on plum pox virus (PPV, *Potyvirus plumpoxi*) have shown that host adaptation is linked to the NIa-Pro activity, but not to specific variations in its sequence (Calvo et al.2014). Their study revealed that the host range, such as adaptation to *Nicotiana*species or *Prunus* species, is influenced by differences in the NIa-Pro target sequence located between 6K1 and CI. This suggests that host-specific processing of NIa-Pro at this site determines viral infection success. Other studies have also highlighted how NIa-Pro influences viral movement and replication in different hosts, reinforcing its role in host adaptation (Calvo et al. 2014).

Interestingly, based on CP sequences, the phylogenetic grouping of PRSV isolates primarily occurred as per their climatic origin, classifiedlargely as HST, Mo, Oc, and WSA (Fig.4).Isolates from Mo and Oc revealed that the grouping was not directly influenced by either host or pathotype, as they did not form distinct sub-groups. The Mo isolates uniquely possessed Asn35 and the Oc isolates possessed Val/ Ile63 while forming their respective clusters. However, the HST and WSA isolates did not form separate single major clusters, but, were divided into smaller sub-clusters either due to some unknown factor (in HST isolates) or due to the influence by host/pathotype. The HST isolates possessed Ala^{3}, Gln^{19}, & Thr^{220} among group I (HST-I) that composed of North Indian HST isolates and possessed Ser/Asn^{180} among group II (HST-II) that composed of South Indian HST isolates. Similarly, the WSA isolates got influenced by host/pathotype also on their sub-grouping. Here, the sub-cluster WSA-I possessed the isolates with deletion of an aa (Thr^{52}) at position 52, WSA-IV possessed Asp^{24}, and deletion of an aa at position 32 (Glu^{32}). Whereas, WSA-IV possessed Gln^{42}, WSA-V possessed $Thr^{28\ \&\ 36}$, WSA-VI possessed several unique amino acids conserved such as $Val^{10\ \&\ 155}$, $Glu^{19\ \&\ 43}$, $Lys^{37\ \&\ 39}$, Pro^{42}, Ala^{44}, Gln^{47}, $Lys/Asp/Gly^{48}$, Glu/Gly^{52}, Thr/Leu^{49}, $Asn/Asp/Ser^{50}$, $Glu^{19\ \&\ 54}$, Met^{66}, $Ile^{98\ \&\ 232}$, $Asp^{108\ \&\ 282}$, Thr^{182}, and Leu^{213} as the conserved unique amino acids. However, the reason for formation of sub-clusters WSA-II and WSA-III was not clear (Fig. 4).

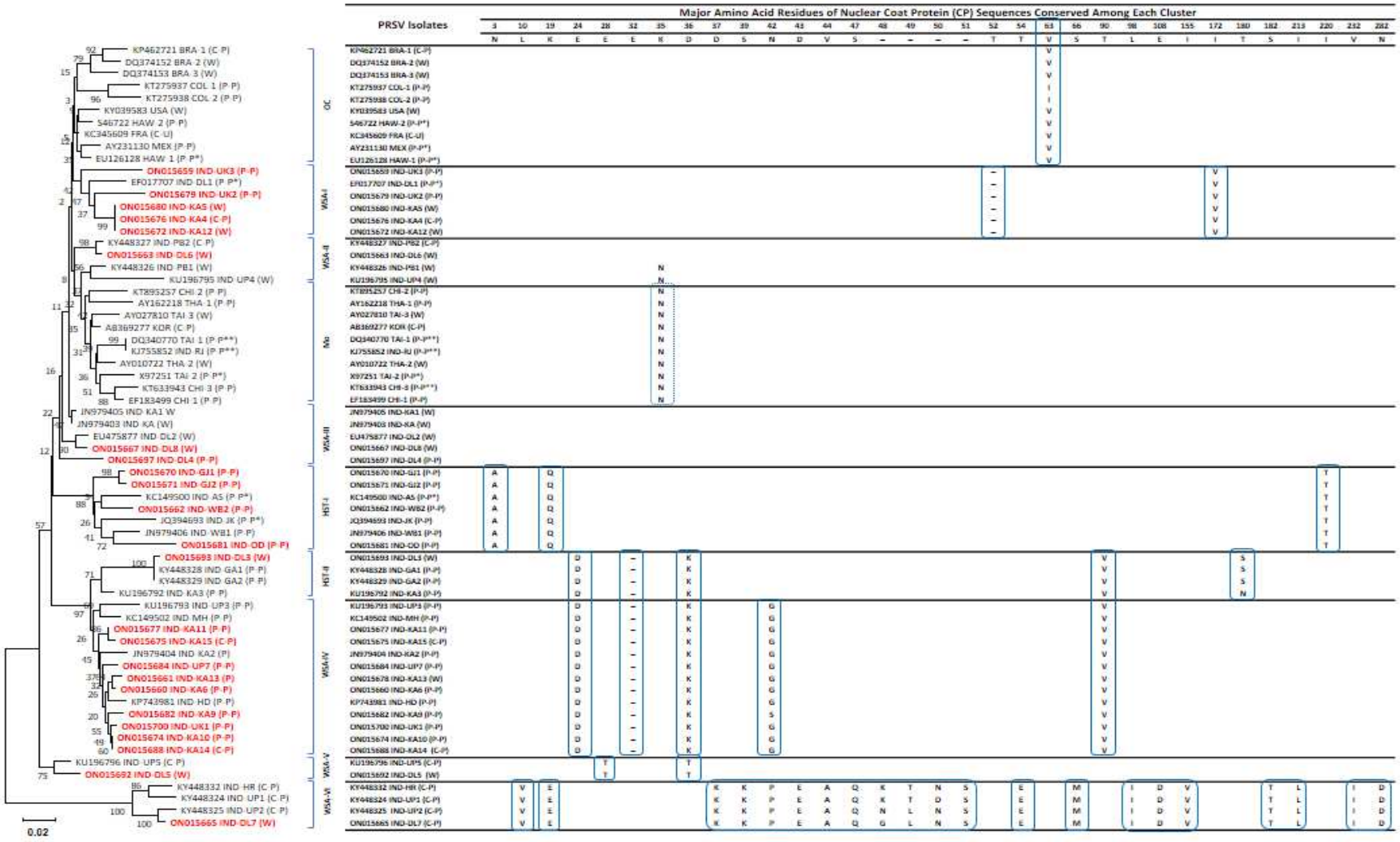


Figure 4: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the CP coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

In theCP sequence-based phylogenetic study, it was evident that the WSA climate significantly contributed to higher variations among PRSV populations during the evolutionary process. This maybe due to the considerable pressure from varying environmental conditions in WSA climatic regions on CP of the virus, which is lying exposed outside of the virion and have direct contact with the external environmental factors.Nevertheless, the Oc climate appears to contribute comparatively less variability. In contrast, PRSV populations among isolates in the Mo climate demonstrated a minimal level of variability. These results make it clear that the evolution of PRSV populations is primarily influenced by both geo-climatic locations and hosts.

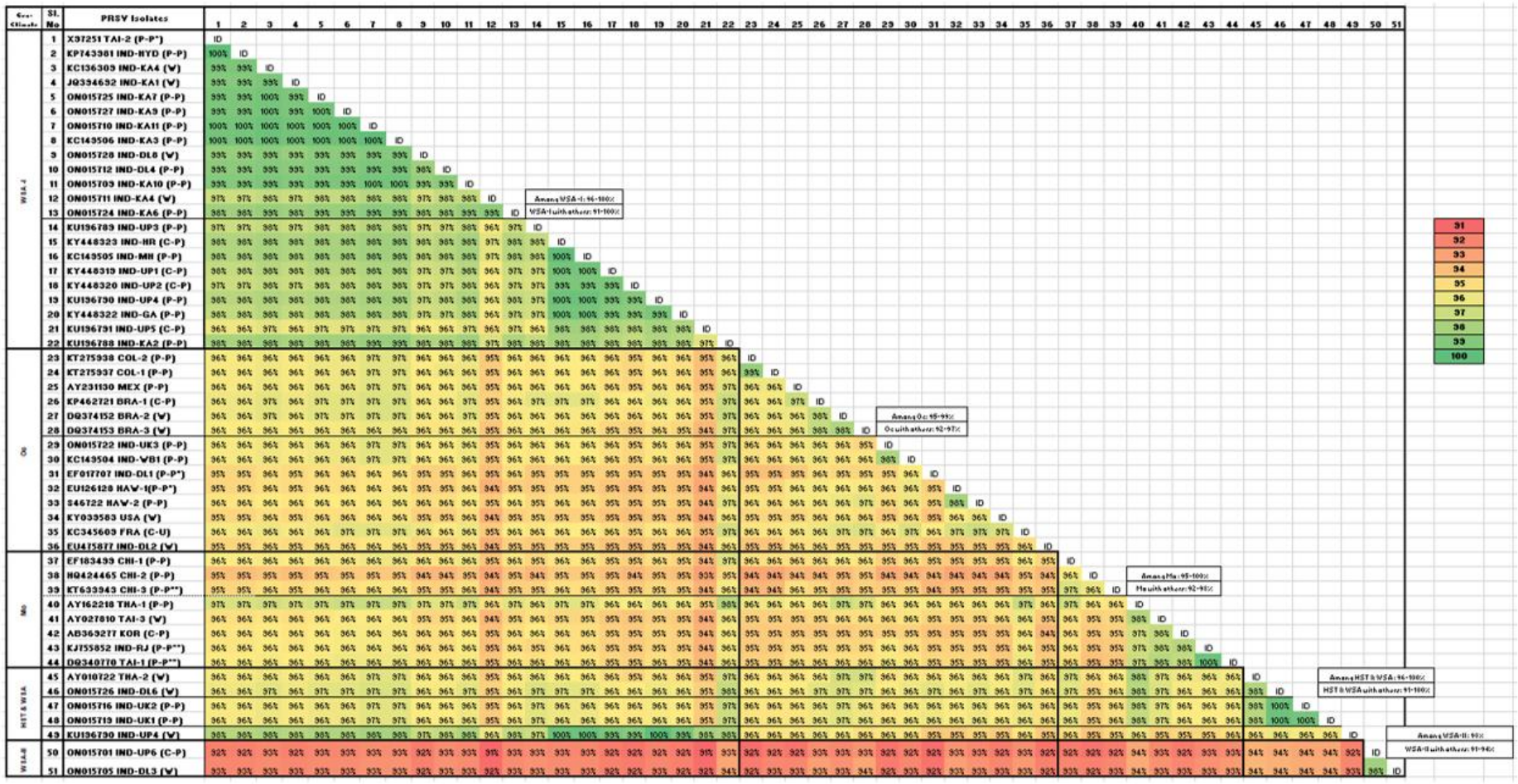
The distinct clustering patterns based on geographical origin in the current investigation also support the hypothesis of regional evolution of the virus and align with the previous findings of Fernández-Rodríguez et al. (2008) and Bateson et al. (2002). In the phylogeny of tomato yellow leaf curl virus (TYLCV, *Begomovirus coheni*) isolates(Lefeuvre et al. 2010) and of the tobacco mild green mosaic virus (TMGMV, *Tobamovirus mititessellati)*isolates (Rodríguez-Cerezo et al. 1991; Moya et al. 1993; Fraile et al. 1996), the datasets were grouped into nine geographical locations, they were grouped based on geographical proximity. Similar studies were also observed on Kennedya yellow mosaic virus (KYMV, *Tymovirus kennedyae*), where the isolates clustered into three groups: one for Desmodium and two for Kennedya, which are separated by latitude (Skotnicki et al. 1996).

PRSV Genetic Diversity in Geo-Climatic Perspectives: In this study, the sequence information of the HC-Pro, NIa-Pro and CP coding region sequences of PRSV isolates originating from four distinct climatic locations: Humid Sub-Tropical (HST), Monsoon (Mo), Oceanic (Oc), and Warm Semi-Arid (WSA) were generated. The sequences were examined for their sequence identities/diversitiesas per their phylogenetic clustering.

Based on the HC-Pro coding sequences, the isolates clustered in the phylogenetic tree as Mo, Oc, HST & WSA, WSA-I, and WSA-II groups representing more or less different geo-climatic regions. These clusters shared 92-98%, 92-97%, 91-100%, 91-100%, and 91-94% aasequence identities, respectively, with rest of the isolates. The results showed that the isolates from WSA-II region were more divergent (with aa sequence identities from as low as 91% and only up to 94%) followed by those from Oc region (with the aa sequence identities from 92-97%), Mo region (with the aa sequence identities from 92-98%)and of HST origin i.e. ON015719 IND-UK1 (P-P) and ON015716 IND-UK2 (P-P) which shared the aa sequence identities from 92-97% 94-98%. These two isolates were grouped along with some WSA isolates and the no reason could be trace for this. Interestingly, the WSA-I isolates were found more stable with 96-100% aa sequence similarities among themselves, but 91-100% with the others (Table S2).

These results reveal that the HC-Pro coding region sequences are highly conserved with a potential variability of only up to 9%. These findings suggest that the correlation between HC-Pro and the prevailing climate is somewhat constrained (Mishra et al. 2019).However, the lack of a relationship between variability and the geographical origin of PRSV isolates was also noted (Bag et al. 2007).

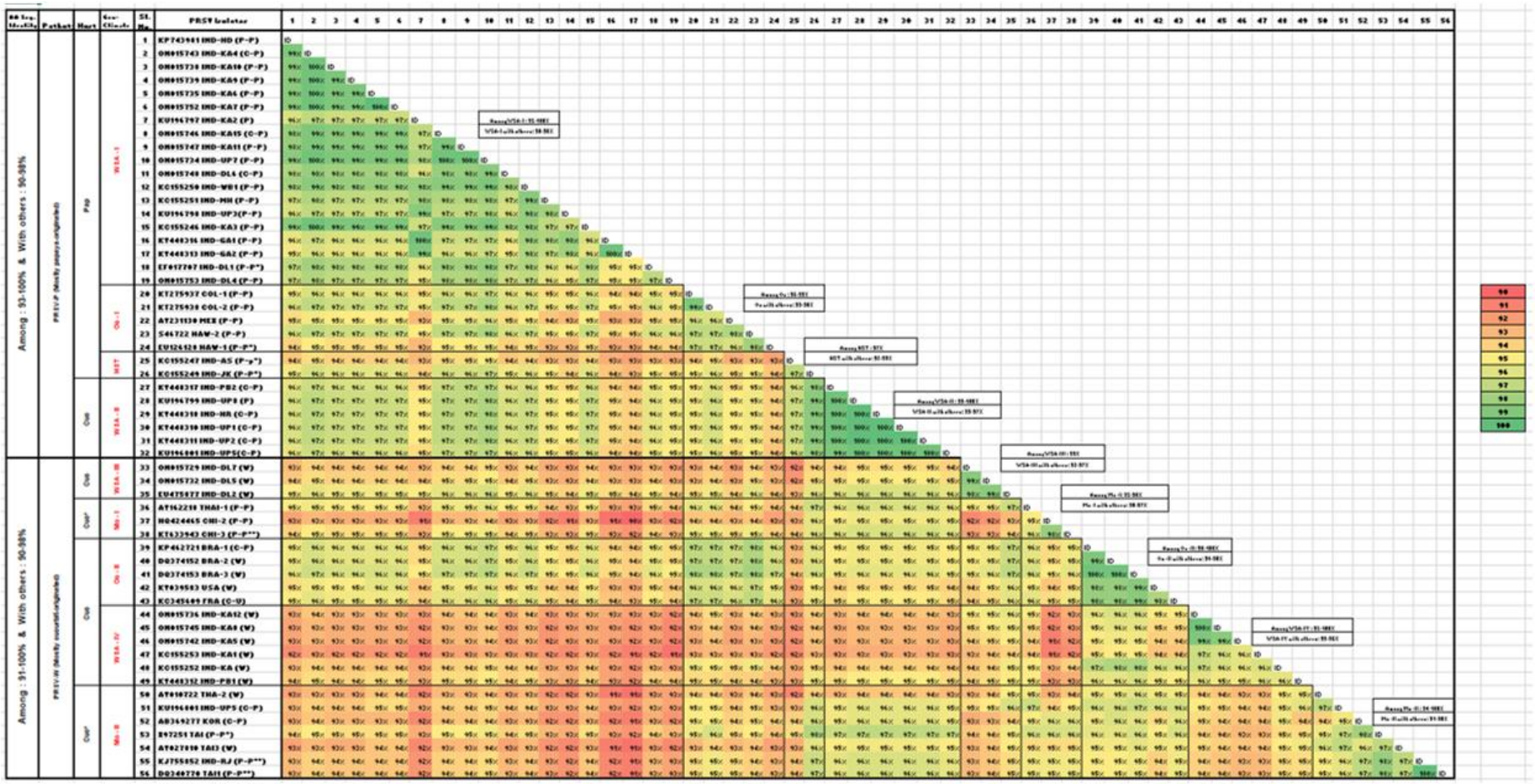
Table S2. Colour-coded pairwise identity matrix generated from the HC-Pro coding region of 51 papaya ring spot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.



Based on NIa-Pro coding sequences, the PRSV isolates under investigation shared a overall 90-100% aa sequence identities (upto 10% variability). However, they formed the two major clustersin the phylogeny, namely PRSV P and PRSV W, which shared 93-100% and 91-100% aa sequence identities, respectively, among themselves. These results showed that the isolates of pathotype P are less variable (only upto 7%) and those of the pathotype W are relatively more variable (upto 9%). Further, the isolates of PRSV P pathotypes were sub-clustered as HST,Oc-I, WSA-I, and WSA-II supporting their geo-climatic origin as well. They shared 92-99%, 93-98%, 90-98%, and 93-99% aa sequence identities, respectively with all the other isolates(Table S3). These results indicated thatthe PRSV P isolates have no significant aa sequence variability among themselves. On the other hand, the isolates of pathotype W have also formed five sub-clusters, namely,Mo-I, Mo-II, Oc II, WSA III, and WSA IV with 91-97%, 91-98%, 94-98%, 92-97%, and 93-96% aa sequence identities, respectively.Among all the pathotype W isolates, those originated from Mo-I region were found comparatively more variable (upto 10%)(Table S3).

Notably, the isolates of the Mo-I group named AY162218 THAI-1 (P-P**), HQ424465 CHI-2 (P-P), andKT633943 CHI-3 (P-P**) are basically available in the NCBI GenBank database as isolates of pathotype P; however, they have been grouped in the present investigation under the major cluster of pathotype W in the phylogeny. Besides, although theypossessLys^{27},thatoccur specifically in the NIa-Procoding region of pathotype P (Chen et al. 2008), but, getting clustered with the pathotypes W, exceptionally. The possible reason for this may be that, although no confirmatory evidences available yet, we hypothesize that these isolates might have been passed through any cucurbitaceous hosts experimentally before generating their sequence information. Thus, they might have acquired some host specific sequence features. This implies that the sequence composition of NIa-Pro is directly linked to the existing climatic conditions, alongside host specificity. Such a connection could drive modifications in the protein sequence composition through natural selection, facilitating adaption/survival and ultimately contributing to genetic diversity within the population.

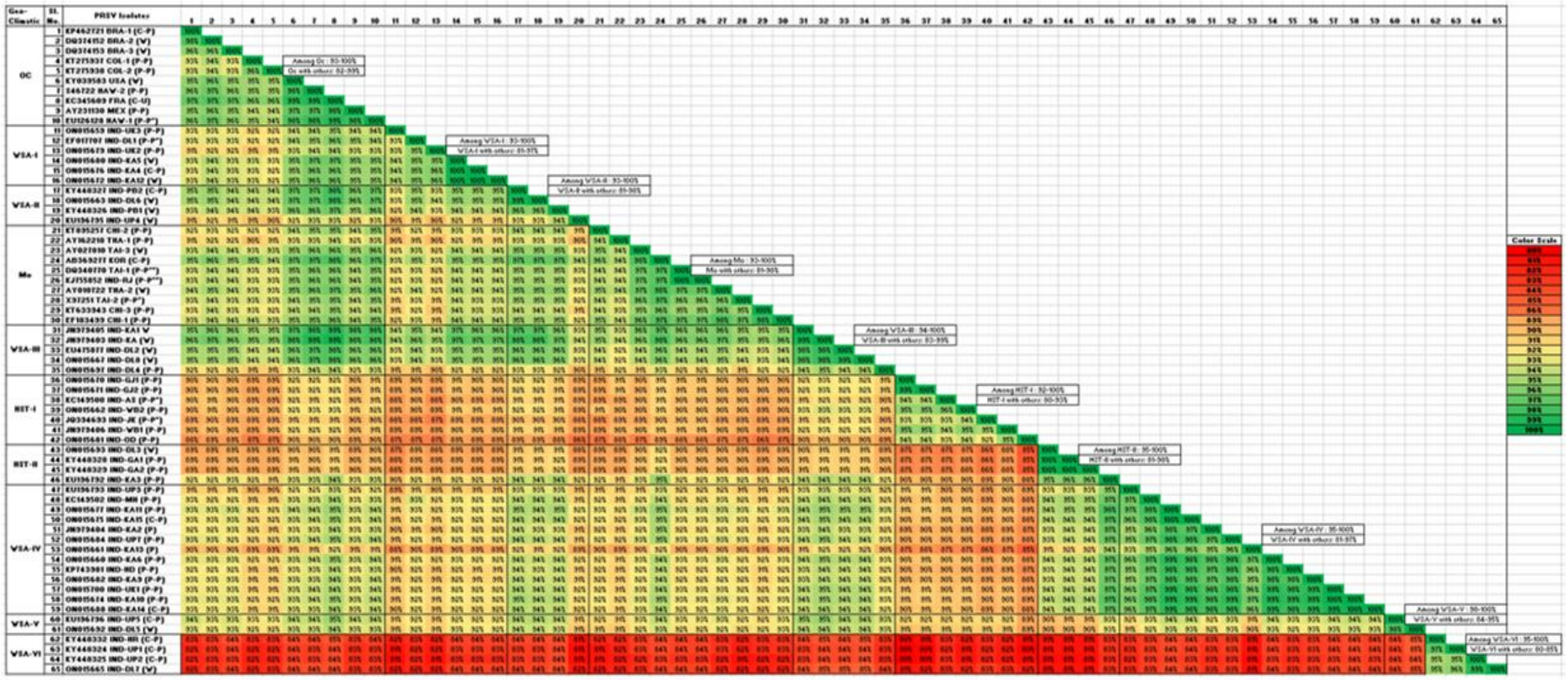
Table S3. Colour-coded pairwise identity matrix generated from NIa-Pro coding region of 55 papaya ringspot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.



Upon examining the CP sequences of PRSV isolates originated from different geo-climates,as clustered under the phylogeny, namely, HST-I, HST II, Mo,Oc, WSA I, WSA II, WSA III, WSA IV, WSA V and WSA VI. It was found that these groups of isolates shared80–93%, 81–98%, 81–98%, 82-99%, 81-97%, 81-98%, 83-99%, 81-97%, 84-95%, and 80-85% aa sequence identities (Table S4). This implies that the genetic makeup of the virus varied across different climates. Notably, when compared across the groups, the WSAand HST isolates demonstrated a significant amino acid diversity of up to 20%. Followed by, the Mo climate population demonstrated the next highest diversity (up to 19%), and the Oc climate isolates exhibited up to 18% diversity. Whereas, when studied the intra population diversity, the HST isolates were found most diverse (upto 8%) among themselves, followed by WSA and Mo isolates (upto 7%). However, when studied based on the host as well as the geo-ecologies from where the isolates have originated, the cucurbit-originated PRSV P (PRSV C-P) isolates of WSA region, although highly conserved among themselves up to only 15% aa diversity, were found highly diverse (up to 20% aa diversity) in comparison to other groups. While the PRSV W isolates of the same geo-ecological regions exhibited lesser comparative aa diversity of up to 19% (Table S4).

The observed variation in CP amino acid sequences among PRSV isolates from different geo-climatic regions indicates substantial geographical structuring of the viral population. The greater inter-population diversity (up to 20%) compared with intra-population diversity (up to 8%) suggests that geographical separation and environmental heterogeneity may contribute to PRSV diversification. Such spatial differentiation in plant viruses can arise through the combined effects of mutation, genetic drift, restricted gene flow, host-associated selection, and environmental factors (García-Arenal and Fraile 2008; McLeish et al. 2019). Notably, cucurbit-origin PRSV-P isolates from the WSA region showed up to 20% amino acid divergence from other groups while remaining comparatively conserved within the population, suggesting possible host-associated selection on viral diversification. In contrast, PRSV-W isolates from the same geo-ecological regions showed comparatively lower divergence (up to 19%). Host adaptation and ecological conditions influence the genetic composition and evolutionary trajectories of plant virus populations (García-Arenal et al. 2001; Elena 2017; Lefeuvre et al. 2019). Collectively, these findings suggest that the genetic differentiation of PRSV may reflect the combined influence of host specificity and geo-climatic conditions.

Table S4: Colour-coded pairwise identity matrix generated from CP coding region of 65 papaya ring spot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.



Deciphering the Genetic Landscape by Exploring Recombination Patterns and its Impacts on Genetic Variability: The *in silico* analysis was performed to examine the presence of possible recombination events in the HC-Pro, NIa-Pro, and CP coding regions of PRSV isolates by employing six recombination detection methods*viz*., RDP, GeneConv, BootScan, MaxChi, Chimaera, SiScan and 3Seq available in the RDP4 software package. The details of the recombination events observed,along with the position of breakpoints and average *p*-values, are mentioned in (Table 2 and Fig. 5).

Table2: Recombination events detected among Papaya ring spot virus (PRSV) isolates based on CP and NIa-pro coding regions.

Recombinant Isolate	Break points	Major Parent	Minor Parent	Detection method@	*p*-Value
(a) Based on coat protein coding region					
ON015697 (IND-DL4-P)	320-784	ON015692 (IND-DL5-W)	EF17707 (IND-DL1-P)	S, **C**& M	3.981×10^{-03}
KU196793 (IND-UP3-P)	699-772	KC149502 (IND-MH-P)	ON015670 (IND-GJ1-P)	B, **M**, T, R, C & G	5.944×10^{-04}
KU196793 (IND-UP3-P)	704-763	KU196795 (IND-UP4-W)	ON015693 (IND-DL3-W)	R, **G**, B, S, & T	7.741×10^{-03}
KU196795 (IND-UP4-W)	246-739	KY448326 (IND-PB1-W)	KU196796 (IND-UP5-C-P)	M, C, & **T**	5.2338×10^{-05}
(b) Based on NIa-pro coding region					
EU475877 (IND-DL2-W)	596-682	ON015731 (IND-DL6-W)	EF17707 (IND-DL1-P)	B, S, T, R,M, **C**, & G	9.543×10^{-08}
ON015745 (IND-KA8-W)	462-625	KC155247 (IND-AS-P-P*)	ON015753 (IND-DL4-P)	R, T, **M**, C, & S	3.271×10^{-03}
ON015736 (IND-KA12-W)	462-625	KC155247 (IND-AS-P-P*)	ON015753 (IND-DL4-P)	R, T, **M**, C, & S	3.271×10^{-03}
ON015742 (IND-KA5-W)	462-625	KC155247 (IND-AS-P-P*)	ON015753 (IND-DL4-P)	R, T, **M**, C, & S	3.271×10^{-03}
KC155246 (IND-KA3-P)	462-625	KC155247 (IND-AS-P-P*)	ON015753 (IND-DL4-P)	R, T, **M**, C, & S	3.271×10^{-03}
KC155252 (IND-KA-W)	462-625	KC155247 (IND-AS-P-P*)	ON015753 (IND-DL4-P)	R, T, **M**, C, & S	3.271×10^{-03}

The method for identifying recombination events with the most significant *p*-value through respective method is highlighted in bold, and the associated *p*-value displayed. These methods include @: RDP (**R**), GeneConv (**G**), BootScan (**B**), MaxChi (**M**), Chimaera (**C**), SiScan (**S**), and 3Seq (**T**).

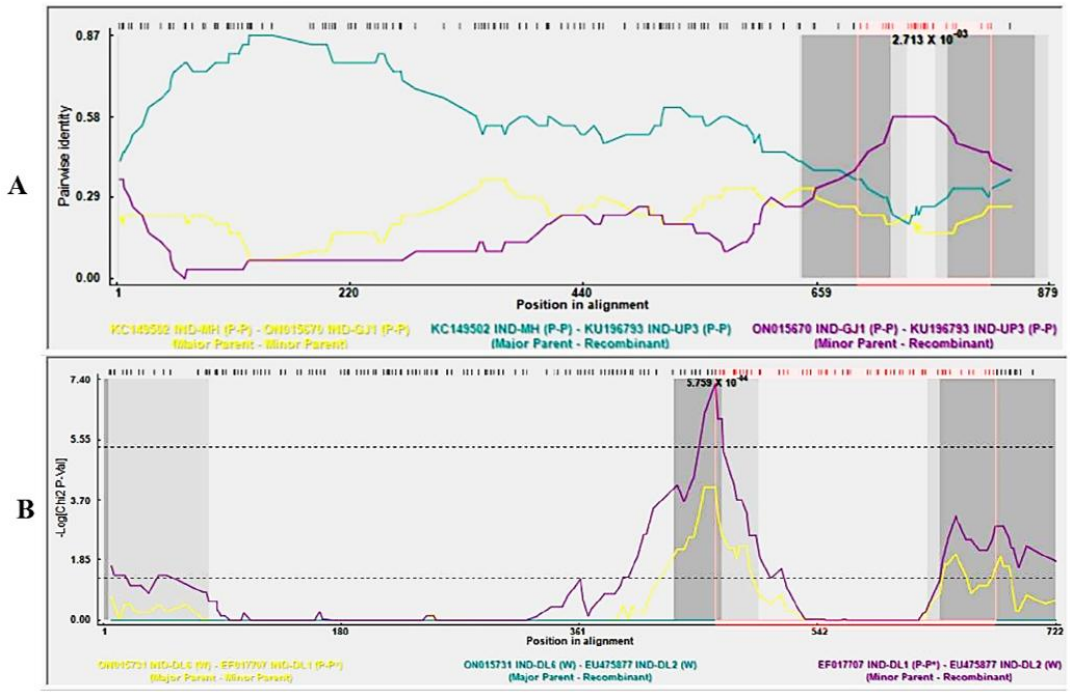


Figure 5. Representative recombination events detected in PRSV isolates using RDP4. (A) Representative recombination event from the CP gene. (B) Representative recombination event from the NIa-Pro gene. Shaded regions indicate inferred recombinant fragments and vertical lines represent predicted breakpoint positions.

Four events were found in the CP coding region of the recombinant isolates namely, [IND-DL4 (P-P), IND-UP3 (P-P), IND-UP3 (P-P), and IND-UP4 (W)] with [(IND-DL5 (W), IND-MH (P-P), IND-UP4 (P-P), & IND-PB1 (W)] as their major parents and[(IND-DL1(P-P), IND-GJ1 (P-P), IND-DL3 (W), & IND-UP5(C-P)] as their minor parents.Possible breakpoint events within the specific isolate were also identified(Table 2 and Fig. 5A).

Based on NIa-Pro coding regions, the analysis revealed six PRSV isolates as recombinants, which largely consisted of PRSV-W isolates named [(IND-DL2 (W), IND-KA8 (W), IND-KA12 (W), IND-KA5 (W), IND-KA3 (P), and IND-KA (W)]. They possessed [(IND-DL6 (W), IND-ASM (P-P)] as their major parents, and [(IND-DL1 (P-P) &IND-DL4 (P-P)] as their minor parents (Table 2 and Fig. 5B).

HC-Pro coding sequences did not reveal any recombination events.This observation provides support for the fundamental reason why the HC-Pro coding region remains highly conserved and stable both in length and sequence homogeneity across the studied PRSV isolates.The attributes of the HC-Pro coding region suggest that it could serve as a more viable target for managing papaya ringspot disease through RNA interference.

Based on the earlier studies, it seems that recombination is a recurrent phenomenon within the genus *Potyvirus*. Consequently, recombination stands out as a crucial factor facilitating the adaptation of potyviruses to diverse hosts and environments, enabling their survival and widespread dissemination (Basavaraj et al. 2019).Homologous recombination is prevalent in plant RNA viruses, such as potyviruses (Chare and Holmes 2006) and bromoviruses (Codonerand Elena 2008), where it plays a critical role in minimizing deleterious mutations in small populations and enabling rapid adaptation to changing environments (Lai 1992; Roossinck 1997; García-Arenal et al. 2001). Similarly, recombination is a key driver in the evolution of single-stranded RNA viruses. Findings from studies on soybean mosaic virus (SMV) (Gagarinovaet al. 2008; Seo et al. 2009) and other potyvirus isolates (Revers et al. 1996) highlight its role in facilitating genetic diversity and promoting

quicker adaptation to selective pressures, such as host resistance and ecological shifts (Andino and Domingo 2015). However, in the case of recombination analysis in the present study,the identification of four recombination events in the CP coding region suggests that recombination is an important evolutionary force contributing to the genetic diversity of PRSV. The involvement of different isolates as major and minor parents indicates active genetic exchange among PRSV populations, leading to the emergence of new viral variants and increased evolutionary adaptability. Similarly, based on NIa-Pro, the host specificity determinant, the recombinants observed were, although W pathotypes, they were not a result of recombination between P & W pathotypes. This observation suggests that in PRSV, recombination may not be playing a crucial role in generating host-specific pathotypes, but mutation must be playing a role.

These findings suggest that recombinant isolates arise naturally when PRSV–P (Papaya) isolates and Cucurbit-P or W isolates co-infect cucurbitaceous hosts.These results indicate thatrecombination has the potential to have significant implications on the genetic diversity, and evolution of viruses and enhance virulence (Escriu et al. 2007). Recombination can create viruses with enhanced capabilities, complicating the prediction of epidemic advancement and the overall behaviour of viral populations. The diversity of PRSV varies by region due to spatial recombination and temporal evolution within provinces (Saleem et al. 2021). This variability demonstrates varying rates of change, potentially influenced by introductions, the movement of plant materials, geographical isolation, and diverse disease management strategies (Chin et al. 2007). These findings emphasize the role of recombination in enhancing genetic diversity, viral adaptability, and virulence, contributing to the evolution and spread of PRSV across regions.

Gene flow and genetic differentiation assessment between PRSV populations: Phylogenetic analysis of HC-Pro sequences did not reveal any clear clustering based on geo-climatic conditions, whereas NIa-Pro sequences exhibited clustering patterns associated with host origin, indicating gene-specific differences in evolutionary constraints. Due to these inconsistent and host-associated clustering patterns, CP sequences were selected for subsequent analyses of gene flow and genetic differentiation.

In this analysis, genetic differentiation among PRSV populations across geo-climatic regions was strongly supported by *Snn*, *Kst**, and *Fst* statistics. Snn values were consistently high (0.88889–1.00000) and significant ($p \leq 0.004$), indicating pronounced population subdivision (Hudson 2000), while *Kst** values ranged from 0.0101–0.2950 (Table 3), Most comparisons showed significant differentiation based on *Snn* and *Kst** ($p<$ 0.05), indicating distinct genetic structuring among the populationsand further confirming significant genetic differentiation among populations (Hudson et al. 1992). Estimates of gene flow based on *Fst* values (0.1899–0.7713) indicated moderate to high genetic structuring. The highest *Fst* was observed between WSA-I and HST-II (0.7713), followed by HST-I–HST-II (0.7125) and WSA-IV–WSA-VI (0.6915).

569 **Table 3.** Statistical tests for genetic differentiation and gene flow between Papaya ringspot virus populations
570 from Oceanic, Humid sub-tropical, Monson, and Warm semi-arid based on the geo-climatic regions.

Population 1	Population 2	Snn	P value	Kst*	P value	Fst
OC	WSA I	0.93750	0.00020	0.06665	0.00000	0.24596
OC	WSA II	0.92857	0.0080	0.06112	0.0000	0.31974
OC	MO	1.00000	0.00000	0.07457	0.00000	0.39992
OC	WSA III	0.96667	0.0020	0.03613	0.00000	**0.18989**
OC	HST I	1.00000	0.0010	0.09705	0.00000	0.52323
OC	HST II	1.00000	0.0000	0.15809	0.00000	0.72317
OC	WSA IV	1.00000	0.00000	0.16518	0.00000	0.66706
OC	WSA V	1.00000	0.0240	0.07127	0.0020	0.53999
OC	WSA VI	1.00000	0.0050	0.12062	0.00000	0.63376
WSA I	WSA II	1.00000	0.0120	0.13811	0.0120	0.38336
WSA I	MO	1.00000	0.00000	0.11170	0.00000	0.46421
WSA I	WSA III	1.00000	0.00000	0.09468	0.0010	0.23877
WSA I	HST I	1.00000	0.00000	0.14591	0.00000	0.55896
WSA I	HST II	1.00000	0.0020	0.29499	0.0060	**0.77129**
WSA I	WSA IV	1.00000	0.00000	0.20316	0.00000	0.71228
WSA I	WSA V	1.00000	0.00000	0.20245	0.00250	0.57605
WSA I	WSA VI	1.00000	0.0060	0.20528	0.0050	0.67323
WSA II	MO	0.92857	0.0020	0.05716	0.0010	0.27095
WSA II	WSA III	0.88889	0.0050	0.04259	0.00000	0.19412
WSA II	HST I	1.00000	0.0060	0.07752	0.0020	0.40226
WSA II	HST II	1.00000	0.0250	0.23307	0.0210	0.56653
WSA II	WSA IV	1.00000	0.0030	0.13648	0.00000	0.49386
WSA II	WSA V	1.00000	0.00000	0.04331	0.00260	0.34776
WSA II	WSA VI	1.00000	0.0010	0.01013	0.00000	0.54239
MO	WSA III	1.00000	0.0010	0.06231	0.00000	0.31880
MO	HST I	1.00000	0.00000	0.09779	0.00000	0.50483
MO	HST II	1.00000	0.00000	0.15336	0.00000	0.68967
MO	WSA IV	1.00000	0.00000	0.15685	0.00000	0.62619
MO	WSA V	1.00000	0.0130	0.07083	0.0070	0.51640
MO	WSA VI	1.00000	0.0030	0.12100	0.0010	0.62401
WSA III	HST I	1.00000	0.0050	0.07725	0.00000	0.40817
WSA III	HST II	1.00000	0.0040	0.20688	0.0040	0.61653
WSA III	WSA IV	1.00000	0.00000	0.13932	0.00000	0.53439
WSA III	WSA V	1.00000	0.0930	0.06567	0.0310	0.41636
WSA III	WSA VI	1.00000	0.0230	0.11526	0.0090	0.58002
HST I	HST II	1.00000	0.0060	0.19257	0.0050	0.71247
HST I	WSA IV	1.00000	0.00000	0.16825	0.00000	0.65339
HST I	WSA V	1.00000	0.08500	0.08065	0.0280	0.53023
HST I	WSA VI	1.00000	0.00000	0.13825	0.00000	0.62339
WSA IV	HST II	1.00000	0.0040	0.12590	0.00000	0.60098
WSA IV	WSA V	1.00000	0.0040	0.14056	0.00000	0.64540
WSA IV	WSA VI	1.00000	0.0010	0.18265	0.00000	0.69151
WSA V	WSA VI	1.00000	0.0010	0.08443	0.00000	0.60745

571 ns: not significant; *0.01 < P < 0.05; **0.001 < P < 0.01; ***P < 0.001
572 Kst*, Snn, and Fst were incorporated into DnaSP 6 for analysis. Population differentiation was evaluated through 1000
573 permutations of the raw data to assess deviations.

574 The high *Fst* values suggest restricted gene flow and limited genetic exchange, likely resulting from
575 geographic isolation, ecological barriers, or differences in vector distribution and host availability across these
576 regions (Lehmann et al. 1999). In contrast, the lowest differentiation was observed between OC and WSA-III
577 (*Fst* = 0.1899), WSA II & WSA III (0.19412) suggesting comparatively greater genetic similarity and higher
578 gene flow between these populations, possibly due to overlapping agro-climatic conditions or anthropogenic

movement of infected plant material (Gao et al. 2020). The interpretation of *Fst* values in this study is consistent with previous reports, where values greater than 0.25 were considered indicative of significant genetic differentiation and limited gene flow (Wu et al. 2018; Moradi et al. 2019). Similar patterns have been observed in other plant viruses; for example, potato virus X (PVX, *Potexvirus ecspotati*) populations showed high *Fst* values (~0.54) between South American and Japanese populations, reflecting restricted gene flow due to geographical separation (Hajizadeh and Sokhandan-Bashir 2017). Likewise, in potato virus Y (PVY, *Potyvirus yituberosi*) and iris yellow spot orthotospovirus (IYSV, *Orthotospovirus iridimaculaflavi*), higher *Fst* values >0.6 have been associated with infrequent genetic exchange and strong population subdivision (Hamid et al. 2019; Iftikhar et al. 2014). These studies collectively highlight the role of genetic differentiation in shaping the evolutionary dynamics of PRSV populations.

The significant *Kst** and Snn values observed in this study further reject the null hypothesis of no genetic differentiation between populations, confirming strong population subdivision. Overall, the high and significant *Snn* and *Kst** values, together with moderate to high *Fst* estimates, demonstrate strong genetic structuring and restricted gene flow among PRSV populations, highlighting the influence of geo-climatic factors on viral population dynamics and suggesting ongoing localized adaptation (García-Arenal et al. 2001; García-Arenal and Fraile 2013).

The estimated gene flow parameter (*Nm*), representing the effective number of migrants per generation, further supported these observations. The low *Nm* value (0.20) indicates limited gene flow among PRSV populations, which is consistent with the observed high levels of genetic differentiation. This aligns with the concept that reduced *Nm* values indicate higher levels of genetic isolation and differentiation due to limited gene flow (Wright 1931). Under such conditions, populations tend to accumulate distinct genetic variations over time, promoting divergence and the formation of genetically distinct subpopulations (Slatkin 1987). Similar patterns of limited gene flow have been reported in plant viruses, where spatial separation and host-specific adaptation restrict viral movement and contribute to population structuring (Ohshima et al. 2010).

Similar patterns of genetic structuring have been reported in other plant viruses. Studies on tobacco mosaic virus (TMV, *Tobamovirus tabaci*) have demonstrated that geographically or biologically isolated populations exhibit higher levels of genetic differentiation, driven by limited migration and adaptation to specific host plants or environmental conditions (García-Arenal et al. 2001). Likewise, in fig mosaic virus (FMV, *Emaravirus fici*), restricted gene flow across countries has been associated with the formation of genetically distinct populations, where genetic drift plays a key role in shaping population structure (Walia et al. 2014). The observed genetic differentiation and limited gene flow among PRSV populations in the present study similarly indicate that these isolates are evolving along largely independent trajectories. Such population isolation is likely to promote the accumulation of distinct genetic variations over time, potentially leading to the emergence of host-specific strains adapted to local environmental conditions. Collectively, these findings suggest that genetic isolation and restricted migration are major drivers of PRSV evolution, facilitating divergence and ecological specialization across different geo-climatic regions.

Population size change of PRSV: Nucleotide diversity (π) is a widely used measure of genetic variability in plant virus populations, reflecting the accumulation of mutations and evolutionary dynamics (García-Arenal et al. 2001). In the present study, nucleotide diversity (π) based on CP sequences was estimated to be 0.11675, indicating a high level of genetic variation within the PRSV population (Fig. 6). This relatively elevated π value

suggests the presence of substantial sequence divergence among isolates, which may be attributed to ongoing adaptive evolution, mutation pressure, and the influence of diverse ecological and host-related factors (Nei 1987). High nucleotide diversity is often associated with populations experiencing heterogeneous selection pressures and long-term evolutionary divergence. Similar findings have been observed in PVY, which exhibits moderate genetic diversity, with nucleotide diversity (π) ranging from 0.039 to 0.059 across populations, reflecting its ability to adapt to varying selective pressures (Chang et al. 2015). The higher nucleotide diversity observed in PRSV suggests a greater degree of genetic heterogeneity, which may enhance its evolutionary potential and adaptability across different geo-climatic regions.

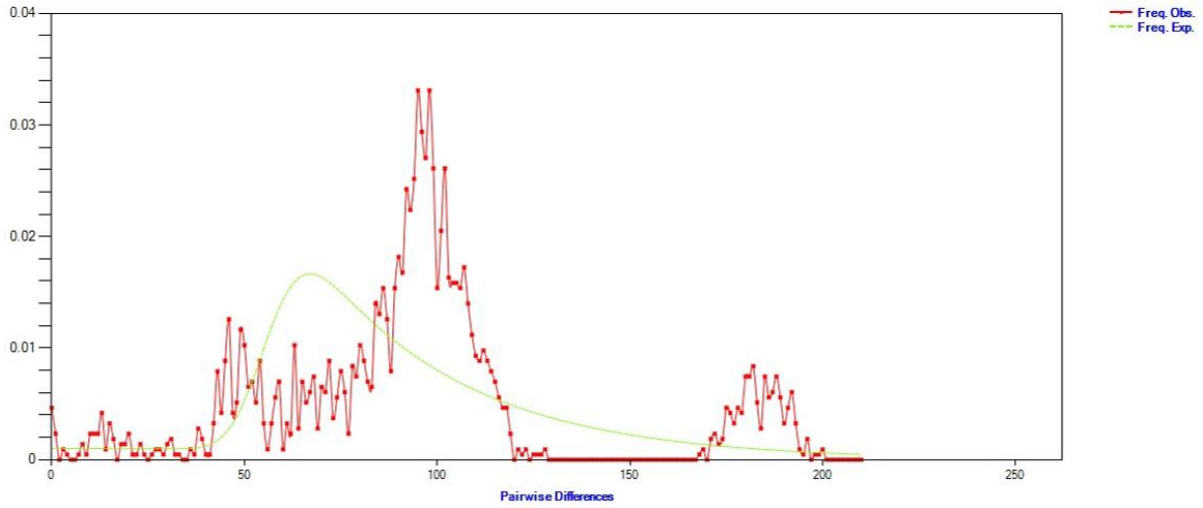


Figure 6: The demographic and genetic diversity dynamics of papaya ring spot virus (PRSV) over time, showing the Comparison of Observed and Expected Frequencies for genetic differences in the PRSV population.

Pairwise nucleotide difference distribution is a widely used approach to infer the demographic history, population expansion, and evolutionary dynamics of viral populations. This analysis helps distinguish between populations that have undergone recent expansion, bottlenecks, or long-term stability by evaluating the pattern and variability of genetic distances among isolates. In the present study, the analysis of pairwise nucleotide differences among PRSV isolates revealed substantial variability, as indicated by a high variance in pairwise differences (1552.8462) and a coefficient of variation (C.V.) of 0.4095. These values reflect pronounced heterogeneity in genetic distances across the dataset, suggesting the presence of multiple divergent lineages or subpopulations within the PRSV population. Such elevated variability is a characteristic feature of RNA virus populations, where high mutation rates, rapid replication, and adaptation to diverse host and environmental conditions contribute to increased genetic divergence (Domingo and Perales 2019). This observation is also consistent with the high nucleotide diversity ($\pi \approx 0.11$) reported in the present study, further supporting the presence of substantial genetic heterogeneity.

Demographic model fitting provided additional insights into the population history of PRSV. The estimated initial effective population size (θ_initial = 38.161) indicates relatively low genetic diversity at the early stage of the population, suggesting that the virus population may have undergone a bottleneck or founder event. Such events are typically associated with a reduction in population size, leading to decreased genetic variation before subsequent expansion (Novella et al. 1995; Li and Roossinck 2004). The scaled time parameter

(τ = 58.431) suggests that the population size change occurred approximately 58.431units of 2N generations ago (where N represents the effective population size), indicating that the demographic expansion is not a very recent event but rather occurred at an earlier stage in the evolutionary history of the population (Fig. 6). This temporal signal points to a historical expansion likely driven by ecological or environmental changes, such as host availability, vector dynamics, or agricultural practices (Tajima et al. 1989).

The model further predicts a large final effective population size (θ_final ≈ 1000.000), representing a substantial increase in population size following the demographic shift. This marked expansion suggests that the PRSV population has successfully established and diversified across different geo-climatic regions. The transition from low initial diversity to high final population size is indicative of a classic expansion scenario, where populations recover from bottlenecks and accumulate genetic variation over time.

Collectively, these results indicate that PRSV populations have undergone historical expansion following an initial phase of reduced diversity, leading to the maintenance of high levels of genetic variability. Such demographic expansion is often associated with increased genetic diversity, which enhances the evolutionary potential of viral populations by providing a broader pool of variants for selection to act upon. This, in turn, facilitates adaptation to new hosts, environments, and selective pressures. Similar patterns have been documented in other viral systems, where demographic expansion following bottlenecks contributes to increased genetic diversity and adaptive potential (Pybus et al. 2007; Susi et al. 2015).

Haplotype diversity and neutrality tests of PRSV populations: In haplotype diversity analysis, a total of 61 active haplotypes were identified among 65PRSV CP sequences of various geographical locations. Altogether,421 segregating sites with a haplotype diversity (*Hd*) of 0.996 were recorded, reflecting diverse genetic backgrounds among the populations.

In Tajima's test, a total of 58.431 mutations were observed with a 96.592average number of nt difference. Moreover, Theta (θ) values provided insights into the genetic mutation rate of 131.11239per sequence and 0.15970per site. Neutrality test statistics, including Tajima's D and Fu and Li's D and F, were estimated to evaluate the evolutionary forces and demographic history of PRSV populations across different geo-climatic regions based on the CP gene (Table 4). All estimated values were statistically non-significant (p> 0.10), indicating that deviations from neutrality are not strong enough to reject the null hypothesis of neutral evolution. However, the direction and magnitude of these statistics still provide useful insights into underlying population trends.

Table 4: Neutrality test statistics (Tajima's D, Fu and Li's D, and Fu and Li's F) for papaya ringspot virus populations across different geo-climatic regions based on coat protein (CP).

PRSV Population	Tajima's D	Fu & Li's D	Fu & Li's F
OC	-0.90455	-0.78013	-0.91719
WSA I	-0.09839	-0.18237	-0.18058
WSA II	-0.28733	-0.05709	-0.10634
MO	-0.69913	-0.58272	-0.69175
WSA III	-0.39396	-0.15321	-0.21919
HST I	-0.65551	-0.48815	-0.58309
HST II	-0.85708	-0.85708	-0.89263
WSA IV	-1.49910	-1.27377	-1.52102
WSA VI	0.23738	0.51601	0.53368

ns: Not significant, P > 0.10 (For the isolates of WSA-V region, the Tajima values are not generated because it has only 2 sequences; at least 4 should be there)

Most PRSV populations, including Oc, HST-I, HST-II, Mo, WSA-I, WSA-II, WSA-III, WSA-IV& WSA-VI, exhibited negative values of Tajima's D, Fu and Li's D, and Fu and Li's F (Table 4). However, the significant negative Tajima's D values strongly indicated that the population is experiencing bottlenecks, where the population is severely limited by some factors and leaving only a few sequences to dominate the gene pool (Tajima 1989). This result might indicate recent population expansion or ongoing purifying selection, which could suppress the accumulation of deleterious mutations, as commonly observed in expanding viral populations adapting to new hosts (Pybus et al. 2007). Among these, WSA-IV (-1.499, -1.273, -1.521) showed comparatively stronger negative values, suggesting a more pronounced signal of expansion or selective constraint in this population. These negative values of neutrality tests, though insignificant, align with patterns seen in viruses like IYSV, where population expansion and purifying selection were inferred to maintain viral fitness in diverse host and environmental conditions (Iftikhar et al. 2014). In contrast, the WSA VI population exhibited positive values (Tajima's D = 0.237, Fu and Li's D = 0.533, Fu and Li's F = 0.516), indicating an excess of intermediate-frequency alleles. This pattern may reflect balancing selection, population subdivision, or demographic stability, suggesting that these isolates may be evolving under different evolutionary pressures compared to other isolates.

The isolates exhibiting positive values of Tajima's D, Fu and Li's D, and F were predominantly associated with cucurbit-origin P pathotypes and displayed variation in CP length (876 bp). The excess of intermediate-frequency alleles in these isolates suggests balancing selection or population structuring, likely driven by host-specific evolutionary pressures. The observed variation in CP length further supports this interpretation, as insertions or deletions in the CP region may contribute to functional diversification, potentially affecting host interaction or vector transmission. Collectively, these findings indicate that cucurbit-originated PRSV isolates represent a distinct, host-adapted subpopulation undergoing structural and evolutionary divergence. Despite the lack of statistical significance, the overall trend of predominantly negative neutrality test values across most PRSV populations is consistent with other population genetic and demographic analyses in this study, including high nucleotide diversity ($\pi \approx 0.11$), elevated *Fst* values, and low gene flow ($Nm \approx 0.20$), all of which support a scenario of historical population expansion followed by genetic diversification. The demographic parameters (θ_initial, θ_final, and τ) and mismatch distribution analysis further corroborate this pattern, indicating that PRSV populations have likely undergone expansion after an initial phase of reduced genetic diversity.

CONCLUSION

The present study highlights the complex evolutionary dynamics of PRSV populations across diverse hosts and geo-climatic regions of India. The combined effects of host association, environmental heterogeneity, recombination, and gene-specific evolutionary constraints contribute significantly to PRSV genetic diversification and population structuring. High haplotype diversity coupled with limited gene flow and strong genetic differentiation indicates ongoing population divergence and local adaptation. These findings provide valuable insights into the adaptive evolution and epidemiology of PRSV, which may aid in the development of effective disease management and resistance breeding strategies.

This study elucidates the intricate evolutionary dynamics governing PRSV populations across heterogeneous hosts and geo-climatic zones in India. Our findings indicate that host-driven selection pressure, environmental heterogeneity, genetic recombination, and gene-specific evolutionary constraints collectively drive the genetic diversification and population structuring of the virus. Furthermore, high haplotype diversity paired with restricted gene flow and pronounced genetic differentiation underscores ongoing population divergence and localized adaptation. Concomitantly, these insights enhance our understanding of PRSV epidemiology and adaptive evolution, offering critical baseline data to inform targeted disease management protocols and marker-assisted resistance breeding programmes.

Acknowledgements

The authors acknowledge the financial support provided by the Department of Science and Technology (DST), Government of India, under the Early Career Research (ECR) scheme. The authors are thankful to all the farmers and landowners who helped during the sample collection and survey at different locations. The authors also thank the Indian Agricultural Research Institute (IARI), New Delhi, for providing the necessary facilities and infrastructure to conduct this research.

Conflicts of interest

Authors have no conflicts of interest.

REFERENCES

Acosta O, Mayo MA (1990) Accumulation of different types of raspberry ringspot nepovirus particle in infected Nicotiana protoplasts. J Gen Virol 71:713–717.

Akhter MS, Basavaraj YB, Pun KB, Sharma S, Jain RK (2013) Molecular characterization and genetic diversity of Papaya ringspot virus isolate from India. Indian J Virol 24:164–172.

Al-Ani RA, Adhab MA, Ali AA, Diwan SNH (2011) Zucchini yellow mosaic virus: characterization and management in Iraq. Int J Curr Res 3:220–224.

Ali A (2017). First complete genome sequence of PRSV-W isolated from a gourd in the United States. Genome Announc 5:e01434–16.

Andino R, Domingo E (2015) Viral quasispecies. Virology 479:46–51.

Attasart P, Kertbundit S, Panyim S, Juricek M (2002) Nucleotide sequence of a Thai isolate of Papaya ringspot virus type W. Acta Virol 46:241–246.

Bag S, Agarwal S, Jain RK (2007) Sequence diversity in the coat proteins of Papaya ringspot virus isolates originating from different locations in India. Indian Phytopathol 60:244–250.

Basavaraj YB, Kumar A, Jain RK, Kumar P, Parameswari B, Prakash J (2019) Molecular diversity of Papaya ringspot virus in India: genetic recombination and mutations between the isolates from different hosts and geo-climatic locations are role players in virus evolution. Indian Phytopathol 72:497–511.

Bateson MF, Henderson J, Chaleeprom W, Gibbs AJ, Dale JL (1994) Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). J Gen Virol 75:3547–3553.

Bateson MF, Lines RE, Revill P, Chaleeprom W, Ha CV, Gibbs AJ, Dale JL (2002) On the evolution and molecular epidemiology of the potyvirus Papaya ringspot virus. J Gen Virol 83:2575–2585.

Bau HJ, Cheng YH, Yu TA, Yang JS, Yeh SD (2003) Broad-spectrum resistance to different geographic strains of Papaya ringspot virus in coat protein gene transgenic papaya. Phytopathology 93:112–120.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. Ecol Lett 15:365–377.

Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. ISME J 12:173–184.

Bru D, Martin-Laurent F, Philippot L (2008) Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Appl Environ Microbiol 74:1660–1663.

Calvo M, Malinowski T, García JA (2014) Single amino acid changes in the 6K1-CI region can promote the alternative adaptation of Prunus- and Nicotiana-propagated Plum pox virus C isolates to either host. Mol Plant Microbe Interact 27:136–149.

Canto T, Aranda MA, Fereres A (2009) Climate change effects on physiology and population processes of hosts and vectors that influence the spread of hemipteran-borne plant viruses. Glob Change Biol 15:1884–1894.

Chang F, Zou WC, Gao FL, Shen JG, Zhan JS (2015) Comparative analysis of population genetic structure of Potato virus Y from different hosts. Yi Chuan 37:292–301.

Chare ER, Holmes EC (2006) A phylogenetic survey of recombination frequency in plant RNA viruses. Arch Virol 151:933–946.

Chen KC, Chiang CH, Raja JA, Liu FL, Tai CH, Yeh SD (2008) A single amino acid of NIa-Pro of Papaya ringspot virus determines host specificity for infection of papaya. Mol Plant Microbe Interact 21:1046–1057.

Chen YK, Shih PJ, Chao HY (2023) First report of mosaic disease associated with zucchini tigre mosaic virus in wax gourd (Benincasa hispida) in Taiwan. Plant Dis 107:591.

Chin M, Rojas Y, Moret J, Fermin G, Tennant P, Gonsalves D (2007) Varying genetic diversity of Papaya ringspot virus isolates from two time-separated outbreaks in Jamaica and Venezuela. Arch Virol 152:2101–2106.

Christopherson C, Sninsky JJ, Kwok S (1997) The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res 25:654–658.

Clark MF, Bar-Joseph M (1984) Enzyme immunosorbent assays in plant virology. Methods Virol 7:51–85.

Codoner FM, Elena SF (2008) The promiscuous evolutionary history of the family Bromoviridae. J Gen Virol 89:1739–1747.

Domingo E, Perales C (2019) Viral quasispecies. PLoS Genet 15:e1008271.

Elena SF (2017) Local adaptation of plant viruses: lessons from experimental evolution. Mol Ecol 26:1711–1719.

Escriu F, Fraile A, García-Arenal F (2007) Constraints to genetic exchange support gene coadaptation in a tripartite RNA virus. PLoS Pathog 3:e8.

Fermin GA, Castro L, Tennant P (2010) CP-transgenic and non-transgenic approaches for the control of papaya ringspot: current situation and challenges. Transgenic Plant J 4:1–15.

Fraile A, Malpica JM, Aranda MA, Rodríguez-Cerezo E, García-Arenal F (1996) Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant Nicotiana glauca. Virology 223:148–155.

Gagarinova A, Babu M, Strömvik M, Wang A (2008) Recombination analysis of Soybean mosaic virus sequences reveals evidence of RNA recombination between distinct pathotypes. Virol J 5:143.

Gao F, Chen C, Li B, Weng Q, Chen Q (2020) The gene flow direction of geographically distinct Phytophthora infestans populations in China corresponds with the route of seed potato exchange. Front Microbiol 11:1077.

García-Arenal F, Fraile A (2013) Trade-offs in host range evolution of plant viruses. Plant Pathol 62:2–9.

García-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157–186.

García-Arenal F, Fraile A (2008) Questions and concepts in plant virus evolution: a historical perspective. In: Plant Virus Evolution. Springer, Berlin, Heidelberg, pp 1–14.

Gibbs AJ, Harrison BD(1976) Plant virology: the principles. Edward Arnold, London. pp. 292.

Gogoi SH, Nath PD, Thakuria N, Gogoi S, Das B, Raj K (2019) Molecular detection and characterization of Papaya ring spot virus (PRSV) disease in Jorhat district of Assam, India. Int J Curr Microbiol Appl Sci 8:1564–1571.

Gong YN, Tang RQ, Zhang Y, Peng J, Xian O, Zhang ZH, Liu Y (2020) The NIa-protease protein encoded by the Pepper mottle virus is a pathogenicity determinant and releases DNA methylation of Nicotiana benthamiana. Front Microbiol 11:102.

Guo D, Rajamäki ML, Saarma M, Valkonen JPT (2001) Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system. J Gen Virol 82:935–939.

Hajizadeh M, Sokhandan-Bashir N (2017) Population genetic analysis of potato virus X based on the CP gene sequence. VirusDis 28:93–101.

Hamid A, Zhai Y, Ramesh SV, Pappu HR (2019) Complete genome characterization and population dynamics of potato virus Y-NTN strain from India. VirusDis 30:252–260.

Hema MV, Prasad DT (2004) Comparison of the coat protein of a South Indian strain of PRSV with other strains from different geographical locations. J Plant Pathol 86:35–42.

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics 155:2011–2014.

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Bio Evol 9:138–151.

Inoue-Nagata AK, Franco CM, Martin DP, Rezende JAM, Ferreira GB, Dutra LS, Nagata T (2007) Genome analysis of a severe and a mild isolate of Papaya ringspot virus-type W found in Brazil. Virus Genes 35:881–882.

Jain RK, Sharma J, Sivakumar AS, Sharma PK, Byadgi AS, Verma AK, Varma A (2004) Variability in the coat protein gene of Papaya ringspot virus isolates from multiple locations in India. Arch Virol 149:2435–2442.

Jain RK, Pappu HR, Pappu SS, Varma A, Ram RD. 1998. Molecular characterisation of papaya ringspot potyvirus isolates from India. Ann Appl Biol 132(3):413-25.

Klein D, Leutenegger CM, Bahula C, Gold P, Hofmann-Lehmann R, Salmons B, Lutz H, Gunzburg WH (2001) Influence of preassay and sequence variations on viral load determination by a multiplex real-time reverse transcriptase-polymerase chain reaction for feline immunodeficiency virus. J Acquir Immune Defic Syndr 26:8–20.

Kumar S, Gautam KK, RajSK (2015) Sequence analysis and genetic diversity of five new Indian isolates of cucumber mosaic virus. Acta Virol 59:398–404.

Kung YJ, You BJ, Raja JA, Chen KC, Huang CH, Bau HJ, Yeh SD (2015) Nucleotide sequence-homology-independent breakdown of transgenic resistance by more virulent virus strains and a potential solution. Sci Rep 5:9804.

Kwok S, Kellogg DE, McKinney N, Spasic D, Godal L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res 18:999–1005.

Lai MM (1992) RNA recombination in animal and plant viruses. Microbiol Rev 56:61–79.

Lefeuvre P, Martin DP, Harkins G, Lemey P, Gray AJ, Meredith S, Heydarnejad J (2010) The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathog 6:e1001164.

Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A (2019) Evolution and ecology of plant viruses. Nat Rev Microbiol 17:632–644.

Lehmann T, Hawley WA, Grebert H, Danga M, Atieli F, Collins FH (1999) The Rift Valley complex as a barrier to gene flow for Anopheles gambiae in Kenya. J Hered 90:613–621.

Li H, Roossinck MJ (2004) Genetic bottlenecks reduce population variation in an experimental RNA virus population. J Virol 78:10582–10587.

Lu YW, Shen WT, Zhou P, Tang QJ, Niu YM, Peng M, Xiong Z (2008) Complete genomic sequence of a Papaya ringspot virus isolate from Hainan Island, China. Arch Virol 153:991–993.

Mangrauthia SK, Parameswari B, Praveen S, Jain RK (2009) Comparative genomics of Papaya ringspot virus pathotypes P and W from India. Arch Virol 154:727–730.

Mangrauthia SK, Parameswari B, Jain RK, Praveen S (2008) Role of genetic recombination in the molecular architecture of Papaya ringspot virus. Biochem Genet 46:835–846.

Maoka T, Hataya T (2005) The complete nucleotide sequence and biotype variability of papaya leaf distortion mosaic virus. Phytopathology 95:128–135.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003.

McLeish MJ, Fraile A, García-Arenal F (2019) Evolution of plant–virus interactions: host range and virus emergence. Curr Opin Virol 34:50–55.

Medina-Salguero AX, Cornejo-Franco JF, Grinstead S, Mowery J, Mollov D, Quito-Avila DF (2021) Genetic characterization of a mild isolate of papaya ringspot virus type-P (PRSV-P) and assessment of its cross-protection potential under greenhouse and field conditions. PLoS One 16:e0241652.
Mishra R, Patil S, Patil A, Patil BL (2019) Sequence diversity studies of papaya ringspot virus isolates in South India reveal higher variability and recombination in the 5′-terminal gene sequences. VirusDis 30:261–268.
Moradi Z, Mehrvar M, Nazifi E (2019) Population genetic analysis of Zucchini yellow mosaic virus based on the CI gene sequence. J Cell Mol Res 10:76–89.
Moya A, Rodríguez-Cerezo E, García-Arenal F (1993) Genetic structure of natural populations of the plant RNA virus tobacco mild green mosaic virus. Mol Bio Evol 10:449–456.
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.
Nivetha T, Renukadevi P, Gayathri M, Indra N, Vidhya D, Rajagopal B (2025) Global scenario of viruses infecting papaya. 3 Biotech 15:208.
Noa-Carrazana JC, González-de-León D, Ruiz-Castro BS, Piñero D, Silva-Rosales L (2006) Distribution of Papaya ringspot virus and Papaya mosaic virus in papaya plants (Carica papaya) in Mexico. Plant Dis 90:1004–1011.
Novella IS, Elena SF, Moya A, Domingo E, Holland JJ (1995) Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. J Virol 69:2869–2872.
Ohshima K, Akaishi S, Kajiyama H, Koga R, Gibbs AJ (2010) Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. J Gen Virol 91:788–801.
Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC (2007) Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol Bio Evol 24:845–852.
Ortiz-Rojas LY, Chaves-Bedoya G (2017) Molecular characterization of two Papaya ringspot virus isolates that cause devastating symptoms in Norte de Santander, Colombia. Eur J Plant Pathol 148:883–894.
Parameswari B, Mangrauthia SK, Praveen S, Jain RK (2007) Complete genome sequence of an isolate of Papaya ringspot virus from India. Arch Virol 152:843–845.
Quiot-Douine L, Lecoq H, Quiot JB, Pitrat M, Labonne G (1990) Serological and biological variability of virus isolates related to strains of papaya ringspot virus. Phytopathology 80:256–263.
Revers F, Le Gall O, Candresse T, Le Romancer M, Dunez J (1996) Frequent occurrence of recombinant potyvirus isolates. J Gen Virol 77:1953–1965.
Rodamilans B, Shan H, Pasin F, García JA (2018) Plant viral proteases: beyond the role of peptide cutters. Front Plant Sci 9:666.
Rodríguez-Cerezo E, Elena SF, Moya A, García-Arenal F (1991) High genetic stability in natural populations of the plant RNA virus tobacco mild green mosaic virus. J Mol Evol 32:328–332.
Romay G, Lecoq H, Desbiez C (2014) Zucchini tigré mosaic virus is a distinct potyvirus in the papaya ringspot virus cluster: molecular and biological insights. Arch Virol 159:277–289.
Roossinck MJ (1997) Mechanisms of plant virus evolution. Annu Rev Phytopathol 35:191–209.
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Bio Evol 34:3299–3302.
Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.
Saeed F, Arshad MU, Pasha I, Naz R, Batool R, Khan AA, Shafique B (2014) Nutritional and phyto-therapeutic potential of papaya (Carica papaya Linn.): an overview. Int J Food Prop 17:1637–1653.
Saleem A, Ali Z, Yeh SD, Saeed W, Binat Imdad A, Akbar MF, Naseem S (2021) Genetic variability and evolutionary dynamics of atypical Papaya ringspot virus infecting papaya. PLoS One 16:e0258298.
Scheffers BR, De Meester L, Bridge TCL, Hoffmann AA, Pandolfi JM, Corlett RT, Watson JEM (2016) The broad footprint of climate change from genes to biomes to people. Science 354:aaf7671.
Seo J, Ohshima K, Lee H, Son M, Choi H et al (2009) Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. Virology 393:91–103.

Silva-Rosales L, Becerra-Leor N, Ruiz-Castro S, Téliz-Ortiz D, Noa-Carrazana JC (2000) Coat protein sequence comparisons of three Mexican isolates of papaya ringspot virus with other geographical isolates reveal a close relationship to American and Australian isolates. Arch Virol 145:835–843.

Singh S, Jangre A, Kumar P (2017) Studies on the molecular variability in Indian isolates of Papaya ringspot virus. Virol Res J 1:10–16.

Skotnicki ML, Mackenzie AM, Gibbs AJ (1996) Genetic variation in populations of kennedya yellow mosaic tymovirus. Arch Virol 141:99–110.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787–792.

Susi H, Barrès B, Vale PF, Laine AL (2015) Co-infection alters population dynamics of infectious disease. Nat Commun 6:5975.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tamura K, Stecher G, & Kumar S (2021). MEGA11: molecular evolutionary genetics analysis version 11. Mol Bio Evol 38(7):3022-3027.

Tripathi S, Suzuki JY, Ferreira SA, Gonsalves D (2008) Papaya ringspot virus-P: characteristics, pathogenicity, sequence variability and control. Mol Plant Pathol 9:269–280.

Verma RK, Mishra R, Gaur RK (2016) Genetic variability of the replicase (NIb) gene of papaya ringspot virus in northern India indicates common ancestry with isolates from China and Taiwan. J Plant Pathol 98:105–110.

Walia JJ, Willemsen A, Elçi E, Çağlayan K, Falk BW, Rubio L (2014) Genetic variation and possible mechanisms driving the evolution of worldwide fig mosaic virus isolates. Phytopathology 104:108–114.

Wang CH, Yeh SD (1997) Divergence and conservation of the genomic RNAs of Taiwan and Hawaii strains of papaya ringspot potyvirus. Arch Virol 142:271–285.

Wei JY, Liu DB, Li XY, Zhou P (2007) Variation in the coat protein gene of Papaya ringspot virus isolates from multiple locations of China. J Integr Plant Biol 49:1062–1069.

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159.

Wu Z, Mo C, Zhang S, Li H (2018) Characterization of papaya ringspot virus isolates infecting transgenic papaya 'Huanong No. 1' in South China. Sci Rep 8:8206.

Yeh SD, Gonsalves D (1994) Practices and perspective of control of papaya ringspot virus by cross-protection. In: Advances in disease vector research. Springer, New York, pp 237–257.

Yeh SD, Gonsalves D, Provvidenti R (1984) Comparative studies on host range and serology of papaya ringspot virus and watermelon mosaic virus 1. Plant Dis 68:100–104.

Yeh SD, Jan FJ, Chiang CH, Doong TJ, Chen MC, Chung PH, Bau HJ (1992) Complete nucleotide sequence and genetic organization of papaya ringspot virus RNA. J Gen Virol 73:2531–2541.

Figure Legends

Figure 1. Various symptoms of Papaya ringspot virus (PRSV) on cucurbits and papaya were observed under field conditions. A. Leaf distortion and Vein banding; B. Vein clearing; C. Vein banding and Inter-veinal chlorosis; D. Leaf distortion and blistering; E. Leaf distortion and vein clearing; F. Leaf distortion, blistering and chlorosis; G. Filiformy and shoe stringing; H. Ring spots on leaf; I. Ring spots on fruits.

Figure 2: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the Hc-Pro coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

Figure 3: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the NIa-Pro coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the

Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

Figure 4: Evolutionary relationship trees were constructed using the Neighbour-joining (NJ) method (with bootstrap consensus 1000 replicates) based on deduced amino acid sequences of the CP coding region of PRSV pathotypes P and W, illustrating the evolutionary clustering of Indian isolates with globally distributed isolates according to conserved amino acid residues. The evolutionary distances were calculated using the Poisson correction method and are represented as the number of amino acid substitutions per site. The evolutionary analyses were performed in MEGA11.

Figure 5. Representative recombination events detected in PRSV isolates using RDP4. (A) Representative recombination event from the CP gene. (B) Representative recombination event from the NIa-Pro gene. Shaded regions indicate inferred recombinant fragments and vertical lines represent predicted breakpoint positions.

Figure 6: The demographic and genetic diversity dynamics of Papaya ring spot virus (PRSV) over time, showing the Comparison of Observed and Expected Frequencies for genetic differences in the PRSV population.

Supplementary Material

Figure S1. Sequence alignment of PRSV coat protein (CP) coding region showing the variation in EK repeat region.

Table S1: Details of primers designed for PCR amplification of HC-pro, NIa-pro, & CP coding regions of PRSV.

Table S2: Colour-coded pairwise identity matrix generated from the HC-Pro coding region of 51 Papaya ringspot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.

Table S3: Colour-coded pairwise identity matrix generated from NIa-Pro coding region of 55 Papaya ringspot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.

Table S4: Colour-coded pairwise identity matrix generated from CP coding region of 65 Papaya ringspot virus isolates. Each colored cell represents a percentage identity score between two sequences (one indicated horizontally to the left and the other vertically at the top). A colored key indicates the correspondence between pairwise identities and the colors displayed in the matrix.