VirusDis.
https://doi.org/10.1007/s13337-019-00512-x

ORIGINAL ARTICLE

Sequence diversity studies of papaya ringspot virus isolates in South India reveal higher variability and recombination in the 5′-terminal gene sequences

Ritesh Mishra[1] · Sharana Patil[2] · Ayyanagouda Patil[2] · Basavaprabhu L. Patil[1,3]

Received: 15 October 2018 / Accepted: 18 January 2019

Abstract Papaya ringspot virus (PRSV) is one of the most devastating viruses which causes huge damage to papaya plantations across the globe. PRSV is a positive sense RNA virus encoding for a polyprotein that is processed into ten proteins. In this study for the first time we analyzed the variability for 15 PRSV isolates from a selected geographical region of a South Indian state Karnataka, which is under intensive papaya cultivation. Variability studies were done for two genes at the 5′ end of the viral genome, namely *P1* and helper component proteinase (*Hc-Pro*) and towards the 3′ end, a 788 nt overlapping region of nuclear inclusion B (*NIb*, 692 nt) and of capsid protein (*CP*, 96 nt), referred as *NIb-CP*. Our studies indicate that the *P1* is most variable region with a wider range of sequence identity, followed by *Hc-Pro*, while the 788 nt of *NIb-CP* was most conserved. *P1* also showed maximum recombination events followed by *Hc-Pro*, whereas *NIb-CP* did not show any recombination. Further, the pattern and number of phylogenetic clusters was variable for each of the three genomic regions of PRSV isolates. Estimation of selection pressure for all the three PRSV genomic regions indicated negative and purifying selection.

Keywords Variability · Recombination · *Papaya ringspot virus* · Potyvirus

Introduction

Papaya (*Carica papaya* L.), a rich source of vitamin A and C, is one of the most widely cultivated fruit crops in the tropics and sub-tropics, for both culinary and industrial uses. Although papaya is native to Mexico, India is the largest producer of papaya accounting for nearly 45% of global papaya production [17]. Andhra Pradesh, Gujarat, Karnataka and Madhya Pradesh are the states that lead in papaya production in India (http://nhb.gov.in/Horticulture%20Crops/Papaya/Papaya1.htm). A number of pests and diseases hamper papaya production across the globe, of which viral diseases such as ringspot and leafcurl are more predominant [17]. Globally papaya ringspot virus (PRSV) is considered to be the most devastating virus that significantly reduces papaya yield and this virus is also widespread in India. It is non-persistently transmitted by various aphid species [23]. PRSV belongs to the genus *Potyvirus* and family *Potyviridae* [28]. It consists of a single stranded RNA genome ranging in size from 10,317 to 10,349 nt, with a virus encoded polyprotein at the 5′ end and a poly-A tail at the 3′ terminus [28, 31]. Subsequently, the single polyprotein of size 350 kDa is proteolytically cleaved into ten final products: P1, HC-Pro, P3, P3N-PIPO, CI, K6, VPG, NIaPro, NIb, and CP [3, 28, 31]. The isolates of PRSV are classified into two groups, those that infect papaya are PRSV-P type and the cucurbit infecting isolates are termed as PRSV-W [29]. However under experimental conditions PRSV-P isolates infect plants of the families Caricaceae, Cucurbitaceae and Chenopodiaceae, whereas

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13337-019-00512-x) contains supplementary material, which is available to authorized users.

✉ Basavaprabhu L. Patil
basavaprabhu.patil@icar.gov.in; blpatil2046@gmail.com

[1] ICAR-National Research Centre on Plant Biotechnology, IARI Campus, Pusa, New Delhi 110012, India

[2] University of Agricultural Sciences, Raichur, Karnataka, India

[3] ICAR-Indian Institute of Horticultural Research, Bengaluru 560089, India

Published online: 28 January 2019

PRSV-W isolates infect plants of the families Cucurbitaceae and Chenopodiaceae [29].

PRSV has a global presence and a recent report suggests that PRSV most likely originated in India about 2250 years ago, which later expanded to Thailand and China about 600 years ago and subsequently about 200 years ago it reached the Americas [19]. However, it is still not clear if PRSV-P emerged from PRSV-W, since papaya was introduced to India about 400 years ago [19]. Hence more sequences have to be analysed to make any concrete conclusion on the origin and evolution of both the PRSV strains. It is also possible that the PRSV-W has evolved from PRSV-P, as the appearance of PRSV-P in Australia was an outcome of mutations in the established PRSV-W that provided the ability to infect papaya [8].

To develop any virus control strategies, it is important to understand the population structure of the viruses, for which sequence variability study is a prerequisite [7, 26]. Hitherto, most of the variability studies done in the case of PRSV isolates were based on the sequence information of coat protein [10, 24, 30] and not much attention was given to the sequences of other important genes/ORFs, such as *P1*, *Hc-Pro* and *NIb*. Thus this study investigated the current population structure of PRSV isolates in north Karnataka (India); we have sequenced major parts of the three important genes, namely, *P1*, *Hc-Pro* and *NIb*, of 15 different PRSV isolates. We analysed these sequences to determine the extent of relative variability among the three genomic regions of 15 PRSV isolates from Karnataka state. We also subject these sequences to recombination and phylogenetic analyses.

Materials and methods

Papaya leaf sample collection

Papaya orchards under cultivation of the papaya variety Taiwan Redlady 786 across North Karnataka (India) were surveyed for viral diseases. Twenty different leaf samples showing characteristic symptoms of PRSV infection, such as ring spots on the papaya fruits and mosaic and shoe string symptoms on the leaves were collected (Suppl. Figure 1). Of these 20 samples, 15 were positive for PRSV by PCR using universal primers for potyvirids [6] and other 5 were infected with leaf curl causing geminiviruses. Of these 15 PRSV infected samples, 4 were from Santhekellur, Kesarhatti and Neeralkera of Raichur district, 3 from Hulihyder in Koppal district, 1 from Gokak in Belagavi (formerly Belgaum) district, 3 from Mudhol and Kaladgi of Bagalkot district, 3 from Muddebihal, Talikote and Bijapur of Vijayapura district (formerly Bijapur) and one from Bheemarayangudi of Yadgir district (Suppl. Figure 2). All these leaf samples were immediately brought to New Delhi and snap frozen in liquid nitrogen and stored in − 80 °C for further use.

RNA extraction and RT-PCR for amplification of PRSV sequences

Total RNA was extracted from symptomatic papaya leaf using Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) following the manufacturer's instructions. The 2 µg of total RNA was reverse transcribed using a cDNA synthesis kit (Verso cDNA synthesis kit, Thermo Fisher Scientific, India) with Oligo (dT) as primer, following the manufacturer's instructions in a Bio-Rad S1000 Thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). The resulting cDNA was diluted 1:10 and 1 µL of the diluted cDNA was used for PCR amplification using a Taq polymerase with proof reading activity from the same kit. The PCR amplified segments were 1011 nt of C-terminal of *P1* (719–1729 nt), 656 nt C-terminal of *Hc-Pro* (2384–3039 nt) and 788 nt of overlapping region of *NIb-CP* (8558–9346 nt) of the 15 PRSV isolates, using specific primer pairs designed from conserved regions of Indian PRSV isolates (Suppl. Figure 3 and Supplementary Table 1).

Cloning and sequencing of PRSV genomic fragments

The RT-PCR amplified products were resolved on 1% TAE (Tris-acetate–EDTA) agarose gels, stained with ethidium bromide and visualized under ultraviolet (UV) light. These DNA fragments were excised from agarose gel and eluted by using a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and were then ligated in the pGEM-T Easy vector (Promega, Madison, WI, USA) as per the manufacturer's instructions. Subsequently the ligations were transformed in DH5α strain of *Escherichia coli*. Recombinant clones were identified by colony PCR, as well as by restriction digestion of the recombinant plasmids, isolated using SureSpin Plasmid Mini Kit (Nucleo-pore, India) and subsequently sequenced (Eurofins Genomics, Bengaluru, India).

Sequence analysis of the three fragments of PRSV isolates

The raw sequences were edited to remove vector sequences and later subjected to homology search by BLASTN (nucleotide blast) sequence analysis tool of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Assembled sequences were submitted to NCBI GenBank and accession numbers were obtained for all the three genomic regions of the 15 PRSV

Table 1 Summary of unique recombination events identified by Recombination Detection Program v.4.16 (RDP4) for the partial *P1* (719–1729 nt) and *Hc-Pro* (2384–3039 nt) gene sequences of *papaya ringspot virus* (PRSV) isolates from India

PRSV isolate (accession no.)	Recombination break point (nt) (relative genomic position)	Parent isolate		*p*-values					
		Major	Minor	RDP	GENECONV	BOOTSCAN	MAXI CHISQUARE	CHIMAERA	SISCAN
P1 (719–1729 nt = 1011 nt)									
Santhekellur S-15 (MF356483)	1078–1515 (359–796)	PRSV DEL	PRSV HYD	NS	**1.632×10^{-09}**	**1.475×10^{-06}**	**1.203×10^{-14}**	**1.179×10^{-14}**	**4.027×10^{-18}**
Kesarahatti S-18 (MF356484)	1519–1672 (800–953)	Bheemarayanagudi K-30	Gokak K-4	**1.957×10^{-12}**	**1.048×10^{-10}**	1.372×10^{-02}	**3.508×10^{-14}**	**9.160×10^{-08}**	**4.942×10^{-10}**
Neeralkera S-21 (MF356485)	844–926 (125–207)	Bheemarayanagudi K-30	Kaladgi K-13	NS	3.304×10^{-02}	**1.506×10^{-08}**	**2.881×10^{-07}**	NS	**4.677×10^{-07}**
Mudhol K-8 (MF356490)	1078–1515 (359–796)	PRSV DEL	PRSV HYD	NS	**1.632×10^{-09}**	**1.475×10^{-06}**	**1.187×10^{-14}**	**7.753×10^{-15}**	**4.027×10^{-18}**
Kaladgi K-13 (MF356492)	1519–1672 (800–953)	Bheemarayanagudi K-30	Gokak K-4	**1.957×10^{-12}**	**1.048×10^{-10}**	1.372×10^{-02}	**3.508×10^{-14}**	**9.160×10^{-08}**	**4.942×10^{-10}**
Muddebihal K-15 (MF356493)	844–926 (125–207)	Unknown	Kaladgi K-13	NS	3.304×10^{-02}	**1.506×10^{-08}**	**2.881×10^{-07}**	NS	**4.677×10^{-07}**
	1577–1670 (858–951)	Bheemarayanagudi K-30	Kaladgi K-13	**3.162×10^{-09}**	**1.395×10^{-08}**	2.778×10^{-04}	**3.682×10^{-07}**	**3.781×10^{-06}**	**2.511×10^{-09}**
Talikoti K-17 (MF356494)	844–926 (125–207)	Unknown	Kaladgi K-13	NS	3.304×10^{-02}	**1.506×10^{-08}**	**2.881×10^{-07}**	NS	**4.677×10^{-07}**
Bijapur K-20 (MF356495)	844–926 (125–207)	Unknown	Kaladgi K-13	NS	3.304×10^{-02}	**1.506×10^{-08}**	**2.881×10^{-07}**	NS	**4.677×10^{-07}**
Hc-Pro (2384–3039 nt = 656 nt)									
Kesarahatti S-18 (MF356469)	2515–2693 (131–309)	Talikoti K-17	Unknown	3.272×10^{-05}	**3.665×10^{-06}**	1.125×10^{-05}	**4.684×10^{-10}**	NS	**1.907×10^{-11}**
Hulihyder S-24 (MF356473)	2479–2728 (95–344)	Gokak K-4	Unknown	1.727×10^{-02}	7.149×10^{-04}	NS	**5.311×10^{-07}**	**5.230×10^{-08}**	**5.543×10^{-07}**
Muddebihal K-15 (MF356478)	2404–2614 (20–230)	Talikoti K-17	Unknown	7.911×10^{-05}	1.044×10^{-04}	NS	**5.166×10^{-07}**	**7.879×10^{-07}**	**1.521×10^{-07}**

The methods used to infer recombination breakpoints were RDP, GENECONV, BOOTSCAN, MAXIMUM CHISQUARE, CHIMAERA, SISCAN and 3 SEQ (the *p* values obtained by 3 SEQ were all non-significant and hence not shown in this table). The *p* values of all the methods are given; non-significant *p* values are indicated as NS

The *p*-values that are significant are shown in bold fonts

isolates (Supplementary Table 2). Subsequently the sequence identity matrixes were calculated for both nucleotide and amino acid sequences of the partial *P1*, *Hc-Pro* and the *NIb-CP* of the 15 PRSV isolates along with other PRSV sequences from India using Bioedit [9].

Recombination breakpoints among the *P1*, *Hc-Pro* and *NIb-CP* segments of the above PRSV isolates were analyzed using the recombination detection program package RDP4 [15], which uses several recombination detection methods, such as RDP, GENECONV, CHIMAERA, MAXIMUM CHISQUARE, BOOTSCAN, SISTER SCAN and 3SEQ. The recombinations were considered to be significant only when the *p* values obtained were less than 1×10^{-6} for at least three methods.

Phylogenetic and preliminary population genetic analysis of PRSV isolates

Equal lengths of nucleotide sequences of the partial *P1*, *Hc-Pro* and *NIb-CP* segments of 15 PRSV isolates from this study, along with published sequences of Indian PRSV isolates from the NCBI GenBank were aligned using MUSCLE and the phylogenetic trees were constructed by the maximum-likelihood, applying the JTT matrix and pairwise gap deletion options embedded within the MEGA6 [25].

Further for preliminary understanding of the population genetics, the degree of selective constraints at the amino acid level was estimated with MEGA6 by analyzing separately the rate of nonsynonymous (dN) and synonymous (dS) substitutions by the Pamilo–Bianchi–Li method [21]. The program DNASP 5.10 was used to evaluate the importance of natural selection by testing the mutation neutrality hypothesis with several statistics, such as Tajima's D, based on the differences between the number of segregating sites and the average number of nucleotide differences; Fu and Li's D test, based on the differences between the number of singletons (mutations appearing only once among the sequences) and the total number of mutations; and Fu and Li's F test, based on the differences between the number of singletons and the average number of nucleotide differences between every pair of sequences [13].

Results

Analysis of PRSV sequence identity and identification of amino acid motifs

Estimation of sequence identity among the three partial gene sequences (*P1*, *Hc-Pro* and *NIb-CP*) of the 15 PRSV isolates from this study along with published sequences from India revealed significant sequence variability (Suppl. Tables 3–5). The nucleotide sequence identities were in the range of 81.3–99.9% for *P1*, 87.5–100% for *Hc-Pro* and 94–99.6% for *NIb-CP*. While, the amino acid sequence identities were in the range of 83–99.4% for P1, 83.3–100% for Hc-Pro and 84–100% for NIb-CP (Suppl. Tables 3–5).

Further, the important amino acid motifs of P1 such as VELI (220–223 aa), GSSG (285–288 aa) and RG (310–311 aa), which may have a role in the protease activity, were found to be conserved in all the 15 PRSV isolates of this study [22] (Suppl. Figure 4). Further the Q/Y motif (334–335 aa) which serves as the internal cleavage point for the protease to cleave P1 and Hc-Pro was also identified in all the above isolates [22]. Similarly analysis of the Hc-Pro amino acid sequences of all the 15 PRSV isolates revealed conservation in the motifs that are involved in aphid transmission (PTK motif at 88–90 aa position), protease activity (NIFLAML at 126–132 aa and AELPRILVDH at 181–190 aa) and cell–cell long distance movement of the virions (LKANTV motif at 211–216 aa position) (Suppl. Figure 4) [1]. Analysis of the NIb-CP sequences of the above PRSV isolates also showed the conservation of the important motifs involved in RNA-dependent RNA polymerase activity such as GNNSGQPSTVVDNTLMV (1–17 aa) [24] and NGDDL (44–48)-AA_{34}-K [1]. The cleavage site at the Q/S (231–232 aa) motif present at the intersection of NIb and CP was also conserved [1]. Additionally, the DAG (238–240 aa) motif involved in aphid transmission was present in all the 15 PRSV isolates (Suppl. Figure 4) [1].

Recombination analysis in PRSV isolates

Significant numbers of recombination break points were detected for *P1*, followed by *Hc-Pro* and none in *NIb-CP*, when all the sequences of the 15 PRSV isolates from this study and other selected PRSV isolates from India were subjected to RDP4 analysis (Table 1). Two PRSV isolates from this study (Kesarahatti S-18 and Muddebihal K-15) showed recombination break points in both *P1* and *Hc-Pro*. Another two isolates (S-15 and K-18) turned out be recombination products of Delhi and Hyderabad isolates at *P1*, while others were recombinations among the isolates from Karnataka itself (Table 1). The major parent could not be determined for recombinations in *P1* of three isolates from Vijayapura (formerly Bijapur) district (K-15, K-17 and K-20), while the minor parent was unknown for all the three isolates (S-18, S-24 and K-15) for which recombinations were detected at *Hc-Pro* (Table 1). No significant recombination breakpoints were observed for any of the 15 isolates of *NIb-CP* sequences.

Phylogenetic and preliminary population genetic analysis of PRSV isolates

Phylogenetic analysis for the three gene sequences of all the 15 PRSV isolates from this study, along with other selected PRSV isolates from India and with *Cowpea aphid-borne mosaic virus* (CABMV) as an out-group resulted in varying clustering pattern. There were six, seven and five clusters formed when the sequences of *P1*, *Hc-Pro* and *NIb-CP* respectively were subjected to phylogenetic analysis (Fig. 1, Supplementary Table 6). There were 1–2 subclusters within the first two clusters and in the case of *P1* phylogenetic tree the PRSV isolate Mudhol (K-8) formed a separate sub-cluster within the second cluster and was the sole member of this sub-cluster. Similarly when the *Hc-Pro* sequences were analysed following three isolates, Muddebihal (K-15), Hulihyder (S-22) and Neeralkera (S-21) formed their own distinct clusters. In the case of *NIb-*

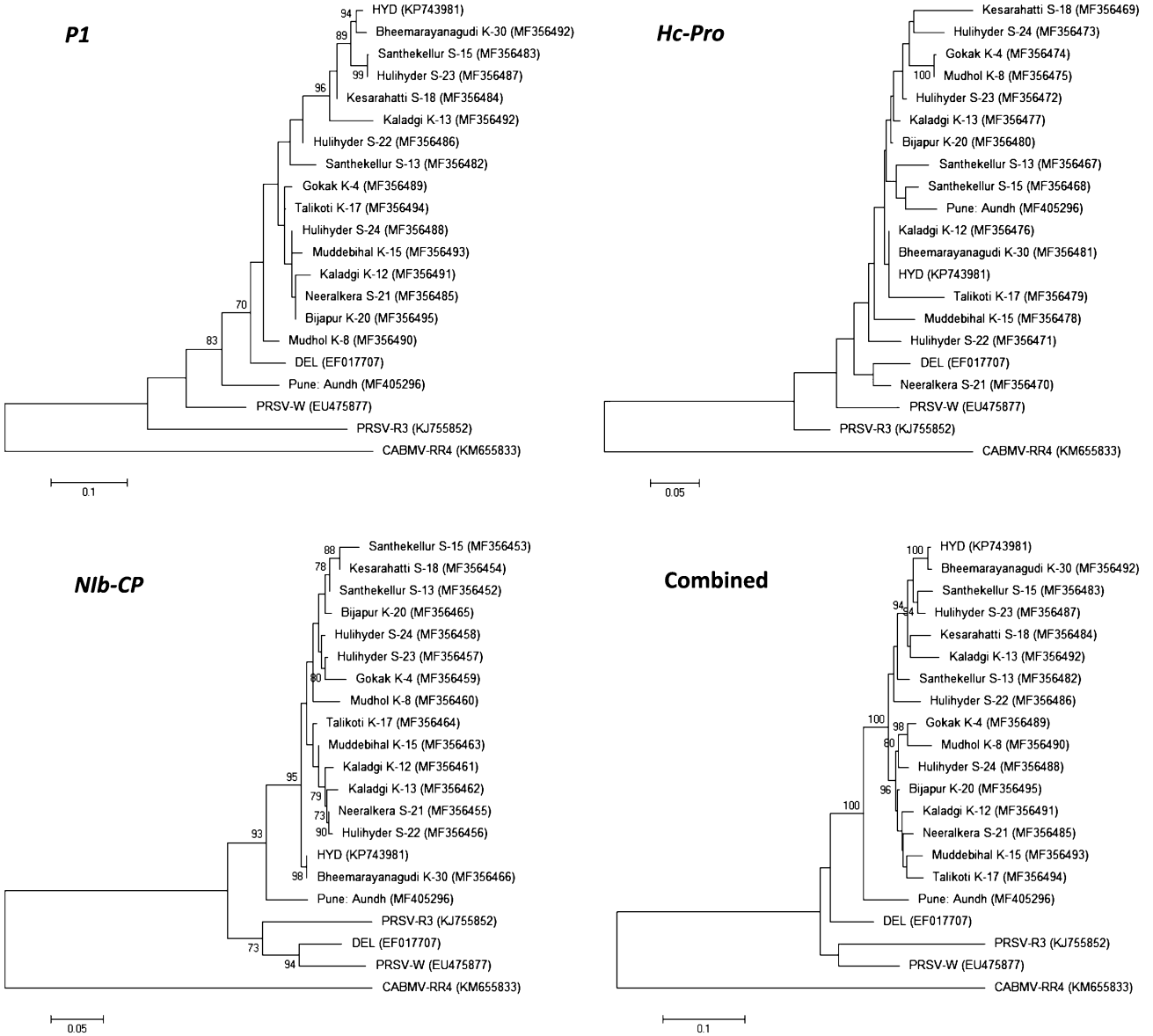

Fig. 1 Phylogenetic analysis of the nucleotide sequences of partial *P1* (719–1729 nt), *Hc-Pro* (2384-3039 nt), *NIb-CP* (8558–9346 nt) and the combination of all the three sequences, of *papaya ringspot virus* (PRSV) isolates from India, along with the corresponding sequence of *cowpea aphid-borne mosaic virus* (CABMV) as an out-group. The GenBank accession number for each sequence is given in parentheses. Phylogenetic trees were constructed using ''MUSCLE'', and trees were drawn by the maximum-likelihood, applying the JTT matrix and pairwise gap deletion options implemented in MEGA6, using 1000 bootstrap replicates. Bootstrap values greater than 70% are given at each node. The scale bar represents 0.1 and 0.05 substitutions per nucleotide position for *P1* and *Hc-Pro*/*NIb-CP* sequences, respectively

CP sequences, all the isolates from this study clustered together, except for one single isolate Bheemarayanagudi (K-30) which clustered with the Hyderabad isolate and it is also noteworthy to note that this isolate always clustered along with Hyderabad isolate even when the *P1* and *Hc-Pro* sequences were analysed. All the 15 isolates clustered together to form two subclusters when all the three gene sequences were combined and subjected to phylogenetic analysis along with other selected PRSV sequences from India.

The population genetic parameters and neutrality tests indicated a higher genetic variation for *P1* sequences with a nucleotide diversity (π) of 0.142, when compared to the sequences of *Hc-Pro* ($\pi = 0.106$) and *NIb-CP* ($\pi = 0.104$) (Table 2). Both the dN and dS values were higher for *P1* when compared to *Hc-Pro* and *NIb-CP* sequences, suggesting a positive selection for *P1* at nucleotide level. However, the dN/dS values for *P1*, *Hc-Pro* and *NIb-CP* were < 1, indicating negative and purifying selection [11].

Discussion

The major objective of our study was to investigate the diversity of PRSV isolates, across its genome, in a selected geographical region with almost similar climatic conditions. Hence we selected the Northern part of the Karnataka state of India, where papaya is widely cultivated and its cultivation is on the increasing trend (http://nhb.gov.in/Horticulture%20Crops/Papaya/Papaya1.htm). Along with the increase in the area under papaya cultivation in this region there is also an increase in the incidence of ring spot viral disease. Previously most of the biodiversity studies across India were restricted to the coat protein (*CP*) gene alone and major portion of the viral genome was ignored [30]. Hence we analyzed the coding regions from both 5′ terminal of PRSV genome, covering *P1* and *Hc-Pro* genes, as well as the 3′ terminal, covering overlapping regions of *NIb* and *CP* (*NIb-CP*). Selected locations known for papaya cultivation were surveyed across the six districts of North Karnataka (Suppl. Figure 2). Further from these six districts of North Karnataka 15 plant samples that were positive for PRSV infection were identified and subjected to molecular/sequence analysis. Nucleotide sequences analysis of the selected three genomic regions, namely *P1*, *Hc-Pro* and *NIb-CP* revealed that *P1* showed a wider range of sequence identity, followed by *Hc-Pro* and very narrow range of sequence identity in the case of *NIb-CP*. However, the range of amino acid sequence identities was similar for all the three genomic regions, indicating that the nucleotide sequence variation was mostly because of synonymous nucleotide substitution coding for the same amino acid due to degeneracy of the genetic code. All the important and well characterized protein motifs of potyviral genome, such as VELI, GSSG and RG in the P1 protein, which may have a role in the protease activity; PTK, NIFLAML, AEL-PRILVDH and LKANTV in the HC-Pro coding region for aphid transmission and symptomatology; GNNSGQPSTVVDNTLMV and GDD in NIb for replicase activity, and DAG motif in the CP coding region for aphid transmission, were identified to be conserved in all the 15 PRSV isolates sequenced in this study [19].

RDP analysis of the 15 PRSV isolates from this study along with other selected PRSV isolates revealed that the *P1* was the most recombinogenic region followed by *Hc-Pro*, and with no detectable recombinations in *NIb-CP* region. Higher levels of *P1* sequence diversity was also reported when the sequences of PVY were analyzed [27]. Presence of a recombination break point at $\sim$ 850 nt was reported to be common in all the American isolates and also this showed a strong statistical support that an Indian isolate of PRSV was the major parent [14, 20]. Our studies also showed recombination breakpoint at a similar nucleotide position of *P1* and also in two more additional sites of *P1*. Putative recombination breakpoints were also traced in the 100–200 nt stretch of *Hc-Pro* of at least three PRSV isolates from this study. The higher level of sequence identity of NIb proteins among different PRSV isolates reflects its high level of conservation. Further, because of the RNA-dependent RNA-Polymerase (RdRp) function of the NIb protein, the important amino acid domains were conserved. Additionally a common selection

Table 2 Population genetic parameters and neutrality tests calculated for the three genomic regions of *papaya ringspot virus* (PRSV) analyzed in this study

Genomic region	N	S	π	DS	DN	DN/dS (ω)	Tajima's D	Fu and Li's D	Fu and Li's F
P1	1202	633	0.142	0.245 ± 0.020	0.187 ± 0.012	0.059	− 1.734	− 2.440	− 2.601
HC-Pro	666	361	0.106	0.196 ± 0.023	0.109 ± 0.008	0.086	− 1.807	− 2.305	− 2.541
NIb-CP	867	410	0.104	0.149 ± 0.019	0.137 ± 0.008	0.011	− 1.953	− 2.360	− 2.62

n = number of sites, S = number of segregating (polymorphic) sites, π = pairwise nucleotide diversity (mean nucleotide differences per site between sequence pairs), dS = frequency of synonymous substitution per site, dN = frequency of nonsynonymous substitution per site

pressure may be in place to maintain a correct conformation of the NIb protein, which plays an important role in viral replication through its RdRp activity. However earlier studies have shown that the CP, particularly its N-terminal is relatively less conserved than the NIb proteins [2].

Phylogenetic analysis of the three gene sequences of the 15 PRSV isolates from this study along with other PRSV isolates from India indicated significant variation in the clustering pattern. Similar such variation in the clustering pattern has also been reported by Zhao et al. [32], when *CP*, *NIb* and *Hc-Pro* sequences of PRSV isolates from Hainan province of China were subjected to phylogenetic analysis. Such a variation in phylogenetic clustering could be because of the differential rates of evolution of these gene/genomic sequences as a result of different selection pressures [5]. The proportion of nucleotide substitution between the non-synonymous and synonymous sites (dN/ds) was in the range of 0.011–0.086, indicating a negative and purifying selection or selection for the amino acid sequence conservation [20]. In general, the selection for virus populations is negative, although exceptionally the diversified selection has been reported for some plant viruses [18].

These diversity studies in a selected geographical region under papaya cultivation indicate that there is a differential variability and recombination across the PRSV genome. The *P1* gene was the most variable while the *NIb* was the most conserved making it a good target for the control of PRSV through RNA-interference technology [16]. Previous studies also have reported significant variability in the size and sequence of *P1* gene of family *Potyviridae* [5, 16, 32], indicating that *P1* is an ideal region for variability studies which may reveal the population differentiation and eventually the incipient speciation.

Although several different strategies have been adapted for the management of PRSV [7, 8] not much success has been accomplished or the virus control is restricted to local viruses [32]. Of all the management strategies, cultivation of virus resistant transgenics remains the most effective control for PRSV, which is obvious from the world's first commercialized transgenic papaya, introduced in Hawaii (USA) in 1998 [4]. However, the success of control of PRSV is mainly determined by the genetic diversity of the PRSV populations in different geographical regions [7, 12, 28]. Recently, the breakdown of *CP*-transgenic has been reported in Taiwan by super virus isolates with a stronger silencing suppressor (*Hc-Pro*) [12, 32]. Hence sequence diversity studies are preliminary and important to effectively design a control strategy for the management of plant viral diseases. This study on PRSV variability should help in identifying appropriate viral target sequences using technologies such as RNA-interference.

Acknowledgements We acknowledge the financial support from ICAR's Network Project on Transgenics in Crops (NPTC) and ICAR-NRCPB. RM acknowledges the post-doctoral fellowship from DST-SERB.

Author's contribution BLP designed and monitored all the experiments, SP and AP collected the leaf samples, RM and SP cloned the PRSV fragments, RM helped in sequence analyses and BLP interpreted the results and wrote the manuscript. All the authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Adams MJ, Antoniw JF, Fauquet CM. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. Mol Plant Pathol. 2005;6:471–87.
2. Aleman-Verdaguer ME, Goudou-Urbino C, Dubern J, Beachy RN, Fauquet C. Analysis of the sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic potyvirus isolates, implications for the intraspecies molecular diversity of potyviruses. J Gen Virol. 1997;78:1253–64.
3. Chung BY, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the Potyviridae. Proc Natl Acad Sci USA. 2008;105(15):5897–902.
4. Ferreira SA, Pitz YK, Manshardt R, Zee F, Fitch M, Gonsalves D. Virus coat protein transgenic papaya provides practical control of papaya ringspot virus in Hawaii. Plant Dis. 2002;86:101–5.
5. Garcia-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. Annu Rev Phytopathol. 2001;39:157–86.
6. Gibbs A, Mackenzie A. A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. J Virol Methods. 1997;63(1–2):9–16.
7. Gonsalves D. Control of Papaya ringspot virus in papaya, a case study. Annu Rev Phytopathol. 1998;36:415–7.
8. Gonsalves D, Tripathi S, Carr JB, Suzuki JY. Papaya ringspot virus. Plant Health Instruct. 2010. https://doi.org/10.1094/PHI-I-2010-1004-01
9. Hall TA. BioEdit, a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41:95–8.
10. Jain RK, Sharma J, Siva kumar AS, Sharma PK, Byadgi AS, Verma AK, Varma A. Variability in the coat protein gene of papaya ringspot virus isolates from multiple locations in India. Arch Virol. 2004;149(12):2435–42.
11. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. PLoS Genet. 2008;4:1–10.
12. Kung YJ, You BJ, Raja JAJ, Chen KC, Huang CH, Bau HJ, Yang CF, Huang CH, Chang CP, Yeh SD. Nucleotide sequence–homology–independent breakdown of transgenic resistance by more virulent virus strains and a potential solution. Sci. Rep. 2015;5:1–10.

13. Librado P, Rozas J. DnaSP v5, a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
14. Mangrauthia SK, Parameswari B, Jain RK, Praveen S. Role of genetic recombination in the molecular architecture of papaya ringspot virus. Biochem Genet. 2008;46:835–46.
15. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4, Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1(1):vev003.
16. Mishra R, Patil BL. First complete genome sequence of a distinct papaya ringspot virus isolate from the north eastern region of India. Genome Announc. 2018;6:e00437-18. https://doi.org/10.1128/genomeA.00437-18.
17. Mishra R, Gaur RK, Patil BL. Current knowledge of viruses infecting papaya and their transgenic management. In: Gaur R, Petrov N, Patil B, Stoyanova M, editors. Plant viruses, evolution and management. Singapore: Springer; 2016. p. 189–203.
18. Moury B, Morel C, Johansen E, Jacquemond M. Evidence for diversifying selection in Potato virus Y and in the coat protein of other potyviruses. J Gen Virol. 2002;83:2563–73.
19. Olarte Castillo XA, Fermin G, Tabima J, Rojas Y, Tennant PF, Fuchs M, Sierra R, Bernal AJ, Restrepo S. Phylogeography and molecular epidemiology of papaya ringspot virus. Virus Res. 2011;159:132–40.
20. Ortiz-Rojas L, Chaves-Bedoya G. Molecular characterization of two papaya ringspot virus isolates that cause devastating symptoms in Norte de Santander Colombia. Eur J Plant Pathol. 2017;148:883–94.
21. Pamilo P, Bianchi NO. Evolution of the Zfx and Zfy genes, rates and interdependence between the genes. Mol Biol Evol. 1993;10:271–81.
22. Pasin F, Simón-Mateo C, García JA. The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses. PLoS Pathog. 2014;10(3):1–16.
23. Purcifull DE, Edwardson JR, Hiebert E, Gonsalves D. Papaya ringspot virus. In: CMI/AAB descriptions of plant viruses. Wellesbourne, UK: Association of Applied Biologists; 1984. p. 292.
24. Silva-Rosales L, Ruiz-Castro S, Teliz-Ortiz D, Noa-Carrazena JC. Coat protein sequence comparison of three Mexican isolates of papaya ringspot virus with other geographical isolates reveal a close relationship to American and Australian isolates. Arch Virol. 2000;145:835–43.
25. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6, molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.
26. Tennant PF, Fitch MM, Manshardt RM, Slightom JL, Gonsalves D. Papaya ringspot virus resist-ance of transgenic rainbow and sunup is affected by gene dosage, plant development, and coat protein homology. Eur J Plant Pathol. 2001;107:645–65.
27. Tordo VMJ, Chachulska AM, Fakhfakh H, Ronancer ML, Robaglia C, Astier Manifacier S. Sequence polymorphism in the 5′NTR and in the P1 coding region of potato virus Y genomic RNA. J Gen Virol. 1995;76:939–49.
28. Tripathi S, Suzuki JY, Ferreira SA, Gonsalves D. Papaya ringspot virus-P, characteristics, pathogenicity, sequence variability and control. Mol. Plant Pathol. 2008;9:269–80.
29. Webb RE, Scott HA. Isolation and identification of Watermelon mosaic viruses 1 and 2. Phytopathology. 1965;55(8):895–900.
30. Wei J, Liu D, Li X, Zhou P. Variation in the coat protein gene of Papaya ringspot virus isolates from multiple locations of China. J Integr Plant Biol. 2007;49:1062–9.
31. Yeh SD, Jan FJ, Chiang CH, Doong TJ, Chen MC, Chung PH, Bau HJ. Complete nucleotide sequence and genetic organisation of papaya virus RNA. J Gen Virol. 1992;73:2531–41.
32. Zhao H, Jia RZ, Zhang Y-L, Zhu YJ, Zeng HC, Kong H, McCafferty H, Guo AP, Peng M. Geographical and genetic divergence among Papaya ringspot virus populations within Hainan Province, China. Phytopathol. 2016;106:937–44.