Genetic variability and evolutionary dynamics of atypical Papaya ringspot virus infecting Papaya.

Anam Saleem
COMSATS University Islamabad
Zahid Ali
COMSATS University Islamabad
Shyi-Dong Yeh
National Chung Hsing University
Wajeeha Saeed
COMSATS University Islamabad
Amna Imdad
COMSATS University Islamabad
Muhammad Akbar
University of Karachi
Saadia Naseem (saadia.naseem@comsats.edu.pk)
COMSATS University Islamabad

Research Article

Keywords: Papaya ringspot virus, isolate Pakistan, evolution, atypical, genetic variability

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-118033/v1

Genetic variability and evolutionary dynamics of atypical *Papaya ringspot virus* infecting Papaya.

Anam Saleem[1], Zahid Ali[1], Shyi-Dong Yeh[2], Wajeeha Saeed[1], Amna bint Imdad[1], Muhammad Faheem Akbar[3], Saadia Naseem[1,*]

1. Department of Biosciences
   COMSATS University Islamabad
   Islamabad, Pakistan
2. Department of Plant Pathology
   National Chung Hsing University
   Taichung, Taiwan
3. Department of Agriculture and Agribusiness management
   University of Karachi
   Karachi, Pakistan

Corresponding author:

[1,*]Saadia Naseem

E. mail: saadia.naseem@comsats.edu.pk

Abstract

The mutation prone RNA genome of *Papaya ringspot virus* could be a driving force behind its geographical spread and dissemination. Here we present the molecular investigations on atypical PRSV–P strain identified from Pakistan with genome sequence phylogenetic and recombination analysis. The PRSV-P, Pakistan outbreak strain showing a geographic spread across India and Bangladesh in phylogenetic lineage. In major recombination events, it has acquired genome variation in amino terminal of PRSV coat protein gene, whereas the gene for helper component-proteinase (HC–Pro), a nonstructural coding region of multi-domain provenance, also evolves at nucleotide and amino acid levels. The phylogenetic analysis of another highly variable P1 region showed evolutionary dynamics with respect to other geographical strains, particularly the Indian isolate from North East region (Meghalaya). PRSV–PK holds high levels of genetic divergence in comparison to American, Australian and Asian isolates. The genetic and phylogeographic analyses indicate that a spatial recombination has occurred from first PRSV, however temporal evolution is within the region of occurrence. The genetic variations and evolutionary dynamics of this virus may challenge the resistance developed in papaya against PRSV and give rise to virus lineage because of its atypical emergence where geographic spread is already occurring.

Key words: ***Papaya ringspot virus*****, isolate Pakistan, evolution, atypical, genetic variability**

Introduction

Plant viruses cause significant losses to quantity and quality of almost all cultivated crop plants, hence always being the greatest threat to the economical food crops. Therefore, on time management of an invading virus is mandatory. The insights into the genome of a virus, study of viral gene functions and fulfillment of Koch's postulates have great importance for efficient management strategy. Unearthing of next generation sequencing (NGS), molecular characterization of viruses and their gene motifs have gained pace in recent decade and ultimately leads to the emergence of many new viruses in International Committee on Virus Taxonomy (ICTV) reports. Therefore, crop health and disease management rely on the development of supportive information from a particular virus genome and its encoded genes for the estimation of overall field conditions and avoid subsequent crop losses to ensure food safety[1].

Papaya ringspot virus (PRSV) is a devastating pathogen for papaya and cucurbits production worldwide. The virus is non–persistently transmitted by aphids in nature and belongs to the genus *Potyvirus*, family *Potyviridae*[2]. There are two serologically indistinguishable biotypes of PRSV which differ in their ability to infect papaya[3], papaya-infecting biotype P (PRSV–P), which also infects cucurbits, and cucurbit-infecting biotype W (formerly known as water melon mosaic virus 1) (PRSV–W)[4] PRSV–P was first isolated from papaya in Hawaii[5]. Later on, it has been reported from several other countries and regarded as an extremely damaging pathogen within initial years of its infection[4,6]. The typical symptoms induced by the virus include the characteristic ringspots on fruits, along with mosaic, chlorosis, distortion on leaves, stunting in growth, flower abortion and water oily streaks on petioles and upper young trunks. The PRSV possess a single-stranded, positive sense RNA genome. The RNA is encapsidated with a coat protein of 36K[7]. The virus under favorable climatic conditions leads to 100% yield losses[8]. The viral genome contains 10,326 nucleotides followed by a poly A tail and have a single open reading frame encoding a polyprotein of 3344 amino acids (nt 86–10120), which is processed by three viral proteases to generate all viral proteins[7].

Mutation, recombination, and reassortment are important mechanisms for genetic variation of RNA viruses and so is the case with PRSV. Therefore, understanding of their roles in the genome dynamics of PRSV is important. Studies on PRSV genome dynamics have been conducted over years with an intention to get insights into the adaptation of the virus to changing environment, evolutionary behavior of the virus, and most importantly highlighting the conserved domains for implementing an effective control strategy such as transgenic resistance against the virus[4,9,10]. The PRSV isolate from Pakistan (PRSV–PK) is evolutionary diverse as compared to other PRSV isolates from different geographical locations, possessing a high genetic variation in its CP gene[11,12]. The PRSV strains from diverse geographical locations possessed genetic variability in their CP genes depicting direct link of genome variation with the geographic origin of the virus[13]. Thus, understanding of origin and evolution of PRSV-PK require further exploration and analysis of its genome. The preliminary evidence on the molecular evolution of PRSV proposed that PRSV-P might be originated from the type W, as PRSV-W appeared in Australia at least 20 years before the PRSV-P[3]. Studies on the CP sequences of most of the PRSV-P and some PRSV–W isolates from different parts of world have been conducted to estimate the relative genetic divergence[3,14]. The data based on the nucleotide and amino acid sequences of the CP

genes of these isolates suggested a sequence divergence of up to 14% and 10% among these isolates[15,16]. However, sequence analysis of the CP genes from USA and Australia showed only marginal variation in their PRSV isolates[14,15]. The CP gene sequences of PRSV isolates from Pakistan and all geographic locations in five continents were studied further for evolutionary divergence and a unified phylogenetic tree was constructed. These findings definite that the PRSV–PK isolates emerge as atypical strain[11,12]. Here we tried to probe and dissect the virus PK, particularly at the whole genome level of the PRSV isolates from the Asian and American countries, which provide a solid base for comparative analysis of PRSV meta-genomes. Moreover, the ambiguity in the earliest identification of PRSV from cucurbits and the non-clarity in the evolutionary and epidemiological studies have also set the rationale.

In this investigation, we describe the comparison of the annotated whole genome sequence of PRSV–PK, with the fragments including PI, CP, HC–Pro and 3' UTR which are prone to high sequence variation, with representative genome segments of PRSV isolates from rest of the world. The derived phylogenetic relationships are helpful to explore the spread, interlinking and complexity of PRSV populations worldwide. In addition, the constant divergent behavior and conserved expression of PRSV are fundamental aspects to analyze the genetic differentiation and gene flow. The diversified evolutionary lineage studied using meta-genomes and the analysis of recombination events in the genome segment of PRSV–PK provide supportive information for testing lately developed resistance in papaya against this virus and for sustainability of this economy boosting crop in many Asian countries and worldwide.

Results

Sequence variability and phylogenetic assessment of CP, P1, HC-Pro, 3'UTR and PRSV whole genome

The whole genome of atypical PRSV–P isolate PK was sequenced through meta-genome sequencing and verification through Sanger sequencing of cloned amplified fragments schematically presented in Figure 1 (Fig.1); annotated genome sequence has been submitted in GenBank (Accession number MT090406). The sequence analysis using CP, P1, HC–Pro, 3'UTR gene fragments and the resulting phylogenetic tree derived from the comparison of these individual sequences is shown in the Figure 2 a, b, c and d respectively (Fig. 2). The highly variable complete CP gene nucleotide sequence of this virus made an independent cluster with former sequences of CP gene of PRSV Pakistani origin, confirming the genetic variants of the same strain of the virus. The next relative becomes JN979399 and JQ394695 lately reported from Bangladesh with 92% homology. The overall clade extends to only one Indian isolate from Meghalaya (MF356497) in the geographic region (Fig. 2a). The CP gene sequences of PRSV strains from America, China, Vietnam, Australia, Thailand, Japan, Europe, Taiwan and rest of India are included in a large separate clade, showing least homology with our representative virus.

The P1 gene of PRSV from Pakistan was found to be most provenance with respect to genetic variation with nucleotide and subsequent amino acids insertions. The Phylogenetic analysis of the complete P1 gene of PRSV PK isolate genome when compared with those of the 11 isolates from

America, China, Thailand, Taiwan and India showed only 81.0% homology with the Indian Isolate from North East region (Meghalaya).

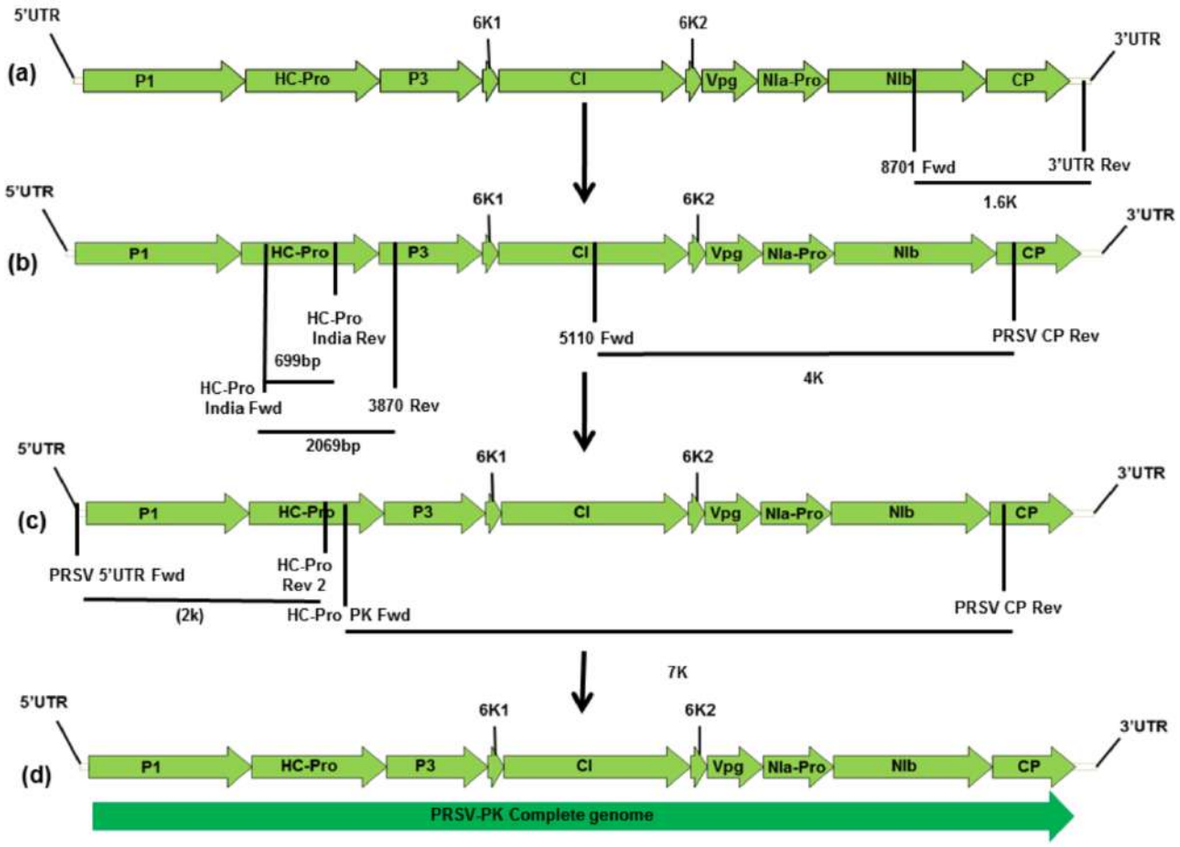

Figure 1: Schematic amplification of PRSV–PK genome **(a)** CP gene amplification; **(b)** 4K, HC–Pro and full HC–Pro amplification; **(c)** 5' terminal end and 7K amplification; **(d)** Full genome amplification.

Open reading frames are indicated as arrows on viral sense RNA. Direction of arrow indicates the direction of translation. PI:; HC-Pro:; P3: ; 6K1:; CI: Cylindrical inclusions; 6K2:; Vpg:; NIa-Pro:; NIb:; CP: coat protein.

This degree of homology in case of P1 (81%), is not as high as those with CP (91.8%), full genome (88.2%) and HC–Pro (87.8%) of the PK isolate with Meghalaya (MF356497). This tree mentioned in (Fig. 2b) further provides the authentic insight of atypical nature of PRSV–PK strain consistently illustrating the contrasting genome variation with the isolates from China, Taiwan and west of India. The phylogenetic of P1 gene generated three clades, first having PRSV–PK and Indian isolate Meghalaya, isolates of USA, Colombia, Hawaii and France origin were grouped in a separate cluster. Yet again another cluster consisted of PRSV isolates from India, Thailand, Taiwan and China (Hainan) (Fig. 2b).

Another, variable region in RNA virus genome is HC–Pro which was analyzed for PRSV genome variability and it showed 87.8% homology of PRSV–P isolate PK with Indian Meghalaya PRSV isolate

whereas, it was divergent from the isolates from southwestern and majority of isolates from northern regions of India (Fig. 2c).

The degree of relatedness of 3'UTR assessed through phylogenetic tree depicts that the 3'UTR of PRSV–PK is closely related to the isolate from PRSV–Meghalaya (MF356497) 85.4% and an Indian isolate MH311882 (90%) from Pune, the region in west of India. Whereas the isolates from America, France, Colombia, India and few isolates from Taiwan showed substantial sequence variation from PRSV–PK. The third divergent clade comprised most of the isolates from China and Taiwan (Fig. 2d).

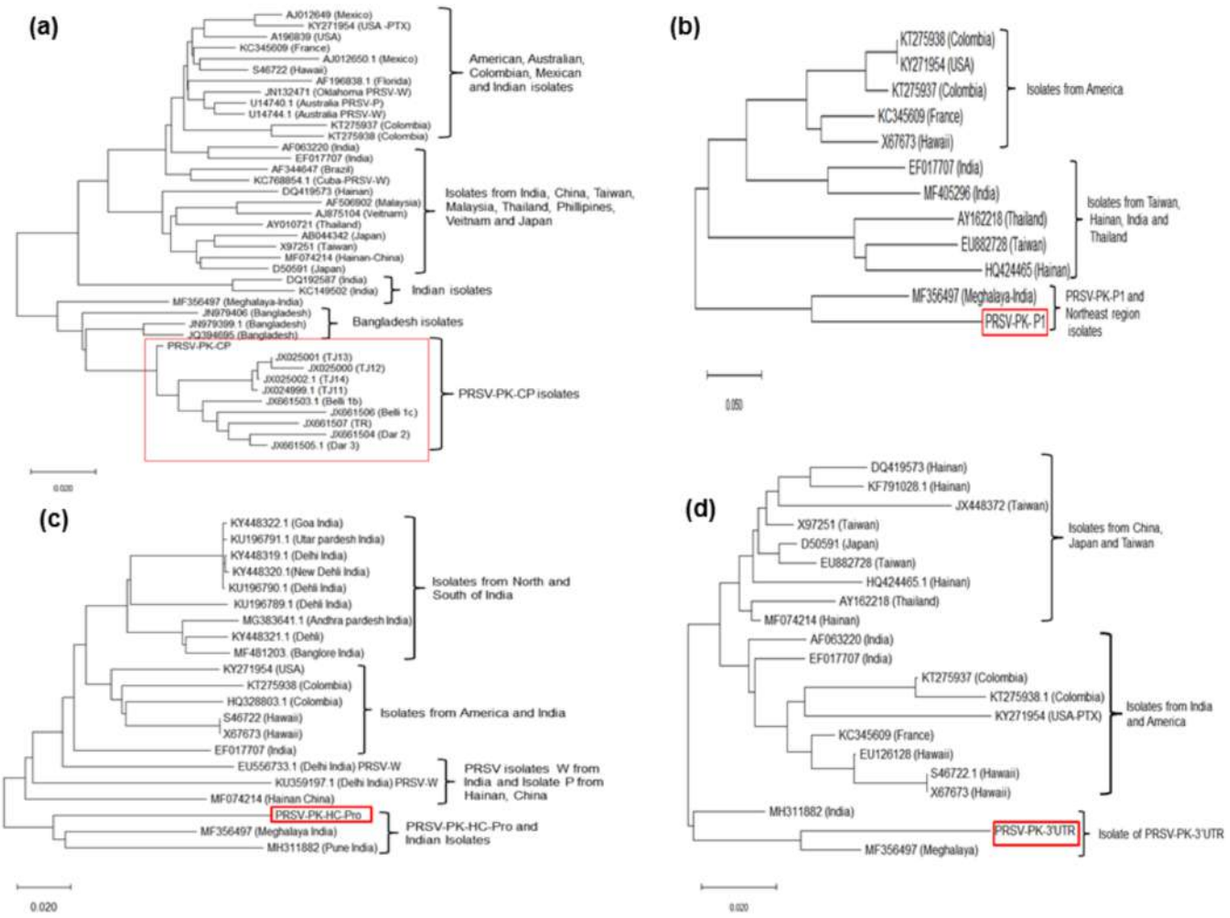

Figure 2: Neighbor-joining tree representing phylogenetic relationships of Papaya ringspot virus-P strain PK (PRSV–PK) **(a)** Coat Protein (CP) sequence **(b)** Protease P1 sequence **(c)** Helper component HC-Pro Sequence **(d)** 3'UTR sequence from other related sequences selected via BLAST search. Upper and lower branch points show bootstrap values (1,000 replicates) supporting a particular phylogenetic group. The scale bar represents nucleotide substitutions per site. All nucleotide sequences are retrieved according to the isolate name and the GenBank accession number**.**

To address PRSV–PK phylogeographical occurrence, the tree based on whole genome sequence comparison of this virus with available complete genome sequences of PRSV isolates worldwide, was constructed conclusively. It has demonstrated convincing overall tree topology results. The complete genome sequences of isolates with genome length up to 10,343 bp from China, Taiwan, Hawaii, USA,

Colombia, Thailand, India, Bangladesh, and France were retrieved from NCBI database subsequently aligned with ClustalX2 and a neighbor joining tree with the bootstrap value of 1,000 replicates. The phylogenetic tree constructed using MEGA–X software revealed that the full length PRSV–PK isolate (in this study) illustrated only 88% and 85% homology with the isolate of Meghalaya (MF356497) and (MH311882) from India (Fig. 3) respectively. However, the closest relative becomes MH397222, lately reported from Bangladesh with 89.7% homology **(Fig. 3).**

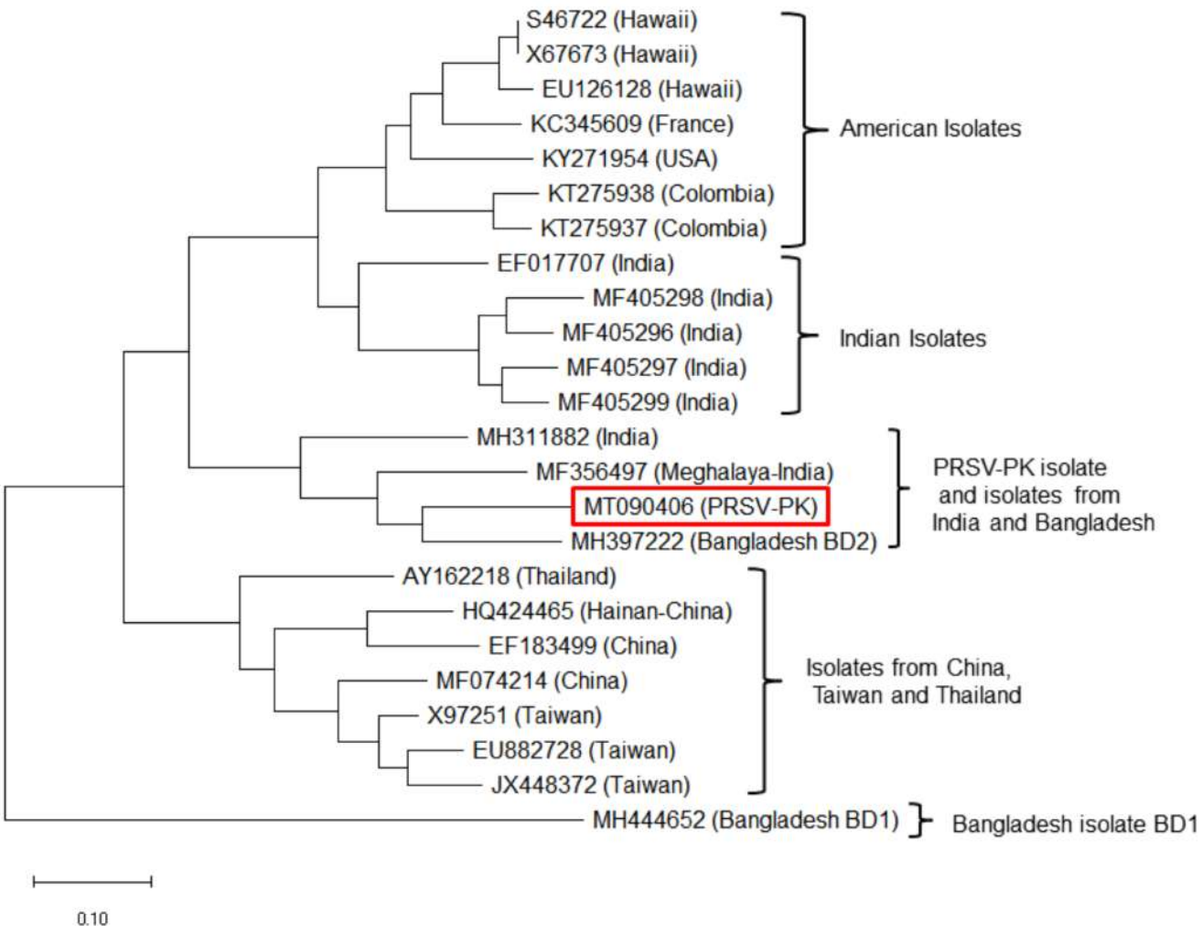

Figure 3: Region specific neighbor joining tree representing phylogenetic relationship between complete genome sequences of PRSV-P isolate of Pakistan and PRSV complete genome sequence from rest of the world. Tree was constructed using ClustalX2 and MegaX Program. The scale bar represents nucleotide substitution per site, Bootstrap value of 1,000.

Genetic Differentiation and gene flow assessment in CP gene and full length genome of PRSV-PK

Three statistical parameters were used to estimate genetic differentiation between PRSV-PK population and PRSV population from other geographical regions based on their CP and full genomic sequences. In the individually aligned CP and full-length genomic sequences, a permutation based statistical test, Kst* (sequence based statistical test) and Snn (measurement of frequency of the nearest neighbor sequences in the same locality) were utilized to measure the genetic differentiation between PRSV population of Pakistan with the population from diverse geographical locations. The genetic differentiation estimation between the subpopulations using CP sequences created in DnaSP showed significant P values of Kst* and Snn (Table 1), which significantly denied the null hypothesis statement that “there is no genetic

differentiation between two subtypes" thereby declaring genetic differentiation. Fst (the interpopulation component of genetic variation or the standardized variance in allele frequencies across populations) was used to access the gene flow level. Two subpopulations were created to compare the differences. Considering the data (Table 1) the Fst values generated between the populations based on the CP based PRSV isolates were above 0.33. The Fst values between population comprising of PK CP isolates vs Bangladesh isolates were (0.40) , PK vs Indian isolates (0.38), PK vs other Asian (China, Taiwan and Thailand) (0.62), PK vs American isolates (0.66) , PK vs Mexican (0.59), PK vs Colombian (0.75) and PK vs Australian isolates (0.75) depicts that there exists a significant genetic differentiation between these populations. The estimation of genetic differentiation and gene flow between populations using CP gene sequences is mentioned in Table 1 and estimation of gene flow between PRSV–PK populations with rest of populations based on full genome sequences is mentioned in Table 2.

Populations	**Kst***	**P value**	**Snn**	**P value**	**Fst**
Pakistan vs Colombian	0.16499	0.0060**	1.00000	0.0360*	0.75400
Pakistan vs Mexican	0.18204	0.0010**	1.00000	0.0010**	0.59729
Pakistan vs American	0.0000	0.0000***	1.00000	0.0000***	0.66495
Pakistan vs Asian	0.17561	0.0000***	1.00000	0.0000***	0.62725
Pakistan vs Bangladesh	0.10821	0.0020**	0.96154	0.0020**	0.40202
Pakistan vs Australian	0.15900	0.0120*	1.00000	0.0320*	0.75895
Pakistan vs Indian	0.13216	0.0000***	0.93333	0.0070**	0.38514

Table 1: Statistical tests for genetic differentiation and gene flow between *Papaya ringspot virus* populations from Pakistan with the populations from Colombia, Mexico, America, Asia, Bangladesh, Australia and India based on the CP gene nucleotide sequences.

ns, not significant; *0.01 < P < 0.05; **0.001 < P < 0.01; ***P < 0.001. *Kst*, Snn and Fst* were implemented in DnaSP 6. The deviation hypothesis from null population differentiation was tested by 1000 permutations of the raw data.

The P–values of Kst* and Snn calculated between PRSV–Populations compared on the basis of full genomes indicating that populations are not well differentiated yet, while P values of Kst* between Pakistan vs Asian populations were significant showing that there exists genetic differentiation between this population (Table 2). However, the Fst values calculated between PRSV-PK population based on full length sequence with populations from India, America, Asia and Colombia are 0.62609, 0.84587, 0.74163 and 0.90632 respectively. The high Fst values revealed infrequent gene flow between these populations. The estimation of genetic differentiation and gene flow between PRSV-PK populations with rest of populations based on full genome sequences is mentioned in Table 2.

Populations	**Kst***	**P-value**	**Snn**	**P-value**	**Fst**
Pakistan vs Indian	0.01114	0.0790^{ns}	0.88889	0.0640^{ns}	0.62609
Pakistan vs American	0.12458	0.0540^{ns}	1.00000	0.0540^{ns}	0.84587

Pakistan vs Asian (China, Taiwan, Thailand)	0.02901	0.0330*	1.00000	0.0510ns	0.74163
Pakistan vs Colombia	1.00000	1.00000ns	1.00000	0.3450ns	0.90632

Table 2: Statistical tests for genetic differentiation and gene flow between *Papaya ringspot virus* populations from Pakistan with the populations from Colombia, Mexico, America, Asia, Bangladesh, Australia and India based on the full length genomic gene sequences.

ns, not significant; *0.01 < P < 0.05; **0.001 < P < 0.01; ***P < 0.001. *Kst*, Snn and Fst* were implemented in DnaSP 6. The deviation hypothesis from null population differentiation was tested by 1000 permutations of the raw data.

Recombination Analysis

Recombination events in CP domain of MT090406-PRSV-PK strain and other 28 geographically distinct PRSV isolates were analyzed by using GENECONV, BOOTSCAN, MaxChi, CHIMAERA, SiSCAN, 3SEQ and Lard within RDP4 package software. The breakpoints at gene positions (232-624) were observed and PK-PRSV-CP isolates identified to be a potential recombinant by at least 4 of 7 methods implemented in RDP4. The major parent detected by the event is JN979399 (Bangladesh). The recombination event along with the breakpoint position and P-values is mentioned in the (Table 3) where recombinant analysis was positive with Chimaera (C), 3Seq (3S), MaxChi (M)and Lard (L), whereas; RDP (R), Gene Conversion (G), Bootscan (B) and Siscan (S) revealed no evidence of recombination.

Event No	Recombinant	Major Parent	Minor Parent	Break point	RDP (R)	Gene Conversion (G)	Boot Scan (B)	MaxChi (M)	Chimaera (C)	Siscan (S)	3Seq (3S)	Lard (L)
1	MT090406 PRSV-PK	JN979399 Bangladesh	Unknown	232-624	NS	NS	NS	6.63E-03	4.23E-04	NS	8.72E-03	2.78E-04

Table 3: Recombination detection in CP gene of atypical PRSV-P Isolate PK (MT090406). The break points listed refer to their position in the alignment; RDP(R), GENECONV(G), BOOTSCAN(B), MaxChi(M), CHIMAERA(C), SiSCAN(S), 3SEQ(3S), and Lard (L).

Recombination events in full length genome of PRSV-P (PK) along with other reported PRSV full length genome sequences from Hawaii, Taiwan, France, USA, Colombia, India, China and Thailand were detected using GENECONV, BOOTSCAN, MaxChi, CHIMAERA, SiSCAN and 3SEQ within RDP4 package software. The target genome was found to be highly recombinant with a total of highly 40 recombination events detected in RDP and plausible breakpoints were also identified in particular event. Results are highlighted for four selected events including 5, 17, 37 and 40 detecting PRSV–PK

isolate to be recombinant by the methods within RDP software. Meghalaya (MF356497) was detected to be the potential parent in at least two events, event 5 and event 37 detected by RDP, Maxchi, Chimaera and Siscan. The potential breakpoints during these two events were (1713-1830) and (586-1046) thereby indicating the recombination towards the 5' terminal region of genome respectively. The RDP event number 17 detected two recombinants; one is PRSV–PK with the major parent from JX448372 (Taiwan) with the break point position of (586-1046) and another recombinant in the same event is MF356409 (Meghalaya) with three major parents including EU882728 (Taiwan), X97251 (Taiwan), MF074214 (Hainan). Another event number 40 with PRSV–PK as a recombinant showed an unknown major parent with the breakpoint position of (42–332). The detail of the recombination events showing PRSV–PK to be recombinant along with the breakpoints position and the average P–value is mentioned (Table 4).

Sr. No/Event No	**Recombinant**	**Major parent**	**Minor parent**	**Break points**	**RDP (R)**	**Gene Conversion (G)**	**Bootscan (B)**	**Maxchi (M)**	**Chimaera (C)**	**Siscan (S)**	**3Seq**
1(5)	MT090406 PK-PRSV	MF356497 Meghalaya	Unknown	1713-1830	$4.64x10^{-14}$	$5.99x10^{-14}$	$9.82x10^{-10}$	$1.87x10^{-6}$	$9.04x10^{-07}$	$1.14x10^{-08}$	$3.01x10^{-08}$
2(17)	MT090406 PK-PRSV	JX448372 Taiwan	Unknown	586-1046	$5.03x10^{-08}$	$7.81x10^{-4}$	$2.02x10^{-3}$	$7.77x10^{-03}$	$2.96x10^{-06}$	$6.9x10^{-09}$	NS
	MF356409 Meghalaya-India	EU882728 Taiwan	Unknown	-	-	-	-	-	-	-	-
		X97251 Taiwan	-	-	-	-		-	-	-	-
		MF074214 Hainan	-	-	-	-	-	-	-	-	-
		EU882728 Taiwan	-	-	-	-	-	-	-	-	-
3(37)	MT090406 PK-PRSV	MF356497 Meghalaya	Unknown	586-1046	$2.99x10^{-02}$	NS	NS	0.04654399	$1.35x10^{-03}$	$1.36x10^{-08}$	NS
4(40)	MT090406 PK-PRSV	Unknown	KT275938 Colombia	42-332	NS	NS	NS	$2.18x10^{-04}$	$1.75x10^{-04}$	$1.05x10^{-05}$	$4.29x10^{-02}$
	MF356409 Meghalaya-India	Unknown	KT275937 Colombia 2	-	-	-	-	-	-	-	-

Table 4: Recombination detection in full length atypical PRSV–P Isolate PK (MT090406). The break points listed refer to their positions in the alignment; RDP(R), GENECONV(G), BOOTSCAN(B), MaxChi(M), CHIMAERA(C), SiSCAN(S), 3SEQ(3).

Discussion

Since the introduction of Papaya in cropping system, incidence and damage caused by aphid transmitted papaya ringspot disease (PRSD) has been increasing over time globally[17]. In the realm of pathogen derived systemic resistance strategies for management of PRSD, transgenic antiviral papaya was effectively utilized to overcome widespread invading PRSV[13,18,19]. Nevertheless, reports of recent outbreaks in China has shown concomitant decline in transgenic resistance because of occurrence of resistant viral strains in Hainan province, thus leading to loss of resistance and massive crop damage[20]. Since the occurrence of the PRSV and subsequent infection of cultivated papaya in Pakistan, the preliminary scientific investigations have shown emergence of an atypical PRSV isolate in Pakistan on the basis of its coat protein gene[12]. Earlier findings have also indicated that PRSV biotype W infecting cucurbits in Pakistan remains atypical in the pre-history of potyviruses on the basis of its divergent CP gene[11]. Our follow up studies, on PRSV infection in Papaya grown at large scale in Pakistan has shown the overall estimated loss of 80-100% in all major orchards particularly in Sindh province. Here, we present the first annotated genome sequence of *Papaya ringspot virus* biotype P isolate Pakistan (PRSV–P isolate PK), of 10,320 nucleotides deposited in the GenBank with accession number MT090406. This sequence corresponds to the viral strain found to be most prevalent (88%) among all isolates of PRSV from Pakistan. The aggressive symptomatic appearance and higher rate of infection in papaya crop plants with this atypical PRSV strain, yield losses up to 100% have been recorded in surveys and analyses. The occurrence and disease tendency required further genetic variability and genome recombination analysis of the virus to estimate its provenance in global perspective. It has already been speculated that emergence of PRSV–P in earlier identification is related to its counterpart PRSV–W recorded five years before in the country[12].

The mutation prone RNA genome of PRSV has exceptionally evolved over time showing evolutionary behavior and diverse nature[20]. The key efforts to find origin and distribution of PRSV targeted CP gene (the position of CP approximates 9258–10117 in complete genome of virus) sequence revealed that the isolates from Mexico corroborates maximum genetic compliance with the USA and Australian isolates; whereas it cluster separately from PRSV isolates of Asia particularly originated from

Indian continent[14,21]. Thereafter, we extended the finding that the sequence diversity based on CP gene is higher among the Asian isolates. The isolates analyzed on the basis of CP gene from India showed divergence of 0-11% which do not differ largely form reported variation with Bangladesh (9-14%), rest of Asian PRSV population (4–14%), and with Australia/America (5–11%) (Fig. 2a). In homology studies on the basis of CP gene, PRSV–PK showed highest relatedness with previous PRSV isolates from Pakistan whereas it is distantly related to PRSV isolates from Bangladesh and India (Meghalaya) (Fig. 2a), It is worth to mention that PRSV isolates from Bangladesh and India (Meghalaya) have occurred after establishment of atypical PRSV in Pakistan. However, PRSV–PK diverged from the other Indian isolates from Pune, Andhra Pardesh and Delhi. Nevertheless, since its inception, PRSV–PK is establishing a distinct subpopulation of the virus. The possibility behind the divergent nature of Pakistani PRSV strain could be due to the occurrence of recombination events in the amino terminal of CP region where some sequence variation in other variant of this virus has been reported[22], however structural and functional integrity of PRSV is maintained[14]. Our study showed the full-length CP gene of PRSV–PK only of 849 nucleotides with the deletions at the 5' terminal of CP (nt position 9318) and at position (9346) with an overall 6% variation in the 5'terminal region corresponding to EK aa repeats deletion. Further pointing to substitutions occurred at amino acid positions 122, 157 and 244 in CP protein causing differentiation of the PRSV–PK as compared to CP protein worldwide.

The P3, CI, Vpg, NIa–Pro appeared relatively conserved at nucleotide and amino acid sequence level in comparison to other isolates of PRSV globally (Fig. 4). Therefore, variable genome components P1 and HC–Pro and 3'UTR of PRSV–PK were selected for individual gene analysis. The regions of variation have been demonstrated in Fig. 4, showing that PRSV-PK isolate possess nucleotide and amino acid variation in P1 and HC-Pro genes as well; eventually, corresponding to divergence of this particular virus from other isolates of regional or distinct geographical locations. The length of P1 gene appears to be 1647 nucleotides with the high level of evenly distributed genetic variation 3% towards 5', up to 10% in the middle, and 6% towards the 3'termini. Variable nature of P1 has also been reported in earlier studies[23]. The HC–Pro gene length is 1370 nucleotides with approximately 5% genetic variation at the immediate 5' terminal region.

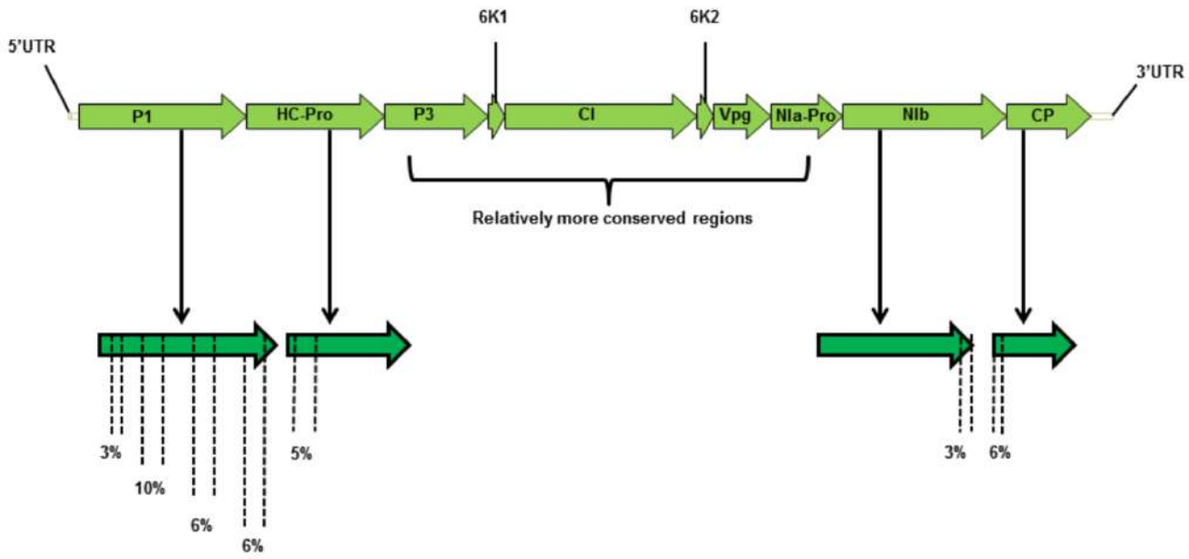

Figure 4: Approximate genetic variation in the individual gene fragments of PRSV-PK.

Despite high genome variation throughout PI gene and HC–Pro of PRSV–PK isolate, it clustered with an isolate from India (Meghalaya) in phylogenetic analysis. the high variation in P1 gene of PRSV-PK as mentioned in (Fig. 4) is the possible reason for its relatively low affinity towards PRSV Meghalaya isolate of India (homology 81%). Similarly, HC-Pro appears to have high genetic divergence from other isolates worldwide with several interlinking clusters that appeared in phylogeographic analysis (Fig. 2c).

Initially the 3'UTR region of PRSV was considered relatively conserved because of its least vulnerability towards recombination[23,24]. However, in our findings, 3'UTR phylogenetic analysis distinctly made three clusters (Fig. 2d). The phylogenetic analysis based on PRSV complete genome sequences revealed 88.2% homology of PRSV–P isolate PK (MT090406) with PRSV–Meghalaya (MF356497), a recently sequenced genome[25]. Nevertheless, the closest relative appeared yet another recently annotated complete genome sequence of PRSV from Bangladesh strain BD2 (MH397222)[23]. The PRSV isolates from northern and western regions of India, reported prior to Meghalaya form genetically distinct cluster (Fig. 3). This distinct clustering complies with the clustering of PRSV isolates of western Indian with those of American[26]. Similar Indian and American isolates clustering appeared in the comparative neighbor joining trees based on full length PRSV genomes at amino acid level concluding that American PRSV strains have originated from India[24]. Thereby, strengthening the postulate of emergence of distinctive virus in Indian subcontinent, it holds the oldest PRSV population and that the actual origin of this virus had already been scientifically speculated in the South Asia[27] where we reinforce that changes in length of CP gene could be potential reason of genetic variation[22,28].

Nevertheless, the biogeographical analysis of complete genome of PRSV–PK fairly prospected that this atypical PRSV strain emerged and evolved in Indian Peninsula (Fig. 5). The dispersal is spreading in Asian region including India, Pakistan, and Bangladesh and may be taking route to rest of the Asia, including China. In the past, major dispersal events of PRSV have provoked devastation to papaya in US mainland to Hawaii, and recently the breakup of transgenic resistance in China. However, there is no known present evidence or correlation that this atypical Asian strain pre-ambling replacement of PRSV Americas-Australia cluster. Furthermore, the PRSV–PK isolate that groups together with designated Indian and Bangladeshi isolates occurred so far from a separate dispersal event; not the same as apparent with rest of the Asian regions including Thailand, China and Taiwan. In another report of only CP based comparative analysis, Indian and Bangladesh isolates are separate from other Asian isolate's cluster therefore depicting that clustering of PRSV could not be directly linked with the geographical origin[28]. We also agree that the movement and dispersal events lead to the occurrence of single and mixed population or subpopulation of the isolates.

Discussing the facts on viral origin that whether the PRSV–P have originated from PRSV–W or not[16], the possibility still resides in the order of virus emergence in a particular geographical region. Just in case of Australia it was PRSV–W which was first to be discovered and identified in Australia and the PRSV-P was recognized much later than isolate W[3]. Similarly, the occurrence of PRSV–W was recorded in 2004[29] PRSV–P emergence reported in 2009 onwards[12]. However, we have studied constantly the emerging PRSV–P in Pakistan to develop management strategies against its widespread and found it stably conserved and devastating in the country.

The mutation prone RNA viruses mark recombination and genetic differentiation which can lead to evolution as biodiversity[20]. The study on genetic differentiation and gene flow between PRSV populations from diverse localities is of critical importance keeping in view this newly emerged atypical PRSV cluster from India, Bangladesh and Pakistan. The gene flow and genetic differentiation between PRSV populations conducted using DnaSP provided comparative analysis of DNA polymorphism suggested functional significance of evolutionary process in genomic regions of PRSV. Analysis of differentiation between phylogroups based on CP sequences revealed complete differentiation as indicated by the significant Kst, Snn and high Fst values >0.3. The previous reports on genetic differentiation analysis between virus population having Fst values above 0.25 designate high genetic differentiation and infrequent gene flow between the populations[20,30]. Consequently, there exist high genetic differentiation between virus population of Pakistan and the virus population from Bangladesh,

India, Mexico, America and Australia. Similar results based on highest Fst (0.54) value between South America and Japanese populations using *Potato virus X*, CP gene sequences suggest infrequent gene flow attributed to the long distances between these geographic regions[31]. Nevertheless, in our study higher Fst values also suggests a correlation between geographical position and genetic distances. In case of genetic differentiation assessment, with the exception of Pakistan vs Asia population all the Kst* values between other PRSV populations based on full length genomes range between 0.015-0.25 and a bit higher than "0" supported by P values >0.05 suggesting that the populations are not well differentiated. Non-significant Snn values also reveal that the populations based on full genome are not well differentiated yet. Similar results were presented showing no genetic differentiation and non-significant Snn values calculated by pairwise comparison of *Zucchini yellow mosaic virus* sequences between populations of Asia vs. Europe and Oceania vs. Africa[30]. Moreover, the Fst values mentioned in Table 2 generated between populations based on full genome of PRSV-PK vs Indian, other Asian (China, Taiwan, Thailand), Colombian and American isolates were 0.62609, 0.74163, 0.90632, 0.84587 respectively. High Fst values between these populations suggests infrequent gene flow as Fst >0.33 is its indicator[32]. Fst values associated with free, high and moderate gene flow have also been described in gene flow studies of *Potato virus Y* , where, higher Fst values >0.6 depicts less frequent gene flow and low genetic exchange between populations[33]. It can be concluded from the above-mentioned data showing higher Fst values and low genetic exchange that there exists a correlation between geographical isolation and gene flow. Similar correlation has been discussed before[30]. Thus geographical isolation is playing a definite part in shaping PRSV population structure. The infrequent gene flow and high genetic differentiation between PRSV–PK populations and others also suggests that PRSV–PK have evolved independently.

Other important factors that strongly influence the virus evolution include host adaptability and geoclimatic conditions. As this is a well-known fact that climatic conditions of particular geographical locations plays a part in shaping the phenotypic and genotypic properties of flora present in that location. Thereby changing climatic conditions will leads to the change in both phenotype and genotype of existing host[34,35]. Similarly the host plants from different geographical locations possessing diverse genetic makeup in-turn will contribute in shaping the invading virus population[36]. Hence the diverse and atypical nature of PRSV–PK is attributed to its adaptions to diverse ecological location and of course the host. In Pakistan the local papaya varieties belongs to Keto Bunder area and was cultivated largely long ago in Coastal area of Sindh Province. Later hybrid papaya varieties were also planted and cultivation

was extended from Southern to central part of the country. Therefore, papaya grown in Pakistan is not relevant with the papaya cultivated in India and Bangladesh which led to the emergence of unique PRSV. Even so, the host adaptability of PRSV population from different geographical locations regions is the major factor influencing plant virus evolution. The analysis of diverse nature of three functional protein coding regions, HC–Pro, NIa–Pro and CP of PRSV isolates from different geoclimatic regions and hosts also suggests that the PRSV population diversity is linked with the geoclimatic conditions as well as the host. However, it was also observed that different geo-climatic conditions effect the PRSV evolution with different levels of variations[37].

Recombination analysis of 29 isolates including the Pakistan PRSV full length CP gene sequence in comparison with CP gene of PRSV from rest of the world performed using RDP4 v 4.97. In our study one break point position was detected by 4 methods including Maxchi, Chimaera, 3Seq and Lard, pointing CP gene of PRSV-PK as potential recombinant with the break points at positions (232-624). Putative breakpoints positions at 208 and 605 in the CP gene of Colombian PRSV isolate was reported earlier[38]. The primary scans used in RDP indicating PRSV–PK CP gene as a recombinant region were also supported by the automated secondary scan implemented in Lard. The recombination probability derived through RDP in our study presents PRSV from Bangladesh as parent entity for PRSV–PK, while PRSV from India has been discussed as a parent entity of PRSV for Bangladesh[23]. The comparative recombination analysis between whole genome sequences clearly revealed that PK–PRSV representative isolate is closest with the Indian Meghalaya isolate from North East region. Furthermore, the recombination analysis of PK–PRSV whole genome along with other PRSV isolates identified a total of 40 recombination events. The recombination sites were detected in the full-length PK–PRSV genomic sequence. The potential breakpoints detected in multiple events were 1713-1830, 58-1046 and 42-332. The highest recombination was detected towards the 5'region of the PRSV genome particularly including the first 1020 nt. The potential regions prone to recombination are 5'UTR and P1, thus playing a major role in PRSV genome dynamics as this information strengthen some preliminary findings[26], followed by HC–Pro gene[39] and consistent with other potyviruses[40] as well. We conclude that PRSV from India (Meghalaya) (MF356497) acts as a major parent in this recent outbreak of recombination. It is noteworthy to mention that Meghalaya itself also appeared to be subjected to recombination event at the same time and both Meghalaya (MF356497) and PRSV–PK have major parents from Taiwan as mentioned in Table 3. The circle of PRSV recombination in globe is yet occurring dramatically. Earlier global prospects on spread of PRSV suggested its origin in India about 2250 years ago and its spread to

Thailand and China about 600 years ago, eventually reaching to America about 200 years ago. At the same time keeping in the fact of PRSV-P origin from W biotype it can also be assumed that initially the virus found a slot in hosts of *Cucurbitaceae* family of particularly American origin thus leading to its introduction to America Probably as Biotype-P or biotype-W via already widespread cultivated cucurbits species in South and Central America[41]. Furthermore, characterization of full genome of a greater number of PRSV–W biotype is required to elucidate and validate the actual origin of this virus as well as its evolutionary linkages. The genetic differentiation, gene flow and recombination analysis highlighted the constant divergent behavior as well as atypical nature of the presented virus, the hotspots of virus origin, interlinking of the virus origin and epidemiological behavior with other isolates of variant geographical locations.

Papaya crop originated in Southern Mexico and Costa Rica, remained cultivated in the USA, India, Brazil, Mexico, Nigeria, Jamaica, Indonesia, China, Taiwan, Peru, Thailand, and Philippines to a status of economy boost up crop[42]. The emergence and spread of atypical PRSV has essentially been pointed out to provide updated knowledge of the recombination and genetic variation required for defining management and prevention strategy before it possess any havoc to regional to global papaya production owing to its position attained as economy boost up crop in many cultivating countries.

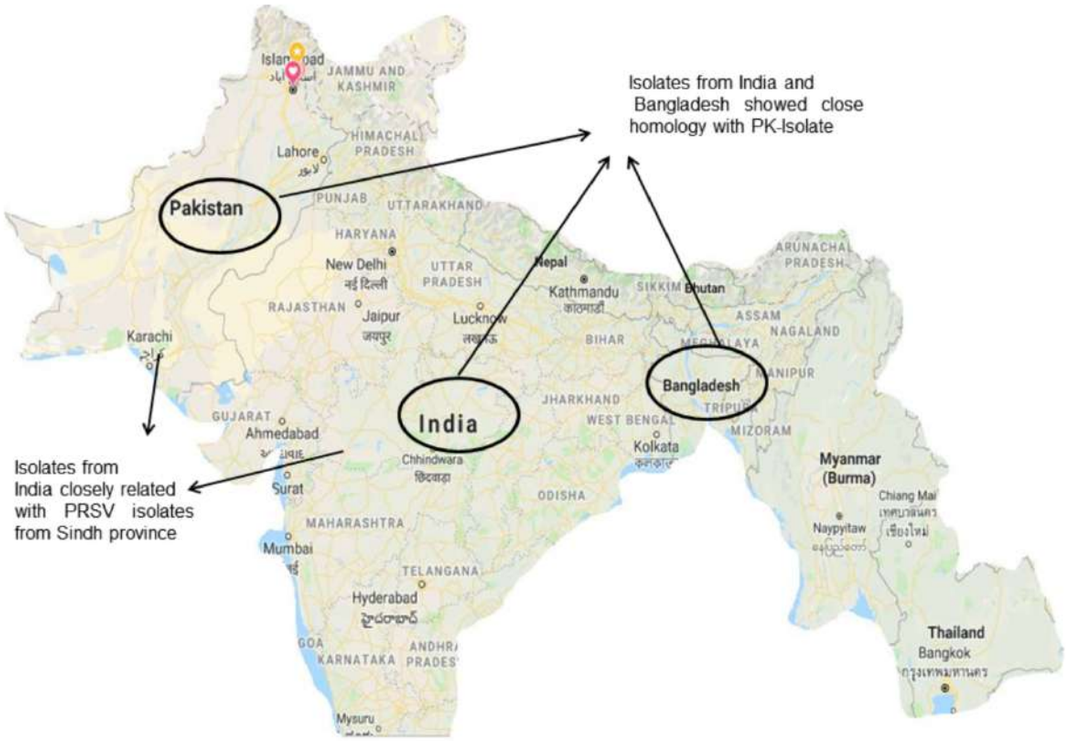

Figure 5: Map showing the isolates of PRSV–P of Pakistan compared on the basis of Coat protein sequence showed close homology to Bangladesh isolates and some of the Indian isolates

Materials and methods:

PRSV Source

The Papaya growing regions in Sindh province, Pakistan were surveyed during 2011-2018 and PRSV infectious source samples showing typical symptoms of mosaic, leaf distortion and mottling were collected. The samples were processed to screen and confirm the virus by ELISA and RT–PCR by essentially following the protocol as described earlier[12]. The samples, nucleic acids and lyophilized infectious material were stored at –80 ◦C for further processing. The virus isolates most pre dominantly found in all infected regions from individual papaya plant were considered as PRSV–P isolate PK (PRSV–PK).

Virus recovery and purification

The virus was recovered by mechanical inoculation of the PRSV infected samples, on plants of *Cucumis metuliferous*, *Chenopodium quinoa* and *Carica papaya* cv. Tainung 2 in glasshouse (28/22 ^{o}C day/night). The typical local lesions were induced on *C. quinoa* upon inoculation by the infected material. Later, the virus pure culture was established by subsequent single lesion transfers on *C. quinoa*. The local lesions were subsequently transferred to *C. quinoa* at least two times aiming to purify the virus and finally inoculated on Papaya and *C. metuliferous* to maintain PRSV–PK in propagative hosts.

RT–PCR, cloning and Genome sequencing of PRSV

Total RNA was extracted from leaves of the *Carica papaya* plants infected with PRSV-PK virus isolates using Trizol Reagent (Ambion, Life technologies). The cDNA was synthesized using Oligo dt (18) Not 1 primer, 2.5 mM dNTPs with M-MLV reverse transcriptase (Invitrogen) and 5X reaction buffer, $MgCl_2$, dTT and RNase block, the mixture was incubated at 42 ^{o}C for 1 hour followed by 10 min at 70 ^{o}C. The meta-genome sequencing was performed on the Illumina platform and analysis of contigs has revealed the first draft genome sequence of PRSV–P from Pakistan. To confirm the sequence draft subsequent amplification and sequencing were performed as detailed. KOD plus Taq polymerase and PRSV specific primers (Supplementary Table S1) were used at specific T_m and optimized PCR profiles. The amplified fragments were analyzed in 1% agarose gel in Tris Borate EDTA (TBE) buffer. The gel was visualized under UV transilluminator and the estimation of product size was made with 10 Kbp ladder. The PCR products were purified with Gene Mark PCR purification Kit and cloned in Topo Blunt

II vector (Invitrogen) using 1unit enzyme/1µg DNA as per manufacturer's protocol. The cloned plasmids were purified through Gene mark Plasmid Purification Kit and confirmed for insert through M13 Forward and M13 reverse primer amplification. The plasmids were subsequently sequenced. The ends of the virus genome were determined by RACE[43].

Full length CP gene was amplified using the primer pairs 8701 Fwd and 3' UTR Rev (Supplementary Table S1) with an amplicon size of about 1.5 Kbp. The sequencing of CP fragment and BLAST analysis enabled to design primer for subsequent amplification. The next amplicon was generated with forward primer from the middle of CIP and one internal CP reverse primer. PCR was performed using KOD polymerase with 94 °C initial denaturation for 2 min, 30 cycles of 15 s melting at 94 °C, 15 s annealing at 54 °C, and 3 min 50 sec extension at 68°C; with a final extension at 68°C for 10 min. Figure. 1 shows the illustration of the amplified fragment. For amplification of rest of the genome random primers forward and reverse, sequences mentioned in Supplementary Table S1, were designed to amplify the segment within HC-Pro with an amplicon of 799 bp. The predetermined sequence of HC-Pro gene was used to design specific primers to amplify the 5' genomic RNA of PRSV-P isolate PK by using 5'UTR forward paired with HC–Pro Rev 2. The final amplification was performed using the Primers HC-Pro forward coupled with PRSV-PK internal CP reverse which yield a big 7 Kbp fragment to complete the genome sequence. The product was directly sequenced via nanopore sequencing. All the primers sequences used in this study are mentioned in Supplementary Table S1and schematic amplification illustration for complete PRSV–P isolate PK is shown in the (Fig. 1).

Phylogenetic analysis of PRSV-P CP isolates and PK genome:

The evolutionary history of PRSV has been inferred with the use of Neighbor-Joining (NJ) method[44] for phylogenetic analysis in MEGAX[45]. In this phylogeographical analysis complete genomes, highly variable coat protein, P1, HC–Pro coding sequences, and highly conserved 3'UTR domain sequences of PRSV were used as the database. The evolutionary distance has been calculated as arm length in tree by the use of the maximum composite likelihood method[46]. A bootstrap test of 1,000 replicates was used to determine the of phylogenetic tree[47]. Alignment of nucleotide sequence was done using ClustalX2.

Primarily, full length individual coat protein gene sequences of PRSV isolates form Pakistan were compared with 30 differential PRSV CP sequences from America, China, Taiwan, India, Bangladesh, Vietnam, Mexico, Japan, Malaysia, Philippines, Thailand and Colombia obtained from GenBank

474 database. Similarly, for the genome sequence analysis of P1, HC-Pro and 3'UTR regions, sequences
475 were retrieved from full length genome sequence retrieved from GenBank, and aligned with individual
476 sequence of atypical PRSV-PK strainand phylogeographic analysis was conducted. For the PRSV
477 complete genome sequence analysis in this study, we have used representative PRSV–PK sequence and
478 compared it with full length genome sequences of 23 PRSV isolates from rest of the world retrieved
479 from GenBank database belonging to different phylogenetic groups on the basis of BLAST search.

480 The sources of CP, P1, HC-Pro, 3'UTR and whole genome sequences of PRSV isolate of Pakistan and
481 other countries used in this study are showed in Supplementary Table S2. The phylogenetic
482 reconstruction has been documented as trees (Fig. 2). The CP gene sequences of the nine PRSV
483 Pakistani isolates are available in GenBank (accession numbers)[12] and the complete genome sequence of
484 PRSV-P isolate PK has been submitted in this study.
485

486 **Genetic differentiation, Gene Flow and Recombination analysis:**
487 The DnaSP 6.12.03 program[48] was implemented to analyze genetic differentiation and the level of gene
488 flow between the populations of PRSV–PK based on CP and complete genome sequences with PRSV
489 populations from various geographical areas with the use of statistic Kst*[49]. Another Snn[50] test statistic
490 was used to measure the frequency of the nearest neighbor sequences in the same locality, whose values
491 may range from 1 (when populations from different localities are genetically distinct) to 1/2 in the case
492 of panmixia[50]. The gene flow assessment between PRSV population based on CP as well as full genome
493 was calculated using Fst[51]. Fst value (fixation index) indicating amount of inter-population diversity and
494 the value ranges from zero (indicating no differentiation between populations) to one (indicating full
495 differentiation between populations)[49]. The coefficient of Fst estimates the extent of inter-population
496 genetic differentiation[52]. An absolute value of Fst > 0.33 suggests gene flow is negligible, while absolute
497 value of Fst < 0.33 indicates gene flow is frequent[32].

498 PRSV Coat protein gene sequences of 40 isolates including Pakistan (10), Bangladesh (3), Indian
499 (5), Australian (2), American (8), Mexican (2), other Asian isolates (8) and Colombian isolates (2) were
500 divided into net number of 8 population. Full length PRSV genome sequences of 23 isolates including
501 the two isolates from Pakistan one is the PK–PRSV (MT090406) and other one modified PK–PRSV
502 sequence with the CP replaced with the previously reported CP sequence (JX024999)[12] and rest of the
503 21 sequences from India, America, Taiwan, China, Thailand and Colombia were analyzed for gene flow
504 assessment. Similarly all full length genomic sequences were grouped into net number of 5 Population

505 comprising Pakistan, Indian, Colombian, American and other Asian isolates (China, Taiwan and
506 Thailand). The PRSV population from Pakistan was compared with rest of the 4 populations
507 individually. Sites with alignment gaps were excluded. After defining populations, the genetic
508 differentiation and gene flow was analyzed using DnaSP6[48].

509 Recombination evidence in the genome sequences of PRSV was searched with the program
510 RDPv4.97 software package[53] that implements a range of recombination-detecting algorithms including
511 GENECONV[54]BOOTSCAN[55,56], MAXCHI[57,58], CHIMAERA[58], SISCAN[59], 3SEQ[60], and RDP[61]. In
512 RDP analysis, every combination of three sequences in the input alignment is sequentially tested to
513 demonstrate two sequences as a parent and third as a recombinant. The analysis of recombination was
514 performed, between and within groups (i.e. countries) of sequences and the events detected by at least
515 four different algorithms were accepted as evidence (breakpoint) of recombination. The recombination
516 effect was considered diligently during selection analysis. The possible recombination breakpoints were
517 also detected in particular event of each RDP method.
518 The first recombination analysis was performed between full length genome of PRSV–PK and 21
519 reported complete PRSV genome sequences from Hawaii, Taiwan, France, USA, Colombia, India,
520 China and Thailand. RDP was also used to check the recombination signals in the core CP region,
521 selecting events detected at least by four different methods as described above.

522

523 **Bibliography**

524 1. Schollaert Uz, S. *et al.* Earth observations and integrative models in support of food and water
525 security. *Remote Sens. earth Syst. Sci.* **2,** 18–38 (2019).

526 2. Harrison, B. D., Finch, J. T., Gibbs, A. J., Hollings, M., Shepherd, R. J., Valenta, V., & Wetter, C.
527 Sixteen groups of plant viruses. *Virology* **45,** 356–363 (1971).

528 3. Bateson, M. F. *et al.* On the evolution and molecular epidemiology of the potyvirus Papaya
529 ringspot virus. *J. Gen. Virol.* **83,** 2575–2585 (2002).

530 4. Gonsalves, D. Control of Papaya ringspot virus in papaya: A Case Study. *Annu. Rev. Phytopathol.*
531 **36,** 415–37 (1998).

532 5. Jensen DD. Papaya virus disease with special reference to papaya ringspot. *Phytopathology* **39,**
533 191–211 (1949).

534 6. Ferreira, S. A. *et al.* Virus coat protein transgenic papaya provides practical control of Papaya
535 ringspot virus in Hawaii. *Plant Dis.* **86,** 101–105 (2002).

536 7. Yeh, S. D. *et al.* Complete nucleotide sequence and genetic organization of Papaya ringspot virus

RNA. *J. Gen. Virol.* **73,** 2531–2541 (1992).

8. Tennant, P. *et al.* Papaya ringspot virus resistance of transgenic rainbow and SunUp is affected by gene dosage, plant development, and coat protein homology. *Eur. J. Plant Pathol.* **107,** 645–653 (2001).

9. Praveen, S., Kushwaha, C. M., Mishra, A. K., Singh, V. & Jain, R. K. Engineering tomato for resistance to tomato leaf curl disease using viral rep gene sequences. *Plant Cell. Tissue Organ Cult.* **83,** 311–318 (2005).

10. Ramesh, S. V, Mishra, A. K. & Praveen, S. Hairpin RNA-mediated strategies for silencing of Tomato Leaf Curl Virus AC1 and AC4 genes for effective resistance in plants. *Oliogonucleotides* **17,** 251–257 (2007).

11. Gibbs, A. J., Ohshima, K., Phillips, M. J. & Gibbs, M. J. The prehistory of potyviruses: Their initial radiation was during the dawn of agriculture. *PLoS One* **3,** e2523 (2008).

12. Naseem, S., Roofi, A., Zafar, Y. & Hafeez, F. Y. Detection and phylogenetic relationship of papaya ringspot virus-P in Pakistan. *J. Plant Pathol.* (2013). doi:10.4454/JPP.V95I3.043

13. Davis, M. J. & Ying, Z. *Development of papaya breeding lines with transgenic resistance to Papaya ringspot virus*. *Plant Disease* **88,** (2004).

14. Silva-Rosales, L., Becerra-Leor, N., Ruiz-Castro, S., Téliz-Ortiz, D. & Noa-Carrazana, J. C. 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 Brief Report. *Arch Virol* **145,** 835–843 (2000).

15. Quemada, H., L'Hostis, B., Gonsalves, D., Reardon, I. M., Heinrikson, R., Hiebert, E. L., Sieu, L. C & Slightom, J. L. The nucleotide sequences of the 3 ' -terminal regions of papaya ringspot virus strains W and P. *J. Gen. Virol.* **71,** 203–210 (1990).

16. Bateson, M. F., Henderson, J., Chaleeprom, W., Gibbs, A. J., & Dale, J. L. Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). *J. Gen. Virol.* **75,** 3547–3553 (1994).

17. Kumari, S., Trivedi, M. & Mishra, M. PRSV resistance in papaya (Carica papaya L.) through genetic engineering: A review. *J. Appl. Hortic.* **17,** 243–248 (2015).

18. Lius, S. *et al.* Pathogen-derived resistance provides papaya with effective protection against papaya ringspot virus. *Mol. Breed.* (1997). doi:10.1023/A:1009614508659

19. Bau, H. J., Cheng, Y. H., Yu, T. A., Yang, J. S., & Yeh, S. D. Broad-spectrum resistance to different geographic strains of Papaya ringspot virus in coat protein gene transgenic papaya. *Virology* **93,** 112–120 (2003).

20. Wu, Z., Mo, C., Zhang, S. & Li, H. Characterization of Papaya ringspot virus isolates infecting transgenic papaya 'Huanong No.1' in South China. *Sci. Rep.* **8,** 1–11 (2018).

572 21. Noa-Carrazana, J. C., González-De-León, D., Ruiz-Castro, B. S., Piñero, D. & Silva-Rosales, L.
573 Distribution of Papaya ringspot virus and Papaya mosaic virus in papaya plants (Carica papaya)
574 in Mexico. *Plant Dis.* **90,** 1004–1011 (2006).

575 22. Shukla, D.D., Ward, W.W & Brunt, A. . The Potyviridae. Wallingford UK: CAB International.
576 (1994).

577 23. Hamim, I. *et al.* Papaya Ringspot Virus isolates from Papaya in Bangladesh : detection ,
578 characterization , and distribution. **103,** 2920–2924 (2019).

579 24. Mangrauthia, S., Balasubramaniam, P., Jain, R. K. & Praveen, S. Role of genetic recombination
580 in the molecular architecture of Papaya ringspot virus. *Biochem. Genet.* **46,** 835–846 (2008).

581 25. Mishra, R. First complete genome sequence of a distinct Papaya Ringspot Virus isolate from the
582 Northeastern Region of India. *Genome Announc.* **6,** (2018).

583 26. Gorane, A. *et al.* Characterization of a severe isolate of papaya ringspot virus from papaya in
584 western India. *J. Plant Pathol.* **101,** 1203–1209 (2019).

585 27. Tripathi, S., Suzuki, J. Y., Ferreira, S. A. & Gonsalves, D. Papaya ringspot virus-P:
586 Characteristics, pathogenicity, sequence variability and control. *Mol. Plant Pathol.* **9,** 269–280
587 (2008).

588 28. Jain, R. K. *et al.* Variability in the coat protein gene of Papaya ringspot virus isolates from
589 multiple locations in India Brief Report. *Arch Virol* **149,** 2435–2442 (2004).

590 29. Ali, A., Natsuaki, T. & Okuda, S. Identification and molecular characterization of viruses
591 infecting cucurbits in Pakistan. *J. Phytopathol.* **152,** 677–682 (2004).

592 30. Moradi, Z., Mehrvar, M., & Nazifi, E. Population Genetic Analysis of Zucchini yellow mosaic
593 virus based on the CI Gene Sequence. *J. Cell Mol. Res.* **10,** 76–89 (2019).

594 31. Hajizadeh, M., & Sokhandan-Bashir, N. Population genetic analysis of potato virus X based on
595 the CP gene sequence. *VirusDisease* **28,** 93–101 (2017).

596 32. Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. DnaSP , DNA polymorphism
597 analyses by the coalescent and other methods. *Bioinfomatics* **19,** 2496–2497 (2003).

598 33. Hamid, A., Zhai, Y., Ramesh, S. V., & Pappu, H. R. Complete genome characterization and
599 population dynamics of potato virus Y-NTN strain from India. *VirusDisease* **30,** 252–260 (2019).

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

602 35. Scheffers, B. R. et al. The broad footprint of climate change from genes to biomes to people.
603 *Science (80-.).* **354,** 6313 (2016).

604 36. Gilbert, G. S. Evolutionary ecology of plant diseases in natural ecosystems. *Annu. Rev.*
605 *Phytopathol.* **40,** 13–43 (2002).

606 37. Basavaraj, Y. B., Kumar, A., Jain, R. K., Kumar, P., Parameswari, B., & Prakash, J. Molecular
607 diversity of Papaya ringspot virus in India : genetic recombination and mutations between the
608 isolates from different hosts and geo-climatic locations are role players in virus evolution. *Indian*
609 *Phytopathol.* **72,** 497–511 (2019).

610 38. Chaves-bedoya, G. & Ortiz-rojas, L. Y. Genetic variability of Papaya ringspot virus isolates in
611 Norte de Santander - Colombia Variabilidad genética del virus de la mancha anular de la papaya
612 en Norte de Santander - Colombia. *Agron. Colomiana* **33,** 184–193 (2015).

613 39. Mishra, R., Patil, S., Patil, A. & Patil, B. L. Sequence diversity studies of papaya ringspot virus
614 isolates in South India reveal higher variability and recombination in the 5 0 -terminal gene
615 sequences. *VirusDisease* **30,** 261–268 (2019).

616 40. Ohshima, K. *et al.* Patterns of recombination in turnip mosaic virus genomic sequences indicate
617 hotspots of recombination. *J. Gen. Virol.* **88,** 298–315 (2007).

618 41. Castillo, X. O. et al. Phylogeography and molecular epidemiology of Papaya ringspot virus. *Virus*
619 *Res.* **159,** 132–140 (2011).

620 42. Azad, M. A. K., Amin, L. & Sidik, N. M. Gene technology for papaya ringspot virus disease
621 management. *The Scientific World Journal* (2014). doi:10.1155/2014/768038

622 43. Li, Z., Yu, M., Zhang, H., Wang, H. Y. & Wang, L. F. Improved rapid amplification of cDNA
623 ends (RACE) for mapping both the 5′ and 3′ terminal sequences of paramyxovirus genomes. *J.*
624 *Virol. Methods* **130,** 154–156 (2005).

625 44. Saitou, N. & Nei, M. The neighbor-joining method : a new method for reconstructing
626 phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425 (1987).

627 45. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary
628 genetics analysis across computing platforms. *Mol. Biol. Evol.* **35,** 1547–1549 (2018).

629 46. Tamura, K., Nei, M. & Kumar, S. Prospects for inferring very large phylogenies by using the
630 neighbor-joining method. *Proc. Natl. Acad. Sci.* **101,** 11030–11035 (2004).

631 47. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N.*
632 *Y).* **39,** 783–791 (1985).

633 48. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins,
634 S. E., & Sánchez-Gracia, A. DnaSP 6 : DNA sequence polymorphism analysis of large data sets.
635 *Mol. Biol. Evol.* **34,** 3299–3302 (2017).

636 49. Hudson, R. R., Boos, D. D., & Kaplan, N. L. A statistical test for detecting geographic
637 subdivision. *Mol. Biol. Evol.* **9,** 138–151 (1992).

638 50. Hudson, R. R. A new statistic for detecting genetic differentiation. *Genetics* **155,** 2011–2014
639 (2000).

640 51. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure.

Evolution (N. Y). **38,** 1358–1370 (1984).

52. Tsompana, M., Abad, J., Purugganan, M., & Moyer, J. W. The molecular population genetics of the Tomato spotted wilt virus (TSWV) genome. *Mol. Ecol.* **14,** 53–66 (2005).

53. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4 : Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1,** 1–5 (2015).

54. Padidam, M., Sawyer, S. & Fauquet, C. M. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265,** 218–225 (1999).

55. Salminen, M. O., Carr, J. K., Burke, D. S., & Mccutchan, F. E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11,** 1423–1425 (1995).

56. Martin, D. P., Posada, D., Crandall, K. A., & Williamson, C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retroviruses* **21,** 98–102 (2005).

57. Smith, J. M. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34,** 126–129 (1992).

58. Posada, D. & Crandall, K. A. Evaluation of methods for detecting recombination from DNA sequences : Computer simulations. *Proc. Natl. Acad. Sci.* **98,** 13757–13762 (2001).

59. Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. Sister-Scanning : a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinfomatics* **16,** 573–582 (2000).

60. Boni, M. F., Posada, D. & Feldman, M. W. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176,** 1035–1047 (2007).

61. Martin, D. & Rybicki, E. RDP: detection of recombination amongst aligned sequences. *Bioinforma. Appl. note* **16,** 562–563 (2000).

Acknowledgements

We are greatly thankful to Higher Education Commission Pakistan (HEC) for funding this work as well as for supporting the research carried out under International Research Support Initiative program (IRSIP). Authors are also thankful to research and technical team at National Chung Hsing University (NCHU) Taichung, Taiwan for hosting IRSIP for 6 months.

Author Contributions

AS and AB conducted the experimental work and compiled the literature. ZA and MFA had conducted the surveys, sampling and disease report data. ZA helped in the execution of the research work, and the revision of the final manuscript. WS helped in revision of final manuscript. SN, SDY and ZA discussed the results. SN designed the strategy, supervised the overall research work, helped in sequence analysis and interpretation of the results. All the authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Data Availability

Supplementary data is available as Supplementary Table S1 and S2.

List of Figures :

1. **Figure 1:** Schematic amplification of PRSV–PK genome) CP gene amplification; b) 4K, HC-Pro and full HC-Pro amplification; c) 5' terminal end and 7K amplification; d) full genome amplification of PK-PRSV.
2. **Figure 2:** Neighbor-joining tree representing phylogenetic relationships of Papaya ringspot virus-P strain PK (PRSV-PK) **(a)** Coat Protein sequence **(b)** Protease P1 sequence **(c)** Helper component HC–Pro Sequence **(d)** 3'UTR sequence from other related sequences selected via BLAST search. Upper and lower branch points show bootstrap values (1,000 replicates) supporting a particular phylogenetic group. The scale bar represents nucleotide substitutions per site. All nucleotide sequences are retrieved according to the isolate name and the GenBank accession number**.**
3. **Figure 3:** Region specific neighbor joining tree representing phylogenetic relationship between complete genome sequences of PRSV-P isolate of Pakistan and PRSV complete genome sequence from rest of the world. Tree was constructed using ClustalX2 and MegaX Program. The scale bar represents nucleotide substitution per site, Boot strap value of 1,000.
4. **Figure 4:** Approximate genetic variation in the individual gene fragments of PRSV–PK
5. **Figure 5:** Map showing the isolates of PRSV–P of Pakistan compared on the basis of Coat protein sequence showed close homology to Bangladesh isolates and some of the Indian isolates

List of Tables:

Table 1: Statistical tests for genetic differentiation and gene flow between *Papaya ringspot virus* populations from Pakistan with the populations from Colombia, Mexico, America, Asia, Bangladesh, Australia and India based on the CP gene nucleotide sequences. ns, not significant; *0.01 < P < 0.05; **0.001 < P < 0.01; ***P < 0.001. Kst*, Snn and Fst were implemented in DnaSP 6. The deviation hypothesis from null population differentiation was tested by 1000 permutations of the raw data.

Table 2: Statistical tests for genetic differentiation and gene flow between *Papaya ringspot virus* populations from Pakistan with the populations from Colombia, Mexico, America, Asia, Bangladesh, Australia and India based on the full length genomic gene sequences. ns, not significant; *0.01 < P < 0.05; **0.001 < P < 0.01; ***P < 0.001. Kst*, Snn and Fst were implemented in DnaSP 6. The deviation hypothesis from null population differentiation was tested by 1000 permutations of the raw data.

Table 3: Recombination detection in full length atypical PRSV-P Isolate PK (MT090406). The break points listed refer to their position in the alignment; RDP(R), GENECONV(G), BOOTSCAN(B), MaxChi(M), CHIMAERA(C), SiSCAN(S), 3SEQ(3)

Table 4: Recombination detection in CP gene of atypical PRSV–P Isolate PK (MT090406). The break points listed refer to their position in the alignment; RDP(R), GENECONV(G), BOOTSCAN(B), MaxChi(M), CHIMAERA(C), SiSCAN(S), 3SEQ(3), and Lard (L).

Supplementary Table S1: Primers used for schematic amplification of PRSV-P Isolate PK genome

Supplementary Table S2: The sources of CP, P1, HC-Pro, 3'UTR and whole genome sequences of PRSV isolate of Pakistan and other countries used in the study

Figures

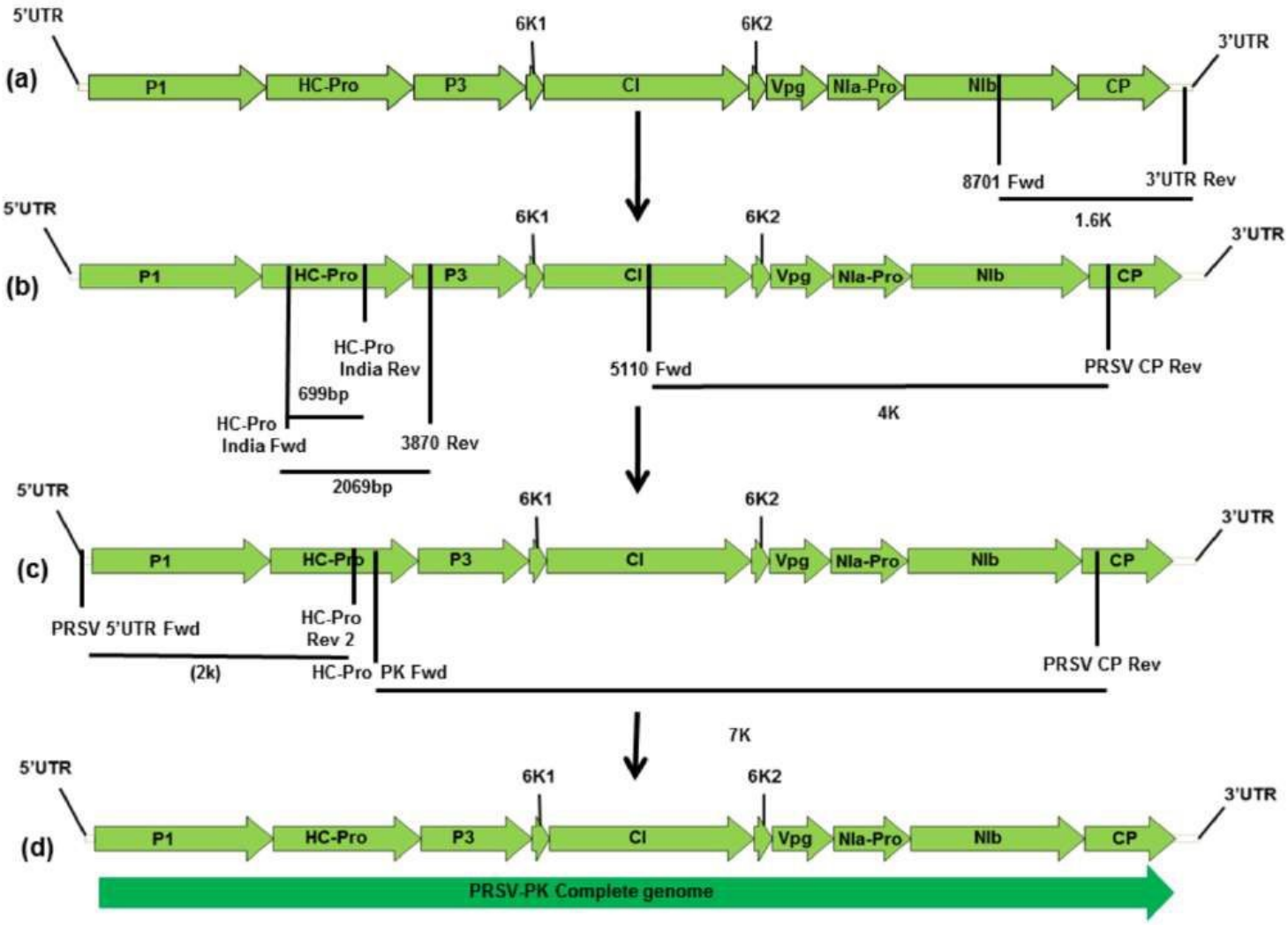

Figure 1

Schematic amplification of PRSV–PK genome (a) CP gene amplification; (b) 4K, HC–Pro and full HC–Pro amplification; (c) 5' terminal end and 7K amplification; (d) Full genome amplification.

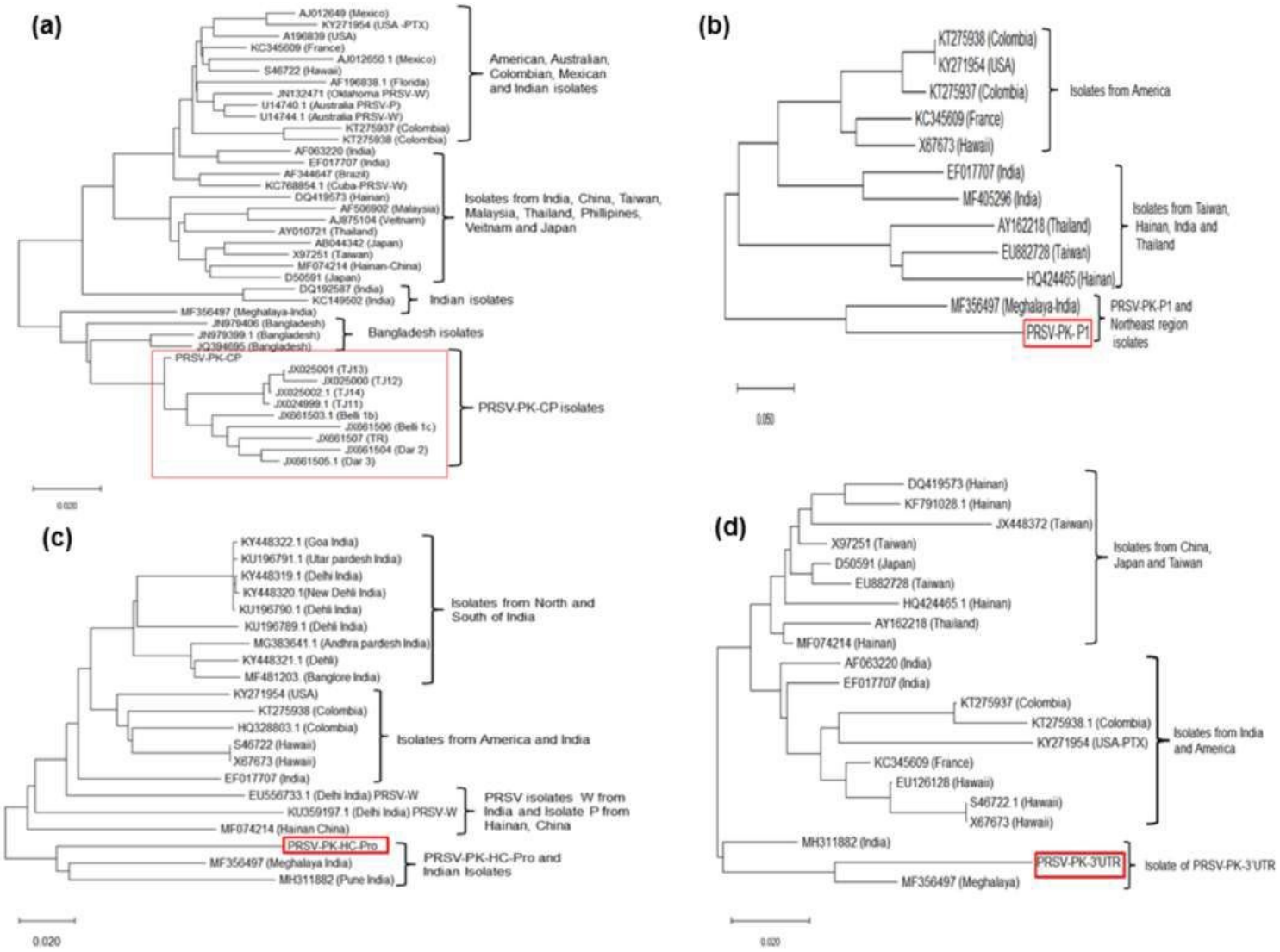

Figure 2

Neighbor-joining tree representing phylogenetic relationships of Papaya ringspot virus-P strain PK (PRSV–PK) (a) Coat Protein (CP) sequence (b) Protease P1 sequence (c) Helper component HC-Pro Sequence (d) 3'UTR sequence from other related sequences selected via BLAST search. Upper and lower branch points show bootstrap values (1,000 replicates) supporting a particular phylogenetic group. The scale bar represents nucleotide substitutions per site. All nucleotide sequences are retrieved according to the isolate name and the GenBank accession number.

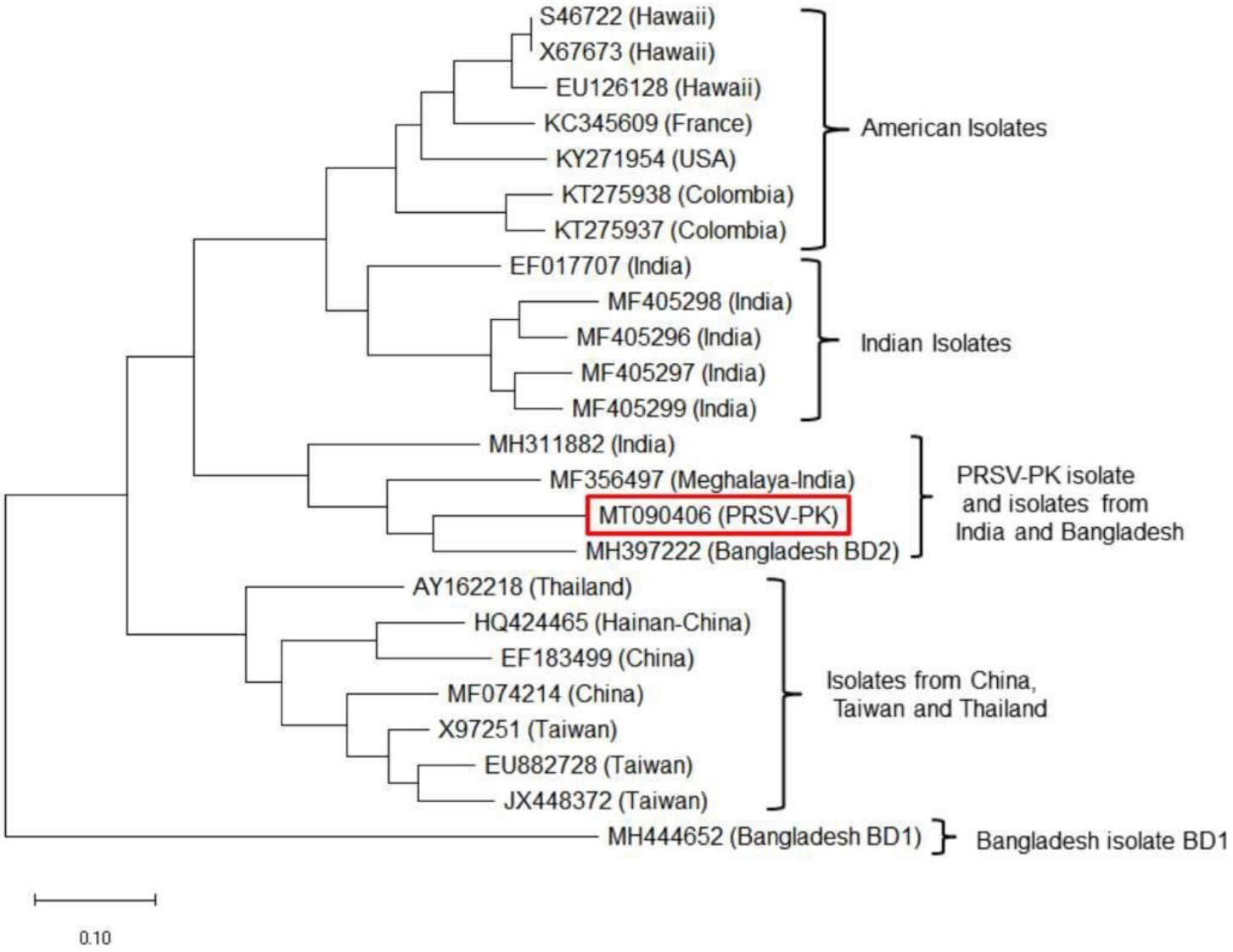

Figure 3

Region specific neighbor joining tree representing phylogenetic relationship between complete genome sequences of PRSV-P isolate of Pakistan and PRSV complete genome sequence from rest of the world. Tree was constructed using ClustalX2 and MegaX Program. The scale bar represents nucleotide substitution per site, Bootstrap value of 1,000.

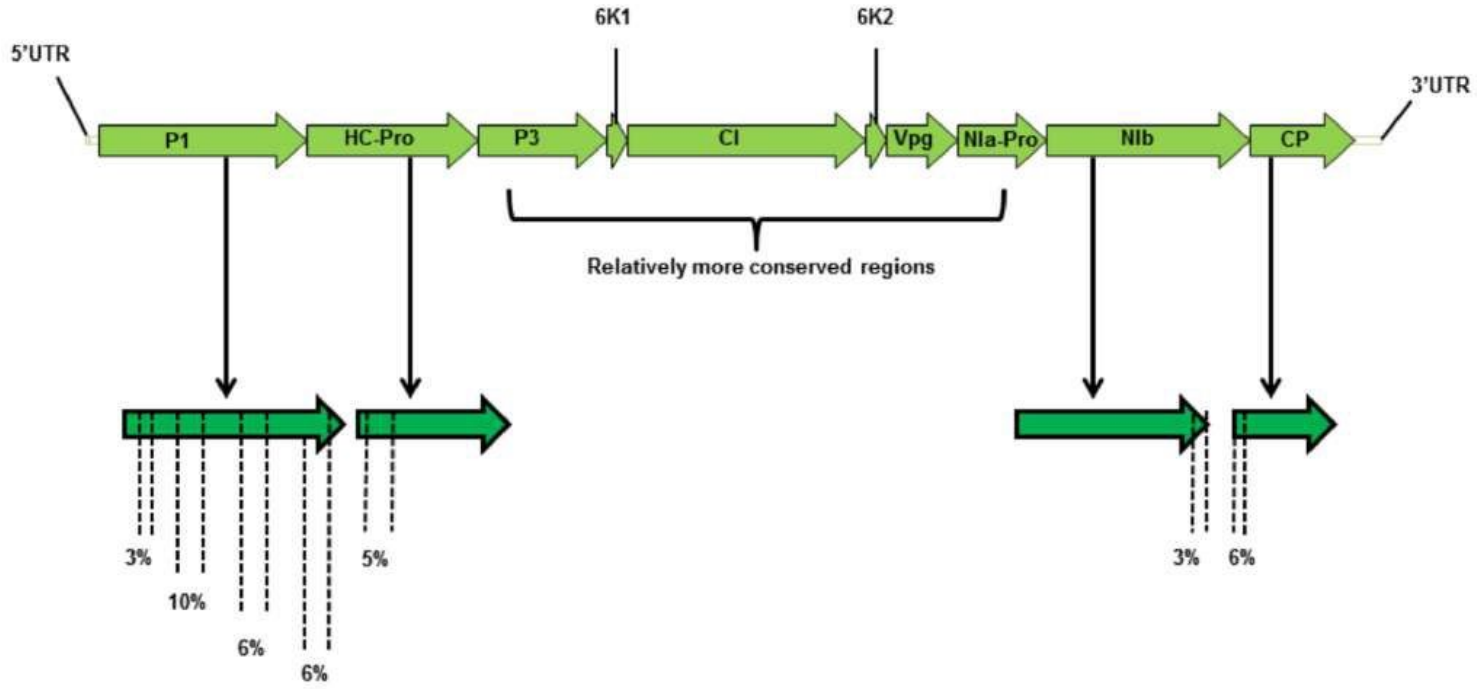

Figure 4

Approximate genetic variation in the individual gene fragments of PRSV-PK.

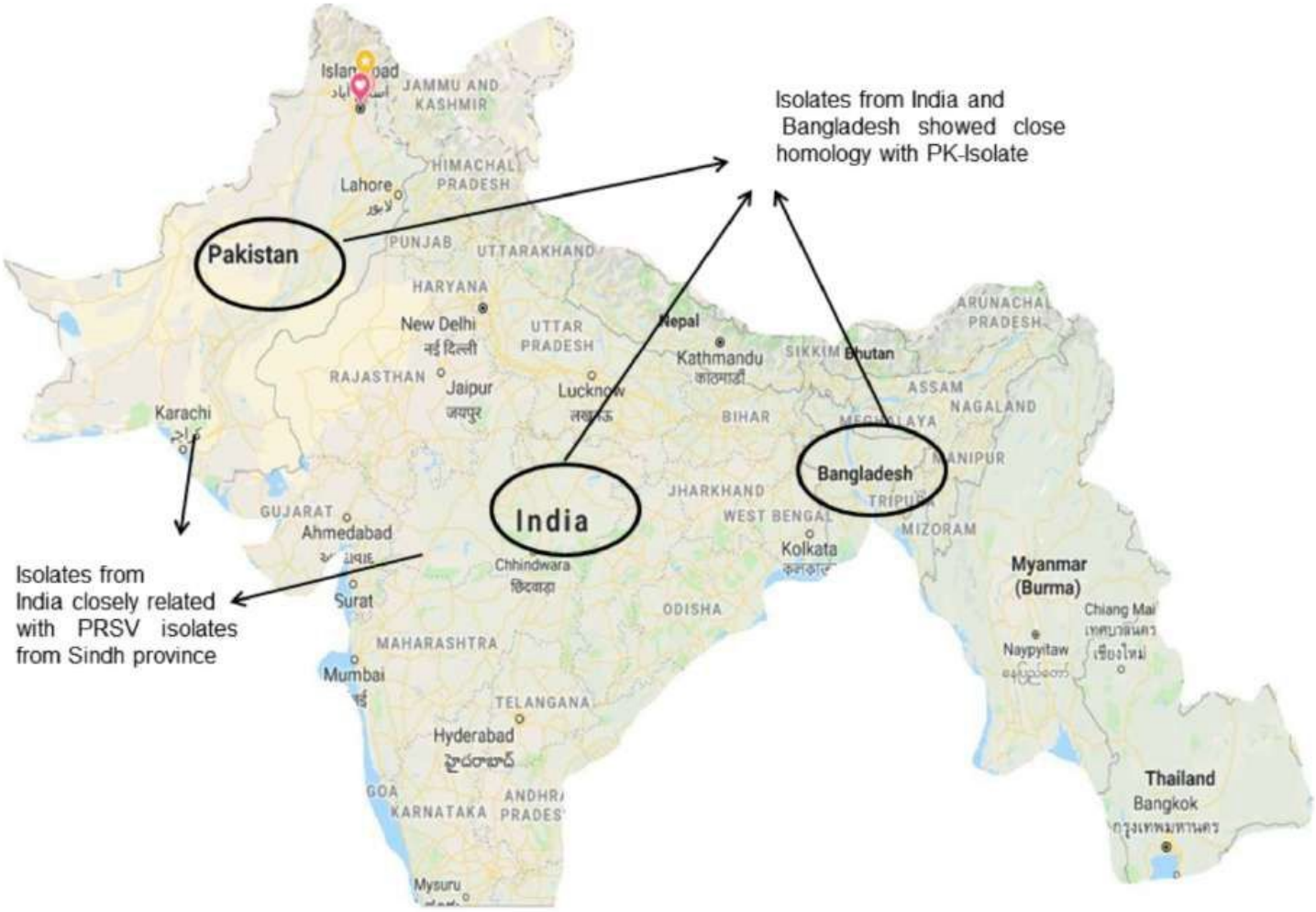

Figure 5

Map showing the isolates of PRSV–P of Pakistan compared on the basis of Coat protein sequence showed close homology to Bangladesh isolates and some of the Indian isolates. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablefilePRSVScientificreports1129.docx