Phylogeny and synonymous codon usage pattern of *Papaya ringspot virus* coat protein gene in sub-Himalayan region of north-east India
Phylogeny and synonymous codon usage pattern of *Papaya ringspot virus* coat protein gene in sub-Himalayan region of north-east India

Authors

1. **Prosenjit Chakraborty**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: nбу.prosen@yahoo.in
2. **Shibu Das**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: nбу.sid88@yahoo.in
3. **Bikram Saha**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: bik_ram2007@yahoo.co.in
4. **Piyali Sarkar**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: piyalisrkr93@gmail.com
5. **Arup Karmakar**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: nбу.arup@yahoo.in
6. **Arnab Saha**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: arnabsaha.svm@rediffmail.com
7. **Dipanwita Saha**, Department of Biotechnology, University of North Bengal, Siliguri-734013, India, e-mail: dsahanbu@yahoo.co.in
8. **Aniruddha Saha (Corresponding author)**, Department of Botany, University of North Bengal, Siliguri-734013, India, e-mail: asahanbu@yahoo.co.in, Ph. No.: +919832372105
Abstract

Sub-Himalayan West Bengal is favorable for production of several fruits and vegetables. Papaya is one of the common plants cultivated in the area. Most of the papaya plants of the area are susceptible to PRSV. Coat protein genes of six *Papaya ringspot virus* (PRSV) isolates of the area have been sequenced following RT-PCR. Phylogenetic study of the PRSV isolates showed about 80-90% similarity with Cuban isolates also. The codon usage pattern of our isolates has also been analyzed along with several other isolates. PRSV isolates of our study showed more preference to 8 putative optimal codons. Correspondence analysis of the genes of different isolates along the first two major axes were done, as the first two axes contributed more in shaping codon usage pattern. In the phylogenetic tree, constructed by neighbour joining method our isolates clustered together with the east Indian, north Indian and Bangladeshi isolates. The diversity and codon usage pattern of the PRSV isolates of different regions were studied, where it has been observed that the codon usage pattern of PRSV isolates is influenced probably by translational selection along with mutational bias.

**Keywords:** Codon usage, mutational bias, translational selection, codon preference, correspondence analysis.
**Introduction**

Papaya (*Carica papaya* L.) is a member of the small family Caricaceae and is cultivated widely for its edible and industrial value. In 2010, the global production of papaya is 11,223,031 metric tonnes produced in 60 countries worldwide in a total area of 3,71,345 hectares, with an average yield of 18,6810 kg/ha. India stands first in the production contributing 38.61 per cent of global production followed by Brazil, Indonesia and Nigeria (Evans and Ballen 2012). *Papaya ringspot virus* (PRSV) infecting papaya was first reported in Oahu in 1938 (Parris 1938). Subsequently the virus was reported from every continent. In India, it was first reported in west India in 1958 (Capoor and Varma 1958). Surekha et al. (1977) recorded incidence of PRSV in Udaipur of Rajasthan and Marathwada region of Maharashtra. Khurana and Bhargava (1970) during their survey observed 75-100 % incidence of PRSV in and around Ghorakhpur district of Uttar Pradesh. Occurrence of PRSV from other regions of India like, Tamil Nadu (Sharma et al. 2005), Andhra Pradesh, Himachal Pradesh, Jharkhand, Karnataka Maharashtra (Jain et al. 2004) have also been reported previously. Although disease incidence was noted from southern part of West Bengal (Jain et al. 2004), but reports from sub-Himalayan West Bengal (Saha et al. 2014) are scanty.

*Papaya ringspot virus* is a member of the genus *Potyvirus* belonging to the family *Potyviridae*. PRSV is a non-enveloped, flexuous rod (760-800 nm×12 nm) consisting of positive sense single stranded RNA (ssRNA) genome with a VPg (a genome-linked protein) covalently bound to the 5' end and a 3' poly-A tail (Gonsalves and Ishii 1980; Wang and Yeh 1997). The coat protein (CP) gene sequence diversity is observed in different classes of the viruses (Dolja et al. 1992). Coat protein molecules of most of the RNA viruses are encoded by only one gene. Synonymous codon usage has been studied in many plant viruses because of their high genetic diversity. According to Adams and Antoniw (2004) the mutational bias
is believed to be one of the major factors that influence the codon usage amongst the plant viruses. Although, the codon usage pattern of virus is affected by both mutational bias and translational selection (Wright 1990). In the present study, synonymous codon usage biasness of the CP gene sequences of PRSV isolates is being reported along with their sequence diversity.

**Materials and methods**

**Collection of pant samples, maintenance of plants and mechanical sap inoculation**

On the basis of disease symptoms (leaf curl, leaf distortion, mosaic and stunted growth), observed in the papaya growing fields of sub-Himalayan region of north-east India (Figure 1), 35 infected leaf samples were collected. The present study area includes six districts (Darjeeling, Jalpaiguri, Cooch Behar, Uttar Dinajpur, Dakshin Dinajpur and Malda districts) of West Bengal. The plants were grown under natural conditions and maintained throughout the growing season. For maintenance of infected and healthy plants in the experimental garden, both healthy and infected plants were covered separately with mosquito net and maintained in the experimental garden of the Department of Botany, University of North Bengal, so that, contamination can be prevented. Mechanical sap inoculation of the infected papaya plants were done following the method of Reddy et al. (2007).

**Extraction of total RNA and Reverse Transcription Polymerase Chain Reaction (RT-PCR)**

Total RNA were extracted from the infected and healthy plants following the method of Ghawana et al. (2011). Reverse Transcription Polymerase Chain Reaction (RT-PCR) were done from the extracted total RNA by ‘One Step M-MuLV RT-PCR kit’ (GeNei, Bangalore) using CPuP (TGAGGATCCTGGTGYATHGARAAYGG) and P9502.
(GCGGATCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT) primer pair (N= T+C+A+G; Y= T+C; H= T+C+A; R= A+G) as reported by van der Vlugt et al. (1999) following manufacturer’s protocol. Agarose gel electrophoresis were done at 5V/cm through 1.2% (w/v) agarose gels in 1X Tris Acetic acid EDTA (TAE) buffer for 60min for visualization of RNA and PCR product under GeNei UV- Transilluminator.

**Cloning and nucleotide sequencing of the PCR product**

All the purified PCR products were cloned into the pGEM T-easy vector (Promega, Madison, USA) following the method of Sambrook and Russel (2001). The cloned inserts were sent to Chromous Biotech Pvt. Ltd. for sequencing (from both in forward and reverse directions).

**Dataset**

The coat protein gene sequences of 60 PRSV isolates were downloaded from GenBank database for sequence analysis (Table 1). GC content is the frequency of guanine (G) and cytosine (C) in a coding gene. GC1, GC2 and GC3 content is the frequency of guanine (G) and cytosine (C) of a codon at the first, second and third position respectively. Whereas, A3, C3, T3 and G3 are the frequencies of adenine (A), cytosine (C), thymine (T) and guanine (G) at the third position of a codon in a coding gene (Xu et al. 2008). Codon adaptation index (CAI) is the measure of high or low expression of a gene. It can range from 0 to 1, where higher value indicates highly expressed genes (Sharp et al. 1986). L_sym and L_aa indicate number of synonymous codons and number of translatable codons respectively. GRAVY is the general average hydropathicity and Aromo is the frequency of aromatic amino acids in hypothetically translated gene product (Wei et al. 2014).
Sequence analysis

The nucleotide and deduced amino acid sequences were compared with the corresponding sequences of other PRSV isolates deposited in the GenBank database (http://www.ncbi.nlm.nih.gov) using the BLAST programme (Altschul et al. 1997). Sequences obtained by using forward and reverse primers were aligned using CLUSTAL W (Thompson et al. 1994) to obtain the correct alignment, sequence identity matrix was generated using BioEdit version 7.2.5 (Hall 1999) and tree was generated by neighbour-joining method and Kimura-2 parameter using MEGA version 6.0 (Tamura et al. 2013). A3, T3, G3, C3, GC and GC3 contents, \( N_C \) values and RSCU values were calculated using CodonW version 1.4.2 (Peden 1999). GC1 and GC2 contents were calculated using Codon O (Angellotti et al. 2007). Correlation analysis was done using SPSS 19.0.

Results and discussion

Symptomatology

Sub-Himalayan region of north-east India have a conducive climate for production of several fruits and vegetables (Saha et al. 2014). Prevalence of PRSV in sub-Himalayan region revealed that, the cultivated papaya varieties of this area are very much prone to virus attack. Several papaya plants with vein clearing, leaf curling, leaf deformation and ring spots (on fruits) were observed throughout the present study area. Stunted growth of the plants and severe losses of produces were experienced (Figure 1). The observed symptoms are very much similar to the previously reported PRSV isolates (Lindner et al. 1945; Bhandari 1952; Dahal et al. 1997; Roy et al. 1999, Saha et al. 2014). Mechanical sap inoculation of PRSV to healthy plants also confirmed the presence of virus in the samples. Similar types of symptoms were also observed after mechanical sap transmission as reported previously (Yemewar and Mali 1980).
RT-PCR detection and sequencing of the coat protein gene

RT-PCR of the infected plants with universal primer pair CPuP and P9502 gave an expected amplicon of ~650bp (Figure 2). Six samples from different locations of sub-Himalayan north-east India were sequenced (Table 1). All the six sequences showed 91% to 95% sequence similarity with *Papaya ringspot virus* during BLAST analysis. These results further confirm the presence of the virus in the present study area (Singh et al. 2007).

Base composition analysis of the coat protein gene

To analyze the base composition variation of PRSV isolates, the base composition of the coat protein genes from different isolates were calculated (Table 2). From the results it is evident that, the Cuban isolates contains much more GC (0.47), GC1 (0.48), GC2 (0.41) content and the GC3 content (0.45) along with higher G3 (0.32) and C3 (0.27) contents. This indicates that the Cuban isolates have special codon usage pattern. The C3 content is also higher in Sri Lankan and Thai (0.27) isolates. The T3 content is higher in Brazilian isolate (0.48), whereas, the A3 content is higher in Bangladeshi isolate (0.38). GC1 content is higher in east Indian isolates (0.49). Our isolates have average GC3 content (0.39) but the GC (0.47), GC1 (0.49) and GC2 (0.41) content is quite high like the Cuban isolates. The T3, C3, A3 and G3 contents of our isolates are 0.44, 0.26, 0.33 and 0.26 respectively like the east Indian isolates. The C3, T3, G3, GC and GC3 contents of our isolates were very much similar to that of isolates of east India and Bangladesh, indicating that the codon usage pattern of our isolates are not much different from other east Indian and Bangladeshi isolates. A relationship between the codon preferences of viruses in the same family or genus, regardless of host or genomic nucleotide content is reported previously by Cardinale et al. (2013).

Factors influencing codon usage
Several factors such as t-RNA abundance (Kanaya et al. 2001), m-RNA and protein structure (Knight et al. 2001), random genetic drift (Bulmer 1991; Sharp and Li 1986), replicational, transcriptional and translational bias (Hershberg and Petrov 2008) and other environmental factors (Behura et al. 2013) play vital roles in shaping codon usage pattern in different organisms. The $N_C$ plot (i.e., a plot of $N_C$ vs GC3 content) was used to investigate the codon usage variations among different isolates (Figure 3(A)). According to Wright (1990) and several other scientists (Adams and Antoniw 2004; Xu et al. 2008; Zhang et al. 2011; Belalov and Lukashev 2013), if the GC3 content of the gene fall on the continuous curve of $N_C$ value and GC3 contents, the codon usage pattern is constrained only by mutation bias. But, if the codon choice is influenced by other factors also, viz., translational selection, gene length, and gene function along with mutation bias the values will lie below the curve. In our study, most of the values lie below the standard curve ($N_C$ values of 50.22-61.00 at the GC3 values 0.36-0.41) indicating an involvement of other factors such as translational bias (Wright 1990; Adams and Antoniw 2004). Therefore, it is obvious that codon usage bias of PRSV isolates is affected by both total nucleotide compositional constraints (i.e., mutational bias) and translational selection and other associated factors (Wright 1990; Guo et al. 2007; Zhang et al. 2011). Xu et al. (2008) proposed that, the selection pressure have major role in shaping mutational bias, as the direct change in the nucleotide sequences may be harmful for the virus itself. Stoletzki and Eyre-Walker (2007) also proposed that, the conserved genes and coding sequences shows higher codon bias. According to Jain et al. (1998; 2004) the CP sequences may be helpful to determine the sequence diversity within PRSV isolates from different geographical locations. Thus, from that point of view the codon usage pattern of the CP gene sequence of PRSV isolates are of great importance.

To examine the correlation among different factors influencing codon usage Pearson correlation test was done (Table 3). It has also been observed that the effective number of
codon \( (N_C) \) is significantly correlated with the gene length \( (L_{aa}) \), where correlation coefficient is \(-0.773\) \((p<0.01)\). Thus, an increasing gene length may increase codon bias (Figure 3(B)). The relationship among GC1, GC2 and GC3 content was further calculated to study the reason behind \( N_C \) variation under same GC3 content. It has been observed that the coat protein genes have always higher GC1 content than GC2 content (Figure 3(C)). It seems that GC2 content and GC3 content is positively correlated (correlation coefficient 0.671, \( p<0.01 \)).

**Correspondence analysis (CA)**

Correspondence analysis is an important and commonly used multivariate statistical analysis where the genes were plotted in a 59-dimensional hyperspace corresponding to the usage of the 59 sense codons (except for Met, Trp and Stop codons) (Greenacre, 1984). The first two coordinates i.e., Axis 1 and Axis 2 contribute to 22.65% and 14.43% of the total variation (Figure 4(A)). Thus, correspondence analysis of the coat protein genes along Axis 1 and Axis 2 may be helpful in analyzing the codon variation (Figure 4(B)). The distribution of the genes along the first two major axes showed that, our isolates clustered with some of the east Indian isolates indicating a common codon usage pattern (Sharp et al. 1988).

**Synonymous codon usage bias**

In order to investigate the codon usage bias in coat protein gene of PRSV isolates, relative synonymous codon usage (RSCU) value was calculated (Table 4). The ‘\( ^{c*} \)’ marked values in Table 4 indicate the more preferred codons than other synonymous codons and the ‘\( ^{d*} \)’ marked codons are the putative optimal codons. In this case, there are 23 preferred codons and 8 optimal codons (UAC, UGC, UCG, CUA, ACC, AAG, GAC and GGG) are acting together to form the bias.
Phylogenetic analysis

The sequence similarity of the coat protein sequences of the PRSV isolates from different geographical locations are shown in Figure 5. They were compared with the other PRSV isolates of India and abroad and a phylogenetic tree was constructed (Figure 6). In the phylogenetic tree our isolates clustered together with the east Indian, north Indian and Bangladeshi isolates.

During phylogenetic analysis, CP gene sequence revealed that, our PRSV isolates and the isolates of sub-Himalayan India and Bangladesh are clustering together. The present study revealed the sequence variation within PRSV population from different geographical origin. Jain et al. (1998; 2004) proposed that the Indian PRSV isolates are the most divergent virus group worldwide. And our results support the proposed hypothesis. Again, the results of our study may have great impact on virus management strategy and coat protein gene derived transgenic (Jain et al. 2004).

In the present study, we analyzed the diversity and codon usage pattern of the PRSV isolates. The codon usage pattern of PRSV isolates are influenced by translational selection and mutational bias, where translational selection along with mutational bias also plays a major role. However, in future the importance of geographical position along with climatic condition on viral sequence diversity and codon usage pattern in addition to the mutational selection need to be studied in a broader spectrum.

Acknowledgements

This work was supported by the University Grants commission, New Delhi under the major research project grant [F.-41 No.-392/2012(SR) dated 16-07-2012].
References


Table 1. Coat protein genes used in the study [Code names of some isolates are indicated within parenthesis].

<table>
<thead>
<tr>
<th>Accession No.</th>
<th>Place of occurrence</th>
<th>Accession No.</th>
<th>Place of occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Isolates</strong></td>
<td></td>
<td><strong>Central Indian Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>KF918721</td>
<td>Darjeeling, West Bengal (S)</td>
<td>DQ354071</td>
<td>Chhattisgarh (CH)</td>
</tr>
<tr>
<td>KF918720</td>
<td>Uttar Dinajpur, West Bengal (I)</td>
<td>AY491011</td>
<td>Chhattisgarh (CH)</td>
</tr>
<tr>
<td>KF906262</td>
<td>Dakshin Dinajpur, West Bengal (B)</td>
<td>KC149501</td>
<td>Madhya Pradesh (MP)</td>
</tr>
<tr>
<td>KP238490</td>
<td>Malda, West Bengal (M)</td>
<td>DQ650651</td>
<td>Madhya Pradesh (MP)</td>
</tr>
<tr>
<td>KP238489</td>
<td>Jalpaiguri, West Bengal (J)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KP238488</td>
<td>Coochbehar, West Bengal (C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>East Indian Isolates</strong></td>
<td></td>
<td><strong>Chinese Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>JX567310</td>
<td>West Bengal (WB)</td>
<td>AF469066</td>
<td>China</td>
</tr>
<tr>
<td>JN979406</td>
<td>West Bengal (WB)</td>
<td>KF033096</td>
<td>China</td>
</tr>
<tr>
<td>KC149500</td>
<td>Assam (AS)</td>
<td>KF033095</td>
<td>China</td>
</tr>
<tr>
<td>EF210196</td>
<td>Bihar (BI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JQ394693</td>
<td>Jharkhand (JK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY458619</td>
<td>Jharkhand (JK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQ354072</td>
<td>Sikkim (SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brazilian Isolates</strong></td>
<td></td>
<td><strong>Taiwanese Isolates</strong></td>
<td></td>
</tr>
<tr>
<td><strong>West Indian Isolates</strong></td>
<td></td>
<td><strong>Cuban Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>KC149502</td>
<td>Maharashtra (MR)</td>
<td>EU350569</td>
<td>Cuba</td>
</tr>
<tr>
<td>AY238881</td>
<td>Maharashtra (MR)</td>
<td>EU350568</td>
<td>Cuba</td>
</tr>
<tr>
<td><strong>North Indian Isolates</strong></td>
<td></td>
<td><strong>Thai Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>AY238883</td>
<td>Delhi (DL)</td>
<td>EU350567</td>
<td>Cuba</td>
</tr>
<tr>
<td>DQ088670</td>
<td>Haryana (HR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY458617</td>
<td>Himachal Pradesh (HP)</td>
<td>U14743</td>
<td>Thailand</td>
</tr>
<tr>
<td>AY238882</td>
<td>Uttar Pradesh (UP)</td>
<td>AF506904</td>
<td>Thailand</td>
</tr>
<tr>
<td>AY458620</td>
<td>Uttar Pradesh (UP)</td>
<td>AF506901</td>
<td>Thailand</td>
</tr>
<tr>
<td><strong>South Indian Isolates</strong></td>
<td></td>
<td><strong>Sri Lankan Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>AY238880</td>
<td>Andhra Pradesh (AP)</td>
<td>U14741</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>EF210197</td>
<td>Andhra Pradesh (AP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY903266</td>
<td>Andhra Pradesh (AP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY238884</td>
<td>Karnataka (KA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY458618</td>
<td>Karnataka (KA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JN979403</td>
<td>Karnataka (KA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JN979404</td>
<td>Karnataka (KA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JN979405</td>
<td>Karnataka (KA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF104919</td>
<td>Tamil Nadu (TN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY687386</td>
<td>Tamil Nadu (TN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQ077175</td>
<td>Tamil Nadu (TN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bangladeshi Isolates</strong></td>
<td></td>
<td><strong>Australian Isolates</strong></td>
<td></td>
</tr>
<tr>
<td>HM778170</td>
<td>Tamil Nadu (TN)</td>
<td>D00595</td>
<td>USA</td>
</tr>
<tr>
<td>HM754218</td>
<td>Tamil Nadu (TN)</td>
<td>KF649236</td>
<td>USA</td>
</tr>
<tr>
<td>HM454197</td>
<td>Tamil Nadu (TN)</td>
<td>KF649233</td>
<td>USA</td>
</tr>
<tr>
<td>HM454196</td>
<td>Tamil Nadu (TN)</td>
<td>KF649243</td>
<td>USA</td>
</tr>
<tr>
<td><strong>US Isolates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Code names of some isolates are indicated within parenthesis.
Table 2. Base composition of the coat protein genes from different of PRSV isolates [Values within parenthesis indicates ±standard errors].

<table>
<thead>
<tr>
<th>Isolates</th>
<th>T3s</th>
<th>C3s</th>
<th>A3s</th>
<th>G3s</th>
<th>CAI</th>
<th>Nc</th>
<th>GC1s</th>
<th>GC2s</th>
<th>GC3s</th>
<th>GC</th>
<th>L_sym</th>
<th>L_aa</th>
<th>Gravy</th>
<th>Aromo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Isolates</td>
<td>0.44</td>
<td>0.26</td>
<td>0.33</td>
<td>0.26</td>
<td>0.25</td>
<td>56.35</td>
<td>0.49</td>
<td>0.41</td>
<td>0.39</td>
<td>0.45</td>
<td>132.33</td>
<td>143.33</td>
<td>-0.77</td>
<td>0.09</td>
</tr>
<tr>
<td>Australian</td>
<td>0.46</td>
<td>0.23</td>
<td>0.37</td>
<td>0.25</td>
<td>0.22</td>
<td>51.97</td>
<td>0.47</td>
<td>0.36</td>
<td>0.36</td>
<td>0.41</td>
<td>262.67</td>
<td>279.00</td>
<td>-0.92</td>
<td>0.07</td>
</tr>
<tr>
<td>Bangladeshi</td>
<td>0.44</td>
<td>0.24</td>
<td>0.38</td>
<td>0.25</td>
<td>0.23</td>
<td>54.66</td>
<td>0.48</td>
<td>0.38</td>
<td>0.37</td>
<td>0.42</td>
<td>267.00</td>
<td>283.60</td>
<td>-0.85</td>
<td>0.07</td>
</tr>
<tr>
<td>Brazilian</td>
<td>0.48</td>
<td>0.23</td>
<td>0.36</td>
<td>0.25</td>
<td>0.21</td>
<td>52.13</td>
<td>0.47</td>
<td>0.37</td>
<td>0.35</td>
<td>0.41</td>
<td>290.33</td>
<td>307.00</td>
<td>-0.97</td>
<td>0.07</td>
</tr>
<tr>
<td>Central Indian</td>
<td>0.44</td>
<td>0.24</td>
<td>0.35</td>
<td>0.28</td>
<td>0.22</td>
<td>52.53</td>
<td>0.48</td>
<td>0.37</td>
<td>0.39</td>
<td>0.42</td>
<td>268.50</td>
<td>286.00</td>
<td>-0.91</td>
<td>0.07</td>
</tr>
<tr>
<td>Chinese</td>
<td>0.46</td>
<td>0.24</td>
<td>0.35</td>
<td>0.24</td>
<td>0.20</td>
<td>50.52</td>
<td>0.47</td>
<td>0.38</td>
<td>0.36</td>
<td>0.42</td>
<td>234.67</td>
<td>252.00</td>
<td>-0.68</td>
<td>0.08</td>
</tr>
<tr>
<td>Cuban</td>
<td>0.41</td>
<td>0.27</td>
<td>0.27</td>
<td>0.32</td>
<td>0.23</td>
<td>61.00</td>
<td>0.48</td>
<td>0.41</td>
<td>0.45</td>
<td>0.47</td>
<td>138.00</td>
<td>151.00</td>
<td>-0.66</td>
<td>0.09</td>
</tr>
<tr>
<td>East Indian</td>
<td>0.19</td>
<td>0.11</td>
<td>0.14</td>
<td>0.12</td>
<td>0.10</td>
<td>23.88</td>
<td>0.49</td>
<td>0.38</td>
<td>0.18</td>
<td>0.19</td>
<td>94.12</td>
<td>100.96</td>
<td>-0.31</td>
<td>0.03</td>
</tr>
<tr>
<td>North Indian</td>
<td>0.44</td>
<td>0.24</td>
<td>0.36</td>
<td>0.26</td>
<td>0.22</td>
<td>53.89</td>
<td>0.48</td>
<td>0.38</td>
<td>0.38</td>
<td>0.42</td>
<td>267.20</td>
<td>283.80</td>
<td>-0.88</td>
<td>0.07</td>
</tr>
<tr>
<td>South Indian</td>
<td>0.45</td>
<td>0.23</td>
<td>0.37</td>
<td>0.26</td>
<td>0.22</td>
<td>51.76</td>
<td>0.48</td>
<td>0.37</td>
<td>0.37</td>
<td>0.42</td>
<td>268.73</td>
<td>285.67</td>
<td>-0.89</td>
<td>0.07</td>
</tr>
<tr>
<td>West Indian</td>
<td>0.45</td>
<td>0.23</td>
<td>0.37</td>
<td>0.25</td>
<td>0.21</td>
<td>50.22</td>
<td>0.48</td>
<td>0.37</td>
<td>0.36</td>
<td>0.42</td>
<td>268.00</td>
<td>285.50</td>
<td>-0.90</td>
<td>0.07</td>
</tr>
<tr>
<td>Sri Lankan</td>
<td>0.41</td>
<td>0.27</td>
<td>0.37</td>
<td>0.26</td>
<td>0.22</td>
<td>52.34</td>
<td>0.46</td>
<td>0.38</td>
<td>0.40</td>
<td>0.43</td>
<td>261.00</td>
<td>278.00</td>
<td>-0.90</td>
<td>0.07</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>0.46</td>
<td>0.23</td>
<td>0.36</td>
<td>0.26</td>
<td>0.21</td>
<td>54.46</td>
<td>0.47</td>
<td>0.38</td>
<td>0.37</td>
<td>0.42</td>
<td>290.00</td>
<td>307.00</td>
<td>-0.93</td>
<td>0.07</td>
</tr>
<tr>
<td>Thai</td>
<td>0.43</td>
<td>0.27</td>
<td>0.37</td>
<td>0.24</td>
<td>0.21</td>
<td>51.45</td>
<td>0.47</td>
<td>0.37</td>
<td>0.38</td>
<td>0.42</td>
<td>261.33</td>
<td>278.00</td>
<td>-0.91</td>
<td>0.07</td>
</tr>
<tr>
<td>US</td>
<td>0.46</td>
<td>0.23</td>
<td>0.34</td>
<td>0.26</td>
<td>0.22</td>
<td>57.50</td>
<td>0.46</td>
<td>0.40</td>
<td>0.36</td>
<td>0.42</td>
<td>137.25</td>
<td>146.50</td>
<td>-0.57</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Table 3. Correlation coefficients of the factors that influence codon bias in coat protein gene [\(^a\) and \(^b\) indicate significant correlations at a level of 0.05 and 0.01, respectively and are highlighted].

<table>
<thead>
<tr>
<th></th>
<th>T3s</th>
<th>C3s</th>
<th>A3s</th>
<th>G3s</th>
<th>CAI</th>
<th>GC1s</th>
<th>GC2s</th>
<th>GC3s</th>
<th>GC</th>
<th>L_sym</th>
<th>L_aa</th>
<th>Gravy</th>
<th>Aromo</th>
<th>Nc</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.868(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3s</td>
<td>0.352</td>
<td>-0.482</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3s</td>
<td>-0.488</td>
<td>0.368</td>
<td>-0.765(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAI</td>
<td>-0.35</td>
<td>0.348</td>
<td>-0.491</td>
<td>0.387</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC1s</td>
<td>-0.174</td>
<td>0.04</td>
<td>-0.216</td>
<td>0.325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC2s</td>
<td>-0.388</td>
<td>0.518(^a)</td>
<td>-0.843</td>
<td>0.509</td>
<td>0.615(^a)</td>
<td>0.174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC3s</td>
<td>-0.820(^b)</td>
<td>0.822(^b)</td>
<td>-0.790(^b)</td>
<td>0.821(^b)</td>
<td>0.496</td>
<td>0.271</td>
<td>0.671(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>-0.672(^b)</td>
<td>0.708(^b)</td>
<td>-0.891(^b)</td>
<td>0.761(^b)</td>
<td>0.685(^b)</td>
<td>0.408</td>
<td>0.849(^b)</td>
<td>0.929(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L_sym</td>
<td>0.259</td>
<td>-0.379</td>
<td>0.803(^b)</td>
<td>-0.408</td>
<td>-0.602(^a)</td>
<td>-0.12</td>
<td>-0.883(^b)</td>
<td>-0.532(^a)</td>
<td>-0.730(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L_aa</td>
<td>0.251</td>
<td>-0.371</td>
<td>0.798(^b)</td>
<td>-0.403</td>
<td>-0.602(^a)</td>
<td>-0.11</td>
<td>-0.882(^b)</td>
<td>-0.523(^a)</td>
<td>-0.723(^b)</td>
<td>1.000(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gravy</td>
<td>-0.118</td>
<td>0.172</td>
<td>-0.677(^b)</td>
<td>0.252</td>
<td>0.225</td>
<td>-0.11</td>
<td>0.695(^b)</td>
<td>0.322</td>
<td>0.501</td>
<td>-0.858(^b)</td>
<td>-0.856(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aromo</td>
<td>0.025</td>
<td>0.138</td>
<td>-0.626(^a)</td>
<td>0.168</td>
<td>0.366</td>
<td>-0.12</td>
<td>0.772(^b)</td>
<td>0.232</td>
<td>0.448</td>
<td>-0.923(^b)</td>
<td>-0.927(^b)</td>
<td>0.870(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nc</td>
<td>-0.342</td>
<td>0.362</td>
<td>-0.827(^b)</td>
<td>0.690(^b)</td>
<td>0.587(^a)</td>
<td>0.222</td>
<td>0.852(^b)</td>
<td>0.665(^b)</td>
<td>0.780(^b)</td>
<td>-0.770(^b)</td>
<td>-0.773(^b)</td>
<td>0.628(^a)</td>
<td>0.644(^b)</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Synonymous codon usage in the coat protein region of PRSV isolates [The ‘c’ marked values indicate the more preferred codons than other synonymous codons and the ‘d’ marked codons are the putative optimal codons].

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>RSCU</th>
<th>Amino acid</th>
<th>Codon</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>1.00c</td>
<td>Tyr</td>
<td>UAU</td>
<td>1.20c</td>
</tr>
<tr>
<td></td>
<td>UUC</td>
<td>1.00c</td>
<td></td>
<td>UACd</td>
<td>0.80</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
<td>0.00</td>
<td>TER</td>
<td>UAA</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>UUG</td>
<td>2.00c</td>
<td></td>
<td>UAG</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>CUU</td>
<td>1.00</td>
<td>His</td>
<td>CAU</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td>1.00</td>
<td></td>
<td>CAC</td>
<td>1.60c</td>
</tr>
<tr>
<td></td>
<td>CUAd</td>
<td>1.00</td>
<td>Gln</td>
<td>CAA</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>CUG</td>
<td>1.00</td>
<td></td>
<td>CAG</td>
<td>1.33c</td>
</tr>
<tr>
<td>Ile</td>
<td>AUU</td>
<td>1.20</td>
<td>Asn</td>
<td>AAU</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>1.80c</td>
<td></td>
<td>AAC</td>
<td>1.11c</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td>0.00</td>
<td>Lys</td>
<td>AAA</td>
<td>0.57</td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>1.00c</td>
<td></td>
<td>AAGd</td>
<td>1.43c</td>
</tr>
<tr>
<td></td>
<td>GUU</td>
<td>0.67</td>
<td>Asp</td>
<td>GAU</td>
<td>1.33c</td>
</tr>
<tr>
<td></td>
<td>GUC</td>
<td>1.33c</td>
<td></td>
<td>GACd</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>GUA</td>
<td>0.67</td>
<td>Glu</td>
<td>GAA</td>
<td>1.56c</td>
</tr>
<tr>
<td></td>
<td>GUG</td>
<td>1.33</td>
<td></td>
<td>GAG</td>
<td>0.44</td>
</tr>
<tr>
<td>Ser</td>
<td>UCU</td>
<td>1.00</td>
<td>Cys</td>
<td>UGU</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>UCC</td>
<td>0.00</td>
<td></td>
<td>UGCd</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>UCA</td>
<td>2.00c</td>
<td>Trp</td>
<td>UGG</td>
<td>1.00c</td>
</tr>
<tr>
<td></td>
<td>UCGd</td>
<td>0.00</td>
<td>Arg</td>
<td>CGU</td>
<td>0.46</td>
</tr>
<tr>
<td>Pro</td>
<td>CCU</td>
<td>2.40c</td>
<td></td>
<td>CGC</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>CCC</td>
<td>0.00</td>
<td></td>
<td>CGA</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>CCA</td>
<td>0.80</td>
<td></td>
<td>CGG</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>CCG</td>
<td>0.80</td>
<td></td>
<td>AGA</td>
<td>2.77c</td>
</tr>
<tr>
<td>Thr</td>
<td>ACU</td>
<td>2.67c</td>
<td></td>
<td>AGG</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>ACCd</td>
<td>0.00</td>
<td>Ser</td>
<td>AGU</td>
<td>3.00c</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>0.89</td>
<td></td>
<td>AGC</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>ACG</td>
<td>0.44</td>
<td>Gly</td>
<td>GGU</td>
<td>2.00c</td>
</tr>
<tr>
<td>Ala</td>
<td>GCU</td>
<td>2.13c</td>
<td></td>
<td>GGC</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>0.00</td>
<td></td>
<td>GGA</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>GCA</td>
<td>1.33</td>
<td></td>
<td>GGd</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>GCG</td>
<td>0.53</td>
<td>TER</td>
<td>UGA</td>
<td>3.00c</td>
</tr>
</tbody>
</table>
Figure 1. Naturally infected papaya plants samples with severe mosaic and leaf distortions collected from sub-Himalayan region of north-east India and their collection spots [S= Siliguri, J= Jalpaiguri, C= Coochbehar, I= Islampur, B= Balurghat and M= Malda].

Figure 2. Amplicons of six positive isolates of PRSV on agarose gel. [L1 and L8= 500bp DNA marker; L2= BLGPRSV, L3= SLGPRSV3, L4= Islampur, L5= PRSV-COB, L6= PRSV-JAL, L7= PRSV-MLD isolates].

Figure 3. (A) Relationship between effective number of codons ($N_C$) and GC3 content of the coat protein genes of PRSV isolates [The continuous curve represents the expected curve between GC3 content and $N_C$ under random codon usage]; (B) Relationship between effective number of codons ($N_C$) and frequency of translatable codons ($L_{_aa}$) of the coat protein genes of PRSV isolates [The regression curve can be described as $y = 8E-05x^2 - 0.064x + 64.89$, $R^2 = 0.449$. The indications of different isolates are given at the top right corner]; (C) Relationship between the GC1/GC2 content and GC3 content of PRSV isolates [GC content at the first, second and third codon position are indicated as GC1, GC2 and GC3. The blue and orange circles represent GC1 and GC2 values respectively and are indicated at the top right corner].

Figure 4. (A) Contributions of 40 axes based on the correspondence analysis; (B) Correspondence analysis of the coat protein genes [X and Y axis corresponds to the Axis 1 and Axis 2. Different isolates are indicated with different markers as shown in the right].

Figure 5. Sequence identity matrix of the 66 PRSV isolates using heat map. Identity percent corresponds to the color matrix is indicated in the lower part of the figure.
Figure 6. Phylogenetic tree generated by neighbour joining of PRSV coat protein (CP) alignments. Values at the nodes indicate percentage of bootstrap support (out of 1000 bootstrap replicates) and are indicated if greater than 50. GenBank accession numbers along with the collection spot of the viruses have been indicated at the end of each branch.
Naturally infected papaya plants samples with severe mosaic and leaf distortions collected from sub-Himalayan region of north-east India and their collection spots [S= Siliguri, J= Jalpaiguri, C= Coochbehar, I= Islampur, B= Balurghat and M= Malda].

135x102mm (300 x 300 DPI)
Amplicons of six positive isolates of PRSV on agarose gel. [L1 and L8= 500bp DNA marker; L2= BLGP, L3= SLGP, L4= Islampur, L5= PRSV-COB, L6= PRSV-JAL, L7= PRSV-MLD isolates].

29x12mm (300 x 300 DPI)
(A) Relationship between effective number of codons (NC) and GC3 content of the coat protein genes of PRSV isolates [The continuous curve represents the expected curve between GC3 content and NC under random codon usage]; (B) Relationship between effective number of codons (NC) and frequency of translatable codons (L_\text{aa}) of the coat protein genes of PRSV isolates [The regression curve can be described as \( y = 8 \times 10^{-5}x^2 - 0.064x + 64.89 \), \( R^2 = 0.449 \). The indications of different isolates are given at the top right corner]; (C) Relationship between the GC1/GC2 content and GC3 content of PRSV isolates [GC content at the first, second and third codon position are indicated as GC1, GC2 and GC3. The blue and orange circles represent GC1 and GC2 values respectively and are indicated at the top right corner].
(A) Contributions of 40 axes based on the correspondence analysis; (B) Correspondence analysis of the coat protein genes [X and Y axis corresponds to the Axis 1 and Axis 2. Different isolates are indicated with different markers as shown in the right].

72x75mm (300 x 300 DPI)
Sequence identity matrix of the 66 PRSV isolates using heat map. Identity percent corresponds to the color matrix is indicated in the lower part of the figure.

70x70mm (300 x 300 DPI)
Phylogenetic tree generated by neighbour joining of PRSV coat protein (CP) alignments. Values at the nodes indicate percentage of bootstrap support (out of 1000 bootstrap replicates) and are indicated if greater than 50. GenBank accession numbers along with the collection spot of the viruses have been indicated at the end of each branch.

64x75mm (300 x 300 DPI)