ORIGINAL RESEARCH ARTICLE OPEN ACCESS

Sequence Divergence of Coat Protein Gene among Indian and non-Indian Isolates of *Rice Tungro Bacilliform Virus*

P. Malathi, D. Krishnaveni and Satendra K Mangrauthia*

ICAR-Indian Institute of Rice research (IIRR), Rajendranagar, Hyderabad

*Corresponding author: skmdrr@gmail.com

Received: 8th March, 2015; Accepted: 27th May, 2015

Abstract

Rice tungro disease (RTD), one of the major constraints to rice production in South and Southeast Asia, is caused by combined action of two viruses: *Rice tungro spherical virus* (RTSV) and *Rice tungro bacilliform virus* (RTBV). The present study was undertaken to describe the sequence divergence and evolution of RTBV isolates present in India and other countries. Phylogenetic analysis based on coat protein (CP) sequences of RTBV generated in this study showed distinct divergence of Indian and non-Indian RTBV isolates into two clusters. Further, Indian RTBV isolates formed two groups- one consisted isolates from Andhra Pradesh and Kanyakumari, and other included isolates from Hyderabad, Punjab, and West Bengal. The results obtained from phylogenetic analysis were further supported with the single nucleotide polymorphisms (SNPs), insertion and deletions (INDELs) and evolutionary distance analysis. Signature sequences and amino acid motifs were identified which showed distinct difference between Indian and non-Indian isolates. This study will help in understanding the geographical evolution and adaptation of RTBV in different rice ecosystems.

Keywords: RTBV, CP, Phylogeny, Diversity, *Oryza sativa*

Introduction

Rice tungro disease (RTD) is one of the most important viral diseases of rice in South and Southeast Asia. When plants are infected with RTD in the early seedling stage, yield losses can be as high as 100 percent (Muralidharan *et al*., 2003). The disease is caused by a complex of two viruses; *Rice tungro bacilliform virus* (RTBV), a DNA virus and *Rice tungro spherical virus* (RTSV), an RNA virus (Jones *et al*., 1991). The viral complex is transmitted between plants by the vector Green Leafhopper (GLH), *Nephotettix virescens* (Hibino and Cabauatan, 1987). RTBV is responsible for symptoms development of tungro disease which includes stunting and yellow to orange discoloration of the infected plants (Hibino *et al*., 1978). RTBV is a plant pararetrovirus having ~8.0 Kb circular double-stranded DNA genome with two discontinuities, one on each strand (Hay *et al*., 1991; Qu *et al*., 1991; Hibino *et al*., 1991). It has four open reading frames (ORFs), potentially capable of encoding proteins of 24, 12, 194 and 46 kDa, respectively (Hay *et al*., 1991; Mangrauthia *et al*., 2012a). The longest ORF3 of RTBV encodes a P194 poly protein that contains four functional domains: a putative movement protein (MP), the coat protein (CP), aspartate protease (PR), and reverse transcriptase (RT) with a ribonuclease H activity (Qu *et al*., 1991; Laco *et al*., 1994).

Considering the importance of tungro virus disease from Indian perspective, efforts to understand the RTBV population structure present in India have been very limited. Recently, we studied the molecular diversity of RTBV by utilizing the DNA sequences generated from ORF-I, II and III (Mangrauthia *et al*., 2012a). It is important to extend this study by comparing the sequences of other important viral proteins such as coat protein, movement protein, and replicase. Coat protein (CP) has been most extensively used in understanding the genetic diversity of viruses due to its crucial role in establishing virus infection and multiplication. The objective of this study was to obtain the nucleotide sequence of the coat protein gene of RTBV, and to assess the sequence diversity and phylogenetic relationship among the RTBV isolates. Further to gain better insight, RTBV isolates were analyzed for single nucleotide polymorphism (SNPs), insertion and deletion (INDELs), evolutionary distance, and Ka/Ks ratio.

Materials and Methods

Virus isolates

Rice plants showing symptoms of tungro disease were collected from Hyderabad (ICAR-IIRR, glasshouse). Virus infected plants were maintained through insect (green leaf hopper)-mediated transmission in separate cages in glasshouse to prevent the cross contamination.

Gene amplification

The total DNA from virus-infected rice leaves was isolated by CTAB method (Murray and Thompson, 1980). Sequences of the forward and reverse primers were designed based on the sequence of RTBV genomes available in NCBI database. Forward primer (CCAGAAGTATCCTCAAAAGAT) and Reverse primer (CTCAGGAAGTCTGTCAAATAG) were used to amplify the sequence coding for coat protein sequence of RTBV. The thermal profile used for these primers was 94ºC for 5 min (initial denaturation), 94ºC for 30s (denaturation), 56ºC for 30s (annealing), 72ºC for 1 min (extension), and 72ºC for 10 min (final extension). The amplification reaction was performed in a reaction mixture containing dNTP mix, primers, 10X PCR buffer with 15 mM $MgCl_2$ and recombinant *Taq* DNA polymerase. Resulting PCR products were analyzed on 1 % agarose gel followed by staining with ethidium bromide.

Cloning and nucleotide sequencing

PCR amplified DNA was cloned into pGEM-T (Promega Corporation, Madison, WI) cloning vector using TA cloning strategy. The PCR product was ligated into pGEM-T vector and transformed into competent *E. coli* strain DH5α using heat shock method. Plasmid DNA from potential recombinants (identified through blue-white screening of bacterial colonies and colony PCR) were purified by Wizard plus SV mini preps DNA purification system (Promega) following manufacturer's instructions. Plasmid isolated from these clones was subjected to PCR confirmation with sequence-specific primers and restriction enzyme digestion. Clones containing desired fragment of ~500 bps corresponding to the CP coding region of RTBV were selected for sequencing. Two clones were sequenced in both the directions to eliminate potential sequence heterogeneity introduced by *Taq* DNA polymerase.

Sequence analysis

Nucleotides as well as translated amino acid sequences representing the complete coat protein gene were compared with the already available sequences of RTBV isolates (Table 1). Sequences were aligned using CLUSTAL W programme. Neighbor-joining tree was generated using CLUSTAL X (Thompson *et al*., 1997). The robustness of the lineages in the phylogenetic tree was assessed from the internodes length in the tree by bootstrapping in CLUSTAL X using 1000 resampling. Sequence identity matrix and sequence difference count matrix was calculated using BioEdit sequence alignment editor version 5.09.04 (Hall, 1999). The evolutionary distance was analyzed by using the bioinformatics tool MEGA version 5 (Tamura *et al*., 2011). SNPs and INDELs were calculated by using DnaSP (version 5.10) by comparing the nucleotide sequence with the Hyderabad isolate as reference sequence. Ka/Ks value was also calculated by using the same tool to analyze the synonymous and non-synonymous mutations at the nucleotide level, which really affect the amino acid sequences of the protein. Protein level changes were analyzed by comparing the deduced amino acid sequences using BioEdit–CLUSTAL W tool.

Results and Discussion

Coat protein gene sequence of RTBV from Hyderabad isolate was deciphered. The complete CP gene was found to be 528 bps long coding for 176 amino acid capsid protein in all isolates tested here. Cluster dendrogram based on the nucleotide sequence of the CP gene and the deduced amino acid sequences were similar and only the phylogenetic tree based on the deduced amino acid sequence is shown (Fig. 1). A distinct divergence of CP of RTBV isolates into two clusters was observed, one cluster included all the isolates from India and the other distinct cluster had RTBV isolates from other countries- Japan, Mal (Malaysia), Phil (Philippines) and Thai (Thailand). Cluster containing all Indian isolates was further divided into two groups; one group consisted isolates from Punjab, WB (West Bengal), CWB (Chinsura, West Bengal) and Hyd (Hyderabad) and another group consisted isolates from AP (Andhra Pradesh) and KK (Kanyakumari).

In order to analyze the polymorphism at sequence level, the CP sequences of RTBV were analyzed for the presence of SNPs and INDELs. All the sequences were compared with Hyderabad-CP sequences (reference sequence), as the analysis was done here and polymorphic data was stored for further analysis. At CP sequence level, WB, KK and AP isolates showed very close similarity with the reference sequence (Hyderabad). These sequences had very less deviation from the reference sequence Hyderabad-CP, as it had 0.044, 0.048 and 0.048 evolutionary distances and 22, 24 and 24 SNPs, respectively. The next closest sequences were from Punjab and CWB with evolutionary distance of 0.049 and 0.053, and 24 and 26 SNPs, respectively, when compared with the reference sequence. The farthest sequences were from Japan, Malaysia, Philippines and Thailand with evolutionary distance of 0.168, 0.171, 0.171 and 0.201, respectively. These non-Indian isolates showed 78, 80, 79 and 90 SNPs, respectively, when compared with the Hyderabad CP sequence. INDELs were not observed in any of the sequences from either Indian isolates or non-Indian isolates (Table 2). Ka/Ks value was calculated

to understand the nucleotide change, which affect the amino acid sequence of the protein. The ratios of non-synonymous to synonymous nucleotide substitution rates (Ka/Ks) ranged from 0.0098 to 0.3863. Malaysian isolate had the highest Ka/Ks value of 0.3863, which indicated the mutations in the nucleotide level as well as protein level (Table 2). Indian isolates showed less similarity with non-Indian isolates both at nucleotide (83–87 %) and amino acid level (89–93 %) of CP gene. Within the Indian origin isolates, 94–100 % similarity at nucleotide level and 97–100 % similarity at amino acid level was recorded. Further, the sequence similarity matrix among the Indian isolates revealed 95% (Hyderasbad *Vs*. KK/Punjab/CWB/AP) to ~100% similarity (KK vs. AP) based on nucleotides. Further, 97% (CWB Vs. Hyderabad/ WB) to ~100% similarity (KK Vs. Punjab/AP and Punjab Vs. AP) based on amino acid sequences was observed (Table 3).

CLUSTAL W multiple alignment of amino acid sequences derived from CP gene revealed some of the differential amino acid sequence motifs which were unique to the Indian or non-Indian isolates. For instance, sequence motifs EVS^{4} and II^{44} present in all Indian isolates of CP were replaced by IEA^{4} and LV^{44} in non-Indian isolates. In addition to the sequence motifs, a number of individual amino acids were also recorded as signature sequence which clearly distinguished the Indian isolates from non-Indian isolates (Fig. 2).

In this study, CP gene of RTBV from Hyderabad isolate was sequenced and analyzed. The isolate was supposed to be an Indian member due to its geographical habitat (Hyderabad, India). The sequence analysis clearly indicated strong similarity of Hyderabad isolate with other Indian isolates of RTBV. Out of the 12 RTBV sequences analyzed here, six isolates, *i.e.* Hyd, KK, Punjab, CWB, WB and AP from the Indian origin formed a cluster and the remaining six *i.e.* Japan, Malaysia, Philippines, Philippines-Ic, Philippines-G1 and Thailand isolates from non-Indian clustered separately (Fig.1). This was further strengthened by the observations that greater identity at nucleotide and amino acid level existed within Indian isolates when compared with non-Indian isolates. Fan *et al*., (1996) reported the existence of two strains of RTBV, which was supported by Nath *et al*. (2002) providing molecular evidence. In the present investigation, much emphasis was given to analyze the genetic variation of RTBV isolates present within India. Recently, a similar study was performed to understand the genetic variation of RTBV and RTSV isolates present in tungro endemic states of India (Mangrauthia *et al*., 2012a and 2012b). Our results indicated the existence of habitat correlation in the evolution of nucleotide and amino acid sequences of RTBV. The genetic diversity was less for the amino acid than the nucleotide for all CPs (Table 3) which is supported by Ka/Ks ratio. In general, the natural selection pressure in genome and its functional evolution are reflected by the differences at the nucleotide level and due to the principle of codon degeneracy; these are reduced at the amino acid residue level (Gojobori *et al*., 1990). Analysis of SNPs, INDELs and evolutionary distance also showed similar pattern wherein Hyderabad isolate showed minimum evolutionary distance and divergence with Indian isolates. Evolutionary distance, the number of substitutions per site separating a pair of homologous sequences since they diverged from their common ancestral sequence, is an extremely important measure in molecular evolution and comparative genomics (Rosenberg, 2005). The data obtained from phylogeny, sequence identity matrix, SNPs, INDELs, and evolutionary distance suggest the diversion of Indian RTBV isolates into two major cluster; one includes isolates of AP and KK while other includes isolates of Hyd, CWB, WB, and Punjab. The presence of Zero INDELs among RTBV CP sequences and very low Ka/Ks ratio suggests that though these sequences had the mutations, they may not alter the protein structure and hence its function. One way to evaluate the selection pressures on protein evolution is to compare the rate of synonymous and non-synonymous nucleotide substitutions. Ks is the estimated number of synonymous changes per synonymous site and corresponds to the rate of amino acid-neutral evolution. Ka, on the other hand, is the number of non synonymous substitutions per non-synonymous site (Roth and Liberles, 2006). Interestingly, Malaysian isolate had the highest Ka/Ks value for CP- 0.386 indicating the significant mutation in the nucleotide level as well as protein level when compared with Hyderabad isolate. It was suggested that recombination as well as substitutions and insertions or deletions, have played a significant role in the evolution of RTBV variants (Cabauatan *et al*.,1999). The amino acid sequence motifs unique to CP of Indian isolates EVS^{4} and II^{44} differed from non-Indian isolates (IEA^{4} and LV^{44}). In addition to sequence motifs, a number of individual amino acids were also noticed which were unique to Indian or non-Indian isolates. It would be interesting to define the role of these unique motifs and amino acid residues in diversification and evolution of RTBV in Indian and non-Indian geographical locations.

In conclusion, sequence analysis of CP gene revealed that Indian isolates of RTBV diverged into two lineages: one included AP and KK and other includes isolates of Hyderabad, CWB, WB, and Punjab. Besides deciphering the molecular diversity, the divergence of two major groups

(Indian and non-Indian) of RTBV was revealed. Sequence difference count matrix analysis of the RTBV CP gene revealed that Indian isolates are significantly different in its molecular genetic composition from rest of the world. Further, though mutation and recombination has occurred at nucleotide level that altered genetic composition of these isolates, the protein sequence is not much affected. It would be interesting to ascertain the role of individual amino acids and sequence motifs (identified in Indian and non-Indian RTBV CP sequences) in diversification and adaptation of RTBV in different rice ecologies.

Acknowledgement

Authors are highly thankful to the Project Director (IIRR) for his kind support and cooperation.

References:

Cabauatan, P.Q., Melcher, U., Ishikawa, K., Omura, T., Hibino, H., Koganezawa, H. and Azzam, O. 1999. Sequence changes in six variants of *rice tungro bacilliform virus* and their phylogenetic relationships. *Journal of General Virology* 80:2229-2237.

Fan, Z., Dahal, G., Dasgupta, I., Hay, J. and Hull, R. 1996. Variation in the genome of *rice tungro bacilliform virus*: molecular characterization of six isolates. *Journal of General Virology* 77: 847-854.

Gojobori, T., Moriyama, E.N. and Kimura, M. 1990. Molecular clock of viral evolution, and the neutral theory. *Proceedings of National Academy of Science* USA 87: 10015-10018.

Hall, T.A. 1999. Bio Edit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Hay, J.M., Jones, M.C., Blakebrough, M.L., Dasgupta, I., Davies, J.W. and Hull, R. 1991. An analysis of the sequence of an infectious clone of *rice tungro bacilliform virus*, a plant pararetrovirus. *Nucleic Acids Research* 19:2615-2621.

Hibino, H. and Cabauatan, P.Q. 1987. Infectivity neutralization of rice tungro-associated viruses acquired by vector leafhoppers. *Phytopathology* 77:473-476.

Hibino, H., Ishikawa, K., Omura, T., Cabauatan, P.Q. and Koganezawa, H. 1991. Characterization of *rice tungro bacilliform* and *rice tungro spherical viruses*. *Phytopathology* 81:1130-1132.

Hibino, H., Roechan, M. and Sudarisman, S. 1978. Association of two types of virus particles with penyakit habang (tungro disease) of rice in Indonesia. *Phytopathology* 68: 1412-1416.

Jones, M.C., Gough, K., Dasgupta, I., Subba Rao, B.L., Cliffe, J., Qu R., Shen, P., Kaniewska, M., Blakebrough, M., Davies, J.W., Beachy, R.N. and Hull, R. 1991. Rice tungro disease is caused by a RNA and a DNA virus. *Journal of General Virology* 72: 757-761.

Laco, G.S. and Beachy, R.N. 1994. *Rice tungro bacilliform virus* encodes reverse transcriptase, DNA polymerase, and ribonuclease H activities. *Proceedings of National Academy of Science* USA 91: 2654-2658.

Mangrauthia, S.K., Malathi, P., Agarwal S., Sailaja, B., Singh, J., Ramkumar, G., Krishnaveni, D. and Balachandran, S.M. 2012a. The Molecular Diversity and Evolution of *Rice tungro bacilliform virus* from Indian perspective. *Virus Genes* 45:126-138.

Mangrauthia, S.K., Malathi, P., Agarwal, S., Ramkukmar, G., Krishnaveni, D., Neeraja, C.N., Madhav, MS., Ladhalakshmi, D., Balachandran, S.M. and Viraktamath, B.C. 2012b. Genetic variation of coat protein gene among the isolates of *Rice tungro spherical virus* from tungro-endemic states of the India. *Virus Genes* 44:482- 487.

Muralidharan, K., Krishnaveni, D., Rajarajeshwari, N.V.L. and Prasad, A.S.R. 2003. Tungro epidemic and yield losses in paddy fields in India. *Current science* 85: 1143–1147.

Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8:4321-4325.

Nath, N., Mathur, S. and Dasgupta, I. 2002. Molecular analysis of two complete *rice tungro bacilliform virus* sequences from India. *Archives Virology* 147(6):1173-1187.

Qu, R.D., Bhattacharyya, M., Laco, G.S., De Kochko, A., Suba Rao, B.L., Kaniewska, M.B., Elmer, J.S., Rochester, D.E., Smith, C.E. and Beachy, R.N. 1991. Characterization of the genome of *rice tungro bacilliform virus*: comparison with *Commelina yellow mottle virus* and caulimo viruses. *Virology*185:354-364.

Rosenberg, M.S. 2005. Evolutionary distance estimation and fidelity of pair wise sequence alignment. *BMC Bioinformatics* 6:102.

Roth, C. and Liberles, D.A. 2006. A systematic search for positive selection in higher plants (Embryophytes). *BMC Plant Biology* 6:12.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology Evolution* 28 (10): 2731-9.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. Nucleic Acids Research 24: 4876-82.

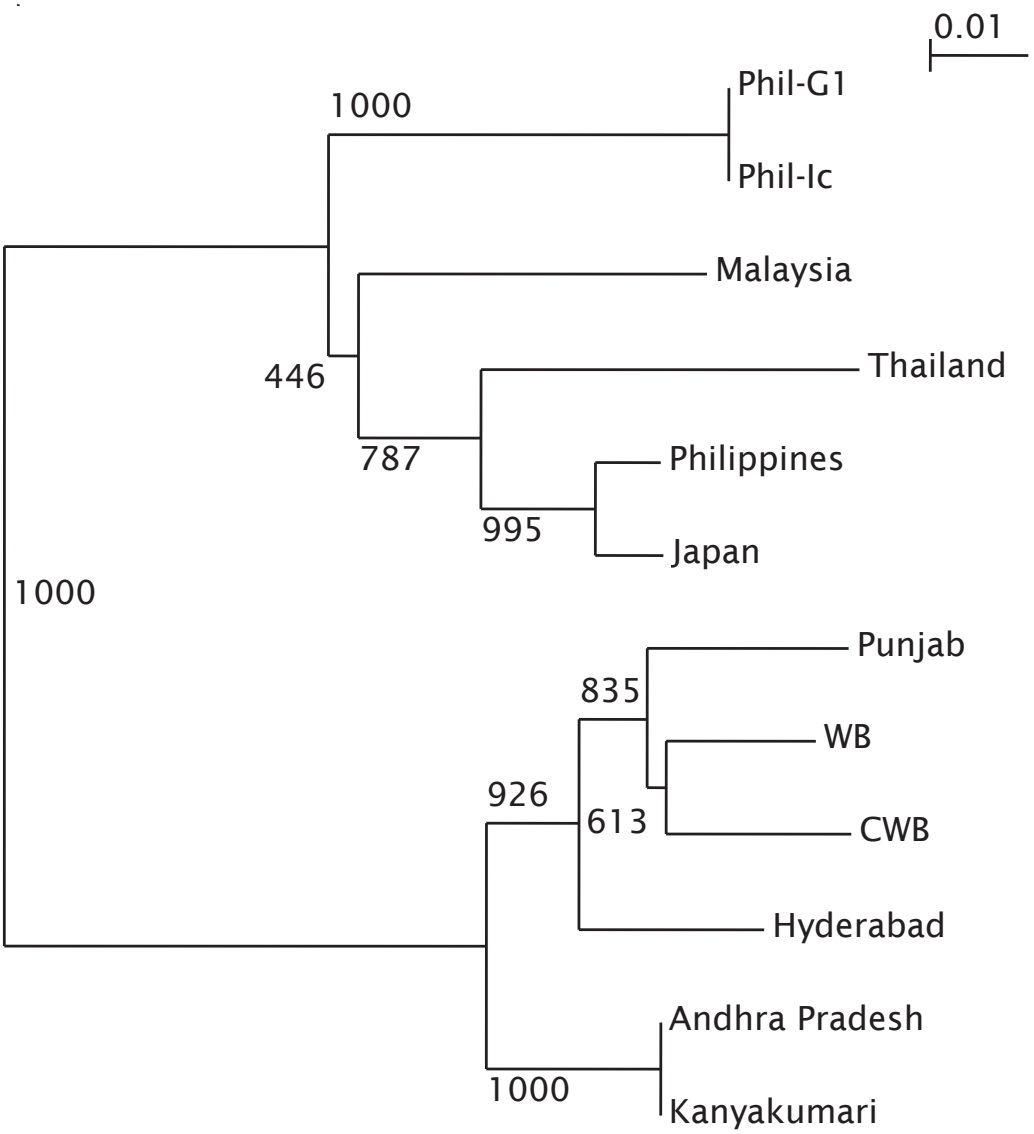

Fig. 1. Neighbor-joining tree based on nucleotide sequences of coat protein of different RTBV isolates. CWB-Chinsura West Bengal, WB-West Bengal

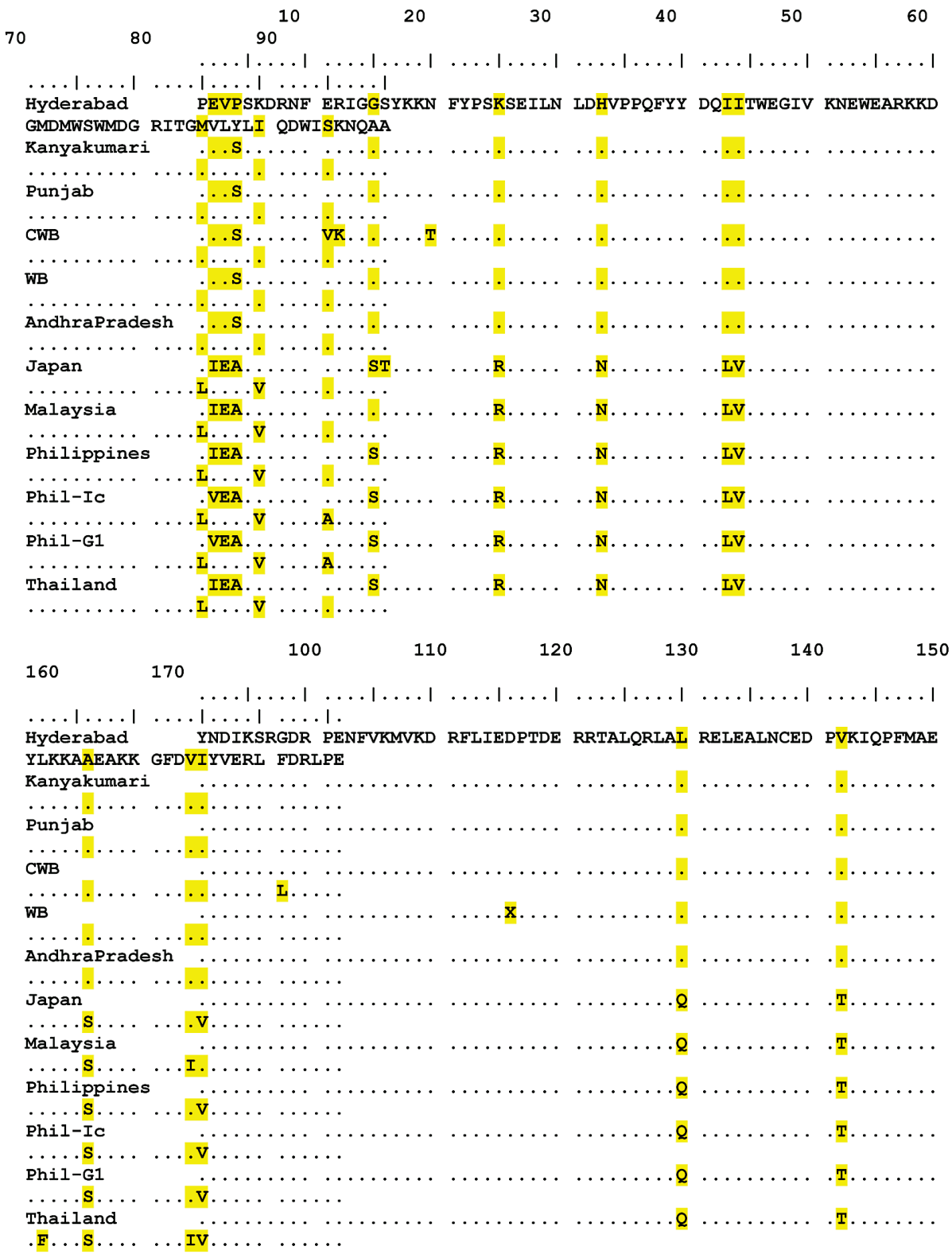

Fig. 2: Amino acid alignment of CP of different RTBV isolates. The CP sequences of Hyderabad, Kanyakumari , Punjab, Chinsura West Bengal (CWB), West Bengal (WB), Andhra Pradesh, Japan, Malaysia, Philippines, Philippines-Ic (Phil-Ic) , Philippines-G1 (Phil-G1) and Thailand isolates were aligned using CLUSTAL W. The motifs and individual amino acids (yellow box) differentiating Indian and non-Indian isolates are highlighted.

Table 1. Sources of coat protein gene sequences of RTBV isolates used in this study for comparison.

S. No.	Origin	Isolates	GenBank accession No.
1	India	Hyderabad	JX644072[a]
2	India	Kanyakumari	HQ385226
3	India	Punjab	JX255736
4	India	Chinsura West Bengal	FN377814
5	India	West Bengal	FN377814
6	India	Andhra Pradesh	AJ292232
7	Japan	Japan	D10774
8	Malaysia	Malaysia	AF076470
9	Philippines	Philippines	NC_001914
10	Philippines	Philippines-Ic	AF113832
11	Philippines	Philippines-G1	AF113830
12	Thailand	Thailand	AF220561

[a] Generated in this study

Table 2. Analysis of evolutionary distance, SNPs, INDELs, and Ka/Ks ratio of coat protein sequence of RTBV isolates

S. No	Name	Evolutionary distance	SE	SNP	INDELs	Ks	Ka	Ka/Ks
1	Hyd	–	–	–	–	–	–	–
2	KK	0.048	0.010	24	0	0.2425	0.0024	0.009896
3	Punjab	0.049	0.010	24	0	0.2412	0.0024	0.009950
4	CWB	0.053	0.010	26	0	0.2173	0.0121	0.055683
5	WB	0.044	0.010	22	0	0.2192	0.0024	0.010948
6	AP	0.048	0.010	24	0	0.2425	0.0024	0.009896
7	Japan	0.168	0.020	78	0	0.9433	0.0469	0.049719
8	Mal	0.171	0.019	80	0	1.0818	0.0418	0.386393
9	Phil	0.171	0.019	79	0	1.0037	0.0444	0.044236
10	Phil-Ic	0.171	0.020	79	0	0.9616	0.0462	0.048044
11	Phil- G1	0.171	0.020	79	0	0.9616	0.0462	0.048044
12	Thai	0.201	0.022	90	0	1.3813	0.0519	0.037573

Hyderabad CP sequence has been taken as reference sequence for this study

Table 3. Nucleotide (Nt) and amino acid (AA) sequence identity matrix showing percent similarity in CP of different RTBV isolates

AA	Nt											
	Hyd	**KK**	**Punjab**	**CWB**	**WB**	**AP**	**Japan**	**Mal**	**Phil**	**Phil- Ic**	**Phil-G1**	**Thai**
Hyd	100	95	95	95	96	95	85	85	85	85	85	83
KK	99	100	94	95	95	100	87	86	87	86	86	84
Punjab	99	100	100	96	96	94	85	84	85	84	84	83
CWB	97	98	98	100	96	95	84	85	85	84	84	83
WB	99	99	99	97	100	95	84	84	84	84	84	83
AP	99	100	100	98	99	100	87	86	87	86	86	84
Japan	91	91	91	89	91	91	100	92	99	94	94	94
Mal	93	93	93	90	92	93	98	100	93	92	92	92
Phil	92	92	92	90	91	92	99	98	100	94	94	95
Phil-Ic	91	91	91	89	91	91	98	97	99	100	100	90
Phil- G1	91	91	91	89	91	91	98	97	99	100	100	90
Thai	91	91	91	89	90	91	98	98	99	98	98	100

Hyd-Hyderabad, KK-Kanyakumari, CWB-Chinsura West Bengal, WB-West Bengal, AP-Andhra Pradesh, Mal-Malaysia, Phil-Philippines, Thai-Thailand