Arch Virol (2002) 147: 1173–1187
DOI 10.1007/s00705-002-0801-y

Archives of
Virology
Printed in Austria

Molecular analysis of two complete rice tungro bacilliform virus genomic sequences from India

N. Nath*, **S. Mathur,** and **I. Dasgupta**

Department of Plant Molecular Biology, University of Delhi South Campus,
New Delhi, India

Received September 13, 2001; accepted January 15, 2002
Published online April 26, 2002 © Springer-Verlag 2002

Summary. The complete genomic sequences of two geographically distinct isolates of rice tungro bacilliform virus (RTBV) from India were determined. Both the sequences showed equal divergence from previously reported Southeast Asian isolates. Numerous insertions, deletions and substitutions, mostly in the intergenic regions, were found. The genome sizes were 7907 and 7934 bp respectively, 95 and 68 residues short of an infectious clone reported earlier. Between them, both the isolates showed high homology all along the genome, except for a 30-nucleotide insertion/deletion close to the 3′ end of ORF III in one of them. Both the isolates indicated an unconventional start codon in ORF I, similar to the type isolate. In addition, as novel features, both the Indian isolates showed an unconventional start codon for ORF IV. Considering the low amounts of genome variability noticed in other RTBV isolates, the Indian isolates show that they have diverged sufficiently from the rest and should be considered belonging to a distinct strain.

Introduction

Rice tungro bacilliform virus (RTBV) is a plant pararetrovirus and belongs to the family *Caulimoviridae*. It is included in the newly formed genus RTBV-like viruses [15], whose members encapsidate a circular double-stranded DNA within a bacilliform particle. RTBV is closely related to the genus Badnavirus, which has *Commelina* yellow mottle virus (CoYMV) as the type member. Other Badnaviruses include Banana streak virus (BSV), Cacao swollen shoot virus (CSSV), Citrus mosaic bacilliform virus (CMBV), *Dioscorea alata* bacilliform virus (DaBV) and Sugarcane bacilliform virus (ScBV). RTBV, along with Rice

*Present address: University of Pittsburgh School of Medicine, Department of Molecular Genetics and Biochemistry, Biomedical Science Tower, Pittsburgh, PA 15261 USA.

tungro spherical virus (RTSV), is the causative agent of the tungro disease of rice in Asia, which is prevalent in almost all rice-growing regions.

Complete sequence analysis of the genome of one infectious clone of RTBV isolated from Philippines [4, 11], henceforth referred to as Phil1, showed that it had a size of 8002 bp. The above genome encoded proteins from four open reading frames (ORFs), which were arranged close to each other with one small and one large intergenic region. The functions of some of the proteins encoded by the ORFs have been determined [12, 16, 18, 19]. The genome of one Indian isolate, IdD1 [6] was subsequently cloned and was found to have a characteristic deletion of 64 residues in the non-coding region, as compared to Phil1. In addition, using DNA cross hybridization, partial nucleotide sequencing and PCR amplification of the genomes of several RTBV isolates, it was concluded that those from Southeast Asia and from South Asia differ significantly from each other. The Southeast Asian types were represented by isolates from Philippines, Malaysia, Thailand, Indonesia, Myanmar and Vietnam and those from South Asia by isolates from India, Bangladesh and Nepal [5, 6]. This and other aspects of RTBV molecular biology have been critically reviewed by Hull [14]. Of the seven complete sequences of RTBV available in the database [3], all are from Southeast Asian region. Their lengths varied from 8000–8016 bp and showed a number of nucleotide substitutions between them, maximum being 394 [3]. Here we report the complete sequence analysis of two geographically distinct isolates of RTBV from India, representing the South Asian region. We present molecular evidence to show that these have diverged sufficiently from the rest of the sequences so as to be classified as belonging to a different strain.

Materials and methods

Collection of virus-infected material

Virus-infected material was collected from experimental farms of Bidhan Chandra Krishi Viswavidyalaya, Kalyani, West Bengal {West Bengal (WB) isolate} and from glasshouse-maintained cultures at Directorate of Rice Research, Hyderabad, Andhra Pradesh {Andhra Pradesh (AP) isolate}. The former was collected in September 1994 and the latter in January 1995. The WB isolate originated from natural infections from neighbouring farms and the AP isolate was earlier collected from West Godaveri district. The two collection sites of the isolates were approximately 1500 km apart.

Cloning and characterization of RTBV DNA

RTBV DNA was cloned using a crude virus preparation as well as from the total DNA of infected plants. For the crude virus preparation, the method of Jones et al., 1991 [17], was followed. The viral DNA was then digested to completion with a number of restriction enzymes and cloned in either pUC18 or pBluescript SK(+/−), using standard procedures. A similar approach was utilized to obtain RTBV clones from the total DNA isolated from infected plants. The clones were selected by hybridizing with a radioactive probe representing Phil1 in the clone pRTBV101 [6]. The identities of the clones were established by constructing restriction maps and checking the arrangement of the derived fragments by hybridization with

defined fragments of pRTBV101 (data not shown). Hybridizing conditions were as described in Sambrook et al. [23].

Sequencing and analysis

A combination of nested deletions and primer walking was utilized to derive the complete sequence of the RTBV clones. Regions of ambiguity were resolved by repeated sequencing. The sequences were assembled and analyzed using the Jotun Hein or Clustal algorithm of MegAlign^{TM} program of LASERGENE software from DNASTAR Inc. at its default settings.

The DNA sequences have been deposited in the EMBL Sequence Database and have been assigned the following accession numbers: AJ292232 (AP isolate) and AJ314596 (WB isolate).

Results

Cloning and restriction analysis of Indian RTBV DNA

WB isolate

Several RTBV-specific clones were obtained on screening a library constructed from crude viral DNA using an RTBV-specific radioactive probe. One clone was found to have an insert size of 8.0 kb on *Bam*HI digestion and was named pRTBV203, following the convention of naming RTBV clones [6]. Several other cross-hybridizing clones to pRTBV203 were also obtained and were seen to have inserts of sizes around 4.0 kb. pRTBV203 was assumed to represent the complete genome of RTBV, based on its size.

AP isolate

A total of approximately 1000 clones representing tungro-infected rice DNA, when screened with the above probe, selected two, which on analysis were seen to have insert sizes of 6.5 kb and 1.5 kb respectively, on *Cla*I digestion. These were restriction mapped and their organization was confirmed by hybridization with defined fragments of pRTBV101 [6], representing Phil1. The clones were named pRTBV204 and pRTBV205, respectively and were found to be contiguous with each other when their sequences from the cloning site inwards were compared to Phil1 and to the WB isolate, which was full-length. It was concluded that, pRTBV204 and pRTBV205, together represented the complete sequence of the AP isolate. This assumption was strengthened by analyzing additional clones from the same infected plants, straddling the cloning sites, which showed restriction and hybridization patterns indicating that the two clones were contiguous (data not shown).

Restriction maps, using twelve restriction enzymes were constructed of both the genomes and compared to that of Phil1. A number of differences were noticed (data not shown), prominent among them being absence of *Sal*I, *Sac*I and *Sph*I sites in both the Indian isolates and absence of *Bam*HI site in the AP isolate. In addition, there were differences in the number of restriction sites in case of *Bst*XI, *Bgl*II, *Cla*I, *Eco*RI, *Hin*dIII, *Kpn*I and *Xho*I between the two Indian isolates.

Sequence analysis

The total length of the WB isolate, represented by pRTBV203, was 7934 bp and that of the AP isolate, represented by pRTBV204 and pRTBV205, was 7907 bp. WB isolate had nucleotide composition of A 40.5%, C 16.1%, G 18.1% and T 25.2% in the positive strand. The AP isolate was similar at A 40.8%, C 16.0%, G 17.9% and T 25.1%. Sequence alignments with Phil1 showed that there were approximately 2000 nucleotide substitutions, mainly in the intergenic regions in the Indian isolates, whereas, between the two, there were only 372. There were 77 locations where insertion/deletions (indels) were detected between the Indian isolates and Phil1, majority being located in the intergenic regions (Fig. 1). Of those located in the coding regions, majority was in multiples of three nucleotides, which did not affect the reading frame. Most of the indels in the intergenic region involved single nucleotides. One of the notable features was the deletion of 30 nucleotide residues in the AP isolate after position 5912. The numbering of the nucleotide residues was done according to the other published RTBV sequences [3]. The arrangement of ORFs, four in number, in both the isolates was the same as that of the other RTBV sequences published earlier, despite such large numbers of nucleotide changes. Figure 2 illustrates the phylogenetic relationships in ORFs I, II, and III at the amino acid level between the two Indian sequences and those of other closely related viruses available from the database using the Jotun Hein algorithm. As ORF IV is present only in RTBV, a similar comparison in it could not be done. It is abundantly clear that both the Indian sequences cluster together and share strongest relationship with Phil1.

ORF I

The non-conventional codon, ATT at position 98, coding for the amino acid isoleucine is the start of ORF I as indicated by sequence alignment. The ORF terminated at position 670, similar to Phil1. The Indian isolates showed nucleotide and amino acid identities around 78% and 79%, respectively with the Phil1 isolate. Between them, the two Indian isolates had a corresponding figure of 96.7% and 99.5%. Both of them had an insertion of a histidine residue at position 38 of the 190 amino acid protein. Out of a total of 40 amino acid substitutions in the Indian isolates, all except one being identical to both, 16 involved a change in the charge of the residue.

ORF II

This ORF initiated at position 667 and continued till 1002, coding for an estimated 111 amino acids, one more than that of Phil1. Nucleotide and amino acid identities of around 81.5% and 87% were found between both the Indian and the Phil1 isolates. Between the two Indian isolates, they had a corresponding figure of 97.3% and 100%. Both showed an insertion of a serine residue in position 82 of this protein. Two out of the 14 changes changed the overall charge of the residues. The consensus sequence PKKGIKRKYP, observed in many members of the genus Badnavirus and other known RTBV was also present in both the isolates.

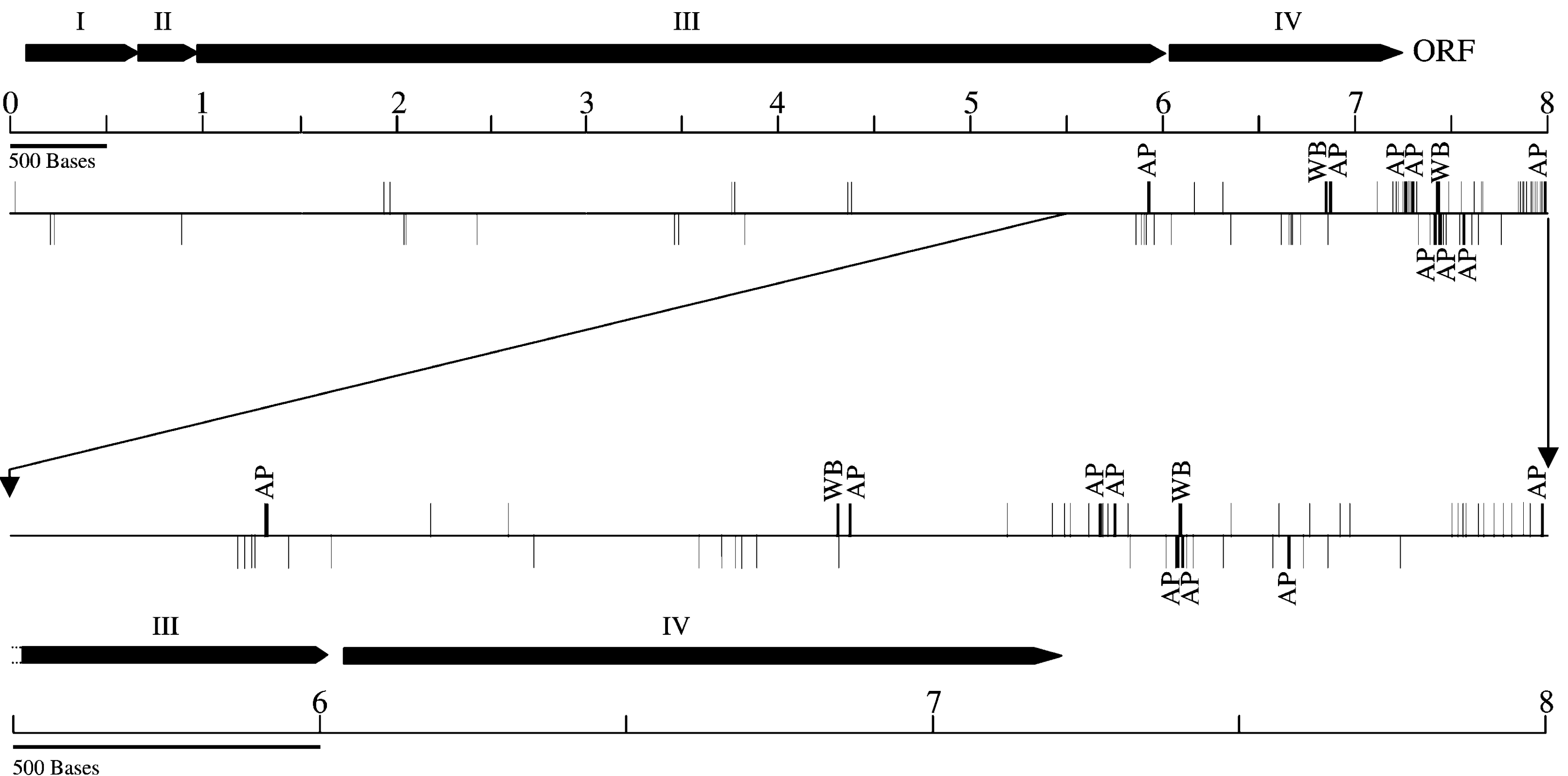

Fig. 1. Diagrammatic representation of indels present in two Indian isolates of rice tungro bacilliform virus as compared to Philippine isolate, Phil1. The solid arrows indicate the four ORFs, numbered at the top, the horizontal straight line below it represents a linearised version of the 8.0 kb Phil1 genome. Each vertical line at the top depicts a deletion and at the bottom an insertion event in both Indian isolates or otherwise specified, being present only in the Andhra Pradesh (*AP*) or West Bengal (*WB*) isolate in thick lines. Lower part shows an enlarged view of the genomic DNA from nucleotide position 5500 to 8000. The indels have been marked on the basis of the alignment report generated after aligning the above three sequences using the Clustal algorithm in MegAlign™ program of LASERGENE software from DNASTAR Inc

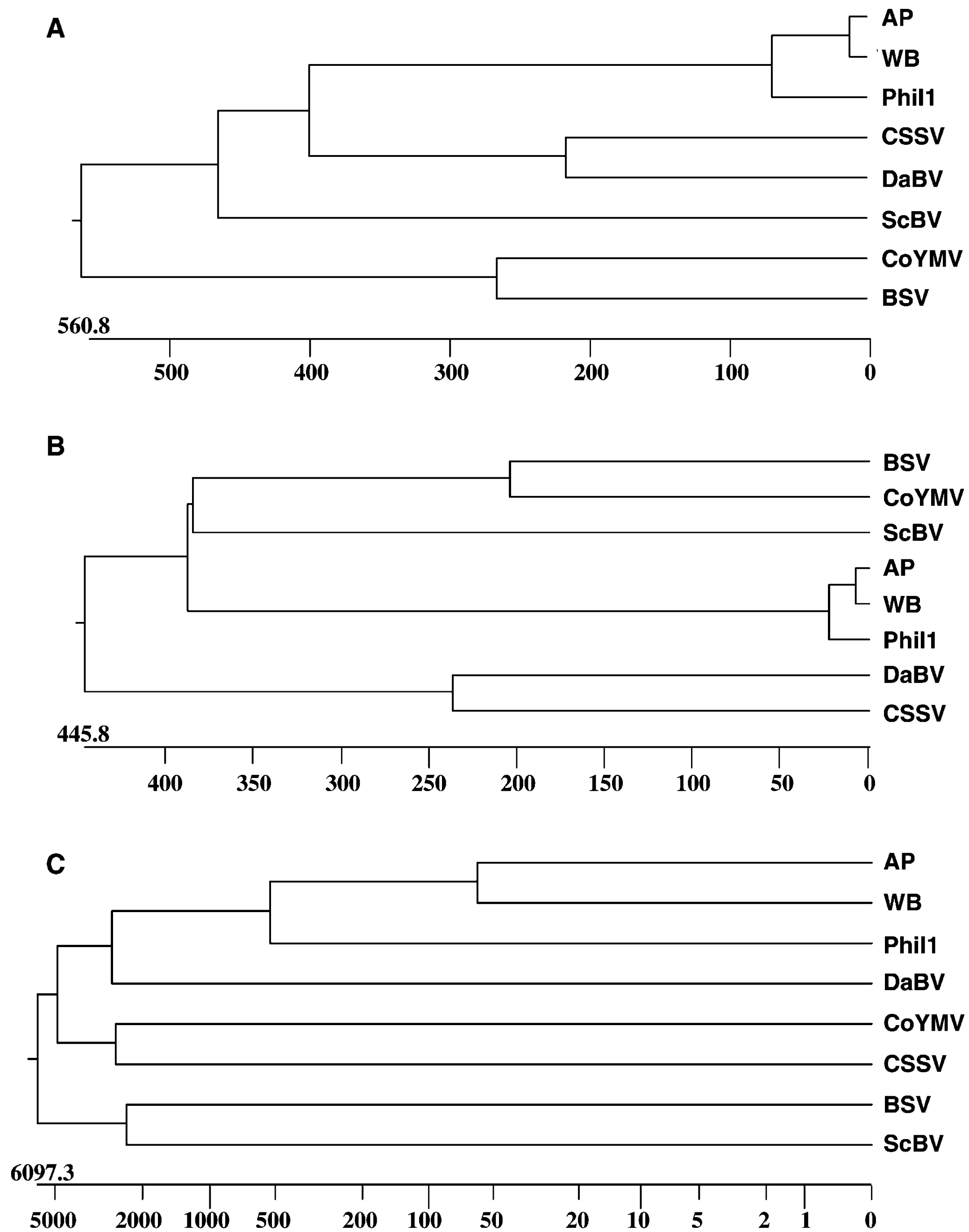

Fig. 2. Phylogenetic analysis of amino acid sequences encoded by ORFs between rice tungro bacilliform virus Andhra Pradesh (AP), West Bengal (WB) and Philippines (Phil1) isolates with members of genus Badnaviruses. **A** Phylogenetic tree of ORFI, **B** Phylogenetic tree of ORFII, **C** Phylogenetic tree of ORFIII. The analysis was done using Jotun Hein algorithm in MegAlign^{TM} program of LASERGENE software from DNASTAR Inc. The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences. Units indicate the number of substitution events. The above sequences were obtained from EMBL database having the following accession numbers: BSV, Banana streak virus (AJ002234); CSSV, Cacao swollen shoot virus (L14546); CoYMV, *Commelina* yellow mottle virus (X52938); DaBV, *Dioscorea alata* bacilliform virus (X94576), Phil1 (X57924) and ScBV, Sugarcane bacilliform virus (M89923)

ORF III

This ORF initiated at nucleotide position 999. In the WB isolate, it terminated at position 6023, but in the AP isolate the end point was shifted to position 5993 because of the deletion of 30 residues following position 5912. ORF III expresses to form a polyprotein, termed P194, coding for an estimated 1675 amino acid residues in Phil1, with distinct domains for coat protein (CP), aspartate protease (PR) and reverse transcriptase/ribonuclease H (RT/RNase H) [11]. In addition, the amino terminal portion exhibits similarities to several viral movement proteins [10]. Thus, the sequences of the Indian isolate in this ORF were analyzed between regions expressed as functional proteins whose boundaries have already been reported [14, 19]. The WB isolate encoded a polyprotein of 1674 amino acid residues whereas, the AP isolate encoded 1664 amino acid residues in P194. In this ORF, overall nucleic acid and amino acid identities of 78.8% and 80.5% with WB isolate and 78.5% and 80.2% with AP isolate were observed, compared to the Phil1 sequence. Between them, they showed corresponding figures of 95.3% and 98.4%. The lowest identity was in the PR region, where both the Indian isolates showed only about 69.2% and 65.2% identities at the nucleotide and amino acid levels, whereas between them, the figures were 96.1% and 99.1% respectively. The rest of the domains showed a consistent figure of approximately 84.0% and 88.7% at the nucleotide and amino acid levels, whereas between them, the four domains were equally identical at around 96% and 99%, respectively.

The N-terminal 300 amino acids of the P194 have residues, which are conserved in several pararetroviruses. Many of them are also characteristically located in the movement protein regions [10]. The above were also conserved in both the Indian isolates. However, there were 51 amino acid substitutions, in them, out of which 11 involved a change in the electrical charge.

The part of the P194 between amino acid residues 476 and 791 is the functional CP [19]. Within this region, the two Indian isolates showed complete identity, except a substitution between an alanine and threonine at position 504. On the other hand, there were 33 replacements between them and Phil1, 5 of which resulted in a change in the electrical charges. The most characteristic feature of this region, observed in all pararetroviruses, is the "cys" motif, with the conserved sequence $CX_2CX_4HX_4C$, found between amino acid residues 775–788. This motif was conserved in both the Indian isolates. A lysine-rich region, observed in Phil1 between resides 743–771 was also conserved in both the Indian isolates. In this region, there were 31% lysine residues and 51.7% lysine + arginine residues. Within this region, there were a total of six amino acid replacements between Phil1 and the two Indian isolates, two involving lysine to arginine and two arginine to lysine. Another consensus sequence, present in several pararetroviruses ($CX_2CX_{11}CX_2CX_4CX_2C$), located between positions 851–877 in the Phil1 was somewhat altered in the AP isolate only. The first two cysteine residues were deleted. The rest of the motif matched perfectly with the consensus sequence, except that the "X_{11}" region had one residue deleted in the AP isolate. In addition, a conserved integrase domain, present between residues 669–680 in Phil1 was also

conserved in both the Indian isolates, except that one valine residue was replaced by an isoleucine in both the Indian isolates.

The PR region of P194, between residues 920–1140 showed 77 substitutions out of 221 including two deletions, the changes being mostly identical in both, except in three cases. Out of these changes, 21 resulted in a change in the electrical charge. Also, there was a deletion of eight amino acid residues in both the Indian isolates, as compared to Phil1, very close to the N-terminal end of the RT-RNase polypeptide. The PR consensus sequence, DSG, was conserved in both Indian isolates.

The C terminal part of P194, encoding the RT/RNase H, shows a consensus sequence for RT at position 1339 of Phil1 (YIDD), which was perfectly conserved in the Indian isolates. In the RT/RNase region, between amino acid residues 1140 till 1619, there were 39 and 40 substitutions between the WB and AP isolates and the Phil1 isolates respectively, 10 of which resulted in a change in the electrical charge. A major difference between the two Indian isolates was noticed after nucleotide position 5912, corresponding to amino acid residue 1635, where a stretch of 30 nucleotides were deleted only in the AP isolate. This resulted in a deletion of 10 amino acid residues close to the C-terminus of P194 (Fig. 3). In the WB isolate, in this region, as compared to Phil1, 16 nucleotides were substituted and one deleted. At the amino acid level, the results were more drastic. There was no significant amino acid identity between the WB isolate and Phil1 in this region. The region of P194 towards the C terminal from position 1624 of Phil1, showed extremely low identity to both the Indian isolates. Both the Indian isolates showed insertion of a block of six amino acid residues, following position 1627 of Phil1. Between the two Indian isolates, other than the 10 amino acid indels, there were replacements in only three residues. In the entire ORF, certain regions stood out as being completely identical in the Indian sequences and Phil1, e.g. amino acid residues 148–199, 421–450, 549–578, 585–633, 715–746, 1084–1109, 1188–1212, 1293–1347, 1363–1402, 1415–1449, 1510–1551 and 1573–1623.

```
      1600               1619
                           ▼
Phil1 KGNKNFLPNFLSREGDFILKCLQNPDST------ESYSIDSSESIPLYIDSKESHSI
WB    KGNKNFLPNFLSREGDFILKCLQNTTSTIESSDSRSSTIESSDSQSSIKESLDSWTP
AP    KGNKNFLPNFLSREGDFILKCLQNTASIIESSDSRSSTIESSDSQS----------P

                          1675
Phil1 ESDDSIPLYRDKLLPLVERLKEKSA
WB    TKESSDSSCRIKLHPLSSRQKVKGE
AP    TKESSDSPCRIKLHPLSSRQKVKGE
```

Fig. 3. Amino acid alignment of P194 between rice tungro bacilliform virus isolates from Philippines (Phil1), West Bengal (WB) and Andhra Pradesh (AP) from position 1600 to 1675. The positions of amino acid residues are indicated. Position 1619 indicates the C-terminal end of RT/RNase H peptide. Residues, which are changed in the Indian isolates as compared to Phil1, are indicated in bold. Residues, which differ between WB and AP, are underlined. Horizontal bars "–" indicate deletions

ORF IV

In Phil1, this ORF starts 20 nucleotides following the stop codon of ORF III. Comparison of the derived amino acid sequences corresponding to nucleotide position 6047 in WB isolate and 6017 in AP isolate, 23 nucleotides downstream of the terminations of their previous ORFs, revealed the presence of another ORF. This ORF started with an isoleucine residue rather than the conventional methionine codon. However, RT-PCR analysis has shown earlier that this ORF in Phil1 to be expressed by a spliced transcript, the splice occurring between sORF1 at position 7504 and position 5973 [7]. The resultant protein starts at the AUG codon of sORF1 and continues from the nucleotide position 5973 into ORF IV, implying that the methionine residue encoded by the codon at position 6042 is not an initiating codon. Thus, the presence of isoleucine residue at that position in the Indian isolates was not surprising. In the AP isolate, the ORF terminated at position 7183 and in the WB isolate, at position 7213. This ORF could be potentially translated into a polypeptide having 388 amino acid residues. The above polypeptide, termed P42 exhibited as many as 169 mismatches (162 substitutions and 7 indels) between the Indian isolates and Phil1, evenly distributed all along the polypeptide. The above substitutions resulted in 48 and 51 residues showing altered electrical charges between the WB and AP and the Phil1 isolates, respectively. There were just 57.3% and 56.6% amino acid identities between the isolates from WB and AP and Phil1 respectively in this ORF (Table 1). On the other hand, the two Indian isolates showed only 8 substitutions, 5 of which were clustered between residues 273–279. Interestingly, three of the five substituted residues, mentioned above in WB isolate and Phil1 were identical.

Intergenic regions

Out of the two intergenic regions, the large one, between ORFs IV and I, contains approximately 890 nucleotides, whereas the small, between ORFs III and IV

Table 1. Percentage nucleotide and amino acid identities between the two Indian and the Phil1 isolates of rice tungro bacilliform virus

ORFs/Domains	AP/Phil1		WB/Phil1		WB/AP	
	Nuc.	AA	Nuc.	AA	Nuc.	AA
ORF I	77.9	79.4	78.8	78.8	96.7	99.5
ORF II	81.4	87.3	81.7	87.3	97.0	100.0
ORF III (overall)	78.5	80.2	78.8	80.5	95.3	98.4
MP	82.6	85.0	83.2	84.7	95.6	98.2
CP	85.1	89.6	84.3	89.6	95.8	99.7
PR	69.6	65.2	68.8	65.2	96.1	99.1
RT/RNase H	84.5	91.7	84.8	91.9	95.8	99.4
ORF IV	63.1	56.6	63.9	57.3	95.0	97.9
Intergenic	65.0	–	65.6	–	94.9	–
Overall	74.7	–	75.2	–	95.4	–

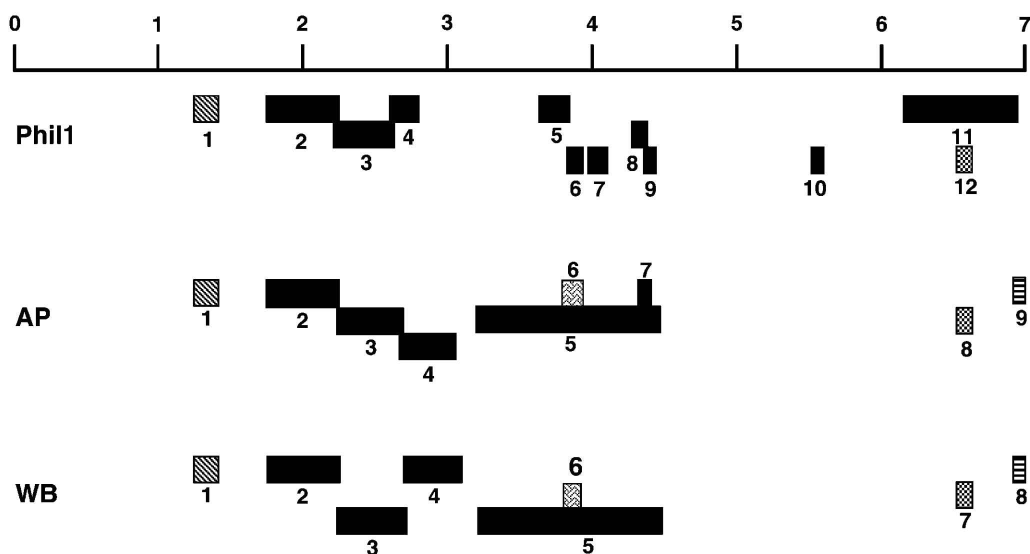

Fig. 4. Schematic representation of the relative positions of short Open Reading Frames (boxes) in the three reading frames, present in the leader sequence of rice tungro bacilliform virus Philippine (Phil1), Andhra Pradesh (AP) and West Bengal (WB) isolates. The scale shows base pairs × 100, the transcription start site being kept at zero position. Similarly shaded boxes depict those showing amino acid sequence identity, whereas filled boxes indicate those showing no identity

contains 23 and 20 nucleotides in the Indian and Phil1, respectively. In Phil1, TATA at position 7373 is involved in the promotion of transcription of the viral genome and AATAAA at 7598 is involved in its termination. These sequences were perfectly conserved in the Indian isolates. The splice donor and acceptor sites, involved in splicing the transcript for expression of ORF IV, were also perfectly conserved in both the sequences. A series of 12 short ORFs (sORFs) are found downstream to the promoter in Phil1. In the Indian isolates, their numbers were 8 (WB) and 9 (AP), respectively, as illustrated in Fig. 4. Only sORF1 was seen to be identical in both Indian and Phil1 isolates at the amino acid level. The amino acid sequence encoded by sORF6 in both the Indian isolates was same. Similarly, sORF12 (Phil1), sORF8 (AP) and sORF7 (WB) encoded the same amino acid sequence. Same was true for sORF8 (WB) and sORF9 (AP). All other sORFs were different from each other.

Discussion

The full-length nature of a viral genome is only confirmed by demonstrating infectivity to the natural host. Several attempts were made to check whether cloned WB isolate DNA was infectious, by agroinoculation [4]. However, the partially repeated viral DNA construct in a binary plasmid, required for the process, could not be obtained in *E.coli*. The full-length nature of the pRTBV203 was, however,

strongly indicated by nucleotide sequencing and restriction mapping of other partial clones derived from the WB isolate.

Out of the two Indian RTBV clones analyzed in this study, the WB isolate, represented by pRTBV203 had an insert size of approximately 8.0 kb, determined by gel electrophoresis, which was similar to the previously reported RTBV clones. However, the AP isolate could not be obtained as a full-length clone. The two partial clones, represented by pRTBV204 and pRTBV205, were concluded to represent the full genome on the basis of the sequence analysis. The boundary sequences of pRTBV204 and pRTBV205 perfectly matched with adjacent nucleotides of pRTBV203, which was a full-length clone. This strongly indicated that the cloned viral DNA represented the complete genome. However, the presence of small additional portions of DNA flanked by *Cla*I sites in the viral genome, which were not cloned could not be completely ruled out, but was considered highly unlikely.

Earlier reports of complete genomic sequences of different RTBV isolates, six from Philippines and one from Malaysia have shown that their lengths varied from 8000–8016 bp [3], the maximum number of nucleotide substitutions being 394 [3]. The only infectious clone reported, Phil1, has a length of 8002 bp. The two Indian RTBV sequences reported here were much shorter, 7907 and 7934 bp respectively and showed approximately 5-fold more substitutions, i.e., approximately 2000, as compared to the Phil1 isolate. This, taken together with the indels, involving one or more nucleotides at 77 locations in the genome, indicated that the two Indian isolates belonged to a different strain of RTBV. This was further strengthened by the observation that nucleotide identities of different ORFs and functional domains with those of the Phil1 were lower than 85% as compared to the earlier reports, which showed mostly above 95%. Comparison of the only available Malaysian isolate sequence with those from the Philippines has shown a minimum identity of 84% [3] in the intergenic region, whereas the same figure for the Indian isolates was as low as 65% (Table 1). This clearly demarcated the sequences into the Indian strain and the rest. Fan et al., 1996 [6], first reported the existence of two strains in RTBV, based on nucleic acid hybridization and partial sequencing. The results presented here provide strong molecular evidence in favour of the above. Analysis of the phylogenetic relationship between the two sequences reported here and other related viruses also supported the above.

The use of isoleucine (ATT), a non-conventional initiation codon by both the Indian isolates for ORF I established it as a general phenomenon in RTBV. In Phil1, two ATT codons are present at positions 87 and 99, out of which, the second one is used as the start codon [8]. Interestingly, in both the Indian isolates, an ACT codon was present at position 86 and thus the conserved ATT codon at position 98 was most likely to act as the initiator.

Table 1 showed that out of all the different domains in P194, the PR domain exhibited significantly lower identities between the Indian isolates and Phil1, but not so between the two Indian sequences. This reflected the close relationship between the two Indian sequences and indicated their evolutionary distinctiveness from the RTBV sequences reported earlier. It is possible that the Indian isolates encode a distinct protease, which needs to be further characterised.

The N-terminal of RT-RNase H polypeptide is formed by the specific cleavage between amino acid residues 1138(Y) and 1139(K), mediated by RT [18]. In both the Indian isolates, eight amino acids, which occupy the third to eleventh position N-terminal to the above cleavage site were deleted. This might reflect a novel recognition sequence of the protease, which anyway shows a high level of amino acid substitution, as mentioned above. The C-terminal end of P194 in both the Indian isolates showed very high levels of substitutions and indels, especially after the residue 1617 and 1614 in WB and AP, respectively, indicating that the above region could tolerate a variety of residues with different physicochemical properties. As the C-terminal end of RT is at position 1619 in Phil1 [18], it would be interesting to determine the role of the rest of the terminal 56 amino acid residues of P194, which showed insertion of six new residues in both the Indian isolates and deletion of ten residues only in the AP isolate (Fig. 3). To confirm the above deletion of 30 nucleotide residues of ORF III in AP isolate, oligonucleotides were designed from sequences flanking the region and the amplified DNA from tissue extracts, infected with AP isolate were analysed. Results (data not shown) indicated the presence of the above deletion in most of the amplified DNA, strongly indicating that this deletion is present in a majority of the viral population in the AP isolate. The above deletion might be linked to some of the symptomatological differences noticeable between the two Indian isolates (F. R. Niazi, Indian Agricultural Research Institute, New Delhi, unpublished observation). The very high substitution rate of amino acid residues (approximately 43%) in ORF IV between the Phil1 and the Indian isolates was intriguing. In contrast, the almost complete conservation of the same between the two Indian isolates strongly pointed to its essential role.

The sORFs have been proposed to have a positive role in the expression of ORF II and III by leaky scanning of the pregenomic RNA by ribosomes [9]. The differences in the organization of sORFs between the Indian isolates and also between them and Phil1 may reflect changes in the control of the expression of ORF II and ORF III. The proposed essential role of sORF1 in translating ORF IV may be one of the factors for the absolute conservation of the amino acid sequence of this ORF across the Indian isolates and Phil1.

The incidence of several different "strains" of "rice tungro viruses" in India has been reported earlier based on symptomatology and transmission characteristics [1, 20–22]. A careful evaluation of the symptoms and transmission characteristics of the isolates described in this report need to be performed in order to determine their relationship to the strains described previously.

Fan et al., 1996 [6], had reported the partial sequencing and analysis of a cloned Indian RTBV isolate, IdD1. The WB isolate described in this report originated from the same region as IdD1 but was collected after a period of several years from the field. One of the important features described was the presence of a 64 bp continuous deletion, starting at the 5′ region of the *Sal*I site at position 7914 till 7977 of Phil1, corresponding to the large intergenic region. Comparison of the available partial sequence of IdD1 showed it to be identical

to pRTBV203. However, alignment of the sequence of the two Indian isolates, using Clustal method indicated that instead of a continuous deletion, a series of smaller deletions, starting at position 7850 and extending till position 7992 of Phil1 is more likely, as illustrated in Fig. 1. The length of the WB isolate reported here along with the nucleotide and amino acid identities with Phil1 agreed with that of an "isolate from India" mentioned earlier [14], whose details were not provided, but which, most likely, represents IdD1. Detailed studies (I. Dasgupta, unpublished data) indicated that there were a few restriction site differences between IdD1 and WB isolate, which might represent accumulated mutations between the two isolates over time or might represent different molecular species, as reported from Philippines [2]. Deletion of 30 nucleotide residues at the 3′ end of ORF III had not been reported in any of the RTBV isolates described earlier [3, 6].

Previous studies [5, 6] had indicated that all the RTBV isolates from within the Indian subcontinent were homogenous and were designated the "South Asian" type. The present study, utilizing just two isolates, points out that important differences can occur within the above group, for example the deletion of 30 nucleotide residues within ORF III, which might be considered to be a "hypervariable" region. The other such region could be the region coding for the sORFs. A thorough survey of field isolates of RTBV in India, as reported from Philippines [2], may give a clearer picture of the true extent of variability.

To meet the threat of a tungro epidemic devastating the rice harvest in India, as occurred in the state of Punjab in North-West India in 1998 [25], it is essential to plan non-conventional resistance strategies against tungro, based on viral transgenes in cultivated rice varieties. Tungro resistance has been engineered in rice by using the CP and replicase of RTSV, the helper virus for RTBV [13, 24], even though the latter causes most of the tungro symptoms. However, as the true nature of the helper function in the tungro virus complex is unknown, it would be desirable to engineer resistance against RTBV as well. This is important because of the high degree of RTBV variability existing at the field level, as reported here and the possibility of emergence of novel strains which might be independent of RTSV for insect transmission. In such a scenario, it is considered extremely important to have a correct picture of the variability of RTBV in the Indian subcontinent, which might then form a basis for deployment of genes as a strategy for developing tungro resistance.

Acknowledgements

The authors are grateful to Prof. S. Mukhopadhyay, Bidhan Chandra Krishi Viswavidyalaya, Kalyani, West Bengal and Dr. A. Ghosh and Dr. E. A. Siddiq of Directorate of Rice Research, Hyderabad for the gift of tungro-infected samples. This work was supported by a research grant to I. D. from the Department of Biotechnology, Government of India. N. N. and S. M. were supported by research fellowships from University Grants Commission and Council for Scientific and Industrial Research, New Delhi.

References

1. Anjaneyulu A, John VT (1972) Strains of rice tungro virus. Phytopathology 62: 1116–1119
2. Arboleda M, Azzam O (2000) Inter- and intra-site genetic diversity of natural field populations of rice tungro bacilliform virus in the Philippines. Arch Virol 145: 275–289
3. Cabauatan PQ, Melcher U, Ishikawa K, Omura T, Hibino H, Koganezawa H, Azzam O (1999) Sequence changes in six variants of rice tungro bacilliform virus and their phylogenetic relationships. J Gen Virol 80: 2229–2237
4. Dasgupta I, Hull R, Eastop S, Poggi-Pollini C, Blakebrough M, Boulton MI, Davies JW (1991) Rice tungro bacilliform virus DNA independently infects rice after *Agrobacterium*-mediated transfer. J Gen Virol 72: 1215–1221
5. Druka A, Hull R (1998) Variation in rice tungro viruses: further evidence of two rice tungro bacilliform virus strains and possibly several rice tungro spherical virus variants. J Phytopathol 146: 175–178
6. Fan Z, Dahal G, Dasgupta I, Hay J, Hull R (1996) Variation in the genome of rice tungro bacilliform virus: molecular characterization of six isolates. J Gen Virol 77: 847–854
7. Fütterer J, Potrykus I, Brau MPV, Dasgupta I, Hull R, Hohn T (1994) Splicing in a pararetrovirus. Virology 198: 663–670
8. Fütterer J, Potrykus I, Bao Y, Li L, Burns TM, Hull R, Hohn T (1996) Position-dependent ATT initiation during plant pararetrovirus rice tungro bacilliform virus translation. J Virol 70: 2999–3010
9. Fütterer J, Rothnie HM, Hohn T, Potrykus I (1997) Rice tungro bacilliform virus open reading frames II and III are translated from polycistronic pregenomic RNA by leaky scanning. J Virol 71: 7984–7989
10. Hagen LS, Jacquemond M, Lepingle A, Lot H, Tepfer M (1993) Nucleotide sequence and genomic organization of cacao swollen shoot virus. Virology 196: 619–628
11. Hay JM, Jones MC, Blakebrough ML, Dasgupta I, Davies JW, Hull R (1991) An analysis of the sequence of an infectious clone of rice tungro bacilliform virus, a plant pararetrovirus. Nucleic Acids Res 19: 2615–2621
12. Herzog E, Guerra-Peraza O, Hohn T (2000) The rice tungro bacilliform virus gene II product interacts with the coat protein domain of the viral gene III polyprotein. J Virol 74: 2073–2083
13. Huet H, Mahendra S, Wang J, Sivamani E, Ong C, Chen L, De Kochko A, Beachy RN, Fauquet C (1999) Near immunity to rice tungro spherical virus achieved in rice by a replicase-mediated resistance strategy. Phytopathology 89: 1022–1027
14. Hull R (1996) Molecular biology of rice tungro viruses. Ann Rev Phytopathol 34: 275–297
15. Hull R (2002) Nomenclature and classification of plant viruses. Matthew's plant virology, 4th edn. Academic Press, San Diego, pp 13–45
16. Jacquot E, Keller M, Yot P (1997) A short basic domain supports a nucleic-acid binding activity in the rice tungro bacilliform virus open reading frame 2 product. Virology 239: 352–359
17. Jones MC, Gough K, Dasgupta I, Subba Rao BL, Cliffe J, Qu R, Shen P, Kaniewska M, Blakebrough M, Davies JW, Beachy RN, Hull R (1991) Rice tungro disease is caused by an RNA and a DNA virus. J Gen Virol 72: 757–761
18. Laco GS, Kent SBH, Beachy RN (1995) Analysis of the proteolytic processing and activation of rice tungro bacilliform virus reverse transcriptase. Virology 208: 207–214
19. Marmey P, Bothner B, Jacquot E, De Kochko A, Ong CA, Yot P, Siuzdak G, Beachy RN, Fauquet CM (1999) Rice tungro bacilliform virus open reading frame 3 encodes a single 37 kDa coat protein. Virology 253: 319–326

20. Mishra MD, Niazi FR, Basu AN, Ghosh A, Raychaudhuri SP (1976) Detection and characterization of a new strain of rice tungro virus in India. Plant Dis Rep 60: 23–25
21. Mukhopadhyay S, Bandopadhyay B (1984) Incidence of different strains of rice tungro virus in West Bengal. Int J Trop Plant Dis 2: 125–132
22. Niazi FR, Basu AN, Mishra MD (1977) Differential transmission of two strains of the rice tungro virus by *Nephotettix nigropictus*. Fitopatol Brasil 2: 247–249
23. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
24. Sivamani E, Huet H, Shen P, Ong CA, De Kochko A, Fauquet C, Beachy RN (1999) Rice plant (*Oryza sativa* L.) containing rice tungro spherical virus (RTSV) coat protein transgenes are resistant to virus infection. Mol Breeding 5: 177–185
25. Varma A, Niazi FR, Dasgupta I, Singh J, Cheema SS, Sokhi SS (1999) Alarming epidemic of rice tungro disease in North-West India. Indian Phytopathol 52: 71–74

Author's address: Dr. I. Dasgupta, Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110021, India.