Journal of Plant Pathology (2018) 100:523–531
https://doi.org/10.1007/s42161-018-0116-9

ORIGINAL ARTICLE

# Molecular characterization of banana bract mosaic virus from India reveals recombination and positive selection in the VPg gene

Chelliah Anuradha[1] · Ramasamy Selvarajan[1]

Published online: 16 July 2018


## Abstract

*Banana bract mosaic virus* (BBrMV), a member of the genus *Potyvirus*, family *Potyviridae,* is an important viral pathogen affecting banana and plantains in India. The present study was undertaken to determine the genetic variation and molecular evolution of BBrMV infecting different cultivars of banana and plantain in the Indian subcontinent, based on the viral genome-linked protein (VPg) gene. Sequence identity of this gene from 29 BBrMV isolates showed a range of nucleotide (nt) and amino acid (aa) identity of 75–100% and 95–100%, respectively. Phylogenetic analysis based on nt revealed that except for two isolates (TN1 and TN2), all the other Indian isolates clustered together. Different functional motifs of the VPg gene previously reported were found to be conserved. A single possible recombination event was detected using a recombination detection programme. A codon-based selection analysis revealed that most of the codons in the VPg gene were under purifying selection except for those at position 46, 47, 71, 107, 149, 153, 156, 175, 176, and 178, which were under positive selection. Gene flow between different Indian populations of BBrMV from banana was relatively low. This is the first report on genetic diversity and evolution of the VPg gene of BBrMV. Since VPg is identified as a virulence determinant in resistance mediated by the eukariotic intiation factor 4E (eIF4E) in several plant-potyvirus interactions, any mutation in the VPg interacting domain may abolish the interaction with eIF4E or its isoforms in vitro and prevent viral infection *in planta*. Therefore, understanding the population structure of BBrMV based on VPg would potentially provide insight into the diversification and infection cycle of this virus.

**Keywords** *Banana bract mosaic virus* · Viral genome-linked protein · Genetic diversity · Recombination · Selection pressure

## Introduction

The genus *Potyvirus* is one of the largest genera of plant RNA viruses and is economically very important as potyviruses cause serious diseases in crop plants (Hull 2002). *Banana bract mosaic virus* (BBrMV), a member of the genus *Potyvirus,* is one of the most destructive viruses infecting banana and plantains in India, Sri Lanka and the Philippines. BBrMV genome is 9711 nucleotides (nt) in length and its Hawaii (HAW) isolate is 9713 nt with the viral genome linked protein (VPg) linked to the 5′ end of the genome and a poly-A tail at the 3′ end. The genome codes for a single long polyprotein that is processed by viral proteinases into 10 mature proteins. Most of the potyviral proteins have multiple functions, and the VPg is one of the most versatile of them. Potyviral VPg has been shown to take part in RNA replication, cell-to-cell and long-distance movement, translation, gene silencing suppression and phloem loading of the virus (Rantalainen et al. 2011). Self association of VPg to form dimers gives rise to a variety of functions (Oruetxebarria et al. 2001; Grzela et al. 2008; Rantalainen et al. 2008). The central domain of VPg interacts with the helper component proteinase (HC-Pro) and any variation in the VPg central domain affects the long distance movement of the virus (Roudet-Tavert et al. 2007). The molecular nature of recessive resistance was elucidated in which the VPg of several potyviruses was found to bind with the eukaryotic translation initiation factor 4E (eIF4E) or to its isoform [eIF(iso)4E] in yeast two-hybrid and in vitro binding assays (Wittmann et al. 1997; Schaad et al. 2000; Leonard et al. 2004). This interaction is very important in the virus life cycle since any mutation in the interacting domain of VPg will abolish the interaction with eIF4E/eIF(iso)4E, thereby preventing viral infection and establishment in plant (Leonard et al. 2004). In addition to its ability to bind eIF4E

✉ Ramasamy Selvarajan
selvarajanr@gmail.com

[1] ICAR–National Research Centre for Banana, Thogamalai Road, Thayanur post, Tiruchirapalli, Tamil Nadu 620 102, India

proteins, the potyviral VPg was also identified as a virulence determinant in eIF4E-mediated resistance in plant–potyvirus pairs such as pea seedborne mosaic virus (PSBMV) and potato virus Y (PVY) infecting *Pisum sativum* and *Capsicum annum* and *Lycopersicon hirsutum*, respectively (Borgstrom and Johansen 2001; Moury et al. 2004). Hence, the knowledge of VPg gene sequences and its diversity will be useful for understanding its role in the initiation of RNA synthesis, resistance breaking, etc. There are several reports on the genetic structure of potyvirus population and these reports have shown that virus populations have been shaped by selection, founder effects and genetic recombination (Tian et al. 2011; Zhang et al. 2011; Nguyen et al. 2013). Most of the codons in VPg are under negative selection which will help with elimination of deleterious mutations, thus assessing the importance of this gene for the virus life cycle as well as for the recent expansion of BBrMV population. To date, a diversity analysis of VPg has been done only for turnip mosaic virus (TuMV) and soybean mosaic virus (SMV), whose occurrence was reported in Europe and USA respectively in the early twentieth century while BBrMV was first reported in the Philippines in 1979 (Magnaye and Espino 1990).

The objective of this study was to characterise the VPg gene sequences of BBrMV isolates from different regions of Southern India where this virus is a major problem, and to assess the sequence diversity and phylogenetic relationship with known BBrMV isolates from other parts of the world. To gain further insight, 29 BBrMV isolates were analysed for SNPs (single nucleotide polymorphism), INDELs (insertion and deletion), evolutionary distance, selection pressure analysis (Ka/Ks ratio) and recombination.

## Materials and methods

**Virus isolates** Banana leaves or bracts showing symptoms of bract mosaic were collected from different locations in Tamil Nadu, Kerala, Karnataka, and Andhra Pradesh states of India. Infected samples were processed immediately or stored at −80 °C. Host plants of some of the isolates were maintained in insect-free glasshouse at the Indian Council of Agricultural Research-National Research Centre for Banana (ICAR-NRCB), Tiruchirapalli, Tamil Nadu.

**Preparation of RNA, cDNA synthesis and VPg amplification** For each isolate, 100 mg of leaf or bract tissues from infected plants were frozen in liquid nitrogen and ground to a fine powder. Total RNA was extracted using the RNeasy Plant Mini Kit according to the manufacturer's instructions (QIAGEN, USA), and the viral RNA was reverse-transcribed using RevertAid™ H Minus first strand cDNA synthesis kit (MBI Fermentas, USA) using oligo (dT) as primer, following the manufacturer's protocol. VPg primers were commercially synthesized: sequences of the forward (5'-GGAAAGCGAAAATT TCAAAAAC-3′) and reverse (5'-CTCAAACTCAACTG GCGTC-3′) primers were based on the sequence of the BBrMV-Try isolate available in NCBI database. Complementary DNA was subjected to PCR amplification using gene-specific primers to amplify 570 bp of VPg gene. The thermo-cycling conditions were as follows: 5 min at 95 °C (1 cycle), 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min (35 cycles), and a final extension at 72 °C for 10 min.

**Cloning and sequencing analysis** The amplified VPg gene was resolved by 1.5% agarose gel electrophoresis, and the fragments were eluted using GenElute Gel Extraction Kit (Sigma, USA), ligated into pTZ57R/T vector (MBI Fermentas, USA), and transformed into competent *E. coli* DH5α cells as per manufacturer's instructions. Plasmid DNA was purified using GenElute Plasmid Miniprep Kit (Sigma-Aldrich, USA) following manufacturer's instructions and the presence of VPg gene was confirmed by digestion with restriction enzymes *Eco*RI and *Hin*dIII. Two independent clones per isolate were sequenced in both directions using an automatic sequencer (Eurofins Genomic India, Bangalore). The complete VPg gene sequences of 25 BBrMV isolates generated in this study were deposited in GenBank as accession numbers KY367359-KY367383. Four VPg gene sequences of BBrMV isolates of banana from India and the Philippines and ornamental ginger from Hawaii were retrieved from GenBank for analysis. VPg sequences of 29 isolates were aligned using the CLUSTALW (Thompson et al. 1994). Sequence identity matrix and sequence difference count matrix were calculated using BioEdit sequence alignment editor version 5.09.04 (Hall 1999). The pairwise nucleotide (nt) and aminoacid (aa) sequence identity scores were represented as color-coded blocks using SDT v.1 software (Muhire et al. 2013). Phylogenetic analysis was performed using Maximum likelihood phylogenetic tree construct in the MEGA 7.0 software. The evolutionary distance was analyzed using the MEGA version 7.0 (Tamura et al. 2011) and SNPs and INDELs were calculated using DnaSP version 5.10 (Librado and Rozas 2009), which compared the nucleotide sequence with the BBrMV-Try isolate (HM131454) as the reference sequence. Ka/Ks value was also calculated using the DnaSP version 5.10 to analyze the synonymous and non-synonymous mutations at the nucleotide level. Recombination sites in the aligned sequences were identified by implementing the RDP4 Beta 4.24 software (Martin et al. 2015). Genetic differentiation between the BBrMV populations was examined by three permutation-based statistical tests, Ks*, Z, and Snn. The extent of genetic differentiation or the levels of gene flow between populations were measured by estimating Fixation index (FST) using DnaSP version 5.10. The absolute value of FST ranges from 0 to 1 for undifferentiated to fully differentiated populations. Normally, an

**Table 1** Analysis of evolutionary distance, SNPs, INDELs and Ks/Ka ratio, and per cent sequence identity of VPg gene of BBrMV isolates

| S.No | Isolate | Accession No. | Evolutionary distance | SNP | Ka | Ks | Ka/ Ks | nt identity % | aa identity % |
|---|---|---|---|---|---|---|---|---|---|
| 1 | TN1# | HM131454 | – | – | – | – | – | 75–98 | 96–97 |
| 2 | TN2 | KT852552 | 0.008 | 9 | 0.0226 | 0.0139 | 1.625 | 75–100 | 96–100 |
| 3 | TN3 | KY367359 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 4 | TN4 | KY367360 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 5 | TN5 | KY367361 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 6 | TN6 | KY367362 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 7 | TN7 | KY367363 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 8 | TN8 | KY367364 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 9 | TN9 | KY367365 | 0.716 | 139 | NA | 0.0488 | NA | 75–99 | 95–99 |
| 10 | TN10 | KY367366 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 11 | TN11 | KY367367 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 12 | TN12 | KY367368 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 13 | TN13 | KY367369 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 14 | TN14 | KY367370 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 15 | TN15 | KY367371 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 16 | KER1 | KY367372 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 17 | KER2 | KY367373 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 18 | KER3 | KY367374 | 0.716 | 139 | NA | 0.0487 | NA | 75–99 | 95–99 |
| 19 | KER4 | KY367375 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 20 | KER5 | KY367376 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 21 | KER6 | KY367377 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 22 | KER7 | KY367378 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 23 | KER8 | KY367379 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 24 | KER9 | KY367380 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 25 | AP | KY367381 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 26 | ASS | KY367382 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 27 | KAR | KY367383 | 0.710 | 138 | NA | 0.0463 | NA | 75–100 | 96–100 |
| 28 | PHI | DQ851496 | 0.006 | 7 | 0.0050 | 0.0147 | 0.340 | 75–98 | 96–98 |
| 29 | HAW | KT456531 | 0.010 | 11 | 0.0113 | 0.0221 | 0.511 | 75–99 | 95–98 |

TN-Tamil Nadu, AP-Andhra Pradesh, ASS- Assam, KER-Kerala, KAR-Karnataka

TN3 (Gros Michel), TN4 (Grand Nain), TN5 (Attu Nendran), TN8 (Kullan) – Coimbatore; TN1 (Nendran), TN2 (Nendran), TN6 (Poovan), TN7 (Ney Poovan), TN10 (Beula), TN11 (Mortman), TN13 (Malbhog), TN15 (H201) –Tiruchirapalli; TN9 (Red Banana) – Theni; TN12 (Nendran) – Mullikarumbur; TN14 (Mortman) – Thottiyam; KER1 (Chenkadali), KER2 (NjaliPoovan), KER3 (Palayankodan),KER4 (Vellaiputabale), KER5 (Nendran), KER6 (Pachabale), KER7 (Rasthali), KER8 (Tongat) and KER9 (Yangambi KM5) – Kannara;AP-Kovur; KAR – Chamraj Nagar; ASS - Kahikuchi * isolates generated in this study

[a] no INDELs were detected

# Used as a reference isolate in this study

absolute value of FST > 0.33 suggests an infrequent gene flow, while an absolute value of FST <0.33 suggests frequent gene flow between the populations compared. DnaSP version 5.10 was used to estimate Tajima's D, Fu, and Li's D* and F* statistical tests and haplotype diversity. Tajima's D, Fu, and Li's D* and F* tests hypothesize that all mutations are selectively neutral. Haplotype diversity refers to the frequency and number of haplotypes in the population. Nucleotide diversity estimates average pairwise differences among sequences. Datamonkey server (http://www.datamonkey.org) was used for the analysis of selection pressure acting on the VPg gene. Conserved domain protein architecture of VPg was modelled using ALL-IN-ONE-SEQ-ANALYSER version 1.35 (http://www-personal.umich.edu/~ino/blast.html).

## Results

Twenty seven VPg gene sequences of BBrMV-infected banana samples from Tamil Nadu, Kerala, Andhra, Karnataka and Assam were analysed together with one sequence each of virus isolates infecting bananas in the Philippines (PHI) and

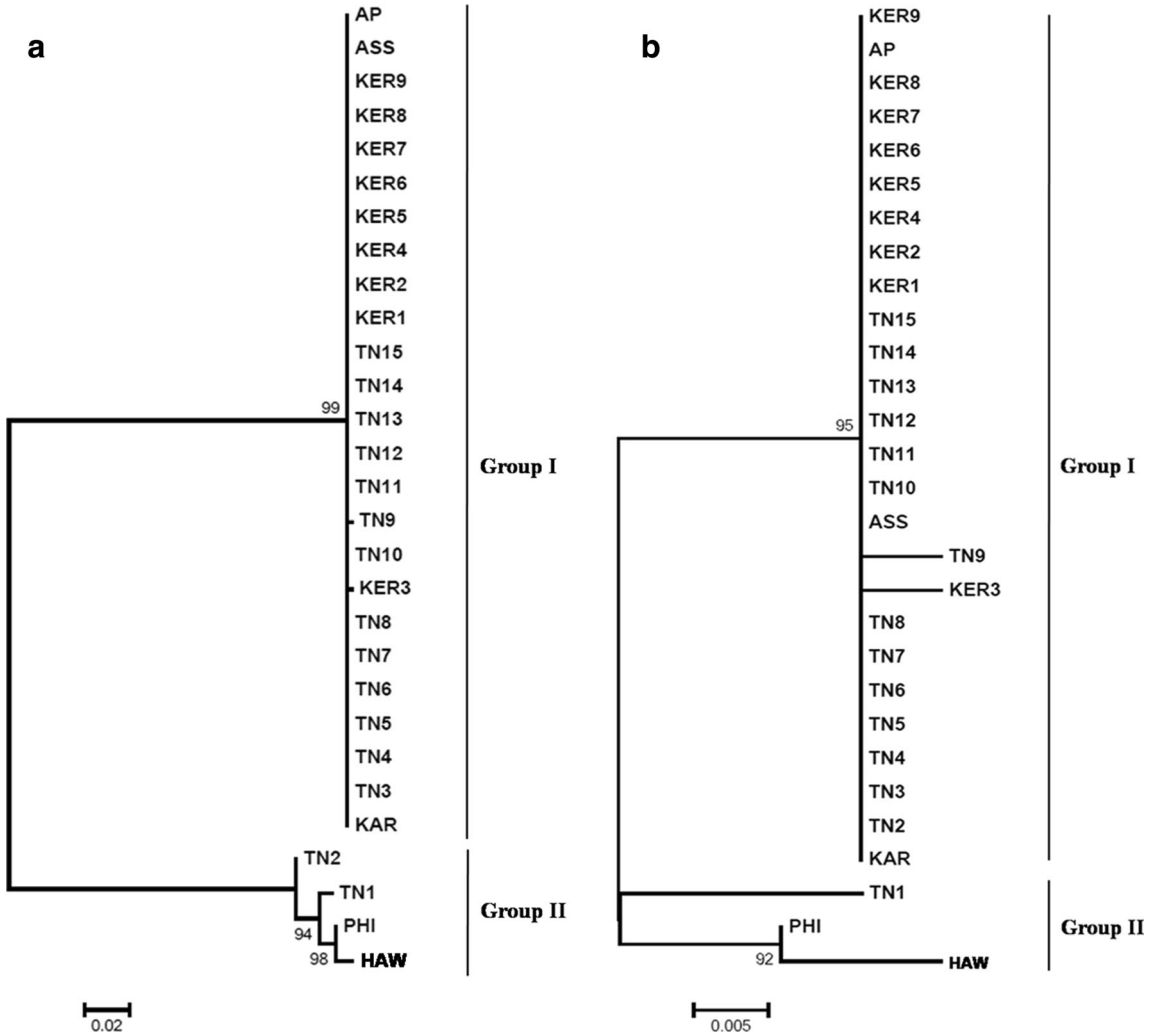


**Fig. 1** Phylogenetic analysis of nucleotide (**a**) and amino acid (**b**) sequences of the viral genome linked protein (VPg) of BBrMV isolates from different parts of the world. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Bootstrap scores above 50% (1000 replicates) are placed at the tree nodes. For the detailed of isolates, refer to Table 1

infecting ornamental ginger in Hawaii (HAW). The VPg gene was found to be 570 nt in size, coding for 190 amino acids (aa) and having a deduced MW of 21.54 kDa, and pI of 9.03. The pairwise sequence identity of the VPg gene among all the 29 BBrMV isolates ranged from 75 to 100% and from 95 to 100% at the nt and aa level, respectively. Interestingly, 23 out of 29 isolates collected from five different Indian states showed 100% identity. Comparative sequence analysis of the Indian isolate with the PHI and HAW isolates revealed 75–98%, 75–99% and 96–98%, 95–98% identity at nt and aa level, respectively (Table 1). Sequence count matrix ranged from 7 to 139 nt and 5–7 aa among the BBrMV isolates used in this study.

Phylogenetic trees were constructed by the maximum likelihood method based on nt and aa sequence alignments revealed that BBrMV isolates TN1, TN2, PHI and HAW are grouped into one cluster (Group II) whereas the remaining isolates formed another cluster (Group I) where, as in the case of aa alignment, TN2 clustered with other Indian isolates (Fig. 1 a, b). Cross species conservation analysis of VPg protein indicated the presence of many conserved domains (Fig. 2a).

The VPg gene is divided into three regions: an N-terminal region (1–50 aa), the central domain (51–153 aa) and a C-terminal domain (154–190 aa) (Yambao et al. 2003). Our study has revealed that the VPg protein is highly conserved among the BBrMV isolates (Fig. 2b). The conserved amino acid $Y^{63}$ is important for uridylylation, thereby helping in viral replication (Murphy et al. 1996; Oruetxebarria et al. 2001; Anindya et al. 2005). The motifs in the central domain between aa 80–99 are involved in VPg self-association and VPg-HC-Pro interaction (Yambao et al. 2003). The functional motifs $GDD^{26}$ and $TRRGKVK^{44}$ act as NTP binding sites (Puustinen and Makinen 2004). The domain GKVKGS(S/T)$KTVG^{51}$ is involved in membrane-associated alpha helix stabilization and also as a nuclear localization signal (NLS)

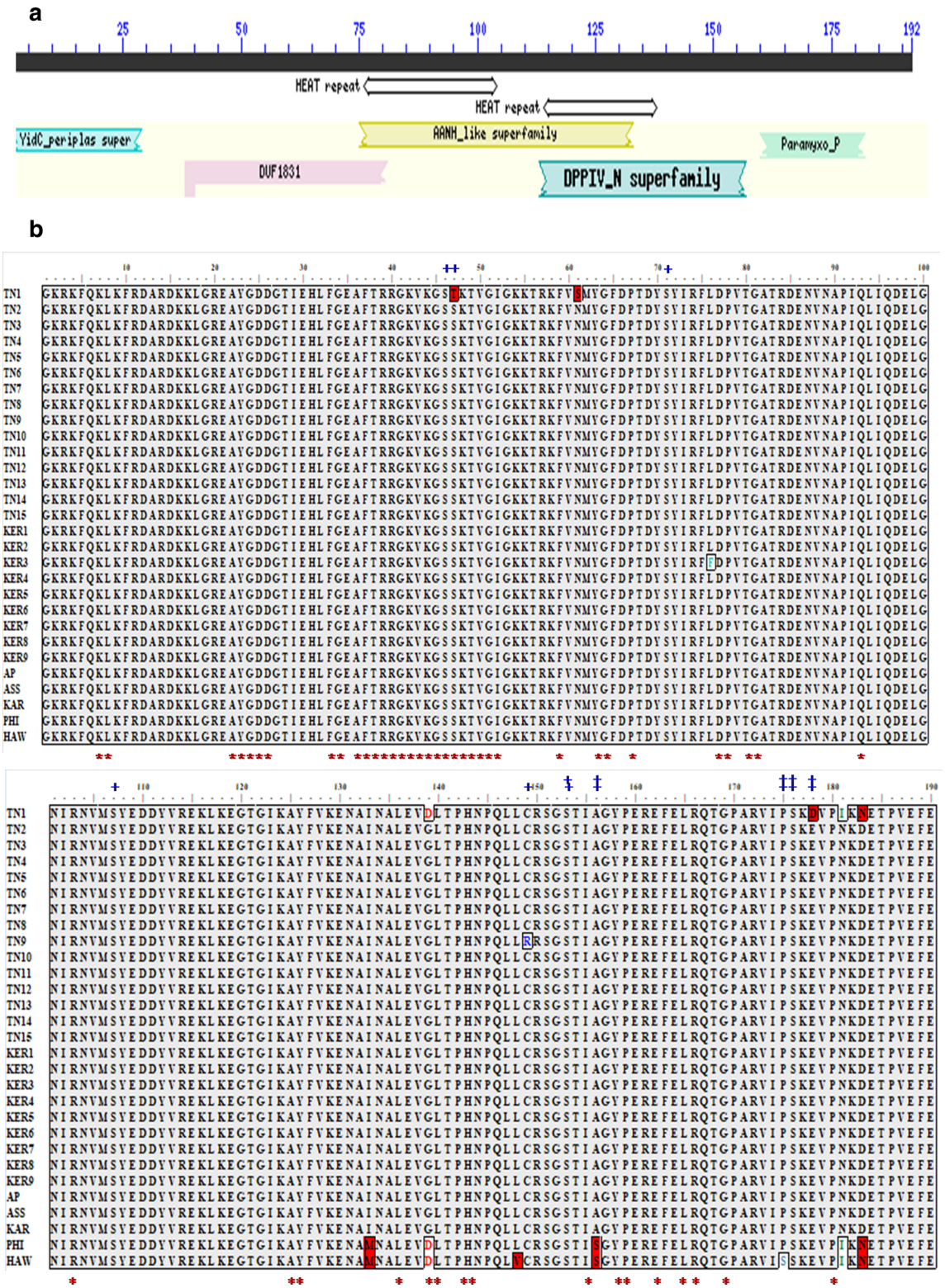


**Fig. 2 a Cross species conservation at the three functional regions of VPg.** The image is generated by ALL-IN-ONE-SEQ-ANALYZER version 1.35 depicting different functional domains of VPg. **b** Multiple amino acid sequence alignment of VPg from different BBrMV isolates. Conserved motifs are shown by red asterisks. Positive selections in VPg codon are shown by single blue crosses and in VPg self-interacting domain by double blue crosses

(Rajamaki and Valkonen 2003). All BBrMV isolates have a unique GKVKGS(S/T)KTVG motif which corresponds to aa positions 41–51 whereas in TN1 residue $S^{47}$ was replaced by T. The motif $Y^{63}$ and $DP^{78}$ are involved in the interaction of

VPg with eIF4E (Roudet-Tavert et al. 2007). The motif $AY^{126}$ may act as an alternative uridylylation site in place of $Y^{63}$ (Rantalainen et al. 2008). One unique observation disclosed that motif $DL^{140}$ was conserved in isolates TN1, PHI and HAW whereas it was replaced by $GL^{140}$ in all the other BBrMV isolates. Many of the conserved motifs studied here may have similar functions as in the case of all potyviruses. However, in the C-terminal region most of the motifs were highly conserved. A number of conserved and divergent functional motifs were identified in this study which may be essential for performing multiple functions.

In order to analyse the polymorphism at the sequence level, the VPg sequences of BBrMV were analysed for the presence of SNPs and INDELs (Table 1). All the sequences were compared with the TN1 sequence GenBank (accession No. HM131454) taken as reference. PHI isolate had an evolutionary distance of 0.006 with seven SNPs from the reference sequence. The next closest were TN2 followed by HAW with an evolutionary distance of 0.008 and 0.010 with nine and 11 SNPs, respectively. Isolate TN9 and KER3 had an evolutionary distance of 0.716 with 139 SNPs, whereas all the other Indian isolates showed an evolutionary distance of 0.710 with 138 SNPs. No INDELs were found in any of the BBrMV-VPg sequences used in this study.

The ratio of non-synonymous (Ka) to synonymous (Ks) nucleotide substitution rates (Ka/Ks) was calculated to understand the nt changes which affect the aa sequence of the VPg protein (Table 1). The value of Ks ranged from 0.0139–0.0487, but the Ka value was not applicable for any isolates except TN2, PHI and HAW.

Recombination has been reported as one of the main forces driving plant virus evolution (Roossinck 2003; Garcia-Arenal et al. 2003) and this may contribute to the evolution of VPg protein and help in the adaptation of members of the family *Potyviridae* to a wide host range. In silico recombination analysis of the available VPg gene sequences using six recombination detection methods revealed a single tentative recombination event in isolate TN2 (major parent was unknown and TN1 appeared as minor parent), suggesting a possible derivation of TN2 from TN1 (Table 2).

Six different codon-based maximum likelihood algorithms SLAC, FEL, IFEL, REL, FUBAR and MEME within the Datamonkey server were used to determine the gene- and site-specific selection pressure acting on BBrMV-VPg (Fig. 2b). Most of the codons were under negative or neutral selection, while codon positions 46, 47, 71, 107, 153, 156, 175, 176 were found under positive selection detected by REL, and codon positions 149, 156, 175, 178 were under positive selection detected by MEME. The codons under negative selection were 72, 73, 13, 103 detected by the methods FEL, IFEL, REL and FUBAR, respectively.

Three independent statistical tests of population differentiation were applied to estimate whether geographical isolates are genetically differentiated populations (Table 3). The statistical tests of genetic differentiation, measurement of gene flow, and population demography of banana groups were investigated. Genetic differentiation between populations was examined by three permutation-based statistical tests, Ks*, Z, and Snn. Analyses of these values showed that the genetic differentiation within isolates from Tamil Nadu vs. Kerala, Tamil Nadu vs. other Indian isolates, Kerala vs. other Indian isolates of banana was non-significant, confirming the low genetic differentiation among this group of isolates. The failure of the tests to differentiate the geographical groups may also be attributed to lack of statistical power as the subpopulation size is small. The FST value between the BBrMV isolates of banana from different geographic locations was smaller than 0.33, indicating a relatively frequent gene flow between them. The results are in agreement with the phylogenetic relationships of each population.

The patterns of molecular diversity were evaluated using Tajima's D, Fu and Li's D* and F* statistical tests at segregating sites and haplotype diversity and nucleotide diversity at all sites (Table 4). All the three statistical tests were negative for populations from Kerala indicating the low frequency polymorphism of BBrMV. The remaining three groups of isolates were negative in Tajima D test and positive in Fu and Li's D and Fu and Li's F test. However, the *P*-values were only significant for Tajima's D test, indicating that the results are convincing. When haplotype and nucleotide diversities of all

**Table 2** Tentative recombination events detected in the VPg gene of BBrMV isolates using RDP

| Recombinant isolate | Recombination break points | Parental isolate [a] | | Recombination detection [b] | |
|---|---|---|---|---|---|
| | | Major | Minor | Methods | P value |
| TN2 | 429–530 | Unknown 1–429, 531–570 (Regions derived from major parent) | TN1 430–530 (Regions derived from minor parent) | M, **S** | $1.907 \times 10^{-05}$ |

The methods whose *P* values are shown are indicated in bold

[a] Minor and Major parents refer to parental isolates contributing the smaller and large fractions of the recombinant's sequence, respectively

[b] The methods used to infer recombination break points were M MAXCHI, S SISCAN

**Table 3** Genetic differentiation measurements for host and geography of BBrMV population

| Parameters | Tamil Nadu ($n$ = 15) vs. Kerala ($n$ = 9) | Tamil Nadu (n = 15) vs. Other Indian ($n$ = 12) | Kerala (n = 9) vs. Other Indian ($n$ = 18) |
|---|---|---|---|
| **Ks*** | 2,103,571 (0.4660) ns | 18.69841 (0.4770) ns | 18.94118 (0.4530) ns |
| **Z** | 138.39892 (0.8100) ns | 175.34348 (0.5180) ns | 176.09292 (0.8930) ns |
| **Snn** | 0.53728 (0.3780) ns | 0.51852 (0.3980) ns | 0.55892 (0.3630) ns |
| **FST** | 0.06836 | 0.06847 | 0.05623 |
| **Nm** | 3.41 | 3.40 | 4.20 |

Ks*, Z, and Snn represent the most powerful sequence-based statistical tests for genetic differentiation and are recommended for use in cases of high mutation rate and small sample size. The Z statistic value results from ranking distances between all pairs of sequences. Snn is the frequency with which the nearest neighbors of sequences are found in the same locality; FST is the coefficient of gene differentiation or fixation index, which measures inter-population diversity; Nm can be interpreted as the effective number of migrants exchanged between demes per generation; *, $0.01 < P < 0.05$

BBrMV populations were compared (Table 4), haplotype diversity was 0.222 to 0.377, indicating a low level of diversity for the locus. In most cases, the haplotype diversity values were high, and nucleotide diversity values were low. Among the isolates, the Tamil Nadu group showed greater diversity than the other genogroups, and Kerala isolates showed less diversity which is consistent with recent population expansion events.

## Discussion

Knowledge on the genetic diversity of the VPg gene of BBrMV is crucial for understanding the viral infection cycle and host-virus interaction. In this study we have collected 27 BBrMV isolates from India and analysed their genetic variability using the VPg gene sequences. The pairwise sequence identity, sequence difference count matrix, SNPs, evolutionary distance and INDELs of VPg gene showed variability at nt level among BBrMV populations, however, the same degree of difference was not observed in the aa sequences of the VPg protein suggesting that the number of substitutions have taken place at the wobble base without altering the aa sequence exploiting the degeneracy. VPg plays multifunctional and essential roles in the virus life cycle in association with various host factors. Therefore, different selective constraints corresponding to various functions seem to act on viral proteins, thus the genetic diversity of viral populations may be limited by such negative selection (Seo et al. 2009). Further, variations at the nucleotide level may also be due to mixed cultivation of different banana cultivars in Southern India (Balasubramanian et al. 2014).

Phylogenetic trees showed two groups but no relationship with the geographical origin and the host from which they were isolated. This indicates that BBrMV is able to move across the banana-growing area fairly freely due to the movement of infected suckers in different states in South India, where no domestic quarantine is enforced to restrict the movement of banana suckers between the states, and to aphid transmission of the same strain to other cultivars. Seo et al. (2009), Tomimura et al. (2004) and Balasubramanian and Selvarajan (2014) have reported that phylogenetic trees did not support geographic and host clustering of SMV, TuMV and BBrMV.

The ratio of non-synonymous (Ka) to synonymous (Ks) nucleotide substitution rates (Ka/Ks) was calculated to understand the nt changes which affect the aa sequence of VPg. TN2 is the only isolate with Ka/Ks ratio greater than 1 indicating positive selection, whereas all the other BBrMV isolates analysed in this study showed a Ka/Ks ratio lower than 1, suggesting negative selection. The Ka/Ks along with the evolutionary distance suggests that though these sequences had deleterious mutations, they are eliminated by purifying selection because this gene is involved in some of the virus essential

**Table 4** Neutrality tests, haplotype and nucleotide diversity of BBrMV population

| Population origin | Tajima's D | Fu and Li's D | Fu and Li's F | Haplotype diversity | Nucleotide diversity |
|---|---|---|---|---|---|
| **All** | −0.38338* | 1.53279 | 1.05930 | 0.377 | 0.05957 |
| **Indian** | −1.84584* | 1.34819 | 0.36028 | 0.279 | 0.03392 |
| **TN** | −0.97536* | 1.34623 | 0.80113 | 0.371 | 0.05881 |
| **KER** | −1.08823* | −1.1899* | −1.28293* | 0.222 | 0.00039 |

Tajima's D test compares the nucleotide diversity with the proportion of polymorphic sites which are expected to be equal under selective neutrality. Fu and Li's D* test is based on the differences between the numbers of singletons (mutations appearing only once among the sequences) and the total number of mutations. Fu and Li's F* test is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences. *$P < 0.10$; *$P < 0.05$

functions such as the interaction with the translation initiation factor 4E. Mutation in the VPg may abolish the interaction with eIF4E, thereby preventing viral infection in plants (Robaglia and Caranta 2006).

Most of the BBrMV VPg codons were under strong purifying selection, indicating the occurrence of a recent BBrMV population expansion, and enhancing the speed of elimination of deleterious mutations in this gene (Kondrashov 1988; Balasubramanian et al. 2014). Mutations at codon positions 156 and 175 may affect the self interaction of VPg protein necessary in homodimer formation (Rantalainen et al. 2008) However, the significance of positive selection in these codons in the host-virus interactions needs to be investigated further. Amino acids within the central and C-terminal parts of PVY and PVA VPg were found to have undergone a positive selection, suggesting that the role of the VPg-encoding region in virulence is determined by the protein and not by the viral RNA (Moury et al. 2004).

A single tentative recombination event was detected in isolate TN2, confined to the central domain and the C-terminal region of the VPg gene which harbours part of the VPg self-interaction domain. Similar recombination events have been detected in the VPg of TuMV and SMV (Tomimura et al. 2004; Seo et al. 2009; Nguyen et al. 2013; Li et al. 2010).

Genetic differentiation between populations was examined by three permutation-based statistical tests, Ks*, Z and Snn (Hudson 2000). The results indicated that the isolates from different regions are closely related and can be considered as the same population from a genetic point of view. The Fst values between isolates from Tamil Nadu vs Kerala, Tamil Nadu vs other Indian isolates and Kerala vs other Indian isolates were smaller than 0.33, indicating the gene flow between BBrMV populations was relatively frequent. This may be due to the frequent movement of suckers between the states, lack of quarantine measures and rapid non-persistent aphid transmission of this virus within these regions. Similar observations underlie the spread of TuMV in China and Japan (Nguyen et al. 2013). Population demographical analyses showed that all three statistical values were negative for the Kerala group suggesting strong negative selection and recent expansion of this population. Tamil Nadu isolates showed greater diversity than the other genogroups, indicating that these populations are relatively older.

In this study we observed 100% identity between 23 of 29 isolates of BBrMV-VPg gene though they were collected from five different states of India. It is likely that strong purifying or negative selection enhances the speed of elimination of deleterious mutations as VPg plays an essential role in the virus life cycle in association with various host factors. The significance of positive selection at codon position 156 and 175 needs further investigation. This study has revealed genetic changes in VPg gene of BBrMV which will help in understanding the genetic structure of the virus population at different geographic scales: country, region and locality. Further, it also provides information on various sequence motifs and domains of VPg of BBrMV to relate its biological functions in understanding its role in plant-virus interaction.

**Acknowledgments** The authors are highly thankful to the Director of ICAR-NRCB for constant support, encouragement and providing necessary facilities to carry out the work. The financial support received from the Indian Council of Agricultural Research, New Delhi is greatly acknowledged.

## References

Anindya R, Chittori S, Savithri HS (2005) Tyrosine 66 of *Pepper vein banding virus* genome-linked protein is uridylylated by RNA-dependent RNA polymerase. Virology 336:154–162

Balasubramanian V, Selvarajan R (2014) Genetic diversity and recombination analysis in the coat protein gene of *Banana bract mosaic virus*. Virus Genes 48:509–517

Balasubramanian V, Sukanya RS, Anuradha C, Selvarajan R (2014) Population structure of *Banana bract mosaic virus* reveals recombination and negative selection in the helper component protease (HC-Pro) gene. Virus Disease 25:460–466

Borgstrom B, Johansen IE (2001) Mutations in *Pea seedborne mosaic virus* genome-linked protein VPg alter pathotype-specific virulence in *Pisum sativum*. Molecular Plant-Microbe Interaction 14:707–714

Garcia-Arenal F, Fraile A, Malpica JM (2003) Variation and evolution of plant virus populations. Int Microbiol 6:225–232

Grzela R, Szolajska E, Ebel C, Madern D, Favier A, Wojtal I, Zagorski W, Chroboczek J (2008) Virulence factor of *Potato virus Y*, genome-attached terminal protein VPg, is a highly disordered protein. J Biol Chem 283:213–221

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics 155:2011–2014

Hull R., 2002. Plant Virology (4th ed.), Academic Press, San Diego, CA, USA

Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–440

Leonard S, Viel C, Beauchemin C, Daigneault N, Fortin MG, Laliberte JF (2004) Interaction of VPg-pro of *Turnip mosaic virus* with the translation initiation factor 4E and the poly(a)-binding protein in planta. J Gen Virol 85:1055–1063

Li K, Yang QH, Zhi HJ, Gai JY (2010) Identification and distribution of *Soybean mosaic virus* strains in southern China. Plant Dis 94:351–357

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Magnaye LV, Espino RRC (1990) Note: banana bract mosaic, a new disease of banana I. Symptomatology. The Philippine Agriculture 73:55–59

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: a flexible and fast computer program for analyzing recombination. Virus Evolution 1. https://doi.org/10.1093/ve/vev003

Moury B, Morel C, Johansen E, Guilbaud L, Souche S, Ayme V, Caranta C, Palloix A, Jacquemond M (2004) Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistance in *Capsicum annuum* and *Lycopersicon hirsutum*. Molecular Plant-Microbe Interaction 17:322–329

Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A (2013) A genome-wide pairwise identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). Arch Virol 158:1411–1424

Murphy JF, Klein PG, Hunt AG, Shaw JG (1996) Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. Virology 220:535–538

Nguyen HD, Tran HT, Ohshima K (2013) Genetic variation of the *Turnip mosaic virus* population of Vietnam: a case study of founder, regional and local influences. Virus Res 171:138–149

Oruetxebarria I, Guo D, Merits A, Mäkinen K, Saarma M, Valkonen JP (2001) Identification of the genome-linked protein in virions of *Potato virus A*, with comparison to other members in genus Potyvirus. Virus Res 73:103–112

Puustinen P, Makinen K (2004) Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. J Biol Chem 279:38103–38110

Rajamaki ML, Valkonen JPT (2003) Localization of a potyvirus and the viral genome-linked protein in wild potato leaves at an early stage of systemic infection. *Molecular Plant-Microbe*. Interaction 12:25–34

Rantalainen KI, Uversky VN, Permi P, Kalkkinen N, Dunker AK, Makinen K (2008) Potato virus a genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. Virology 377:280–288

Rantalainen KI, Eskelin K, Tompa P, Makinen K (2011) Structural flexibility allows the functional diversity of potyvirus genome linked protein VPg. J Virol 85:2449–2457

Robaglia C, Caranta C (2006) Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci 11:40–45

Roossinck MJ (2003) Plant RNA virus evolution. Curr Opin Microbiol 6: 406–409

Roudet-Tavert G, Michon T, Walter J, Delaunay T, Redondo E, Le Gall O (2007) Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPro. J Gen Virol 88:1029–1033

Schaad MC, Anderberg RJ, Carrington JC (2000) Strain-specific interactions of the *Tobacco etch virus* NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. Virology 273:300–306

Seo JK, Ohshima K, Lee HG, Son M, Choi HS, Lee SH, Sohn SH, Kim KH (2009) Molecular variability and genetic structure of the population of *Soybean mosaic virus* based on the analysis of complete genome sequences. Virology 393:91–103

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tian YP, Liu JL, Zhang CL, Liu YY, Wang B, Li XD, Guo ZK, Valkonen JPT (2011) Genetic diversity of *Potato virus Y* infecting tobacco crops in China. Phytopathology 101:377–387

Tomimura K, Spak J, Katis N, Jenner CE, Walsh JA, Gibbs AJ, Ohshima K (2004) Comparisons of the genetic structure of populations of *Turnip mosaic virus* in west and east Eurasia. Virology 330:408–423

Wittmann S, Chatel H, Fortin MG, Laliberte JF (1997) Interaction of the viral protein genome linked of turnip mosaic potyvirus with the translational eukaryotic initiation factor (iso) 4E of *Arabidopsis thaliana* using the yeast two-hybrid system. Virology 234:84–92

Yambao ML, Masuta C, Nakahara K, Uyeda I (2003) The central and C-terminal domains of VPg of clover yellow vein virus are important for VPg–HCPro and VPg–VPg interactions. J Gen Virol 84:2861–2869

Zhang CL, Gao R, Wang J, Zhang GM, Li XD, Liu HT (2011) Molecular variability of *Tobacco vein banding mosaic virus* populations. Virus Res 158:188–198