Analysis of synonymous codon usage bias and phylogeny of coat protein gene in banana bract mosaic virus isolates

Atul B. Patil¹ · Vijayendra S. Dalvi¹ · Akhilesh A. Mishra¹ · Bal Krishna² · Abdul Azeez²

Received: 10 January 2017 / Accepted: 20 April 2017 / Published online: 18 May 2017
© Indian Virological Society 2017

Abstract The banana is one of the world's most important livelihood crops. Banana plants are principally infected by four virus species, *Banana bunchy top virus* (genus *Babuvirus*), *Cucumber mosaic virus* (genus *Cucumovirus*), *Banana streak virus* (genus *Badnavirus*) and *Banana bract mosaic virus* (genus *Potyvirus*). The objective of this study is to understand the codon usage pattern and phylogeny of coat protein gene in different banana bract mosaic virus (BBrMV) isolates. The BBrMV Coat Protein (CP) gene was amplified from BBrMV infected banana plant samples collected from different districts of Tamil Nadu and Karnataka, India. Six new BBrMV isolates were submitted to National Center for Biotechnology Information. Phylogenetic analysis and codon usage indices were studied along with other isolates of BBrMV. Phylogenetic analysis of CP genes shows that

most of BBrMV isolates are closely related to each other except KF385484.1 and KF385478.1. Relative codon usage patterns among different BBrMV isolates were calculated by software CodonW version 1.4.2. In BBrMV, codons with A-ended or U ended are the most preferential except the Leu and Gln whose optimized codons are CAG and UUG ending by G. The codon usage patterns of BBrMV isolates are principally influenced by mutational bias; however, compositional constraints along with mutational bias also play a major role.

Keywords Codon usage · Mutational bias · Translational selection · Correspondence analysis

Introduction

Banana bract mosaic disease was initially known as Kokkan disease in Kerala, India since 1966 [27] and the similar disease was also noted in Philippines in 1979 [17]. Subsequently, banana bract mosaic virus (BBrMV) a member of the genus *Potyvirus* of the family Potyviridae was confirmed as the causal agent of banana bract mosaic disease [24]. The name BBrMV was documented for the first time at the INIBAP/FAO meeting on Bunchy Top held during 1988 in Los Banos, the disease was named as Banana bract mosaic on the basis of the outstanding symptoms on the bract of the inflorescence. Till 1999, BBrMV has been reported from only five countries viz. India, Philippines, Western Samoa, Vietnam and Thailand [24]. This virus is reported from Kerala, Tamil Nadu, Karnataka and Andhra Pradesh states of India [24, 28, 33]. BBrMV is a flexuous filamentous particle having size of 660–760 × 12 nm [4, 23]. Bateson and Dale in 1995 isolated a 38 kDa protein thought to be the BBrMV coat protein (CP) that reacted

Electronic supplementary material The online version of this article (doi:10.1007/s13337-017-0380-x) contains supplementary material, which is available to authorized users.

✉ Akhilesh A. Mishra
virology@jains.com

Atul B. Patil
atulpatil9096@gmail.com

Vijayendra S. Dalvi
virology1@jains.com

Bal Krishna
balkrishna.yadav@jains.com

Abdul Azeez
dr.azeez.abdul@jains.com

¹ Plant Virology Lab, Jain R&D, Jain Hills, Jain Irrigation Systems Limited, Jalgaon 425001, India

² Plant Molecular Biology Lab, Jain R&D, Jain Hills, Jain Irrigation Systems Limited, Jalgaon 425001, India

with broad spectrum potyvirus antiserum in western blots. BBrMV was confirmed as a potyvirus, based on high sequence identity of the open reading frame (ORF) of up to 150 amino acids and the 3' UTR of up to 190 nucleotides with analogous regions of other potyviruses [4]. The complete genome of the BBrMV-TRY isolate (NCBI acc. no. HM131454.1) from India reported to be 9711 nucleotides (nt) long excluding the 3' poly (A) tail. The 5' untranslated region (UTR) has a GC content of 42.2% and AT content of 57.8%. The 5' UTR is assumed to be 128 nt and the 3' UTR (without the poly (A) tail) is 208 nt. Like other potyviruses, the polyprotein of BBrMV-TRY is predicted to cleave into ten putative mature proteins. In descending order (5'–3'), these proteins are the first protein being P1 followed by helper component protein (HC-Pro), P3, 6K1, cylindrical inclusion protein (CI), 6K2, viral protein genome-linked (VPg), NIa-Pro (major protease of small nuclear inclusion protein -NIa), NIb (large nuclear inclusion protein) and coat protein (CP). The amino acid sequence positions of each predicted protein in BBrMV-TRY genome are as follows: P1, 1–329; HC-Pro, 330–786; P3, 787–1133; 6K1, 1134–1185; CI, 1186–1819; 6K2, 1820–1872; VPg, 1873–2062; NIa-Pro, 2063–2305; NIb, 2306–2825 and CP, 2826–3125 [2].

Balasubramanian and Selvarajan [3] studied the genetic diversity and showed a recombination analysis of the coat protein (CP) gene of BBrMV using codon-based maximum likelihood methods. Their study revealed that the most of the codons in CP gene were under negative or neutral selection except for codons 28, 43, and 92, which were under positive selection pressure [3]. The coat protein is a more characterized potyviral protein, and has three domains (N, Core and C). The N domain is highly variable and comprises the major virus-specific epitopes; the core and C domains are more conserved. The difference in the core region is similar to that of the whole genome so it can be a reliable index to study genetic relatedness [32].

It is well known that the genetic code chooses 64 codons to synthesize 20 standard amino acids and three stop signals. The alternative codons for the same amino acid are termed as synonymous codons. Synonymous codons are generated by mutations at the third base location, without altering the primary sequence of the polypeptide product. Codon usage bias (CUB) refers to differences in the frequency of occurrence of synonymous codons in coding DNA. Codon usage biases are affected by nucleotide composition [20], translation processes [29], tRNA abundance [13], gene function [10], gene length [18], protein structure [19] and hydrophobicity [25]. Adams and Antoniw [1] suggested that mutational bias rather than translational selection was the major determinant of codon usage variation amongst plant viruses. They also suggest that structural considerations related to the secondary structure

of the RNA and the informational constraints in plant viral genomes having completely overlapping gene in different reading frames are clearly involved in codon usage bias among plant viruses and could be potential areas for further studies [1]. To the best of our knowledge, there is no information available on codon usage pattern of BBrMV coat protein gene including the relative synonymous codon usage (RSCU) and codon usage bias (CUB). In this study, the genetic determinants of codon usage index of coat protein gene in BBrMV are studied.

Materials and methods

Plant material and virus isolates

Based on the information about banana virus diseases from our agronomist at Jain Irrigation Systems Ltd most BBrMV infected areas were selected for collection of samples. On the basis of characteristic symptomatology, banana leaf samples (14 numbers) were collected from various parts of Tamil Nadu (TN) and Karnataka (KA) states of India. District Sentrupur, Thutukudi, Kanyakumari, Tirunelveli, Coimbatore, Viridhunagar, Theni and Pudukottai were selected from Tamil Nadu and from Karnataka only Chamarajnagar district was selected. The details of sample collection is shown in Supplementary Table 1.

Primer designing

Primers used in the study were designed using OligoPerfect™ Designer with BBrMV TRY CP as a template to amplify CP gene region of BBrMV. The forward primer BBMV CP-F (5' TCTGGAACGGAGTCAACCAATGATG 3') and reverse primer BBMV CP-R (5' AGGTGTGCACGCCAGCTAGATACA 3') were used to amplify the band of 1107 bp.

Total RNA isolation, cDNA synthesis and cloning of CP genes

Total RNAs from 100 mg of infected leaf tissues were isolated using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions. First strand complementary DNA (cDNA) was synthesized using a RevertAid™ First Strand cDNA Synthesis Kit (Fermentas INC, USA) with oligo (dT) primer according to the manufacturer's instructions. The reaction mixtures (20 µl) were incubated at 65 °C for 5 min and 42 °C for 60 min and subsequently incubated at 70 °C for 10 min to inactivate the RT enzyme.

PCR amplifications of CP fragments were done using *Pfu* DNA polymerase (Thermo Scientific) in reaction

volume of 50 µl containing 5 µl 10X *Pfu* buffer, 5 µl of cDNA, 1 µl of each primer (each 0.5 µM), 5 µl dNTPs (0.2 mM each) and 2.5U of enzyme. PCR amplification was performed in a Thermo cycler (T100, BIORAD, USA) and thermocycling conditions consisted of an initial denaturation step at 94 °C for 2 min, followed by 30 cycles involving 30 s of denaturation at 94 °C, 30 s of annealing at 55 °C, 1 min elongation at 72 °C and single cycle of a final elongation at 72 °C for 10 min.

The amplified PCR products were resolved by electrophoresis on 1.5% agarose gels, and desired DNA bands from gel were eluted and purified using a MinElute Gel Extraction Kit (QIAGEN Inc. Valencia, CA, USA). The purified DNA fragments were cloned into pTZ57R/T cloning vector (Thermo Scientific, USA) according to the manufacturer's instructions and transformed into the *Escherichia coli* strain DH5α using standard molecular biology methods [26]. Plasmid was isolated from an overnight culture using PureLink® Quick Plasmid Mini-prep Kit (Invitrogen, USA). Clones were sequenced from both directions using an ABI PRISM BigDye Terminator Kit at Amnion sequencing services, Bangalore, and the sequences were aligned using CLUSTALW for sequence analysis.

Sequence analysis

The partial coat protein gene sequences of 42 BBrMV isolates (including our six isolates) containing 237 codons each were obtained from National Center for Biotechnology Information (NCBI), genome database. The information regarding the serial number (SN), Accession numbers of gene sequences of BBrMV are listed in Supplementary Table 2.

Codon usage analysis

Codon usage indices and correspondence analysis calculations were done by software CodonW version 1.4.2 [21]. GC content is the frequency of guanine (G) and cytosine (C) in a coding gene. GC3 content is the frequency of guanine (G) and cytosine (C) of a codon at the third position. Whereas, A3, C3, U3 and G3 are the frequencies of adenine (A), cytosine (C), uracil (U) and guanine (G) at the third position of a codon in a coding gene. CBI is codon bias index (CBI), Fop is frequency of optimal codons, CAI is codon adaptation index, Nc is effective number of codon values, L_{sym} and L_{aa} indicate number of synonymous codons and number of translatable codons respectively. General average hydropathicity (GRAVY) and aromaticity (Aromo) is the frequency of aromatic amino acids in hypothetically translated gene product [22]. CAI is the measure of high or low expression of gene [30]. To

examine the correlation among different factors influencing codon usage the Pearson product-moment correlation coefficient (Pearson's *r*) is calculated.

Phylogenetic analysis and similarity matrix

The phylogenetic tree of coat protein based on the amino acid sequences of 42 BBRMV isolates was constructed using the Maximum Parsimony method, supported with a bootstrap test of 500 replicates in MEGA6 software [34]. Similarity matrix of coat protein amino acid sequences from 42 BBrMV isolates was drawn using open-source freeware software MatGAT [8].

Results

Isolation of coat protein genes

Fourteen samples were collected on the basis of symptoms from different location (as described in materials and methods). Infection of BBrMV was confirmed using coat protein specific primers yielding 1107 bp band, cDNA prepared by using RNA, isolated from leaves of healthy banana plant was used as negative control in PCR amplification. DNA band from gel was eluted, cloned in pTZ57R/T vector and sequenced. Sequencing result indicated that all bands are specific to coat protein of BBrMV. Identity within banana isolates of BBrMV ranged from 95–100% to 97–100% at nucleotide and amino acid level respectively. Out of fourteen collections six of them were having different sequences than the database thus they were considered as new isolate and NCBI Accession numbers are mentioned in Supplementary Table 2.

Phylogenetic analysis and similarity matrix

The phylogenetic tree (Fig. 1) analysis shows that most of the BBrMV isolates are closely related to each other except two isolates KF385484.1 and KF385478.1 which exhibited higher genetic distance from other isolates. All the identified six BBrMV isolates from present study are shown with black circle in front of accession number. In phylogenetic tree our isolates KM061426.1 and KM06424.1 from Tamil Nadu are closely related with the isolate HM348782.1 (Andhra Pradesh), isolate KM061427.1 from Tamil Nadu is closely related with isolates KF385471.1 (Tamil Nadu) and KF385470.1 (Tamil Nadu), isolates KM061425.1 and KM061428.1 from Tamil Nadu are placed separately; also isolate KM061429.1 from Karnataka is placed separately. There exists minimum 89% and maximum 100% amino acid sequence similarity between the BBrMV isolates in this study.

Fig. 1 Molecular Phylogenetic analysis of BBrMV isolates by Maximum Parsimony analysis method. The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length = 127 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The analysis involved 42 amino acid sequences. All positions containing gaps and missing data are eliminated. There were a total of 237 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. In fig abbreviations used are TN for Tamil Nadu, KN for Karnataka, KL for Kerala, AP for Andhra Pradesh, PHI for Philippines

Relative synonymous codon usage in BBrMV coat protein gene

To determine the variation in base composition of different BBrMV CP genes, the base composition in different

BBrMV CP genes was calculated using software CodonW and the values are shown in Supplementary Table 3. Comparative values of A3, U3, C3 and G3%, indicates clearly that U3% was highest, and C3% was the lowest of all, and Nc values of these BBrMV coat protein genes fluctuated from 51.18 to 58.4, with a mean value of 54.19 and S.D. of 1.66. The Nc value ranges from 20 (when only one synonymous codon is chosen by the corresponding amino acid) to 61 (when all synonymous codons are used equally) [35]. The RSCU is the observed frequency of a codon divided by the expected one. RSCU value equal to 1 means the codon is chosen equally and randomly, however greater than 1 or less than 1 RSCU value indicates the particular codon is used with more or less frequency than expected. The details of the overall RSCU values of 237 codons in 42 BBrMV isolates coat protein genes were analyzed are presented in Table 1. The most preferentially used codons in BBrMV are A or U ended codons except the Leu (L) and Glu (E) ending by G for which optimized codons are CAG and UUG respectively. The Pearson product-moment correlation coefficient (Pearson's r) values of different codon usage indices are calculated in Excel function, the result is given in Supplementary Table 4. The Pearson product-moment correlation coefficient (Pearson's r) is a degree of the linear correlation among two variables. The values more than 0.5 and less than -0.5 to recognize the strong positive and negative correlation respectively among the factors influencing BBrMV coat protein gene codon usage has been considered for this study. There is positive correlation between GC with C3, GRAVY and Aromo with A3, GC3 with G3, CBI and Fop with CAI, Fop with CBI. Similarly there is positive correlation between Fop and CBI, GC and Nc, GC and GC3, Aromo and Gravy. There is negative correlation between C3 and U3, GC and U3, GC3 and G3 with A3, L_{sym} and CAI. The results of negative correlation clarifies that frequency of adenine at the third position of a codon in a coding gene and frequency of uracil at the third position of a codon in a coding gene of BBrMV coat protein gene didn't have any influence on the BBrMV coat protein gene codon usage bias.

Correspondence analysis of RSCU

Correspondence analysis generates a series of orthogonal axes to identify the trends that explain the variation in data [6]. The correspondence analysis of RSCU was performed for different BBrMV CP gene and their intra genomic variation of 59 informative codons. In correspondence analysis all the genes were plotted in a 59 dimensional hyperspace, according to the usage of the 59 sense codons. A plot of correspondence analysis of RSCU patterns in BBrMV CP genes on the basis of codon position by first and second principle axes is shown in Fig. 2. Codons UCC,

Table 1 Overall relative synonymous codon usage

Amino acid	CODON	RSCU	Amino acid	CODON	RSCU	Amino acid	CODON	RSCU
Ser	UCU	1.309405	Phe	UUU	1.318286	Ala	GCU	1.632286
	UCC	0.030167		UUC	0.681714		GCC	0.51731
	UCA	2.566762					GCA	1.826143
	UCG	0.068429		UUA	1.63281		GCG	0.02419
	AGU	1.251167		UUG	2.197405			
	AGC	0.774071		CUU	1.254714		CAU	1.201786
Pro	CCU	2.139524	Ile	CUC	0.408143	Gln	CAC	0.798214
	CCC	0.507048		CUA	0.138643			
	CCA	0.866071		CUG	0.368048		CAA	0.949048
	CCG	0.486643					CAG	1.050952
Thr			Val	AUU	1.407786	Asn	AAU	1.275571
	ACU	1.457048		AUC	0.681762			
	ACC	0.285286		AUA	0.910333		AAC	0.724429
	ACA	1.768738						
Glu	ACG	0.488762	Gly	GUU	1.212881	Arg	GAU	1.58319
				GUC	1.159738		GAC	0.41681
	GAA	0.6715		GUA	1.095167			
	GAG	1.3285		GUG	0.532286		CGU	1.526071
Cys							CGC	0.27931
	UGU	1.214262		GGU	1.696952		CGA	0.907786
	UGC	0.785738		GGC	0.079714		CGG	0.853024
Tyr				GGA	1.312548		AGA	0.951167
	UAU	1.306452		GGG	0.91069		AGG	1.482595
	UAC	0.693548						
			Lys	AAA	1.299			
				AAG	0.701			

Relative Synonymous Codon Usage (RSCU) values were calculated by computing all the 42 BBrMV CP genes, each containing 237 codons. The preferentially used codons for each amino acid are described in bold

UCG coding for serine, codon CUA coding for leucine and codon GCG coding for alanine are placed distantly than all other codons. RSCU values for codons UCC, UCG, CUA, and GCG are 0.030167, 0.068429, 0.138643, 0.02419 respectively; these values are much less as compared with

RSCU values of other codons used by different BBrMV isolates. Correspondence analysis was also performed on the RSCU values based on the BBrMV coat protein genes from 42 BBrMV isolates and the positions of all 42 BBrMV isolates along the first two major axes (Fig. 3).

Fig. 2 Correspondence analysis of RSCU patterns in BBrMV CP genes based on position of codon by first and second principle axes

Fig. 3 Correspondence analysis of RSCU patterns in BBrMV CP genes based on position of each coat protein gene by first and second principle axes

The first two main dimensional coordinates, Axis 1 and Axis 2 show 58.36 and 21.31% of the total variance, respectively, while Axis 3 and Axis 4 show 12.24 and 8.08% of variance, respectively, indicating that the first axis is the major contributor to codon bias. It is clear from the Fig. 3 that four isolates AF071590.1, KF385475.1, KF385484.1 and KF385478.1 are placed distantly than other 38 BBrMV isolates.

Nc Plot to study factors influencing codon usage

The Nc plot was used to investigate the codon usage variations among different BBrMV isolates. In Fig. 4, solid line represents the curve if codon usage is only determined by the GC3 content. If GC3 content is the only determinant of codon usage variation among the genes then the Nc value would fall on the continuous curve between Nc value and GC3 content also if the Nc value would fall below the

Fig. 4 A Plot of Nc versus GC3 content. More Nc values are either on or below standard curve indicating an involvement of GC3 content of BBrMV coat protein genes and mutational bias as main factor influencing codon usage in BBrMV. Presence of some Nc values above the standard curve indicates compositional constrained in BBrMV coat protein gene

continuous curve between Nc value and GC3 content the codon usage pattern is constrained mainly by mutational bias and several other factors such as translational selection, gene length and gene function [35]. Nc value of NC plot study of BBrMV lie on and below standard curve indicating an involvement of GC3 content of coat protein genes and mutational bias as main factor influencing codon usage in BBrMV.

Discussion

Earlier studies on plant viruses reveal that plant viruses has not been studied comprehensively by means of codon usage bias. Adams and Antoniw [1] studied the codon usage biases in plant viruses and they found large differences in the codon preferences of plant viruses, these differences are primarily related to the nucleotide compositional restraints of the genome. They concluded that if the mutational bias is the major factor responsible for the codon usage variation it would be expected that genes of the same virus have similar coding strategies, provided open reading frames should not be overlapped. They also found that some synonymous codons are consistently used with low frequency by both plants and viruses but they could not find any evidence to support the statement that viruses are generally adopted to use (or avoid) those codons most frequently used by their host plants. Their study suggests that mutational bias is the major determinant of codon usage variation amongst the plant viruses [1]. Cardinale et al. 2013 concluded that complex secondary structures, genomic variation and host shifting can be the possible reasons of codon usage bias at specific sites in ssRNA viruses. They also suggest that base composition and translational selection cannot reveal the CUB in begomoviruses [9]. Belalov and Lukashev 2013 address the questions that the origin of nucleotide bias remains unknown, also they suggest that boot scanning

consisting alignment of multiple genomes and other corresponding criteria's are needed to analyze local CUB. They also propose an algorithm provides the information regarding synonymous variation in sixfold degenerate codons and RNA secondary structure can reveal the role of translational pressure in codon usage bias in RNA viruses. In the study they analyze the relative impact of the nucleotide/dinucleotide composition on codon usage bias in RNA viruses. They found that total GC content was deficient to explain viral genome composition and causes of codon usage bias [5]. In the present study, we analyzed the diversity and codon usage pattern of the BBrMV isolates on the basis of coat protein gene.

Codon usage difference could be the result of natural selection and or mutation pressure for correct and efficient translation in different organisms [29]. The coat protein gene is of great importance for the identification and classification of potyviruses due to its structural properties, also the coat protein is the single key gene product in the virion and accounts for 95% of the potyvirus particle [31]. Coat protein gene was amplified and sequenced successfully from the samples collected from different places of Tamil Nadu and Karnataka. Six new BBrMV isolates were identified on the basis of coat protein gene from 14 BBrMV infected banana samples from different banana growing areas of Tamilnadu and Karnataka. Nucleotide content study of coat protein gene of BBrMV clarifies that U3 content is higher in BBrMV coat protein gene and C3 is lowest of all this explains well that most of the preferentially used codons are ended by U and codon with cytosine at third position are lowest of all. Pearson product moment correlation coefficient (Pearson's r) was studied to get the knowledge about positive and negative correlation among the factors influencing BBrMV coat protein gene codon usage. The positive correlations are substantial evidences found in this study which shows that GC3 content and compositional constraints are the main factors influencing BBrMV codon usage along with mutational bias. These same results of factors influencing BBrMV codon usage are obtained in Nc plot study where Nc values are either on or below standard curve indicating an involvement of GC3 content of BBrMV coat protein genes and mutational bias as main factor influencing codon usage in BBrMV. It is generally observed that RNA viruses have higher mutational rate than DNA viruses [11] so this higher mutational rate can be consider as one of the reason that mutational bias is one of the determinant of codon usage bias in BBrMV. The conclusions from Nc plot were made in accordance with the interpretations made by Wright in 1990. Some more Nc values are present above the standard curve between Nc value and GC3 content; Presence of these Nc values above the standard curve indicates that

there exist compositional constraints in BBrMV coat protein gene. Compositional biases are confined variations in amino acid or nucleotide frequencies in biological sequences. Compositional constraints affectes largely the structure and stability of the genome, stability of the transcripts and proteins as well as codon usage [7]. The recent studies on codon usage pattern on porcine circovirus and Equine influenza viruses also report that their codon usage pattern was influenced by compositional constraint [15, 16]. Correspondence analysis study of RSCU of BBrMV revealed that four codons UCC, CUA, GCG and UCG are used rarely by all the BBrMV isolates also the four BBrMV isolates AF071590.1 from Philippines, KF385475.1 from Tamil Nadu, KF385484.1 from Andhra Pradesh and KF385478.1 from Assam have more different trend in codon usage than other BBrMV isolates. From these results of correspondence analyses, it could be said that different geographical locations of the isolates could be one of the reasons for mutational bias or compositional constraints that exists among them. The evolutionary history of BBrMV isolates inferred by using the Maximum Parsimony analysis method specifies that our isolates KM061426.1 and KM06424.1 from Tamil Nadu are closely linked with the isolate HM348782.1 (Andhra Pradesh), isolate KM061427.1 from Tamil Nadu is closely associated with isolates KF385471.1 (Tamil Nadu) and KF385470.1 (Tamil Nadu), from these results it can be said that phylogenetically related isolates possess the different codon usage patterns. Similarity matrix calculated using matgat software showed the minimum 89% and maximum 100% amino acid sequence similarity among the BBrMV isolates considered in this study.

In conclusion the effective number of codon value (ENC value) calculated for BBrMV indicated that a fairly less bias for codon usage exists in BBrMV. Some of the previous results also showed that there is less bias in overall extent of codon usage in RNA viruses and there is less variation in bias between genes [12, 14, 36]. Codon usage analyses of BBrMV can be further supportive to develop novel diagnostics methods or to narrate the evolutionary origin of BBrMV with other potyviruses. It can also complement to study its impacts on transcription and translation of BBrMV which can give information regarding the high or low expression of genes that may be selected positively in the process of evolution and may possibly give rise to new virus strain. Basic information on codon usage pattern and factors affecting the synonymous codon usage of BBrMV revealed in present study will be useful to elucidate the process governing the evolution of BBrMV; still more extensive study is required for in-depth characteristics of synonymous codon usage in BBrMV genome.

Acknowledgement The authors gratefully acknowledge financial support from Jain Irrigation Systems Ltd.

References

- Adams MJ, Antoniw JF. Codon usage bias amongst plant viruses. *Arch Virol*. 2003;149(1):113–35.
- Balasubramanian V, Selvarajan R. Complete genome sequence of a banana bract mosaic virus isolate infecting the French plantain cv. Nendran in India. *Arch Virol*. 2012;157(2):397–400.
- Balasubramanian V, Selvarajan R. Genetic diversity and recombination analysis in the coat protein gene of Banana bract mosaic virus. *Virus Genes*. 2014;48(3):509–17.
- Bateson MF, Dale JL. Banana bract mosaic virus: characterisation using potyvirus specific degenerate PCR primers. *Arch Virol*. 1995;140(3):515–27.
- Belalov IS, Lukashev AN. Causes and implications of codon usage bias in RNA viruses. *PLoS ONE*. 2013;8(2):e56642.
- Benzécri JP. Correspondence analysis handbook. New York: Marcel Dekker; 1992.
- Bernardi G, Bernardi G. Compositional constraints and genome evolution. *J Mol Evol*. 1986;24(1–2):1–11.
- Campanella JJ, Bitincka L, Smalley J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinform*. 2003;4(1):1.
- Cardinale DJ, DeRosa K, Duffy S. Base composition and translational selection are insufficient to explain codon usage bias in plant viruses. *Viruses*. 2013;5(1):162–81.
- Chiapello H, Lisacek F, Caboche M, Henaut A. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene*. 1998;209(1):GC1–38.
- Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Natl Acad Sci*. 1999;96(24):13910–3.
- Gu WJ, Zhou T, Ma JM, Sun X, Lu ZH. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res*. 2004;101:155–61.
- Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*. 1981;151(3):389–409.
- Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*. 2003;92:1–7.
- Kumar N, Bera BC, Greenbaum BD, Bhatia S, Sood R, Selvaraj P, Virmani N. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PloS one*. 2016;11(4):e0154376.
- Liu XS, Zhang YG, Fang YZ, Wang YL. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virol J*. 2012;9(1):68.
- Magnaye LV, Espino RR. Note: banana bract mosaic, a new disease of banana, I Symptomatology. *Philipp Agric*. 1990;73(1):55–9.
- Moriyama EN, Powell JR. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucl Acids Res*. 1998;26(13):3188–93.
- Orešič M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. *J Mol Biol*. 1998;281(1):31–48.
- Osawa S, Ohama T, Yamao F, Muto A, Jukes TH, Ozeki H, Umeson K. Directional mutation pressure and transfer RNA in choice of the third nucleotide of synonymous two-codon sets. *Proc Natl Acad Sci*. 1988;85(4):1124–8.
- Peden J. CodonW version 1.4. 2.
- Peden JF. Analysis of codon usage: University of Nottingham. UK: Nottingham; 1999.
- Rodoni BC, Ahlawat YS, Varma A, Dale JL, Harding RM. Identification and characterization of banana bract mosaic virus in India. *Plant Dis*. 1997;6:669–72.
- Rodoni BC, Dale JL, Harding RM. Characterization and expression of the coat protein-coding region of banana bract mosaic potyvirus, development of diagnostic assays and detection of the virus in banana plants from five countries in southeast Asia. *Arch Virol*. 1999;144(9):1725–37.
- Romero H, Zavala A, Musto H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucl Acids Res*. 2000;28(10):2084–90.
- Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd ed. New York: Cold Spring Harbor Laboratory; 1989.
- Samraj J, Menon MR, Christudas SP, Satyarajan PK. Kokkan, a new disease of banana (*Musa paradisiaca* Linn.). *Agric Res J Kerala*. 1966;4(1):116.
- Selvarajan R, Balasubramanian V, Rao GP, Myrta A, Ling KS. Banana viruses. *Charact Diagn Manag Plant Viruses Hortic Crops*. 2008;2:109–24.
- Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res*. 1986;14(19):7737–49.
- Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res*. 1986;14(13):5125–43.
- Shukla DD, Ward CW, Brunt AA. The Potyviridae. Wallingford: Cab International; 1994.
- Shukla DD, Ward CW. Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Adv Virus Res*. 1989;36:273–314.
- Singh SJ, Selvarajan R, Singh HP. Detection of bract mosaic virus (kokkan disease) by electron microscopy and serology. In: *Banana-improvement, production and utilization*. Proceedings of Conference in AIPUB, Tiruchirappali, India 2000;381–383.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
- Wright F. The effective number of codons used in a gene. *Gene*. 1990;87(1):23–9.
- Zhou T, Gu W, Ma J, Sun X, Lu Z. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems*. 2005;81(1):77–86.