Virus Genes (2014) 48:509–517
DOI 10.1007/s11262-014-1056-x

Genetic diversity and recombination analysis in the coat protein gene of *Banana bract mosaic virus*

V. Balasubramanian · R. Selvarajan

Received: 12 January 2014 / Accepted: 6 March 2014 / Published online: 2 April 2014

Abstract Banana bract mosaic virus (BBrMV), a member of the genus *Potyvirus*, family *Potyviridae*, is the causal agent of the bract mosaic disease (BBrMD) that causes serious yield losses in banana and plantain in India and the Philippines. In this study, global genetic diversity and molecular evolution of BBrMV based on the capsid protein (CP) gene were investigated. Multiple alignments of CP gene of 49 BBrMV isolates showed nucleotide (nt) and amino acid (aa) identity of 79–100 and 80–100 %, respectively. Phylogenetic analysis revealed that except two Indians isolates (TN14 and TN16), all isolates clustered together. Eleven recombination events were detected using Recombination Detection Program. Codon-based maximum-likelihood methods revealed that most of the codons in the CP gene were under negative or neutral selection except for codons 28, 43, and 92 which were under positive selection. Gene flow between BBrMV populations of banana and cardamom was relatively frequent but not between two different populations of banana infecting isolates identified in this study. This is the first report on genetic diversity, and evolution of BBrMV isolates based on recombination and phylogenetic analysis in India.

Keywords Banana bract mosaic virus · Coat protein · Genetic diversity · Variants · Recombination

V. Balasubramanian · R. Selvarajan (✉)
Molecular Virology Lab, Division of Crop Protection, National Research Centre for Banana, Thogamalai Road, Thayanur Post, Tiruchirapalli 620102, Tamil Nadu, India
e-mail: selvarajanr@gmail.com

Introduction

Bananas and plantains (*Musa* spp.) which are staple foods due to their high starch and protein content and nutritional value. These are the main source of livelihood for millions of resource poor farmers in many developing countries in the world [1]. India is the largest producer of banana and plantains with the production of 29.78 million metric tons of fruits per year, contributing nearly one-third of global banana production [2]. Viral diseases of banana cause significant yield loss around the world [3]. Banana bract mosaic disease (BBrMD) was first reported in the Philippines in 1979 [4] and subsequently, reported from India, Sri Lanka, Western Samoa, Thailand, Vietnam, and Ecuador [5–7]. In India, bract mosaic disease was first recorded in 1966 in a plantain cv. Nendran as *Kokkan* disease in Kerala by Samraj et al. [8], but the etiology was not established. Later, the casual agent of *Kokkan* disease was authentically confirmed as banana bract mosaic virus (BBrMV) [9, 10]. BBrMD has now been recorded from several banana growing states viz. Kerala, Tamil Nadu, Karnataka, and Andhra Pradesh [10–12] and is identified as one of the diseases of national importance in India. BBrMV is a member of the genus *Potyvirus* of the family *Potyviridae* and have flexuous filamentous particles of approximately 725-nm long which encapsidate a monopartite ssRNA genome of 9711 bp in length. The virus contains a typical large open reading frame (ORF) coding for a poly protein of 3,125 amino acids. It contains nine protease cleavage sites, yielding ten matured functional proteins that have motifs conserved among homologous protein of other potyviruses [12, 13]. In addition, a small ORF termed *pipo* encoding a 7 kDa protein [14] also been found in BBrMV isolates. The main host of BBrMV is *Musa* and recently, BBrMV is reported in small cardamom (*Elettaria cardamomum*) in India [15]. BBrMV

also infects abaca in Philippines [16] and flowering ginger, *Alpinia purpurata* (Vieill.) K. Schum. in Hawaii [17]. BBrMV is transmitted non–persistently through several aphid species viz. *Pentalonia nigronervosa*, *Rhopalosiphum maidis*, and *Aphis gossypii* [4, 18]. The coat protein (CP) of potyviruses is multifunctional and involved in cell-to-cell and systemic movement [19, 20], regulation of viral RNA amplification [19], encapsidation of the RNA, vector transmission [19, 21], and also host specificity [19, 20]. Plant RNA viruses are highly variable enabling them to adapt rapidly to new or resistant hosts [22, 23]; hence, studies on genetic variability of populations of plant RNA viruses are important for understanding the evolution of viruses and their interactions with hosts [24–26]. Sequence comparison of CP genes can be used to identify and differentiate distinct potyviruses and their strains [27]. Numerous studies have established the phylogenetic relationship between strains of potyviruses based on CP gene sequences, notably on bean yellow mosaic virus (BYMV) [28], potato virus Y (PVY) [29–32], papaya ring spot virus (PRSV) [33], sugarcane mosaic virus (SCMV) [34], soybean mosaic virus (SMV) [35], tobacco vein banding mosaic virus (TVBMV) [36], turnip mosaic virus (TuMV) [37], and zucchini yellow mosaic virus (ZYMV) [38]. Complete genome of BBrMV isolates from India [12] and Philippines [13] have been sequenced and reported. The information currently available on molecular variability of BBrMV isolates in India is quite limited. A better knowledge of genetic diversity would provide an opportunity to understand the evolution and genetic structure of BBrMV. In this study, we report the molecular diversity of BBrMV isolates from India based on the CP gene sequences which might be helpful in developing sustainable disease management approaches.

Materials and methods

Virus isolates

Banana leaves or bracts showing symptoms of bract mosaic disease were collected from different locations in Tamil Nadu, Kerala, Karnataka, and Andhra Pradesh states of India during 2006–2013. The infected samples were processed immediately or stored at −86 °C. Host plants of some of the isolates were maintained in insect free glass house at National Research Centre for Banana (NRCB), Trichy, Tamil Nadu.

Preparation of RNA, cDNA synthesis, and reverse transcription PCR

For each sample, 100 mg of leaf or bract tissues of infected plants were frozen in liquid N_2 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 first strand complementary DNA (cDNA) was synthesized as described earlier [12]. The reverse transcription (RT)-PCR was carried out in a reaction mixture containing dNTP mix, virus specific forward primer (RSR10FP: 5′ATAGGATCCTCTGGAACGGAG TCAACC3′) and reverse primer (RSR10RP 5′TTCATGT TTCATCCCAAGCAGAG3′), and 10× PCR buffer with $MgCl_2$ and *Pfu* DNA Polymerase (Thermo Fisher Scientific, USA) to obtain amplified product of CP gene of BBrMV. The thermo-cycling conditions were as follows: 2 min at 95 °C (1 cycle), 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min (35 cycles), and a final extension at 72 °C for 10 min.

Cloning and sequencing analysis

The synthesized RT-PCR products were resolved in 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 also confirmed by digestion with restriction enzymes EcoRI and Hind III. Two independent clones per isolate were sequenced in both directions using an automatic sequencer (Eurofins Genomic India Pvt Ltd., Bangalore). The complete CP gene sequences of BBrMV isolates generated in this study were deposited in NCBI GenBank as accession numbers KF385470–KF385491. In addition, 27 CP gene sequences of BBrMV isolates of banana and cardamom from India and Southeast Asia were retrieved from NCBI for analysis. Alignments of 49 nucleotide (nt) sequences were done using CLUSTALW [39]. Sequence identity matrix and sequence difference count matrix were calculated using Bioedit sequence alignment editor version 5.09.04 [40]. The evolutionary distance and phylogenetic analyses were conducted using Maximum-likelihood phylogenetic tree constructed in the MEGA 5.0 software [41], and single nucleotide polymorphisms (SNPs) and insertion and deletion (INDELs) were calculated using DnaSP (version 5.10) [42] by comparing the nt sequence with the Trichy isolate (Accession number HM131454) as reference sequence. K_a/K_s value was calculated using the same tool to analyze synonymous and non-synonymous mutations at nt level, which really affect the amino acid (aa) sequences of the protein. Recombination sites in the aligned sequences were identified by implementing the RDP4 Beta 4.24 software [43]. We assessed the possible association between recombination break point location and predicted RNA

Table 1 Analysis of evolutionary distance, SNPs, INDELs** and K_s/K_a ratio, and per cent sequence identity of CP gene of BBrMV isolates

S. no.	Isolates	Accession number	Evolutionary distance	SNP	K_a	K_s	K_a/K_s	Nucleotide identity (%)	Amino acid identity (%)
1	TN5	HM131454#	–	–	–	–	–	80–98	84–99
2	TN4	EU009210	0.019	17	0.0044	0.0697	0.0631	81–99	84–99
3	TN6	KF385470*	0.038	33	0.0213	0.0935	0.2278	81–99	83–98
4	TN7	KF385471*	0.031	27	0.0124	0.0938	0.1321	81–99	85–99
5	TN8	KF385472*	0.017	15	0.0073	0.0490	0.1489	81–99	85–100
6	TN9	KF385473*	0.023	20	0.0073	0.0748	0.0975	81–99	85–100
7	TN10	KF385474*	0.025	22	0.0117	0.0698	0.1676	80–99	84–99
8	TN11	KF385475*	0.049	42	0.0332	0.0992	0.3346	79–96	80–95
9	TN12	KF385476*	0.020	18	0.0132	0.0441	0.2993	81–98	84–99
10	TN13	KF385477*	0.032	28	0.0088	0.1123	0.0783	81–98	83–99
11	TN14	KF385478*	0.257	177	0.0864	1.0758	0.0803	79–94	80–91
12	TN15	KF385479*	0.030	26	0.0117	0.0904	0.1294	83–98	85–99
13	TN16	KF385484*	0.209	152	0.0468	1.0619	0.0440	82–94	88–92
14	TN17	KF385485*	0.022	19	0.0073	0.0696	0.1048	80–100	85–100
15	TN18	KF385486*	0.038	33	0.0147	0.1179	0.1246	81–98	86–98
16	TN19	KF385487*	0.030	26	0.0132	0.0852	0.1549	79–100	84–100
17	TN21	KF385489*	0.043	37	0.0221	0.1118	0.1976	79–97	83–98
18	AP7	KF385480*	0.030	26	0.0087	0.1018	0.0854	80–98	84–99
19	KAR2	KF385481*	0.028	25	0.0117	0.0852	0.1373	80–99	84–100
20	TN22	KF385482*	0.034	29	0.0206	0.0750	0.2746	79–98	82–96
21	TN23	KF385483*	0.031	27	0.0117	0.0963	0.1215	80–98	84–98
22	KER3	KF385488*	0.037	32	0.0266	0.0697	0.3816	80–98	81–95
23	KAR3	KF385490*	0.031	27	0.0221	0.0591	0.3739	79–98	82–97
24	KER2	KF385491*	0.027	24	0.0117	0.0798	0.1466	81–98	84–99
25	KAR	EF654655	0.027	24	0.0124	0.0776	0.1598	81–98	85–99
26	TN20	AY494979	0.020	18	0.0088	0.0591	0.1489	80–99	85–100
27	AP1	AY953427	0.033	29	0.0146	0.0965	0.1513	80–100	83–100
28	AP2	HM348778	0.037	32	0.0191	0.0963	0.1983	80–100	82–99
29	AP3	HM348779	0.034	30	0.0146	0.1019	0.1433	80–99	83–100
30	AP4	HM348780	0.030	26	0.0102	0.0963	0.1059	81–98	85–99
31	AP5	HM348781	0.036	31	0.0146	0.1070	0.1364	80–98	84–98
32	AP6	HM348782	0.026	23	0.0073	0.0907	0.0805	80–98	84–99
33	I1	AF071582	0.019	17	0.0088	0.0540	0.1630	81–98	94–100
34	I2	AF071583	0.028	25	0.0147	0.0746	0.1971	80–98	83–99
35	I3	AF071584	0.052	45	0.0117	0.1990	0.0588	80–100	84–100
36	I4	EU699770	0.049	42	0.0088	0.1932	0.0455	80–100	84–100
37	Card1	HQ709165	0.028	25	0.0132	0.0799	0.1652	80–98	84–98
38	Card2	HQ709166	0.031	27	0.0199	0.0668	0.2979	80–98	83–97
39	Card3	HQ709164	0.023	20	0.0088	0.0696	0.1265	80–99	84–100
40	Card5	HQ709163	0.029	25	0.0176	0.0643	0.2737	80–99	83–98
41	Card6	HQ709162	0.024	21	0.0102	0.0695	0.1468	80–99	84–99
42	P4	DQ851496	0.047	41	0.0088	0.1871	0.0470	80–99	85–99
43	P5	EU414267	0.049	42	0.0088	0.1932	0.0455	80–100	84–100
44	P1	AF071590	0.057	49	0.0154	0.2089	0.0737	80–98	84–98
45	P2	AF071585	0.056	48	0.0162	0.1989	0.0814	79–98	83–98
46	P3	AF071586	0.051	44	0.0088	0.2056	0.0428	79–98	84–100
47	WS1	AF071587	0.057	49	0.0169	0.2026	0.0834	79–98	83–98

Table 1 continued

S. no.	Isolates	Accession number	Evolutionary distance	SNP	K_a	K_s	K_a/K_s	Nucleotide identity (%)	Amino acid identity (%)
48	VT1	AF071588	0.051	44	0.0117	0.1932	0.0606	80–99	84–100
49	TH1	AF071589	0.052	45	0.0102	0.2052	0.0497	79–96	84–100

TN-Tamil Nadu, AP-Andhra Pradesh, KER-Kerala, KAR-Karnataka, TN4, TN5, TN11, TN15, TN16, TN17, TN18, TN19, TN22, TN23 and I3-Tiruchirapalli, TN6-Pudukottai, TN8 and TN9-Theni, TN10-Karur, TN12-Cuddalore, TN13 and TN14-Tanjore, TN21, I1 and I2-Coimbatore, AP7-Kovur, KAR2-Bangalore, KAR3-Arabhavi, KER2-Kasargod, KER3-Kayankulam, TN, AP, KER, KAR, I are belongs to Indian origins. P-Philippines, WS-Western Samoa, VT-Vietnam, THI-Thialand. S. no 37–41 are the sequences of coat protein gene of BBrMV infecting cardamom

* isolates generated in this study

** no INDELs were detected

Used as a reference isolate in this study

secondary structures using mFOLD (Version 3.2) [44]. Genetic differentiation between the BBrMV populations was examined by three permutation-based statistical tests, K_s*, Z, and Snn [45]. The extent of genetic differentiation or the level of gene flow between populations were measured by estimating Fixation index (F_{ST}) using DnaSP version 5.10 [42]. The absolute value of F_{ST} ranges from 0 to 1 for undifferentiated to fully differentiated populations. Normally, an absolute value of $F_{ST} > 0.33$ suggests an infrequent gene flow, while an absolute value of $F_{ST} < 0.33$ suggests frequent gene flow between the populations compared. DnaSP version 5.0 [42] 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. To determine the site-specific selection pressures acting on CP gene, four different codon-based maximum-likelihood methods implemented on Datamonkey server (http://www.datamonkey.org) [46] were used.

Results

In RT-PCR, 900-bp-long CP gene encoding 300 aa was amplified from all 22 isolates of BBrMV. Size of CP gene was similar to the previously published sequences [6, 12]. Based on CP gene sequence analysis, BBrMV isolates shared identity of 79–100 % at nt and 80–100 % at aa levels (Table 1). Out of 22 isolates used in this study, two distinct isolates viz. TN14 and TN16 had a divergence of 21 and 18 % at nt and 20 and 12 % at aa level, respectively, with other isolates. Among the isolates of non-Indian origin, 96–100 % similarity at nt and aa level was recorded. Except isolates TN14 and TN16, all the Indian isolates shared 94–100 % and 92–100 % identity at nt and aa level. Sequence difference count matrix was ranged 3–188 nt and 0–59 aa among the BBrMV isolates used in this study (data not shown).

The phylogenetic trees were constructed from the alignment of nt and deduced aa sequences of CP gene of 45 isolates excluding 4 putative recombinants (Fig. 1a, b). The BBrMV isolates were grouped into two distinct major groups based on genetic distances. Two isolates Trichy (TN14) and Tanjore (TN16) were clustered in group I, and the remaining 43 isolates formed into group II which includes most of the Indian and all Southeast Asian isolates. Group II was further divided into two sub-groups IIa and IIb; sub-group IIa consists of all Southeast Asian isolates (Philippines, Thailand, Vietnam, and Western Samoa) together with two Indian isolates (I3 and I4), and the remaining isolates were grouped in sub-group IIb which includes isolates TN6, TN7, and KAR. Within the sub-group IIb, four isolates of cardamom clustered together, and six isolates of Andhra Pradesh have clustered together. The phylogenetic tree generated based on alignment of deduced aa sequences is presented in Fig. 1b. The isolates TN6, TN7, and KAR were clustered together in sub-group IIa as against their positions in nt tree (sub-group IIb). Grouping of isolates TN14 and TN16 and the clustering of the other isolates were similar to that of nt tree (Fig. 1a).

In order to analyze the polymorphism at sequence level, the CP sequences of BBrMV were analyzed for the presence of SNPs and INDELs. All the sequences of 48 isolates were compared with sequence of TN5 isolate (Accession number HM131454) taken as reference sequence. Isolate TN8 had an evolutionary distance of 0.017 from the reference sequence with 15 SNPs. The next closest sequences were of isolates TN4 and I1 with an evolutionary distance of 0.019 and 17 SNPs. Isolates of TN14 and TN16 showed evolutionary distance of 0.257 and 0.209 with 177 and 152 SNPs, respectively, which were the highest among BBrMV isolates. The farthermost sequence among the non-Indian isolates, namely P1 and

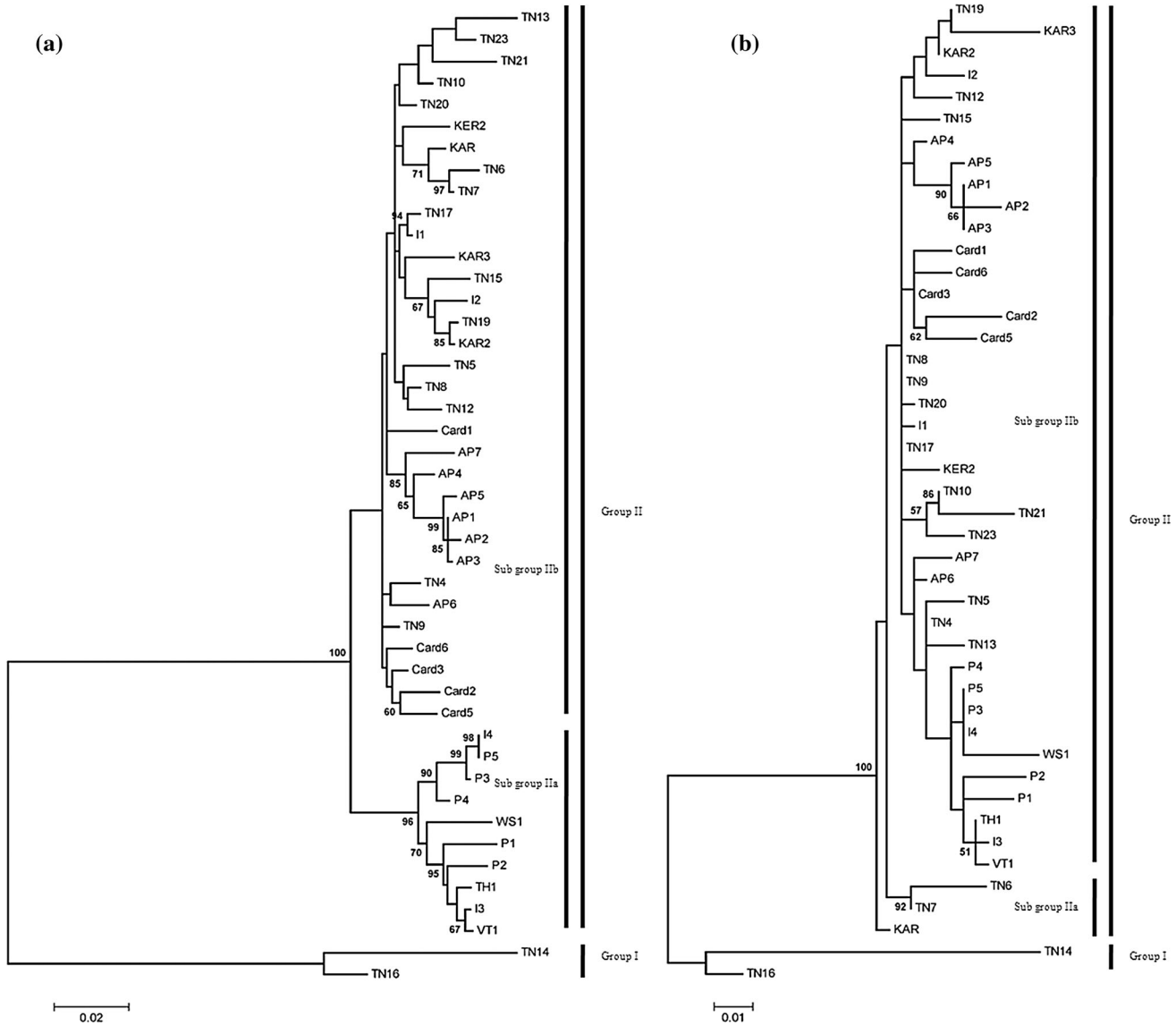

Fig. 1 Phylogenetic analysis of coat protein (CP) of nucleotide (**a**) and amino acid (**b**) banana bract mosaic virus (BBrMV) isolates from different parts of the world. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Bootstrap scores above 50 % (1,000 replicates) are placed at the tree nodes. The scale bar represents the number of nucleotide substitutions per site. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. For the detailed of isolates, refer Table 1

WS1 with an evolutionary distance of 0.057 and 49 SNPs was recorded when compared with the reference sequence. INDELs were not observed in any of the sequences used in this study (Table 1). The ratio of non-synonymous (K_a) to synonymous (K_s) nucleotide substitution rates (K_a/K_s) was calculated to understand the nt change, which affect the aa sequence of the protein. The values of K_a and K_s ranged from 0.0044 to 0.0864 and 0.0441 to 1.0758, respectively. Isolates TN14 and TN16 had the highest K_a and K_s values (Table 1). The value of K_a/K_s ranged from 0.0440 to 0.3816. The isolates TN11, TN12, TN22, Card2, Card5, KER3, and KAR3 had comparitively higher K_a/K_s value ranging 0.2737–0.3816 showed high mutations both in nucleotide and protein level when compared to the reference isolate.

To gain insights into potential recombinations in CP gene of BBrMV, all the available sequences (49 isolates) were examined with various recombination detection algorithms to identify putative recombinants and recombination breakpoints. In this study, a total of eleven potential recombinants were detected (Table 2). Out of 49 BBrMV isolates analyzed, isolate KER3 showed evidence of recombination by five recombination detection algorithms. Recombinants TN11 and TN22 were detected by four recombination detection

Table 2 Recombination events detected in the cp gene of BBrMV isolates using RDP methods showing the parental and recombinant sequences

Recombinant isolate	Recombination break points	Parental isolate[a]		Recombination detection[b]	
		Major	Minor	Methods	*P* value
TN11	317–479	P1	Card3	M, C, **S**, 3S	3.090×10^{-06}
KER3	106–294	KAR3	TN18	M, C, **S**, 3S, L	6.065×10^{-06}
TN22	146–658	TN17	TN6	**M**, C, 3S, L	5.242×10^{-04}
TN21	485–663	TN23	TN6	S	1.644×10^{-03}
TN18	146–454	TN6	TN8	M, C, **S**	1.318×10^{-04}
I3, THI and P1	318–509, 318–505, 318–507	AP7	TN13	S	6.380×10^{-02}
WS1	44–432	AP7	TN13	S	6.380×10^{-02}
TN6and TN7	148–672	TN4	Card5	S	1.516×10^{-01}

The methods whose *P* values 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* MAXIMUM, CHISQUARE, *C* CHIMAERA, *S* SISCAN, *3S* 3SEQ, *L* LARD

Table 3 Genetic differentiation measurement for host and geography of BBrMV population

Parameters	Banana and Cardamom	India and other countries	TN14, TN16 and others	TN14, TN16 and cardamom
K_s*(*P* value)	3.56862 (0.2770)	3.44665 (0.0000)	3.44062 (0.0000)	2.88357 (0.0490)
Z (*P* value)	586.06641 (0.3290)	478.19048 (0.0000)	540.05042 (0.0000)	4.50000 (0.0500)
Snn (*P* value)	0.89796 (0.0720)	0.86735 (0.0150)	1.00000 (0.0000)	1.00000 (0.0500)
F_{ST}	0.09994	0.37968	0.74139	0.78727
Nm	2.25	0.41	0.09	0.07

K_s*, *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 [45]. The *Z* statistic value results from ranking distances between all pairs of sequences. Snn the frequency with which the nearest neighbors of sequences are found in the same locality; F_{ST}, 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

algorithms. SISCAN algorithm alone detected four putative recombinants which were contributed by parents AP7 and TN13. The relationship of putative recombinants to their contributing major and minor parental isolates could not be attributed to either geographical locations or the host cultivar. However, these recombinants would have occurred due to mixed infection of isolates in a particular banana variety, and later it would have been transmitted through vector to other banana varieties. Three isolates namely AP7, TN13, and TN6 were involved as major or minor parental isolates for majority of the recombinants. Eight recombinants were detected from Indian BBrMV isolates, and three were from the Southeast Asian isolates for which the putative parents are from India. It was interesting to record that Indian isolates appeared as major and minor parents for isolates of Philippines, Thailand, and Western Samoa. This suggests a possible derivation of Southeast Asian isolates by genetic exchange involving Indian isolates. To the best of our knowledge, this is the first report on evidence of recombination in Indian BBrMV isolates. We have assessed the possible association between recombination break point location and predicted RNA hairpin loop structures using mFOLD (Version 3.2) [44] (data not shown). Three hairpin loop structures were detected in each of the parental isolates of recombinant TN11, and the position of such structures was invariable. Whereas, the two hairpin loops detected in each of the parental isolates of recombinant KER3 were variable, and a hairpin loop structure detected in one of the parental isolates was also identified at position 106–294 in KER3. Sixteen and ten hairpin loop structures were detected in the parental isolates of recombinants TN22 and TN18 at positions 146–658 and 146–454, respectively.

To determine the gene and site-specific selection pressures acting on BBrMV CP, four different codon-based maximum-likelihood algorithms SLAC, FEL, IFEL, and REL within the HyPhy software package as implemented in Datamonkey server with significance levels set at $P = 0.01$ for the first three programs and Bayes factor = 50 for REL, respectively, were used to estimate the value of dN/dS at each codon site (data not shown). Most of the codons were under negative or neutral selection, while the codon at position 92 was found under positive selection detected by IFEL and REL, and codons 28 and 43 under positive selection was detected by REL. Total number of codons found under negative selection in CP genes was 61, 81, 84, and 171 detected by the methods

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

Host and Geography	Tajima's D	Fu and Li's D	Fu and Li's F	Haplotype diversity	Nucleotide diversity
All	−1.93090	−2.08402	−2.42460	0.999	0.04844
Banana and cardamom	−1.83329	−1.76572	−2.13951	0.999	0.05097
India and other countries	−2.03205	−1.81043	−2.25808	1.000	0.04612

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

SLAC, FEL, IFEL, and REL respectively. The new method FUBAR available in Datamonkey also ensured that codon 28 position was under positive selection (data not shown).

Three independent statistical tests of population differentiation were applied [45] to estimate whether geographical isolates are genetically differentiated populations (Table 3). Thus, the statistical tests of genetic differentiation, measurement of gene flow, and population demography of cardamom and banana groups were investigated. Genetic differentiation between populations was examined by three permutation-based statistical tests, K_s*, Z, and Snn [45]. Analyses of these values (Table 3) showed the genetic differentiation within the isolates of banana, and those from cardamom were significant. The F_{ST} value between the BBrMV isolates from banana and cardamom was smaller than 0.33 indicating a relatively frequent gene flow between them. However, the F_{ST} values within the two banana infecting BBrMV subgroups were bigger than 0.33, indicating infrequent gene flow between them. The results are in agreement with the phylogenetic relationships of each population.

The pattern of molecular diversity was 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). The statistical tests are expected to have negative values for background selection, genetic hitchhiking, and demographic expansion and also indicate that a population maintained low frequency polymorphism [22, 47, 48]. For the majority of groups across the genome values of Tajima's D, Fu, and Li's D* and F* statistical tests were negative, indicating the occurrence of recent population expansion of BBrMV. Haplotype and nucleotide diversities of all BBrMV populations were compared (Table 4). In most cases, the haplotype diversity values were high, and nucleotide diversity values were low. Overall, the deviations from the neutral equilibrium model for the CP gene were analyzed, together with the combination of haplotype diversity and overall lack of nucleotide diversity within individual geographical groups are consistent with a model of recent population expansion events.

Discussion

Studies of the genetic diversity of several potyviruses have shown details of genetic variations at regional or global scales [28–38, 49]. The CP gene of *potyvirus* is multifunctional and plays important roles in virus life cycle and virus-vector interaction; therefore, it could be a target of selection [50] by both host plants and vectors. In this study, we compared CP gene sequence of 49 BBrMV isolates including 27 previously reported isolates originating from different countries. Phylogenetic trees were constructed from CP gene showed two monophyletic clusters in the world population of BBrMV (Fig. 1a, b). However, the Indian isolates did not show any relationship according to geographical origins and the hosts from which they were isolated. Rodoni et al. [6] proposed that the movement of BBrMV into Philippines occurred as one event and that BBrMV persists in India for a longer period of time. A possible explanation for the geographical distribution of BBrMV is that the virus has moved as a separate event, perhaps through different infected cultivars of banana. In India, BBrMV was first noticed in 1966; because of prolonged presence, high divergence of BBrMV populations may have occurred. Though BBrMV was noticed first in southern parts of Kerala, India has moved to three neighboring states in five decades either through infected planting material or through aphid vector. There is no domestic quarantine enforced to restrict the movement of banana suckers between the states. This virus has recently been reported to infect small cardamom which is grown along with banana in Western Ghat region of Kerala and Karnataka.

In this study, eleven putative recombinants were detected among 49 BBrMV isolates. Recombination events are capable of generating considerable genetic diversity [51, 52] and important source of genetic variation for potyviruses [25]. Recombination at the CP sequences has been reported in many potyviruses including BYMV [28], yam mosaic virus [53], plum pox virus [54], PRSV [55], SCMV [34], PVY [29–32], and these frequent recombination footprints indicate the importance of this process in

shaping the fitness of potyvirid populations [56]. Four of the strong putative recombinants detected in our study showed hairpin loop structures in mFold analysis which would have played a role in the higher recombination rates. Our results corroborate with the findings of Galli et al. [57] who have assessed the correlation between the break point clusters and RNA hairpins in the pol gene of HIV type1. Genetic differentiation analyses indicated the frequent gene flows between BBrMV populations of banana and cardamom and infrequent gene flow within banana subgroups. The three neutrality tests of statistical values were negative, suggesting that BBrMV population was under negative or purifying selection. In our study, codons of BBrMV CP gene at position 28, 43, and 92 were found under positive selection. Codon positions of 25 and 68 in PVY CP cistron were detected under positive selection and the mutations at codon positions 25 and 68 significantly modified virus accumulation in different hosts and its transmissibility by aphids, respectively [58]. In case of SCMV, codon 27 of CP gene formed under positive selection [34], but the significance of this change is not known. In the present study, significance of positive selection of BBrMV CP cistron at codon positions 28, 43, and 92 could not be related, and further investigation is needed to test their role in the host-virus-vector interactions.

Evolutionary distance, the number of substitutions per site separating a pair of homologous sequences since they diverged from their common ancestral sequence, is an extremely important measure in molecular evolution and comparative genomics [59]. In this study, only two isolates of TN14 and TN16 showed the highest evolutionary distances and more SNPs compared to the other Indian BBrMV isolates. The data obtained from phylogeny, sequence identity matrix, SNPs, and evolutionary distance suggest the diversion of BBrMV isolates into two major clusters; one that includes an isolate naturally infecting a triploid banana cultivar of Cavendish sub-group Robusta (AAA) (isolate, TN14) collected from a farmer's field, and another isolate was from banana accession Beula (ABB) (isolate, TN16) collected from the field gene bank of NRCB, Trichy. More than ten banana cultivars are cultivated in Trichy and its adjoining areas, and the incidence of BBrMV has been recorded in all these varieties (data not shown). In our study, the variability between the isolates collected in and around Trichy is ranging from 1 to 21 and 1 to 20 % nt and aa, respectively. This might be due to the spread of virus infection from disease area. Existence of variability in BBrMV isolates in this region might be due to population bottlenecks such as switching over to varieties and selection pressure due to vector transmission. Finally, to evaluate the importance of natural selection to shape the population structure of BBrMV, Tajima's D, Fu and Li's D and Fu, and Li's F statistics were used to test the mutation neutrality hypothesis using DnaSP 5.10 program. The three statistical values were significantly negative ($P > 0.05$), suggesting a strong negative or purifying selection.

Overall, this study on BBrMV diversity based on CP gene sequence analysis revealed high variability among the Indian isolates and emergence of recombinants. The divergence of BBrMV in India into two separate groups was shown that emphasis to be given for complete genome sequencing of more isolates including two distinct isolates (TN14 and TN16) identified in this study to have a clear picture on the variability and evolution. As CP gene sequences are frequently used to develop pathogen derived resistance against potyvirus by means of genetic engineering, our results on BBrMV diversity could help in predicting the risk of breakdown of resistance, if BBrMV resistant transgenic banana lines are developed for cultivation. On the basis of the analyses described above, there is a risk of loss-of resistance of CP mediated virus resistant transgenic banana plants due to the presence of new variants. Sequencing of the RNA silencing suppressor helper component protease (HC-Pro) gene of those isolates would further reveal the utility in developing transgenic banana resistant against BBrMV using a hairpin structure targeting to the RNA silencing machinery. As shown in this work, the introduction of new isolates and appearance of new genetic types may occur not only by recombination but also by mutation. This study reports, for the first time, the genetic diversity of BBrMV population in India and Southeast Asia and illustrate that, when designing efficient long term management strategies, it is important to understand the genetic structure of virus population at different geographic scales, country, region, and locality.

Acknowledgments This is a part of the Ph.D work of the first author, and authors are grateful to Director, NRCB, Trichy, for encouragement, and providing necessary facilities to carry out the work. We thank Dr C. Anuradha, for technical discussion.

References

1. G. Aurore, B. Parfait, L. Fahrasmane, Trends Food Sci. Technol. **20**, 78–91 (2009)
2. FAOSTAT (2011), http://faostat.fao.org. Accessed 2011
3. R. Ploetz, Infomusa **13**, 11–16 (2004)
4. L.V. Magnaye, R.R.C. Espino, Philipp. Agric. **73**, 55–59 (1990)
5. M.F. Bateson, J.L. Dale, Arch. Virol. **140**, 515–527 (1995)
6. B.C. Rodoni, J.L. Dale, R.M. Harding, Arch. Virol. **144**, 1725–1737 (1999)
7. D.F. Quito-Avila, M.A. Ibarra, R.A. Alvarez, M.F. Ratti, L. Espinoza, J.M. Cevallos-Cevallos, E.L. Peralta, Plant Dis. **97**(7), 1003 (2013)
8. J. Samraj, M.R. Menon, S.P. Christudas, P.K. Satyarajan, Agric. Res. J. Kerala **4**, 116 (1966)
9. B.C. Rodoni, Y.S. Ahlawat, A. Varma, J.L. Dale, R.M. Harding, Plant Dis. **81**, 669–672 (1997)

10. S.J. Singh, R. Selvarajan, H.P. Singh, in *Banana-improvement, production and utilization*, ed. by H.P. Singh, K.L. Chadha, 2000, pp. 381–383
11. R. Selvarajan, V. Balasubramanian, K.J. Jeyabaskaran, S.D. Pandey, M.M. Mustaffa, Indian J. Hort. **66**(3), 301–305 (2009)
12. V. Balasubramanian, R. Selvarajan, Arch. Virol. **157**, 397–400 (2012)
13. C. Ha, S. Coombs, P.A. Revil, R.M. Harding, M. Vu, J.L. Dale, Arch. Virol. **153**, 25–36 (2008)
14. B.Y.W. Chung, W.A. Miller, J.F. Atkins, A.F. Firth, An overlapping essential gene in the Potyviridae. Proc. Natl. Acad. Sci. USA **105**, 5897–5902 (2008)
15. A. Siljo, A.I. Bhat, C.N. Biju, M.N. Venugopal, Phytoparasitica **40**, 77–85 (2012)
16. M. Sharman, C.F. Gambley, E.O. Oloteo, R.V.J. Abgona, J.E. Thomas, Aus. Plant Pathol. **29**, 69 (2000)
17. I.C. Wang, D.M. Sether, M.J. Melzer, W.B. Borth, J.S. Hu, Plant Dis. **94**(7), 921 (2010)
18. A.R. Munez, Faculty of Graduate School, Los Baños: University of Philippines,1 – 57(1992)
19. S. Urcuqui-Inchima, A.L. Haenni, F Bernardi. Virus Res. **74**, 157–175 (2001)
20. D. Hofius, A.T. Maier, C. Dietrich, I. Jungkunz, F. Bornke, E. Maiss, U. Sonnewald, J. Virol. **81**, 11870–11880 (2007)
21. D.D. Shukla, M.J. Frenkel, C.W. Ward, Canadian J. Plan. Pathol. **13**, 178–191 (1991)
22. M. Tsompana, J. Abad, M. Purugganan, J.W. Moyer, Mol. Ecol. **14**, 53–66 (2005)
23. K. Ohshima, S. Akaishi, H. Kajiyama, R Koga, A.J. Gibbs. J. Gen. Virol. **91**, 788–801 (2010)
24. F. García-Arena, A Fraile, J.M. Malpica. Ann. Rev. Phytopathol. **39**, 157–186 (2001)
25. A.J. Gibbs, K Ohshima. Ann. Rev. Phytopathol. **48**, 205–223 (2010)
26. A.J. Gibbs, M. Gibbs, K. Ohshima, F. García-Arenal, Origin and Evolution of Viruses, ed. By E. Domingo, C.R Parrish, J.J.Holland (2nd ed. Elsevier Academic Press, London, 229–250(2008)
27. D.D. Shukla, P.M. Strike, S.L. Tracy, K.H. Gough, C.W. Ward, J. Gen. Virol. **69**, 1497–1508 (1988)
28. S.J. Wyile, R.A.C. Jones, Phyopathol. **99**, 512–518 (2009)
29. J.M. Cuevas, A. Delaunay, J.C.Visser, D.U. Bellstedt, E. Jacquot, S.F. Elena. PLoS ONE **(75)**, e37853 (2012). doi: 10.1371/journal.pone.0037853
30. A.V. Karasev, X. Hu, C.J. Brown, C. Kerlan, O.V. Nikolaeva, J.M. Crosslin, S.M. Gray, Phytopathol. **101**, 778–785 (2011)
31. T. Ogawa, A. Nakagawa, T. Hataya, K. Ohshima, J. Phytopathol. **160**, 661–673 (2012)
32. Y.P. Tian, J.L. Liu, C.L. Zhang, Y.Y. Liu, B. Wang, X.D. Li, Z.K. Guo, J.P.T. Valkonen, Phytopathol. **101**, 377–387 (2011)
33. R.K. Jain, J. Sharma, A.S. Sivakumar, P.K. Sharma, A.S. Byadgi, A.K. Verma, A. Varma, Arch. Virol. **149**, 2435–2442 (2004)
34. Y. Li, R. Liu, T. Zhou, Z. Fan, Virus Res. **171**, 242–246 (2013)
35. J.K. Seo, K. Ohshima, H.G. Lee, M. So, H.S. Choi, S.H. Lee, S.H. Sohn, K.H. Kim, Virology **393**, 91–103 (2009)
36. C.L. Zhang, R. Gao, J. Wang, G.M. Zhang, X.D. Li, H.T. Liu, Virus Res. **158**, 188–198 (2011)
37. H.D. Nguyen, H.T.N. Trana, K. Ohshimaa, Virus Res. **171**, 138–149 (2013)
38. H. Lecoq, C. Wipf-Scheibel, C. Chandeysson, A. Lê, Van, F Fabre, C. Desbiez, Virus Res. **141**, 190–200 (2009)
39. J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. **22**, 4673–4680 (1994)
40. T.A. Hall, Nucleic Acids Sym. Ser. **41**, 95–98 (1999)
41. K. Tamura, D. Peterson, N Peterson, G. Stecher, M. Nei, S. Kumar. Mol. Biol. Evol. **28**, 2731–2739 (2011)
42. P. Librado, J. Rozas, Bioinformatics **25**, 1451–1452 (2009)
43. D.P. Martin, P. Lemey, M. Lott, V. Moulton, D. Posada, P. Lefeuvre, Bioinformatics **26**, 2462–2463 (2010)
44. M. Zuker, Nucleic Acids Res. **31**, 3406–3415 (2003)
45. R.R. Hudson, Genetics **155**, 2011–2014 (2000)
46. W. Delport, A.F.Y. Poon, S.D.W. Frost, S.L.K. Pond, Bioinformatics **26**, 2455–2457 (2010)
47. J. Hey, E. Harris, Mol. Biol. Evol. **16**, 1423–1426 (1999)
48. F. Tajima, Genet. **123**, 585–595 (1989)
49. F. Revers, O. Le Gall, T. Candresse, M. Le Romancer, J. Dunez, J. Gen. Virol. **77**, 1953–1965 (1996)
50. B. Moury, C. Morel, E. Johansen, M. Jacquemond, J. Gen. Virol. **83**, 2563–2573 (2002)
51. Y. Suzuki, T. Gojobori, O. Nakagomi, FEBS Lett. **427**, 183–187 (1998)
52. A. Varsani, E.van der Walt, L. Heath, E.P. Rybicki, A.L. Williamson, D.P. Martin. J. Gen. Virol. **87**, 2527–2531 (2006)
53. M. Bousalem, E.J.P. Douzery, D. Fargette, J. Gen. Virol. **81**, 243–255 (2000)
54. M.T. Cervera, J.L. Riechmann, M.T. Martin, J.A., Garcia. J. Gen. Virol. **74**, 329–334 (1993)
55. S.K. Mangrauthia, B. Parameswari, R.K. Jain, S. Praveen, Biochem. Genet. **46**, 835 (2008)
56. J. Sztuba-Solin′ ska, A. Urbanowicz, M. Figlerowicz, J.J Bujarski, Ann. Rev. Phytopathol. **49**, 415–443 (2011)
57. A. Galli, A. Lai, S. Corvasce, F. Saladini, C. Riva, L. Deho, L. Caramma, M. Franzetti, L. Romano, M. Galli, M. Zazzi, C. Balotta, J. Gen. Virol. **89**, 3119–3125 (2008)
58. B. Moury, V. Simon, Mol. Biol. Evol. **28**, 2707–2717 (2011)
59. M.S. Rosenber, Bioinformatics **6**, 102 (2005)