Molecular characterization and sequence analyses of Banana bunchy top virus infecting banana cultivar Jahaji (Dwarf Cavendish) in Assam, India

Gajendra Mohan Baldodiya¹ · Geetanjali Baruah¹ · Basanta Kumar Borah¹ · Mahendra Kumar Modi¹ · Palash Deb Nath²

Received: 14 September 2018 / Accepted: 13 February 2019 / Published online: 28 February 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Several isolates of Banana bunchy top virus (BBTV) have been reported worldwide. They are members of either the Pacific Indian Ocean (PIO) or the South East Asian (SEA) group. However, there is only one completely sequenced isolate published from the northeastern part of India till date. Therefore, we obtained the complete sequences of all the six genomic components of a BBTV isolate from the northeastern Indian state of Assam. The isolate was named as BBTV-As-JOR, and its genome showed the presence of the reported conserved motifs. Nevertheless, like other Indian BBTV isolate, the major common regions in DNA-R and DNA-U3 of BBTV-As-JOR had deletions of 26 and 36 nucleotides, respectively. Phylogenetic analysis based on 312 sequences of BBTV DNA-R classified BBTV-As-JOR as a member of the PIO group; similar phylogenetic patterns were also found with the other genomic segments. Analysis with Recombination Detection Program revealed two intra-segment recombination events involving DNA-C of geographically distinct BBTV isolates. On the other hand, DNA-U3 and DNA-N were found to be involved in few inter-segment recombination events in BBTV-As-JOR. This is the first report of a BBTV isolate from Assam and also of another PIO isolate from the region (the other isolate, BBTV-Umiam, was much closer to the SEA group). The detected possible recombinants could emerge as a major future threat for the banana cultivations in the country considering the asexual nature of propagation of banana crop.

Keywords Banana bunchy top nanovirus · BBTV · Pacific Indian Ocean · South East Asian · Recombination

Introduction

Banana bunchy top disease, caused by *Banana bunchy top virus* (BBTV), is the most destructive viral disease of Banana (*Musa* spp.). The disease is considered as a serious threat for banana cultivation worldwide. The disease has been reported from the Pacific region, South and South East Asia and Africa (Dale 1987; Thomas et al. 2000; Kagy

et al. 2001; Kumar et al. 2011). The symptoms of Banana bunchy top disease includes foliar chlorosis, vein clearing, dwarfing of plant and bunchy top appearance followed by yield reduction. The transmission of the virus occurs either by vegetative propagation or the aphid vector, *Pentalonia nigronervosa* (Allen 1978). To understand the Banana bunchy top disease and to prevent its negative impact on banana productivity worldwide, it is very important to know the molecular diversity of *Banana bunchy top virus* across the banana growing regions.

BBTV is a member of the genus, *Babuvirus*, and the family, *Nanoviridae*, comprising single-stranded DNA (ssDNA) genome. The genome of the virus is multipartite in nature and consists of six circular DNA-segments of approximately 1.1 kb each (Harding et al. 1991, 1993; Burns et al. 1995; Karan et al. 1997). All the six segments are encapsidated within separate icosahedral virions, each about 18–20 nm in diameter (Wu and Su 1990). Each ssDNA segment of BBTV holds an open reading frame (ORF) which encodes

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1636-5>) contains supplementary material, which is available to authorized users.

✉ Basanta Kumar Borah
basantabora@gmail.com

¹ Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam 785013, India

² Department of Plant Pathology, Assam Agricultural University, Jorhat, Assam 785013, India

a functional protein, viz., the R segment encodes the replication initiation protein (Rep), the U3 encodes a protein of unknown function, the S segments encodes the coat protein (CP), the M segment encodes the movement protein (MP), the C segment encodes the cell cycle link protein (Clink) and the N segment encodes the nuclear shuttle protein (NSP) (Beetham et al. 1997, 1999; Wanitchakorn et al. 1997, 2000, 2000; Tian et al. 2005). A highly conserved major common region (CR-M) and a stem-loop common region (CR-SL) are found in all these segments (Burns et al. 1995). A satellite DNA component is also found in some BBTV isolates, which encodes a homologous protein to Rep of DNA-R (Horser et al. 2001).

BBTV have been reported from all over the world. Earlier, on the basis of phylogenetic relationship among the DNA-R segments, isolates of the virus were categorized into two groups, viz., the South Pacific group and the Asian group (Karan et al. 1994). Later, the categorization was modified into two different lineages, viz., the Pacific Indian Ocean (PIO) and the South East Asian (SEA) groups, based on their geographical distribution. Today, BBTV isolates reported from Australia, Egypt, Hawaii, India, Myanmar, Pakistan, Sri Lanka and Tonga are considered as the members of the PIO group, and the isolates reported from China, Indonesia, Japan, Philippines, Taiwan and Vietnam are considered as the members of the SEA group (Yu et al. 2012).

Several BBTV isolates have been reported, mostly from the northern and southern parts of India and all these isolates come under the PIO group considering their genetic makeup. However, an isolate of BBTV, recently characterized from Meghalaya (BBTV-India-Umiam), showed distinctness from the PIO group and displayed closer relationships with the SEA group (Banerjee et al. 2014). As the northeastern part of India is considered as a bio-diversity hotspot that harbours several banana cultivars and wild varieties, the finding of an alien type of BBTV in the region may not be a sporadic occurrence. Moreover, the exchange of genetic information among various isolates also leads to generation of a new or distinct virus isolate. In a virus like BBTV, which has a multipartite genome, both the inter- and intra-segment recombination events are equally important so far as exchange of genetic information is concerned.

The present study mainly focuses on the identification, molecular characterization and genetic diversity and recombination analysis of a BBTV isolate infecting the banana cultivar Jahaji (Dwarf Cavendish) in the state of Assam, India. All the six segments of the isolate collected from Jorhat district, Assam, were isolated and characterized; and their complete DNA sequences were analysed. The characterized isolate was named as BBTV-As-JOR, and was compared with others isolates reported from India as well from other parts of the world. The putative recombination events were also studied *in silico* for understanding the evolutionary

relationship and diversity of BBTV-As-JOR with other related isolates.

Materials and methods

Virus source, DNA isolation and PCR amplification

The major banana cultivating areas of Assam were surveyed and from the symptomatic plants, i.e. the plants showing crowning, stunting and bunchy appearance on the top, leaf sample were collected along with non-symptomatic plant samples as negative control. The samples were stored at -70°C ; later, the total genomic DNA was extracted from the leaf samples using the CTAB method (Doyle and Doyle 1990).

The BBTV infection in the collected samples was confirmed by PCR amplification for all the BBTV segments using six primer sets for DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N (Islam et al. 2010). The PCR amplification was performed in 15 μl of reaction volume which consisted of approx. 100 ng of total DNA as template (1 μl), 1.5 μl of 10X *Taq* buffer with MgCl_2 , 2 μl of dNTP mix (2 mM), 0.5 μl *Taq* DNA polymerase (3U/ μl) and 0.5 μl each of the forward and reverse primers (10 pmol/ μl). The mixture was subjected to an initial denaturation at 94°C for 3 min followed by 30 cycles of denaturation (94°C for 1 min), annealing (1 min at respective T_m , Online Resource 1) and extension (72°C for 2 min). The reaction was followed by a final extension at 72°C for 10 min. The reactions were performed in a thermocycler (Gene Amp® PCR System 9700). The amplified products were resolved in 1% agarose gel along with 1 kb Plus DNA ladder (Invitrogen) and were visualized under UV in a gel-documentation system (BIO-RAD Molecular Imager Gel Doc XR+).

Sequencing of full-length BBTV genome

The full-length PCR-amplified products for all six segments were eluted from gel using Gen elute gel extraction kit (Sigma) according to manufacturer's protocol and were sequenced at Bioserve Biotechnologies (I) Pvt. Ltd, Hyderabad, using both forward and reverse primers. The generated sequences were assembled in Codon Code Aligner 4.2.7 to obtain the contigs. Although full-length amplicons were obtained by PCR, often, the full-length sequences could not be obtained due to capillary sequencing not being able to read through the first several bases downstream of the primer. Therefore, to get the full-length sequence of all the segments, walking primers were designed from both the ends of the generated sequences, whenever necessary (Online Resource 2). The PCR conditions used for walking primers were the same as described previously. Finally,

the full-length sequences of all six segments of BBTV-As-JOR isolate were obtained by joining the contigs. The final sequences were submitted to NCBI database with accession no. KX868958, KX868959, KX592198, KX592199, KX868960 and KX592200 for DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N, respectively.

Comparative analysis of all BBTV DNA-segments

Complete sequences of all the six genomic segments of the characterized virus isolate were analysed for sequence similarity and the possible ORFs were searched using BLASTn and ORF finder tool of NCBI (<http://www.ncbi.nlm.nih.gov/orffinder/>). The nucleotide sequences of the genomic components were further analysed for presence and location of TATA boxes, reported conserved regions (viz., CR-M and CR-SL) and variations along the sequences. The known coding regions (CDs) were found to be intact in all the six genomic components. Subsequently, the corresponding proteins encoded by the ORFs, their sizes and location of the poly-A tails were also obtained.

Multiple sequence alignment and phylogenetic analysis

To understand the putative evolutionary origin of BBTV-As-JOR isolate, multiple sequence alignment and phylogenetic analysis were performed with all the previously reported isolates. Those isolates were categorized into PIO and SEA group on the basis of DNA-R sequences. Therefore, the complete sequences of DNA-R available in NCBI database (<http://www.ncbi.nlm.nih.gov>) were retrieved including the isolates from both the PIO and the SEA groups. All the sequences were rearranged to start with the 'origin of replication' (TATTAC) as the first nucleotide. A total of 312 full-length sequences along with the complete sequence of the DNA-R of the BBTV-As-JOR were aligned with ClustalW algorithm in MEGA6 software. The tree was generated using neighbour-joining (NJ) method with 1000 bootstrap replicates and the results are presented in the circular unrooted form.

The other segments, viz., DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N were also evaluated for their genetic grouping through multiple sequence alignments followed by phylogenetic analyses. The complete sequences of each segment (DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N) were aligned with other 20 representative complete sequences selected from both the PIO and the SEA groups. Two ABTV (Abaca bunchy top virus) isolates were also included to distinguish out-groups in the phylogenetic analyses. The phylogenetic tree for all these segments were also generated using NJ method with 1000 bootstrap replicates. An online tool iTOL (interactive tree of life; Letunic and

Bork 2016) was used to refine the images of the generated phylogenetic trees.

Recombination analysis

Both inter- and intra-segment recombination events were analysed using the Recombination Detection Program v.4.79 (RDP4). To check the recombination breakpoints and the major and minor parents, the available methods, viz., RDP, GENCOVE, BOOTSCAN, MaxChi, Chimaera, SISSCAN and 3Seq, were used with their default settings (Martin et al. 2015). For inter-segment recombination analysis, only those isolates were selected which are available with all six segments in NCBI database. A total of 121 BBTV isolates were selected for inter-segment recombination analysis (Online Resource 3).

Results

Virus amplification and sequencing

The presence of BBTV in eight banana samples collected from Jorhat, Assam, showing symptoms of crowning, stunting and bunchy appearance was confirmed by PCR. Full-length amplifications (approx. 1 to 1.1 kb) were obtained with all the six primer-pairs specific for the six BBTV genome segments (Fig. 1). There was no amplification found in non-symptomatic plant samples with any of the BBTV-specific primer sets (Fig. 1). Because all the eight BBTV-positive samples were collected from the same site, so only one sample was subjected to complete sequencing of all the six DNA-segments using the segment-specific primers and other walking primers, whenever needed. Finally, sequences for all the six DNA-segments were obtained for one virus isolate which was tentatively named as BBTV-Assam-Jorhat (BBTV-As-JOR); the sequences have also been deposited in the GenBank with the accession numbers given above.

Sequence analysis of BBTV-As-JOR confirmed the presence of the two major conserved regions (CR-M and CR-SL), the TATA box and the poly-A signal in all the six segments. Each segment of BBTV-As-JOR contains single TATA box, except DNA-U3 segment, which has two TATA boxes, upstream of the coding sequence. Similarly, all the segments harbour single poly-A signals (AATAAA) except in DNA-N, which has two poly-A signals. The detailed genomic features of all the segments are summarized in Table 1.

Analysis of genomic features

The CR-M and CR-SL are the most conserved regions in all six BBTV DNA-segments (Burns et al. 1995). Both these

Fig. 1 PCR amplification of six genomic segments of BBTv from eight different samples from Assam. **a** DNA-R, **b** DNA-M, **c** DNA-U3, **d** DNA-C, **e** DNA-S and **f** DNA-N. 1–8: different samples col-

lected from Jorhat, Assam, *M* DNA size marker, *N* negative control (non-symptomatic banana plant)

Table 1 Characteristics of DNA segments of BBTv-As-JOR

DNA segment	Accession no.	Size (base pairs)	Predicated ORF size (location)	Predicated protein (no. of amino acids/ kDa)	Location of TATA box (exact sequence)	Location of poly-A tail signal	Size of CR-M (location)	Size of CR-SL (location)
DNA-R	KX868958	1111	861 (87–947)	Rep (286/33.58)	36–44 (CTA TAAATA)	926–931	65 (985–1049)	69 (1072–29)
DNA-U3	KX868959	1059	267 (128–394)	U3 (88/10.34)	95–103 (CAATAATTA) 241–249 (CTATAAATA)	891–896	55 (747–801)	53 (1035–28)
DNA-S	KX592198	1075	528 (197–724)	CP (175/20.13)	173–181 (CTATAAATA)	964–969	91 (770–860)	69 (1036–29)
DNA-M	KX592199	1046	354 (267–620)	MP (117/13.79)	246–254 (CTATATATA)	935–940	90 (742–831)	69 (1007–29)
DNA-C	KX868960	1019	486 (225–710)	Clink (161/18.93)	173–181 (CTATTTAAA)	930–935	92 (808–899)	69 (980–29)
DNA-N	KX592200	1089	465 (262–726)	NSP (154/17.37)	207–215 (CTATTAATA)	749–754 and 970–975	91 (786–876)	60 (1059–29)

(*Rep* replication initiation protein, *U3* unknown protein, *CP* coat protein, *MP* movement protein, *Clink* cell cycle link protein, *NSP* nuclear shut-tle protein)

regions and three iterated regions were present in BBTv-As-JOR as reported previously (Herrera-Valencia et al. 2006; Vishnoi et al. 2009). Generally, the CR-M region is around 90 nucleotides (nt) long; however, it may vary among isolates (Burns et al. 1994). In BBTv-As-JOR, the CR-M is 65 nt and 55 nt long in DNA-R and DNA-U3, respectively,

whereas in DNA-S, DNA-M, DNA-C and DNA-N, it is approx. 90 nt long. Analysing the CR-M region of all the segments, the ‘majority sequence’ was calculated (Fig. 2). Compared to the majority, DNA-R and DNA-U3 have deletions of 26 nt and 36 nt, respectively, in the initial part of the CR-M. The deletion in the CR-M of DNA-R was similar

			GC-rich region	
Majority:		AACATACAACACTATAAAATACAACACACATATAACAATGTACGGGTATCTGATTGGGCTATATAACCCCTTAA	GGGCCGAAGGCCCGT	
DNA R:	985	-----*****-*****T***C***GA-CT*****		1049
DNA U3:	747	-----*****A*-*****G*-*****C*****		801
DNA S:	770	*****G***G*****G**G***G*****A*T-CC***GA-CT*****		860
DNA M:	742	*****G***G***G*-***G*****A*T-CC***GA-CT*****		831
DNA C:	808	*****-*****-*****-*****-*****-*****-*****-*****		899
DNA N:	786	*****G***G*****G**G*-***G*****A**A*T*T*G**T*G*-*****C*****		876

Fig. 2 Multiple sequence alignment of the CR-M of each segment of BBTV-As-JOR. The majority nucleotides calculated are shown above the alignment. The deleted regions are shown as ‘-’ and identity of

the nucleotides to the consensus are denoted by ‘*’. The starting and ending nucleotide positions of CR-M with respect to the origin of replication are also denoted for each genomic segment

to that of BBTV-Lucknow as well as BBTV-India-Umiam. However, unlike in these isolates, the deletion in CR-M of DNA-U3 was unique. Meanwhile, other deletions were also found (viz., one, at position 32 of DNA-R, DNA-U3, DNA-M, DNA-C and DNA-N, and another, at position 74 of DNA-R, DNA-U3, DNA-S, DNA-M and DNA-N). Overall, a total of 47 nucleotides (none in DNA-C, three in DNA-U3, six in DNA-R, 12 in DNA-M, 13 each in DNA-S and DNA-N) were found substituted in the CR-M including all the six segments, and notably, most substitutions were identical. Earlier, substitution of 11 nt was reported in BBTV-Lucknow, and 44 nt in BBTV-India-Umiam (Vishnoi et al. 2009; Banerjee et al. 2014); however, BBTV-As-JOR did not show the same substitutions like either of them. Furthermore, a 15-nt GC-rich region (GGGCCGTAGGCCCGT; 82% GC) was found conserved in the CR-M of all the six segments with a single exception (Fig. 2).

In all the segments of BBTV-As-JOR, the stem-loop (CR-SL) was detected. It is 69 nt long in DNA-R, DNA-S, DNA-M and DNA-C; in DNA-U3 and DNA-N, it is 53 nt and 60 nt long, respectively. Our analysis revealed that the CR-SL of each segment has a conserved stem-loop (TAT TATTAC) and three iterated regions, viz., R, F1 and F2 (Fig. 3; Online Resource 4). In the CR-SL in all the six segments; the reported hairpin structure was also found intact. Moreover, in the DNA-M of BBTV-Lucknow, the first

nucleotide ‘G’ of the F1 iteron was replaced by ‘A’ which was also the case in DNA-S and DNA-C of BBTV-As-JOR. The absence of the R iteron in DNA-U3 also indicates similarity of our isolate with BBTV-Lucknow.

The ORFs in DNA-R, DNA-S and DNA-N showed 99% identity with corresponding ORFs of BBTV-Lucknow isolate, whereas in case of DNA-U3, DNA-M and DNA-C, 91%, 97% and 84% identity, respectively, were found with the same.

Multiple sequence alignment and phylogenetic analyses

Genetically, all the BBTV isolates were clustered into two different groups, namely, the SEA and PIO group. In our analyses, the genetic grouping on the basis of DNA-R sequences (retrieved from NCBI database) (<http://www.ncbi.nlm.nih.gov>), clearly indicates two independent clusters (PIO and SEA) of all the BBTV isolates reported worldwide with 100% bootstrap support. The clustering of the Indian isolates with the PIO group supports their independent evolution pattern from the other SEA group members. Considering the DNA-R, the current isolate was found clustered with the PIO group along with the published BBTV-Lucknow isolate (Vishnoi et al. 2009) and other unpublished Indian isolates, including the unpublished one from Assam

		R	Stem loop	F1	F2	
Majority:		ATSTCCG	GAGTTAGTGC	CCACGTAAGCGCTGGGGCT	TATTATTAC	CCCCAGCGCTCGGGACGGGACAT
DNA R:	1072	*****	*****	*****	*****	*****
DNA U3:	1035	-----	***-TG	*****	A	*****
DNA S:	1036	*****	*****	*****	A	*****
DNA M:	1007	*****	*****	*****	A	*****
DNA C:	980	*****	A	*****	A	*****
DNA N:	1059	*****	-----	*****	A*G	AC*****C*T*****

Fig. 3 Sequence alignment of the CR-SL of each segment of BBTV-As-JOR. The majority nucleotides calculated are shown above the alignment. The deleted regions are shown as ‘-’ and identity of the nucleotides to the consensus are denoted by ‘*’. The starting and end-

ing nucleotide position of CR-SL with respect to the origin of replication are also denoted for each genomic segment. The stem-loop region (TATTATTAC) and three iterated regions (R, F1, and F2) are shown in empty boxes

in the Genbank (KP876497) and, and the isolates from Kanpur (India; BBTV-Kanpur; DQ656118), Lucknow (India; BBTV-Lucknow; DQ256267) and Arunachal Pradesh (India; BBTV-AR4-gn; KP876491 and BBTV-AR7-gc; KP876490). Similarly, the evolutionary relationship of the other segments (DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N), based on nucleotide sequences, was also worked out and it was found that BBTV-As-JOR always grouped with the PIO group. In all those cases, the ABTV isolate lied as an out-group, as expected (Online Resource 5). As discussed previously, the isolate reported from Meghalaya (BBTV-India-Umiam) showed the most distinctiveness with the PIO group isolates, and showed closeness to the SEA group members, in phylogenetic analysis. It is known that the PIO group includes isolates reported from India, Tonga, Samoa, USA, Taiwan, Sri Lanka, Egypt, Pakistan, Australia, Fiji, Rwanda, Burundi, Malawi, Democratic Republic of Congo and Myanmar. On the other hand, the SEA group includes BBTV isolates of Philippines, Indonesia, Japan, Taiwan, Egypt, Bali, Vietnam and China (Yu et al. 2012).

Interestingly, in our analysis, it was found that among four isolates reported from Egypt, three clustered with the PIO group and the other clustered with the SEA group. Similarly, out of the 12 reported isolates from Taiwan, 11 isolates were found to cluster with the SEA group, while the remaining one clustered with the PIO group. The clustering pattern of BBTV-As-JOR indicates that it shares high similarity with other Indian isolates (Fig. 4).

Recombination analysis

The role of BBTV-As-JOR in creating divergence of other BBTV isolates, or vice versa, was analysed by inter- and intra-segment recombination analysis. The analysis of intra-segment recombination was accomplished with available complete sequences of each segment in database. The overview of all breakpoint distributions involving

Fig. 4 Phylogenetic tree of DNA-R of BBTV-As-Jorhat based on nucleotide sequence similarity with previously reported BBTV isolates of both PIO and SEA groups. DNA-R of BBTV-As-JOR is represented in orange coloured box. The isolates from the same country are represented with the same colours as indicated in far right

BBTV-As-JOR, either as recombinant or as parental isolate, is presented in Online Resource 6 and 7.

The intra-segment recombinant analysis detected a total of 15 recombination events among which two, involving the DNA-U3 and DNA-C of BBTV-As-JOR, were found significant. In DNA-U3, single recombination was detected involving BBTV isolate TOS93_TO-2010 (Tonga, KM607859) as major and BBTV isolate Q553_LK_1995 (Sri Lanka, KM607809) as minor parent. The recombination break point reveals that nucleotides (nt) between position 736 and 954 of BBTV isolate TOS93_TO-2010 was replaced by BBTV isolate Q553_LK_1995 (Online Resource 6). The DNA-C of our isolate (BBTV-As-JOR) was involved in three recombination events; however, only one event involving BBTV-As-JOR as recombinant was identified as significant. There, nt position 458 and 973 of BBTV-China (AY606085) was replaced by BBTV isolate Q529_4_CN_1990 (China, KM607099) leading to the evolution of DNA-C of the BBTV-As-JOR.

Meanwhile, the inter-segment recombination analysis was done with the complete genomes of 121 BBTV isolates reported (from both PIO and SEA group). The breakpoints identified, and the parent isolates are summarized in Online Resource 7. A total of 37 recombination breakpoints were identified involving DNA-R, DNA-U3 and DNA-N of BBTV-As-JOR in eight events; however, only the recombination events in DNA-U3 and DNA-N were noted significant where BBTV-As-JOR was involved as recombinant. For DNA-U3, BBTV isolate Haikou 2 (China, HQ616075) was found to be the major parent and BBTV isolate MP2_TW_1996 (Taiwan, KM607223) to be the minor parent with the breakpoint mapped between nt position 786 and 871. For DNA-N, a significant breakpoint at nt position 810–848 was identified involving BBTV isolate MS16_PH_2008 (Philippines, KM607371) and BBTV isolate Q278_TO_1989 (Tonga, KM607527) as the major and the minor parents, respectively (Fig. 5).

Both intra- and inter-segment recombination analyses for DNA-R were interesting, where a single isolate was found responsible for recombination in more than one isolate at various positions. We identified three recombination events involving BBTV-As-JOR as minor parent. The sequences at nt position 240–671, 106–1001 and 108–1001 of BBTV isolate MP2_TW_1996 were replaced by BBTV-As-JOR leading to creation of three isolates, viz., BBTV-Vietnam (Boun Ma Thout, AF416476), BBTV-Vietnam (Da Nang, AF416477) and BBTV-Vietnam (Ho Chi Minh City, AF416478), respectively. Six inter-segment recombination were detected in DNA-R; their breakpoints were 984–0, 986–1009 and 983–1007 involving BBTV isolate Kandy (Sri Lanka, JN250597) as major and BBTV-As-JOR (KX868958) as minor parent. This led to evolution of the recombinant BBTV isolate Q624_TW-1996 (Taiwan,

KM607105), BBTV isolate Q623_TW-1996 (Taiwan, KM607104), BBTV isolate 8_150510_EG_2010 (Egypt, KM607033), BBTV isolate 765_TW-1996 (Taiwan, KM607031), BBTV isolate Mvz-80 (Democratic Republic of Congo, KU759891) and BBTV isolate MP2_TW-1996 (Taiwan, KM607081). The bootstrap plot of significant intra-segment recombination (DNA-U3 and DNA-C) defines the interactions between the recombinants and their parents (Fig. 5).

Discussion

Several BBTV isolates have been reported from India, including one (BBTV-India-Umiam; Banerjee et al. 2014) from the northeastern region. Analysis of the full-length sequence of BBTV-As-JOR confirmed its similarity with BBTV isolates from the PIO group. BBTV-Lucknow (Vishnoi et al. 2009) was considered as the standard, as it was the first completely characterized Indian BBTV isolate. BBTV-As-JOR shares high sequence similarity with BBTV-Lucknow, but stands out in few features. The BBTV isolate reported from a close proximity (BBTV-India-Umiam from Meghalaya), however, shares less similarity with BBTV-As-JOR. It is noteworthy that although located closely, due to higher altitude, climate of Meghalaya is considerably different from that of Assam. The insect, *Pentalonia nigronervosa*, cannot efficiently transmit BBTV in higher altitude (Anhalt and Almeida 2008). Besides, as banana is vegetatively propagated, movement of germplasm could be another factor effecting variations in BBTV.

The CR-M and CR-SL are conserved motifs in BBTV genomes (Burns et al. 1995) that indicate their possible functional inevitability. Meanwhile, the conserved GC-rich region in BBTV isolates had been earlier identified as an essential rightward promoter element in *Wheat dwarf virus* (Burns et al. 1995). Although similar studies have not been done in BBTV, the occurrence of this region also could indicate its similar functionality in BBTV (including BBTV-As-JOR). The presence of the highly conserved stem-loop in the CR-SL of BBTV-As-JOR indicates the origin of replication like in other nanoviruses (Vishnoi et al. 2009). Meanwhile, the three iterated regions (R, F1 and F2) have been reported as specific binding motifs for the Rep (Herrera-Valencia et al. 2006; Vishnoi et al. 2009).

The deletions found in the CR-M and CR-SL showed interesting patterns. First, most of the substitutions were identical in all the segments except DNA-C, which could indicate its different origin in BBTV-As-JOR. Second, compared to BBTV-Lucknow (11 nos) and BBTV-India-Umiam (44 nos), more substitutions (47 nos) were found in BBTV-As-JOR. Still, BBTV-As-JOR did not show biased inclination towards either of them. Meanwhile, though the deletions

Fig. 5 Intra-segment analysis of DNA-U3 and DNA-C. **a** Bootstrap support-plot showing the recombination events in DNA-U3 of BBTV-As-JOR between positions 736 and 972. **b** Bootstrap support-plot showing the recombination event in DNA-C between positions 458 and 973. The red highlighted box (on the X-axis) shows the break-

points of both the recombination events. The left and right boundaries of the highlighted box represent the start and the end breakpoints, respectively. The interactions between recombinant-major parents, recombinant-minor parents and major-minor parents are indicated with different colours

in the CR-M of DNA-R were similar in BBTV-As-JOR as well as in both the references, DNA-U3 had a unique 36-nt deletion. Despite such differences, other genomic features shown to be important for functionality have been found intact in BBTV-As-JOR. The ORFs of the current isolate also showed high similarity with the corresponding ORFs

of BBTV-Lucknow (Vishnoi et al. 2009) indicating their closeness.

The genetic diversity of BBTV is determined based on nucleotide sequence similarity of DNA-R that produces the SEA and the PIO group (Karan et al. 1994). BBTV-As-JOR clustered with the PIO isolates and showed high similarity

with few earlier isolates from India. BBTv isolates from India are known members of this group (Wardlaw 1972); this group also includes BBTv isolates from Polynesia, Asia and Africa. Interestingly, in our analysis, several isolates did not follow the expected pattern. However, BBTv-As-JOR, as expected, clustered with the PIO group. BBTv-India-Umiam (Banerjee et al. 2014) also clustered in this group although these two did not share the highest similarity. Again, the vegetative mode of banana propagation, and the distinct geoclimatic condition of Assam and Meghalaya could certainly be responsible for such variation. It was suggested that BBTv was introduced to southern India (then, to the rest of India) from Australia via Fiji and Sri Lanka by infected propagules (Wardlaw 1972); this also justifies the observed clustering patterns.

Recombination plays a major role in creating variation among viruses and leads to evolution of new variants. Recombination analysis among BBTv isolates was performed to understand their evolution, based on their nucleotide sequences. The intra-segment analysis revealed that only DNA-U3 and DNA-N each have one significant recombination involving BBTv-As-JOR. The DNA-U3 of BBTv is the most recombinogenic segment (Hyder et al. 2011; Stantion et al. 2012). We detected single recombination event in DNA-U3 of BBTv-As-JOR, involving KM607859-Tonga isolate as the major and KM607809-Sri Lanka isolate as the minor parent. But, in DNA-C, two BBTv isolates from China, AY606085 and KM607099, were identified as major and minor parents, respectively. Another interesting outcome was that DNA-R of BBTv-As-JOR was found responsible for recombination breakpoints in more than one BBTv isolates from Vietnam which are members of the SEA group. On the other hand, inter-segment recombination analysis identified the events for DNA-U3 and DNA-N to be significant.

The identified breakpoints by both intra- and inter-segment analyses, did not include BBTv-Lucknow, although the genomic features, sequence alignment and phylogeny of BBTv-As-JOR revealed high sequence similarity with it. This may be because of more number of isolates included in the recombination analysis leading to increasing diversity among the BBTv isolates. The CR-M was reported as major recombination breakpoint hotspot and CR-SL as a minor hotspot (Stantion et al. 2012). Furthermore, the intra-segment recombination breakpoints identified in DNA-R mostly covered the genome of the recombinants; but in case of inter-segment analysis, the recombined region resides in CR-M only. The detected intra-segment recombination events involved isolates from both PIO and SEA group; this suggests involvement of diverse BBTv isolates in the evolution of BBTv-As-JOR.

We put forward BBTv-As-JOR as the first reported and completely characterized BBTv isolate from the

northeastern Indian state, Assam. Although, the sequence of DNA-R of another isolate As1-gc (KP876497) exists; it is uncharacterized and unpublished. The presence of a distinct isolate, BBTv-India-Umiam, of the PIO group, in a neighbouring but geo-climatically different state (Meghalaya), supports the hypothesis of several isolates existing in the region and evolutionary lineage of BBTv coming from both the PIO and the SEO groups.

Acknowledgements This work was funded by Department of Science and Technology, Govt of India (GoI; grant number: SERB/F/5181/2012-13). Research in the Department of Agricultural Biotechnology (ABT), Assam Agricultural University (AAU) is also funded by Department of Biotechnology (DBT), GoI; Indian Council of Agricultural Research, GoI and DBT-AAU Centre. The authors thank the Distributed Information Centre, ABT, for providing computational facility. GMB acknowledges University Grants Commission, GoI for providing Rajiv Gandhi National fellowship (RGNF-2015-17-SC-RAJ-6919). The authors thank Dr K.M. Bujarbaruah, Vice Chancellor and Dr B.K. Sarmah, DBT-AAU Centre for their suggestions at various stages.

Author contributions BKB and GMB conceptualized and designed the study; GMB, with the help of GB, performed the sample collection, wet lab experiments and in silico analysis. PDN and MKM helped BKB to design the study. GMB prepared and BKB finalized the manuscript. All the authors reviewed the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving animal and human participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Allen RN (1978) Epidemiological factors influencing the success of rouging for the control of bunchy top disease of bananas in New South Wales virus-like pathogen transmitted by the banana aphid (*Pentalonia nigronervosa*). Aust J Agri Res 29:535–544
- Anhalt MD, Almeida RPP (2008) Effect of temperature, vector life stage, and plant access period on transmission of *Banana bunchy top virus* to banana. Phytopathol 98:743–748
- Banerjee A, Roy S, Behere GT, Roy SS, Dutta SK, Ngachan SV (2014) Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. Virus Res 183:41–49
- Beetham PR, Hafner GJ, Harding RM, Dale JL (1997) Two mRNAs are transcribed from banana bunchy top virus DNA-1. J Gen Virol 78:229–236
- Beetham PR, Harding RM, Dale JL (1999) Banana bunchy top virus DNA-2–6 are monocistronic. Arch Virol 144:89–105
- Burns TM, Harding RM, Dale JL (1994) Evidence that banana bunchy top virus has a multiple component genome. Arch Virol 137:371–380

- Burns TM, Harding RM, Dale JL (1995) The genome organization of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76:1471–1482
- Dale JL (1987) Banana bunchy top: an economically important tropical plant virus disease. *Adv Virus Res* 33:301–305
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13
- Harding RM, Burns TM, Dale JL (1991) Virus-like particles associated with banana bunchy top disease contain small single stranded DNA. *J Gen Virol* 72:225–230
- Harding RM, Burns TM, Hafner GJ, Dietzgen RG, Dale JL (1993) Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J Gen Virol* 74:323–328
- Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL (2006) An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87:3409–3412
- Horser CL, Karan M, Harding RM, Dale JL (2001) Additional re-encoding DNAs associated with banana bunchy top virus. *Arch Virol* 146:71–86
- Hyder MZ, Shah SH, Hameed S, Naqvi SMS (2011) Evidence of recombination in the banana bunchy top virus genome. *Infect Genet Evol* 11:1293–1300
- Islam MN, Naqvi AR, Jan AT, Mohd Q, Haq R (2010) Genetic diversity and possible evidence of recombination among *Banana Bunchy Top Virus* (BBTV) isolates. *Int Res J Microbio* 1(1):001–012
- Kagy V, Thomas JE, Sharman M, Mademba-Sy F (2001) First record of banana bunchy top disease in New Caledonia. *Aust Plant Pathol* 30:71
- Karan M, Harding RM, Dale JL (1994) Evidence for two groups of banana bunchy top virus isolates. *J Gen Virol* 75:3541–3546
- Karan M, Harding RM, Dale JL (1997) Association of banana bunchy top virus DNA components 2–6 with bunchy top disease. *Mol Plant Pathol On-line*:1–16
- Kumar PL, Hanna R, Alabi OJ, Soko MM, Oben TT, Vangu GH, Naidu RA (2011) Banana bunchy top virus in sub-Saharan Africa: investigations on virus distribution and diversity. *Virus Res* 159:171–182
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1):W242–W245
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:ev003
- NCBI database. National center for biotechnology information, U.S. National library of medicine. <http://www.ncbi.nlm.nih.gov>. Accessed on 16 Aug 2017
- ORF finder tool. National Center for Biotechnology Information, U.S. National Library of Medicine. <http://www.ncbi.nlm.nih.gov>. Accessed on 15 Aug 2017
- Stainton D, Krabberger S, Walters M, Wiltshire EJ, Rosario K, Halafih M, Lolohea S, Katoa I, Faitua TH, Aholelei W, Taufa L, Thomas JE, Collings DA, Martin DP, Varsani A (2012) Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* 93:1103–1119
- Thomas JE, Iskra-Caruana ML, Jones DR (2000) Diseases caused by viruses bunchy top. In: Jones DR (ed), *Disease of Banana, Abaca and Enset*. CABI Publishing, Wallingford 241–283
- Tian E, Zhuang Liu ZX (2005) Cloning and sequencing of DNA components of Banana bunchy top virus Hainan isolate. *Chinese J Agri Biotech* 2(2):91–97
- Vishnoi R, Raj SK, Prasad V (2009) Molecular characterization of an Indian isolate of banana bunchy top virus based on six genomic DNA components. *Virus Genes* 38(2):334–344
- Wanitchakorn R, Harding RM, Dale JL (1997) Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch Virol* 142:1673–1680
- Wanitchakorn R, Hafner GJ, Harding RM, Dale JL (2000a) Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *J Gen Virol* 81:299–306
- Wanitchakorn R, Harding RM, Dale JL (2000b) Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch Virol* 145:593–602
- Wardlaw CW (1972) The virus diseases: bunchy top. Mosaic, infectious chlorosis and other virus diseases. In: Wardlaw CW (ed) *Banana diseases including plantains and abaca*, 2nd edn. Longman, London, pp 68–145
- Wu RY, Su HJ (1990) Purification and characterization of banana bunchy top virus. *J Phytopathol* 128:153–160
- Yu NT, Zhang YL, Feng TC, Wang JH, Kulye M, Yang WJ, Lin ZS, Xiong Z, Liu ZX (2012) Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* 44:488–494