



# A New Method to Obtain the Complete Genome Sequence of Multiple-Component Circular ssDNA Viruses by Transcriptome Analysis

Nai-tong Yu<sup>1,2\*</sup>, Yu-liang Zhang<sup>1</sup>, Jian-hua Wang<sup>1,2</sup> and Zhi-xin Liu<sup>1,2\*</sup>

<sup>1</sup> Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, China, <sup>2</sup> Hainan Key Laboratory of Tropical Microbe Resources, Haikou, China

## OPEN ACCESS

### Edited by:

Ekaterina Shelest,  
German Centre for Integrative  
Biodiversity Research (iDiv), Germany

### Reviewed by:

Arthur Gruber,  
University of São Paulo, Brazil  
Renata Carmona Ferreira,  
Federal University of São Paulo, Brazil

### \*Correspondence:

Nai-tong Yu  
yunaitong@163.com  
Zhi-xin Liu  
liuzhixin@itbb.org.cn

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Bioengineering and  
Biotechnology

Received: 24 February 2020

Accepted: 29 June 2020

Published: 21 July 2020

### Citation:

Yu N, Zhang Y, Wang J and Liu Z  
(2020) A New Method to Obtain  
the Complete Genome Sequence  
of Multiple-Component Circular  
ssDNA Viruses by Transcriptome  
Analysis.  
Front. Bioeng. Biotechnol. 8:832.  
doi: 10.3389/fbioe.2020.00832

Circular single-stranded DNA (ssDNA) viruses are widely distributed globally, infecting diverse hosts ranging from bacteria, archaea, and eukaryotes. Among these, the genome of *Banana bunchy top virus* (BBTV) comprises at least six circular, ssDNA components that are ~1 kb in length. Its genome is usually amplified and obtained at the DNA level. However, RNA-based techniques to obtain the genome sequence of such multi-component viruses have not been reported. In this study, transcriptome sequencing analysis showed that the full-length of BBTV each genomic component was transcribed into viral mRNA (vmRNA). Accordingly, the near-complete genome of BBTV B2 isolate was assembled using transcriptome sequencing data from virus-infected banana leaves. Assembly analysis of BBTV-derived reads indicated that the full-length sequences were obtained for DNA-R, DNA-U3, DNA-S, DNA-M, DNA-N, News2, and Sat4 components, while two gaps (73 and 25 nt) missing in the DNA-C component which was further filled by reverse transcription-PCR (RT-PCR). The RT-qPCR analysis indicated that the vmRNA levels of coding regions were 3.19–103.53 folds higher than those of non-coding regions, implying that the integrity of genome assembly depended on the transcription level of non-coding region. In conclusion, this study proposes a new approach to obtain the genome of nanovirids, and provides insights for studying the transcriptional mechanism of the family *Nanoviridae*, *Genomoviridae*, and *Geminiviridae*.

**Keywords:** multi-component, DNA virus, nanovirids, transcriptomic sequencing, genome assembly

## INTRODUCTION

Circular single-stranded DNA (ssDNA) viruses are widely distributed worldwide, infecting diverse hosts from all three domains of bacteria, archaea, and eukaryotes (Lefkowitz et al., 2008). Currently, 11 families of ssDNA viruses have been established by the International Committee on Taxonomy of Viruses (ICTV, 2019 Release). These viruses are *Microviridae* and *Inoviridae* infecting bacteria, *Bacilladnaviridae* infecting diatoms, *Pleolipoviridae* and *Spiraviridae* infecting archaea, *Anelloviridae*, *Geminiviridae* and *Nanoviridae* infecting plants, *Circoviridae* infecting birds and mammals, *Genomoviridae* infecting plants and animals, and *Smacoviridae* that infects both animals and humans (Malathi and Dasgupta, 2019; Venkataraman and Selvarajan, 2019;

Li et al., 2020). In addition to the aforementioned viruses, the satellite ssDNA molecules of nanovirids, and geminiviruses, which are not necessary for infection, have been reclassified into *Alphasatellitidae* and *Tolecusatellitidae* families (Zubair et al., 2017; Briddon et al., 2018).

Viruses that are members of the families *Nanoviridae*, *Genomoviridae*, and *Geminiviridae* have small genomes that replicate ssDNA using the rolling circle replication mechanism (Schmid et al., 2014; Weitzman and Fradet-Turcotte, 2018; Li et al., 2020). The family *Nanoviridae* comprises plant viruses that have very small virions with six to eight components, ranging in size from 1.0 to 1.1 kb; each component is separately encapsulated inside an individual virion (Yu et al., 2019b). *Banana bunchy top virus* (BBTV), belonging to the genus *Babuvirus* within family *Nanoviridae*, is known to infect all members of family *Musaceae* and is transmitted by aphids of the genus *Pentalonia* in a circulative manner (Watanabe et al., 2013, 2016; Kumar et al., 2015). The genome of BBTV contains six components, namely, DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, and DNA-N. Except DNA-U3, each component encodes a protein with known function. Some isolates may carry 1–3 distinct satellite DNA molecules (Yu et al., 2011). The replication initiator protein (Rep), also referred to as the master Rep encoded by DNA-R, is known to be solely necessary and sufficient for replicating all genomic components, whereas the replication initiator assistant protein (RepA) encoded by satellite DNA molecules cannot replicate any genomic components except its own DNA sequence (Venkataraman and Selvarajan, 2019). DNA-S encodes capsid protein (CP), which forms isometric particles, and aids in genome packaging (Ji et al., 2019). DNA-C, -N, and -M encode the cell-cycle link (Clink) protein, nuclear shuttle protein (NSP), and movement protein (MP), respectively, (Amin et al., 2011; Zhuang et al., 2019; Yu et al., 2019b). The function of DNA-U3 remains unknown for BBTV and other nanovirids. Rep and CP are evolutionarily more conserved than the other proteins and thus serve as useful markers for the identification and classification of ssDNA viruses (Yu et al., 2012). Furthermore, evolutionary analysis of concatenated amino acid sequences of Rep and CP is representative for ssDNA virus classification.

Considering the typical circular ssDNA of *Nanoviridae*, characteristic of multiple-component viruses, their complete genome can be directly obtained at the DNA level (Yu et al., 2012; Mukwa et al., 2016). In addition, rolling circle amplification (RCA) technology has also been applied to amplify the circular ssDNA genome in recent years (Jeske, 2018). However, these methods present some drawbacks for obtaining the viral genomes with two or more different isolates or reassortment components of the viral genome. Meanwhile, no RNA-based techniques and methods for obtaining the ssDNA virus genome have been reported so far. In this study, we explored a novel approach to obtain the complete genome of nanovirids based on transcribed vmRNA, which will extend the knowledge of transcriptional activity of the families *Nanoviridae*, *Genomoviridae*, and *Geminiviridae*.

## MATERIALS AND METHODS

### Plant Samples

Samples of BBTV-infected banana leaves (B2) and healthy banana leaves (H4) were collected from Haikou, Hainan, China in October, 2013. Banana, the main planting line in Hainan province, is a species of *Musa* AAA Cavendish subgroup cv. Brazil. The BBTV-infected banana plant showed bunchy top with narrow and yellowing leaves. No symptoms were observed in the healthy banana sample. All banana leaves were collected and stored at  $-80^{\circ}\text{C}$  in the laboratory.

### RNA Extraction, Library Preparation, and Transcriptome Sequencing

Total RNA extraction, cDNA library construction, and transcriptome sequencing of banana leaves were performed as described previously (Yu et al., 2019b). The high-quality of total RNA samples from banana leaves ( $\text{CONC} \geq 108 \text{ ng}/\mu\text{L}$ ;  $\text{OD}_{260/280} = 2.01 \sim 2.12$ ; and  $\text{RIN} \geq 6.6$ ) were obtained and sequencing libraries were generated by using NEBNext<sup>®</sup> Ultra<sup>™</sup> RNA Library Prep Kit. Accordingly, 100-bp paired-end reads were generated on an Illumina HiSeq 2000 platform. All downstream transcriptome analyses were based on high-quality clean data (Yu et al., 2019b).

### BBTV Reads Mapping and Genome Assembly

Reference BBTV genomes and gene model annotation files were downloaded from the National Center for Biotechnology Information (NCBI) website<sup>1</sup>. The dataset of downloaded BBTV genomes was indexed using Bowtie v2.0.6 and paired-end clean reads were aligned to the reference genomes using TopHat v2.0.9 (Kim et al., 2013). HTSeq v0.5.4p3 was used to count the number of viral reads that mapped to each component of the BBTV. RPKM values, Reads Per Kilobase of the exon model per Million mapped reads, a calculation of sequencing depth and gene length for read counts, were used for estimating gene expression levels as described (Mortazavi et al., 2008). In this study, the RPKM of BBTV each genomic component was calculated based on the length and reads numbers. The mapped BBTV reads were initially

<sup>1</sup><https://www.ncbi.nlm.nih.gov/nuccore/?term=banana%20bunchy%20top%20virus>

**TABLE 1** | Reads coverage and BBTV genomic components of B2 sample.

Isolate	Component	GenBank nos.	ORF (bp)	Read Counts	RPKM
B2	DNA-R	MG545610	104–964	1022	42961.81
	DNA-U3	MG545611	143–409	4164	564461.46
	DNA-S	MG545612	227–742	8956	631875.55
	DNA-M	MG545613	282–32	1085	111881.29
	DNA-C	MG545614	239–724	44	3276.81
	DNA-N	MG545615	277–741	10953	852540.40
	NewS2	MG545617	62–919	1007	42628.32
	Sat4	MG545616	49–903	398	16789.22

assembled into contigs in TopHat v2.0.9 and were further used to identify the possible BBTV genotypes in CodonCode Aligner (CodonCode, Centerville, MA).

## Complete Genomic Assembly of BBTV

Assembly analysis of the RNAseq reads, derived from the BBTV genome, indicated that the full-length of BBTV DNA-R, DNA-U3, DNA-S, DNA-M, DNA-N, NewS2, and Sat4 components were transcribed into viral mRNA (vmRNA), but not for the DNA-C component. To further reveal the full-length transcription of BBTV DNA-C component, the two primer pairs of DNA-C 61F/63R and DNA-C 903F/904R were designed based on the two gap regions (**Supplementary Table 1**). Following RT-PCR amplification, the fragments were cloned into the pMD19-T vector (Takara, Beijing, China), and then transferred into *Escherichia coli* DH5 $\alpha$  competent cells (2nd Lab, Shanghai, China). Three positive clones were randomly selected and sent to Invitrogen (Guangdong, China) for Sanger sequencing.

Subsequently, the complete genome sequence of BBTV B2 was assembled from the RNA level. To further verify the accuracy of the assembled BBTV B2 genome, full-length nucleotide sequences of the eight genomic components of BBTV B2 were amplified using full-length specific primers (**Supplementary Table 1**). Total DNA of B2 and H4 was extracted by using DNasecure Plant Kit (TIANGEN Biotech, Beijing). PCR amplification was performed as described previously (Yu et al., 2012).

## Phylogenetic Analysis of BBTV B2 Isolate

Respective amino acid (aa) sequences of Rep and CP from the Pacific-Indian Oceans group (PIO) and Southeast Asian group (SEA) were downloaded from the GenBank database in NCBI (**Supplementary Table 2**). The Rep and CP protein sequences of each BBTV isolate were concatenated sequentially, and multiple alignments of 14 sequences were performed using ClustalW in BioEdit (version 7.0.9.0). To study the relationship between BBTV B2 isolate and BBTV isolates from other parts of the world, a phylogenetic tree was constructed based on the concatenated amino acid sequences. Evolutionary analysis was conducted in MEGA6 and the evolutionary history was inferred using the Maximum Likelihood method based on the JTT matrix-based model (Tamura et al., 2013); *Abaca bunchy top virus* (ABTV) was used as an outgroup (Sharman et al., 2008).

## Verification of Full-Length Transcription of BBTV Genomic Components

To determine whether the full-length of BBTV genomic components was transcribed into vmRNA, RT-PCR was performed to identify the BBTV transcripts from the transcriptome sequencing data. In this study, the ORF vmRNA and vmRNA over the full-length of DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, DNA-N, NewS2, and Sat4 components were used to verify the BBTV genomic components transcription (**Supplementary Table 1**). The BBTV B2 cDNA (after removing DNA contamination) was used as a template, and PCR amplification was performed as described above. Total RNA

and total DNA of BBTV B2 were used as negative and positive controls, respectively.

## Transcripts Levels Between Coding and Non-coding Regions of BBTV Genomic Components

Total RNAs of B2 leaves were extracted and subjected to cDNA synthesis as described above. To measure the transcription levels between coding and non-coding regions of BBTV genomic components, standard curves of each BBTV DNA component were constructed as Yu et al. (2019b). The eight standard plasmids were prepared previously (Yu et al., 2019a). Quantitative PCR (qPCR) was used to measure the vmRNA amounts between the coding and non-coding regions of BBTV genomic components including DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, DNA-N, NewS2, and Sat4. The qPCR was carried out using the SYBR green real-time PCR master mix reagents kit (CW BIO, China) on the StepOne real-time PCR system (Applied Biosystems, United States) with cycle conditions of 95°C for 5 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. The vmRNA level between coding and non-coding regions of BBTV genomic components in per  $\mu$  g of banana leaf was measured three times independently with qPCR using the first stranded cDNAs as templates. Student's *t*-test was used to evaluate the differences.

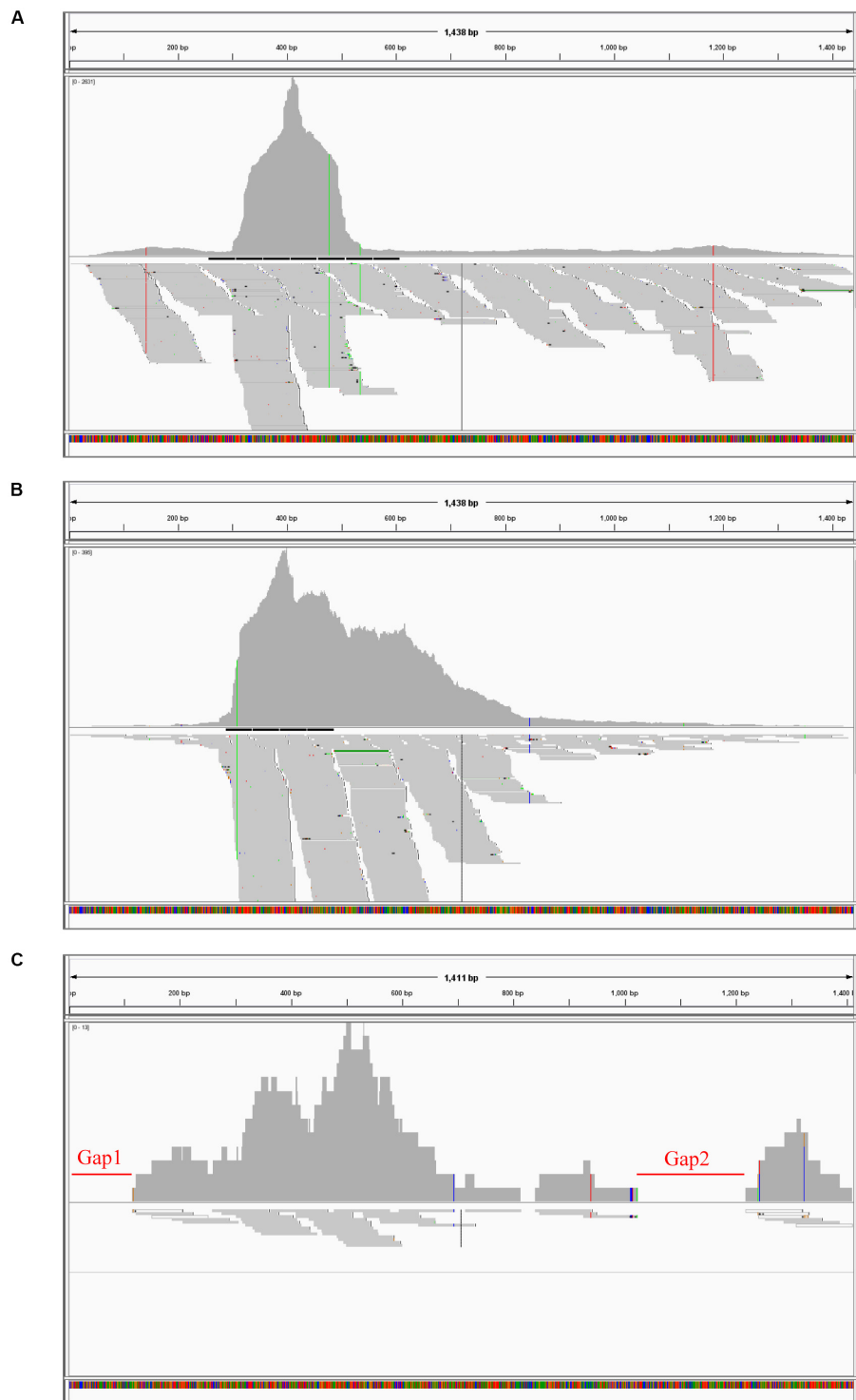
## RESULTS

### BBTV Genome Assembly From Transcriptome Sequencing Data

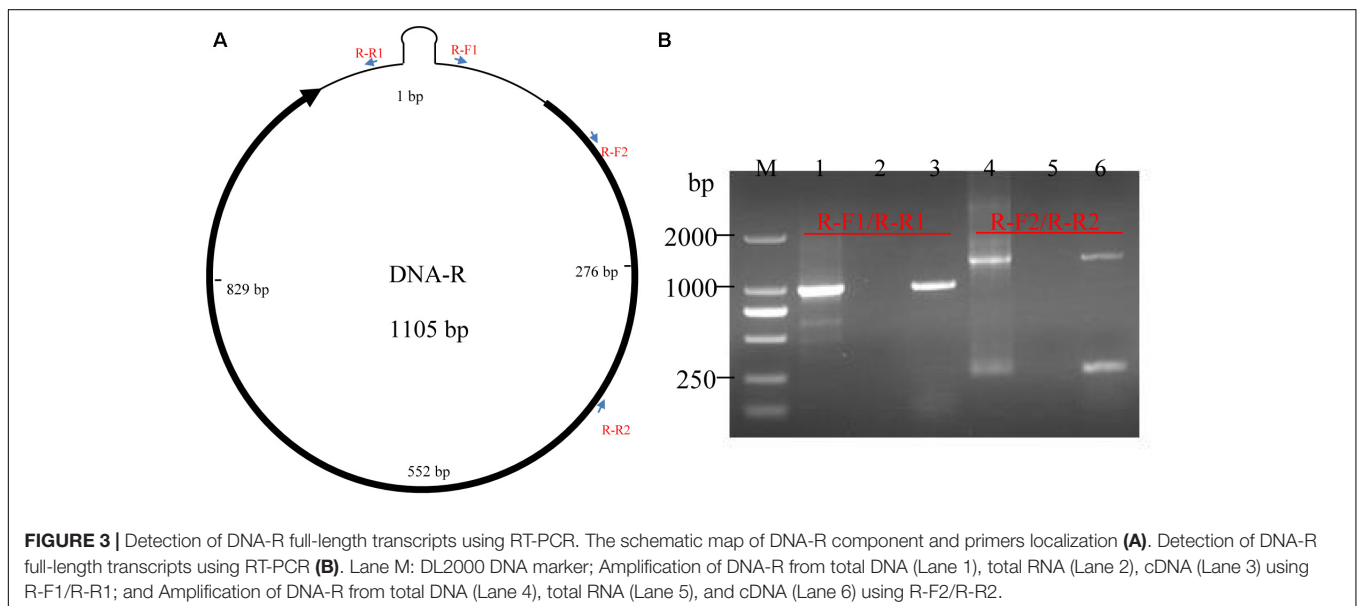
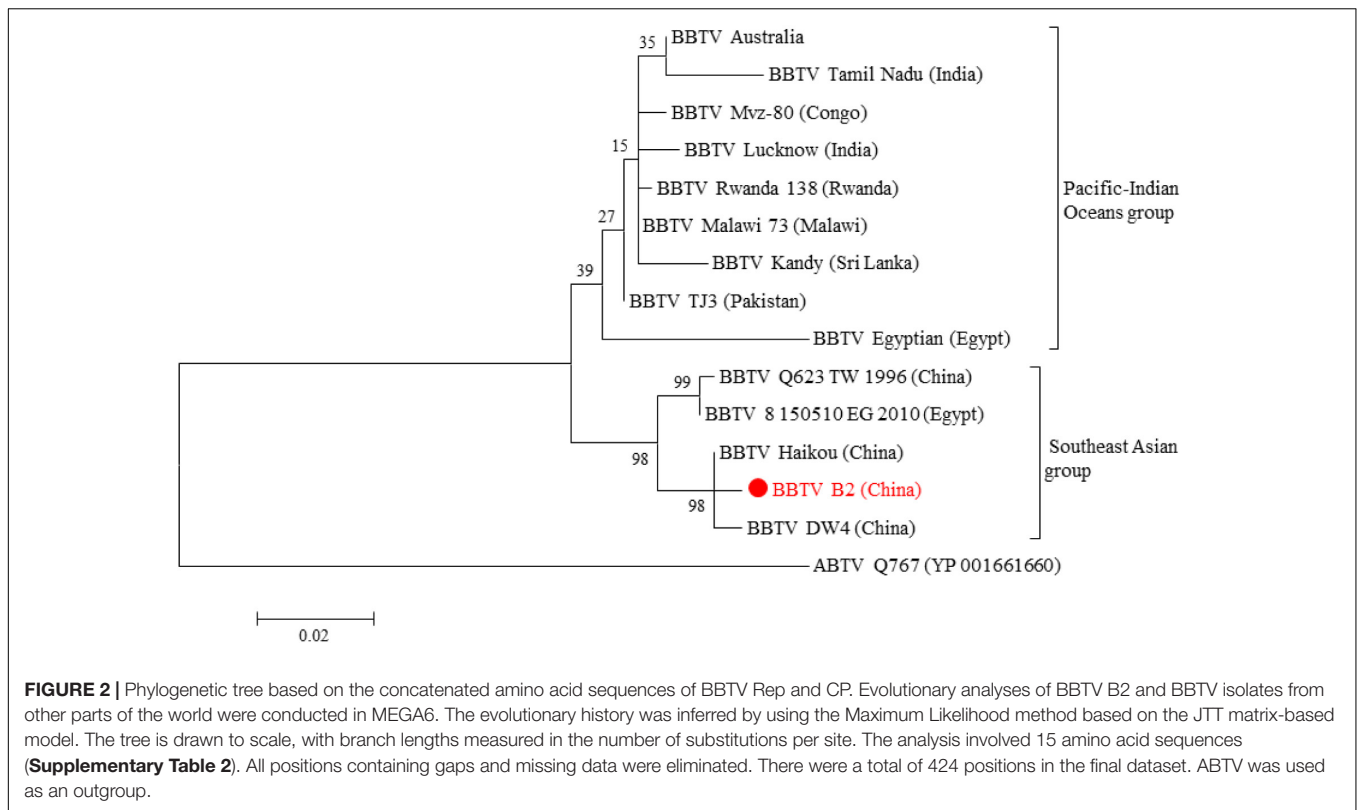
The raw data of B2 and H4 are available in the NCBI database under accession number SRP129855 (Yu et al., 2019b). In total, 59.28 and 56.13 million raw reads were generated from B2 and H4 banana leaves, respectively, and then, 56.22 and 53.34 million clean reads were generated from these samples (**Supplementary Table 3**).

The genome of the BBTV B2 isolate was assembled from transcriptome sequencing data. In detail, contigs were assembled by mapping the reads onto the downloaded BBTV genomes, and these assembled contigs were used to search possible viruses and to remove banana genome sequences as much as possible. The clustering contigs were further used to assemble and identify the BBTV genotypes in CodonCode Aligner. The result showed that the full-lengths of DNA-R, DNA-U3, DNA-S, DNA-M, and DNA-N components of BBTV B2 were assembled from the BBTV transcribed reads, but not for the DNA-C component. Further analysis indicated that the sequences at 48–121 nt (gap1) and 814–839 nt (gap2) of DNA-C were not covered. Interestingly, the full-lengths of two satellite DNA components, designated as NewS2 and Sat4, were also obtained (**Table 1**).

To further clarify the BBTV assembly mechanism, the transcribed vmRNAs in the untranslated region (UTR) and the open reading frame (ORF) of three representative components (DNA-U3, DNA-M, and DNA-C) are analyzed in CodonCode



**FIGURE 1** | BBTv transcripts from the transcriptome sequencing data and mapping profiles of three representative components in BBTv B2 isolate. BBTv-derived reads and mapping profiles for DNA-U3 (A), DNA-M (B), and DNA-C (C). The transcriptional reads were assembled by CodonCode Aligner version 3.7. Gap1 and gap2 represent nucleotide sequences with uncoverage regions.



Aligner. Briefly, a large number of vmRNA reads was transcribed from DNA-U3, DNA-M, and DNA-C ORFs, whereas a relatively low amount vmRNA reads was transcribed from their UTR regions, and some regions were not even covered, such as the sequences at 48–121 nt (gap1), and 814–839 nt (gap2) of the DNA-C (Figure 1). Therefore, except for the DNA-C component, these results showed that all DNA components from BBTV B2 were assembled, although the

amounts of the transcribed reads of UTRs and ORFs were in different levels.

### Genomic Verification of BBTV B2

In order to fill the two gap regions of the DNA-C component of BBTV B2, two RT-PCR amplifications were conducted. Sequencing of three random clones of amplified fragment confirmed the assembled vmRNA sequences and the two gaps in



the DNA-C component were filled. Subsequently, the complete genome sequence of BBTV B2 was obtained.

To further verify the accuracy of the BBTV B2 genome, the full-length nucleotide sequence of eight BBTV genomic components was amplified by regular PCR using Phusion High-Fidelity DNA Polymerase (NEB, United States), and the results confirmed the assembled vmRNA sequences of BBTV B2 genome. As shown in **Table 1**, there was only one BBTV genotype that was highly homologous to the BBTV Haikou isolate (Feng et al., 2010). Interestingly, its genome contained two satellite DNA components, NewS2, and Sat4. The assembled full-length nucleotide sequences of DNA-R (MG545610), DNA-U3 (MG545611), DNA-S (MG545612), DNA-M (MG545613), DNA-C (MG545614), DNA-N (MG545615), DNA-Sat4 (MG545616), and DNA-NewS2 (MG545617) from the BBTV B2 isolate were submitted to GenBank in NCBI.

### Phylogenetic Analysis of BBTV B2 Isolate

Genetically, all BBTV isolates were clustered into two different groups, the SEA and PIO, confirming the BBTV classification proposed by Yu et al. (2012). The phylogenetic analysis showed that BBTV B2 was closer to other BBTV members isolated from China such as the isolates DW4 and Haikou (**Figure 2**). Meanwhile, clustering of the BBTV B2 isolate with the SEA group supports its evolution pattern. Interestingly, our analysis showed that the two isolates reported from Egypt, one (BBTV Egyptian) clustered with the PIO group whereas the other (BBTV 8150510EG2010) clustered with the SEA group. In all the cases, the ABTV isolate was an outgroup, as expected.

### Verification of the Full-Length Transcription of BBTV Genomic Components

According to the RPKM values determined from the transcriptome sequencing data, three components of DNA-N (852540.40), DNA-S (631875.55), and DNA-U3 (564461.46) were transcribed at high levels, followed by DNA-M component (111881.29) with moderate transcriptional level. The DNA-R component (42961.81) was transcribed at a low level, whereas the DNA-C component (3276.81) was even lower, more than 250-fold less than that of DNA-N component (**Table 1**). The two gaps represented the lowest amount of viral reads in the DNA-C component, but the reads of these regions was still transcribed and the vmRNA were further confirmed by the RT-PCR. The two satellite DNA components, NewS2 (42628.32), and Sat4 (16789.22), were transcribed at low levels as well.

To verify the full-length transcription of other BBTV genomic components, the RT-PCR was performed on the vmRNA of DNA-R, DNA-U3, DNA-S, DNA-M, DNA-N, NewS2, and Sat4 components. The result showed that expected DNA bands (1.0–1.1 kb) were obtained, indicating the full-length transcription of these DNA components. In addition, two primer-pairs targeting the coding and non-coding region of DNA-R were designed (**Figure 3A**). The RT-PCR showed that a specific band containing the DNA-R ORF was successfully amplified by the R-F1/R1 primers, whereas two bands were amplified by the R-F2/R2

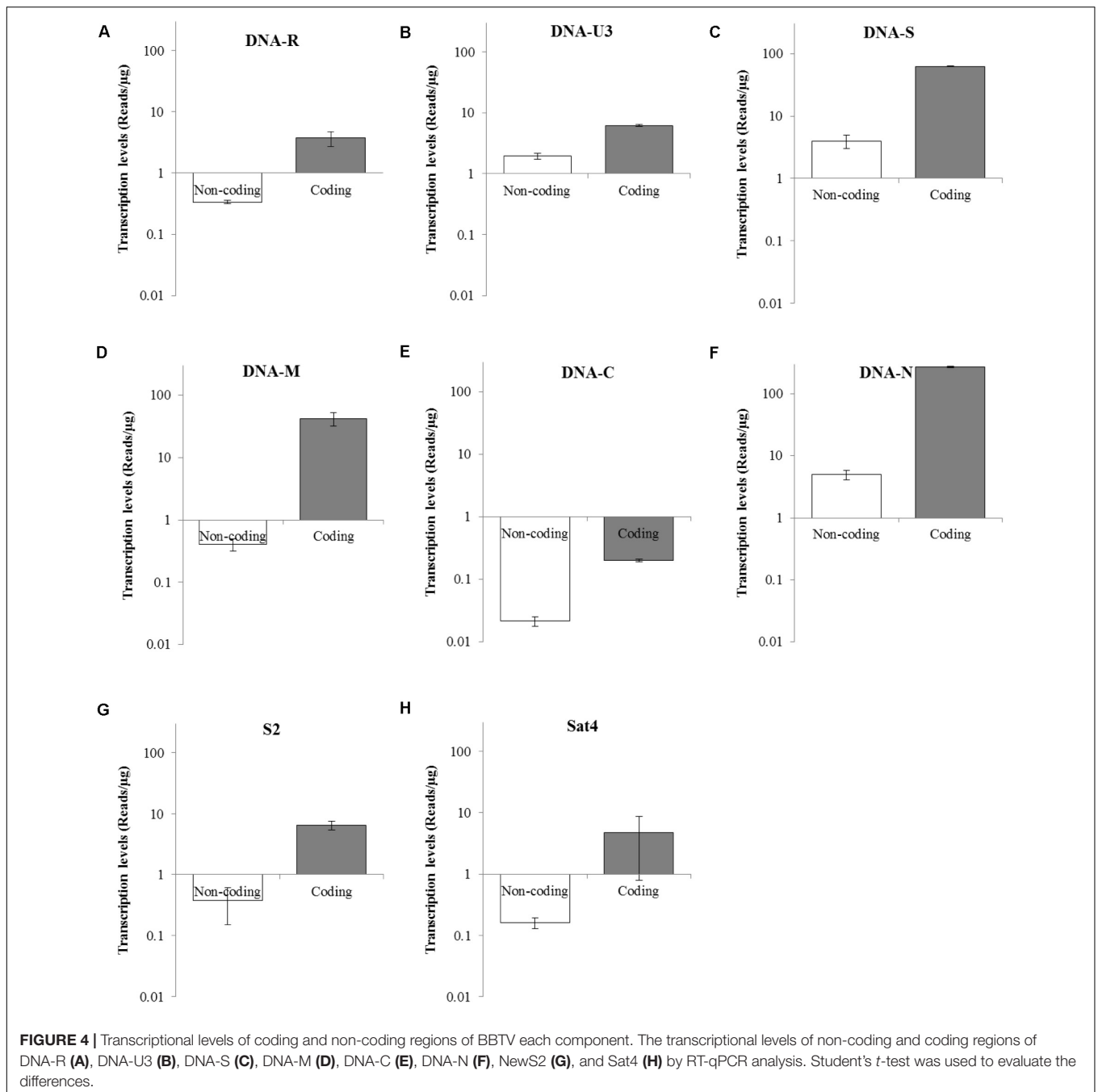
primers. To confirm the two bands amplified by R-F2/R2, the fragments were subjected to Sanger sequencing and the results indicated that the smaller DNA band was about 300 bp whereas the large band was ~1400 bp, which was over the full-length of the DNA-R component (**Figure 3B**). As controls, the PCR obtained from the total DNA sample was positive, whereas a negative result was obtained in the RNA sample, indicating the RNA was not contaminated with DNA. These results indicated that BBTV genomic components can be transcribed into vmRNA with over the full-length (one copy) of their circle DNA molecules, which can be further used to obtain the BBTV genome by transcriptome sequencing.

### Transcriptional Levels of Coding and Non-coding Regions of Each BBTV Genomic Component

To determine the transcriptional levels of the coding and non-coding regions of each BBTV genomic component, primer-pairs targeting specific regions were used for RT-qPCR (**Table 2**). In detail, the coding region of DNA-N was highly transcribed

**TABLE 2** | Primers used for RT-qPCR in this study.

Primer name	Sequences (5'–3')	Usage
DNA-R ncF	GGGAAAAGCAAAGATTCGGG	DNA-R
DNA-R ncR	AGTGTACGATTGGCATAGCG	RT-qPCR
DNA-R cF	GCACACCTTGAGAAACGAAAG	
DNA-R cR	TCAACTCTGCTTGCACTCTG	
DNA-U3 ncF	CCCCTAAGTAGCGTGATGATG	DNA-U3
DNA-U3 ncR	GACCATCTACCTTGACCTTCG	RT-qPCR
DNA-U3 cF	GCTCTCGCTCTTCTGTCAAAG	
DNA-U3 cR	AATACCCGATATACGCTTCGC	
DNA-S ncF	CATATGTCCCAGTAGTGCGG	DNA-S
DNA-S ncR	GCCCCAAAACCCATCTTAACG	RT-qPCR
DNA-S cF	AGGAAGTATGGAAGCAAGGC	
DNA-S cR	TGTAGTCGGCTGGTTGATTC	
DNA-M ncF	AGTTAGTGCGCCACGTAAG	DNA-M
DNA-M ncR	GATACGATAGACCAATCAACCCC	RT-qPCR
DNA-M cF	CAGAGCGGGTGAACAATTC	
DNA-M cR	TTCTGCATCCATACACGTCG	
DNA-C ncF	GGCCCGTTTAAATATGTGTTGG	DNA-C
DNA-C ncR	CACTAACTCGGGACATGGAC	RT-qPCR
DNA-C cF	TGGAATTCTGGGAATCGTCTG	
DNA-C cR	ATCCTTCTGACACAGCACTTC	
DNA-N ncF	CGGGACATGACGTAAGCATAG	DNA-N
DNA-N ncR	TTTATTCGGCAGAGACAGGC	RT-qPCR
DNA-N cF	TGGCTGTGATTGGAAGACG	
DNA-N cR	AGTACTGGACGCATTGTTC	
NewS2 ncF	CCTTACACCTCTGCCTTACAC	NewS2
NewS2 ncR	TGCCACCGACATCCTTATTG	RT-qPCR
NewS2 cF	TGGCGTCTCTAAATGGTG	
NewS2 cR	ATTCTGTTCGTCTGTTCTCTCT	
Sat4 ncF	CAACACCTTTAACCTCTGCG	Sat4
Sat4 ncR	TCCCCTTTTGTATAGCGTAC	RT-qPCR
Sat4 cF	TGGCAGGTACAATTAACGGAG	
Sat4 cR	CTCCCCCTGTGTAGAACAATC	



with reads number of 268.50 copies/ $\mu\text{g}$ , followed by the medium level of DNA-S (62.67 copies/ $\mu\text{g}$ ), and DNA-M (42.24 copies/ $\mu\text{g}$ ), whereas DNA-C was the lowest transcribed region (0.2 copies/ $\mu\text{g}$ ). For non-coding regions, DNA-N (4.99 copies/ $\mu\text{g}$ ), and DNA-S (3.99 copies/ $\mu\text{g}$ ) were transcribed at medium-low levels, but DNA-C had the lowest value of reads number at 0.02 copies/ $\mu\text{g}$  (Figure 4). Accordingly, the relative transcriptional level of coding region was 3.19–103.53 fold higher than that of non-coding region for each BBTV component. Therefore, the complete genome of BBTV assembly depended on the vmRNA content of non-coding regions, especially for

the DNA-C component. This further verified the full-length transcription of BBTV genomic components.

## DISCUSSION

The genomes of most ssDNA viruses, particularly nanovirids, genomoviruses, and geminiviruses, are obtained using DNA-based methods such as PCR and RCA (Bashir et al., 2012; Yu et al., 2012). In recent years, high-throughput DNA sequencing technology has also been successfully applied to obtain the

ssDNA virus genome (Dayaram et al., 2015; Gallet et al., 2017). However, RNA-based techniques to obtain the genome sequence of such multi-component viruses have not been reported. In this study, we explored a novel approach to obtain the complete genome of BBTV, a multiple-component circular ssDNA virus. The full-lengths of DNA-R, DNA-U3, DNA-S, DNA-M, DNA-N, NewS2, and Sat4 components of BBTV B2 were assembled from the Illumina reads, but the *de novo* assembly from the BBTV transcribed reads failed to obtain the full-length nucleotide sequences of DNA-C component. This is most likely due to the rare viral reads mapped onto the gapped regions. Therefore, the BBTV transcribed reads from the transcriptome sequencing data would not completely cover these regions. Similar results were observed in potyvirus, enamovirus, and nucleorhabdovirus (Cao et al., 2019; Yu et al., 2020). Considering the cost involved, we chose to fill the gap regions by RT-PCR. However, these results also indicated the whole genome of BBTV would likely be obtained by increasing the transcriptomic sequencing data coverage. To the best of our knowledge, this study is the first report using the transcriptome sequencing data to assemble the whole genome of BBTV.

Complete genome assembly using BBTV RNA reads revealed that a large amount of vmRNA in the coding region was transcribed in each component, whereas a relatively low amount of vmRNA was transcribed in the non-coding region (UTR). Although the assembled DNA-C component had two small gaps from the Illumina reads, RT-PCR was conducted to fill the gap regions and allowed to determine the full-length of DNA-C, implying that the integrity of the genome assembly depended on the transcriptional levels of non-coding regions, especially for the DNA-C component. The RT-qPCR further confirmed this result. Sequence analysis revealed that the poly(A) signal sequence was found downstream of the ORFs in each component, but the vmRNA from the UTR was still transcribed. This may be due to the incompletely terminated transcriptional regulation of BBTV and it is very common that the partial sequence of 3' UTR transcribes (Hilgers, 2015; Zhang et al., 2015). Based on this transcriptional feature, the whole genome of BBTV might be obtained by transcriptome high-throughput sequencing with sufficient deep-sequencing data.

For multi-component ssDNA viruses, there are some drawbacks to obtain the viral genomes by using the conventional PCR-based or RCA methods. Firstly, the plant may be infected with two or more different viral isolates, and thus, when using PCR amplification, primers may bind to non-predominant target sequences and amplify the non-dominant viral isolates. Secondly, as a multi-component virus, the obtained genome sequences may not represent an isolate or the sequence may come from different isolates. Thirdly, the inter-reassortment component from two BBTV isolates or intra-reassortment component in an isolate could not be recognized (Grigoras et al., 2014; Kraberger et al., 2018). Lastly, some BBTV isolates may contain 1–3 distinct satellite molecules. The nucleotide sequences of these satellite components have large differences, and degenerate primers often fail to obtain all the satellite molecules. However, the aforementioned disadvantages can be overcome using transcriptome high-throughput sequencing. In this study, there

was apparently only one BBTV genotype, BBTV B2, excluding the possibility of contamination by different isolates in the infection. We also identified a new satellite molecule, NewS2, which has not been reported yet.

In conclusion, transcriptome sequencing analysis showed that the full-length of BBTV genomic components was transcribed into vmRNA. Based on this characteristic, its complete genome could be assembled and the obtained sequences were identical to the DNA method. This study found the new characteristic of a full-length transcription rather than one copy of BBTV each component and provides a new approach to obtain the complete genome of BBTV, which will extend the knowledge of transcriptional mechanism of the nanovirids, genomoviruses and geminiviruses.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank MG545610–MG545617.

## ETHICS STATEMENT

The authors declare that ethical standards have been followed and that no human participants or animals were involved in this research.

## AUTHOR CONTRIBUTIONS

NY conceived and designed the research. YZ and JW collected the samples. NY and ZL analyzed and wrote the data. All authors contributed to the article and approved the submitted version.

## FUNDING

The study was funded by the National Key R&D Program of China (2019YFD1000500), the National Natural Science Foundation of China (31401709 and 31261140363), the Hainan Provincial Natural Science Foundation (No. 20153130), and Guangdong Key Laboratory of Tropical and Subtropical Fruit Tree Research, Institute of Fruit Tree Research, and Guangdong Academy of Agricultural Sciences (No. 2017B030314113).

## ACKNOWLEDGMENTS

We would like to thank Editage ([www.editage.com](http://www.editage.com)) for English language editing. We thank Dr. Zhongguo Xiong (University of Arizona) for analyzing the partial transcriptome sequencing data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.00832/full#supplementary-material>



## REFERENCES

- Amin, I., Ilyas, M., Qazi, J., Bashir, R., Yadav, J. S., Mansoor, S., et al. (2011). Identification of a major pathogenicity determinant and suppressors of RNA silencing encoded by a South Pacific isolate of *Banana bunchy top virus* originating from Pakistan. *Virus Genes* 42, 272–281. doi: 10.1007/s11262-010-0559-3
- Bashir, R., Javed, F., Ahmed, R., and Mansoor, S. (2012). Use of rolling circle amplification for the identification of unknown components of *Banana bunchy top virus* from Pakistan. *Pakistan J. Life Soc. Sci.* 10, 91–97.
- Bridson, R. W., Martin, D. P., Roumagnac, P., Navas-Castillo, J., Fiallo-Olivé, E., Moriones, E., et al. (2018). *Alphasatellitidae*: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Arch. Virol.* 163, 2587–2600. doi: 10.1007/s00705-018-3854-2
- Cao, M., Zhang, S., Li, M., Liu, Y., Dong, P., Li, S., et al. (2019). Discovery of four novel viruses associated with flower yellowing disease of green Sichuan pepper (*Zanthoxylum Armatum*) by virome analysis. *Viruses* 11:696. doi: 10.3390/v11080696
- Dayaram, A., Potter, K. A., Pailles, R., Marinov, M., Rosenstein, D. D., and Varsani, A. (2015). Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (*Insecta: Odonata*) of Arizona and Oklahoma, USA. *Infect. Genet. Evol.* 30, 278–287. doi: 10.1016/j.meegid.2014.12.037
- Feng, T., Wang, J., and Liu, Z. (2010). Cloning and sequencing of genome of *Banana bunchy top virus* Haikou isolate. *Acta Phytopathol. Sin.* 40, 40–50.
- Gallet, R., Fabre, F., Michalakakis, Y., and Blanc, S. (2017). The number of target molecules of the amplification step limits accuracy and sensitivity in ultra-deep-sequencing viral population studies. *J. Virol.* 91:e00561-17.
- Grigoras, I., Ginzo, A. I., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., et al. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J. Gen. Virol.* 95, 1178–1191. doi: 10.1099/vir.0.063115-0
- Hilgers, V. (2015). Alternative polyadenylation coupled to transcription initiation: insights from ELAV-mediated 3' UTR extension. *RNA Biol.* 12, 918–921. doi: 10.1080/15476286.2015.1060393
- Jeske, H. (2018). Barcoding of plant viruses with circular single-stranded DNA based on rolling circle amplification. *Viruses* 10:E469.
- Ji, X. L., Yu, N. T., Qu, L., Li, B. B., and Liu, Z. X. (2019). *Banana bunchy top virus* (BBTV) nuclear shuttle protein interacts and re-distributes BBTV coat protein in *Nicotiana benthamiana*. *3 Biotech* 9, 121.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36.
- Kraberger, S., Kumari, S. G., Najar, A., Stainton, D., Martin, D. P., and Varsani, A. (2018). Molecular characterization of *Faba bean necrotic yellows viruses* in Tunisia. *Arch. Virol.* 163, 687–694. doi: 10.1007/s00705-017-3651-3
- Kumar, P. L., Selvarajan, R., Iskra-Caruana, M. L., Chabannes, M., and Hanna, R. (2015). Chapter seven – biology, etiology, and control of virus diseases of banana and plantain. *Adv. Virus Res.* 91, 229–269. doi: 10.1016/bs.aivir.2014.10.006
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2008). Virus taxonomy: the database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Res.* 46, 708–717.
- Li, P., Wang, S., Zhang, L., Qiu, D., Zhou, X., and Guo, L. (2020). A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Sci. Adv.* 6:eaay9634. doi: 10.1126/sciadv.aay9634
- Malathi, V. G., and Dasgupta, I. (2019). Insights into the world of ssDNA viruses. *VirusDisease* 30, 1–2. doi: 10.1007/s13337-019-00529-2
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Mukwa, L. F., Gillis, A., Vanhese, V., Romay, G., Galzi, S., Laboureau, N., et al. (2016). Low genetic diversity of *Banana bunchy top virus*, with a sub-regional pattern of variation, in democratic republic of congo. *Virus Genes* 52, 900–905. doi: 10.1007/s11262-016-1383-1
- Schmid, M., Speiseder, T., Dobner, T., and Gonzalez, R. A. (2014). DNA virus replication compartments. *J. Virol.* 88, 1404–1420. doi: 10.1128/jvi.02046-13
- Sharman, M., Thomas, J. E., Skabo, S., and Holton, T. A. (2008). *Abacá bunchy top virus*, a new member of the genus *Babuvirus* (family *Nanoviridae*). *Arch. Virol.* 153, 135–147. doi: 10.1007/s00705-007-1077-z
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Venkataraman, S., and Selvarajan, R. (2019). Recent advances in understanding the replication initiator protein of the ssDNA plant viruses of the family *Nanoviridae*. *VirusDisease* 30, 22–31. doi: 10.1007/s13337-019-00514-9
- Watanabe, S., Borthakur, D., and Bressan, A. (2016). Localization of *Banana bunchy top virus* and cellular compartments in gut and salivary gland tissues of the aphid vector *Pentalonia nigronervosa*. *Insect Sci.* 23, 591–602. doi: 10.1111/1744-7917.12211
- Watanabe, S., Greenwell, A. M., and Bressan, A. (2013). Localization, concentration, and transmission efficiency of *Banana bunchy top virus* in four asexual lineages of *Pentalonia* aphids. *Viruses* 5, 758–776. doi: 10.3390/v5020758
- Weitzman, M. D., and Fradet-Turcotte, A. (2018). Virus DNA replication and the host DNA damage response. *Annu. Rev. Virol.* 5, 141–164. doi: 10.1146/annurev-virology-092917-043534
- Yu, N. T., Cai, Z. Y., Xiong, Z., Yang, Y., and Liu, Z. X. (2020). Complete genomic sequence of Noni mosaic virus (NoMV) associated with a mosaic disease in *Morinda citrifolia* L. *Eur. J. Plant Pathol.* 156, 1005–1014. doi: 10.1007/s10658-020-01948-4
- Yu, N. T., Feng, T. C., Zhang, Y. L., Wang, J. H., and Liu, Z. X. (2011). Bioinformatic analysis of BBTV satellite DNA in Hainan. *Virol. Sin.* 26, 279–284. doi: 10.1007/s12250-011-3196-7
- Yu, N. T., Liu, F., Ji, X. L., Wang, J. H., Xiong, Z., and Liu, Z. X. (2019a). Full-length infectious clones of *Banana bunchy top virus*. *Phytopathology* 109(Suppl.), S2.186–S2.182.
- Yu, N. T., Xie, H. M., Zhang, Y. L., Wang, J. H., Xiong, Z., and Liu, Z. X. (2019b). Independent modulation of individual genomic component transcription and a cis-acting element related to high transcriptional activity in a multipartite DNA virus. *BMC Genomics* 20:573. doi: 10.1186/s12864-019-5901-0
- Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., et al. (2012). Cloning and sequence analysis of two *Banana bunchy top virus* genomes in Hainan. *Virus Genes* 44, 488–494. doi: 10.1007/s11262-012-0718-9
- Zhang, H., Rigo, F., and Martinson, H. G. (2015). Poly(A) signal-dependent transcription termination occurs through a conformational change mechanism that does not require cleavage at the poly(A) site. *Mol. Cell* 59, 437–448. doi: 10.1016/j.molcel.2015.06.008
- Zhuang, J., Lin, W., Coates, C. J., Shang, P., Wei, T., Wu, Z., et al. (2019). Cleavage of the babuvirus movement protein B4 into functional peptides capable of host factor conjugation is required for virulence. *Virol. Sin.* 34, 295–305. doi: 10.1007/s12250-019-00094-4
- Zubair, M., Zaidi, S. S., Shakir, S., Amin, I., and Mansoor, S. (2017). An insight into *Cotton leaf curl Multan betasatellite*, the most important component of cotton leaf curl disease complex. *Viruses* 9:E280.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yu, Zhang, Wang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.