JOURNAL OF VIROLOGY, Feb. 2007, p. 1746–1761 Vol. 81, No. 4
0022-538X/07/$08.00+0 doi:10.1128/JVI.01390-06

Reassortment and Concerted Evolution in *Banana Bunchy Top Virus* Genomes▽

Jer-Ming Hu,[1]* Hui-Chuan Fu,[2] Chia-Hua Lin,[2] Hong-Ji Su,[2] and Hsin-Hung Yeh[2]*

Institute of Ecology and Evolutionary Biology, National Taiwan University,[1] and Department of Plant Pathology and Microbiology, National Taiwan University,[2] Taipei 106, Taiwan

Received 3 July 2006/Accepted 16 November 2006

The nanovirus *Banana bunchy top virus* (BBTV) has six standard components in its genome and occasionally contains components encoding additional Rep (replication initiation protein) genes. Phylogenetic network analysis of coding sequences of DNA 1 and 3 confirmed the two major groups of BBTV, a Pacific and an Asian group, but show evidence of web-like phylogenies for some genes. Phylogenetic analysis of 102 major common regions (CR-Ms) from all six components showed a possible concerted evolution within the Pacific group, which is likely due to recombination in this region. The CR-M of additional Rep genes is close to that of DNA 1 and 2. Comparison of tree topologies constructed with DNA 1 and DNA 3 coding sequences of 14 BBTV isolates showed distinct phylogenetic histories based on Kishino-Hasegawa and Shimodaira-Hasegawa tests. The results of principal component analysis of amino acid and codon usages indicate that DNA 1 and 3 have a codon bias different from that of all other genes of nanoviruses, including all currently known additional Rep genes of BBTV, which suggests a possible ancient genome reassortment event between distinctive nanoviruses.

Banana and plantain (*Musa* spp.) have been cultivated in more than 120 countries in the tropical and subtropical regions and are the major staple food crop for approximately 400 million people (35). Many viruses are known to infect banana and plantain and cause severe damage to the plants. Among banana viral diseases, banana bunchy top disease caused by *Banana bunchy top virus* (BBTV) is the most common and destructive viral disease in Asia and the Pacific regions and has been considered to be one of the most important plant viral diseases around the world (10, 50). BBTV is the sole member of the genus *Babuvirus* in the family *Nanoviridae*, with an approximately 18-to 20-nm isometric virion; it is transmitted by vegetative planting materials or by an aphid vector, *Pentalonia noigronervosa* (1, 21, 45, 87).

The BBTV genome consists of at least six integral single-stranded circular DNA (ssDNA) components (BBTV DNA 1 to 6) (6, 22, 37). BBTV DNA 1 encodes a replication initiation protein (Rep). An open reading frame (ORF) of DNA 2 was proposed by Beetham et al. (3) on the basis of Northern blot hybridization and sequences obtained by 3′ rapid amplification of cDNA ends from an Australian isolate, but the function of this gene is still unknown (22). DNA 3 encodes a viral coat protein for encapsulation (83). The products of DNA 4 and DNA 6 genes are very similar to BC1 and BV1 of the bipartite *Begomovirus*, respectively, and both genes have been suggested to be involved in virus movement (53, 62, 84). DNA 6 possibly encodes a nuclear shuttle protein, and DNA 4 encodes a protein that can redirect the products of DNA 6 to the cell periphery (53, 62, 84). DNA 5 has been shown to contain an LXCXE motif, and binding between DNA 5 products and retinoblastoma (Rb), demonstrated by yeast two-hybrid analysis, may be involved in host-cell cycle manipulation (82). In addition, more than one Rep-encoded component may be associated with virus infection in nanoviruses (22, 29, 30, 39, 63). Only one of these Rep genes, master Rep, can directly replicate non-self-replicable components, and remaining Rep-encoded components are considered to be satellite viruses (29, 78, 80). Among BBTV, DNA 1 encodes the master Rep, and several additional Rep-encoded components have been reported in the Asian strains (29, 30, 88, 91).

All BBTVs and associated components have a stem-loop structure with a conserved sequence TA(G/T)TATTAC in the loop region that is common to all nanoviruses. In addition, a stretch of 69 nucleotides (nt) flanking the stem-loop region shares at least 62% homology among BBTV DNA 1 to 6 and has been defined as the stem-loop common region (CR-SL) (6). Another stretch of 66 to 92 nt located 5′ to CR-SL and sharing at least 76% sequence homology among BBTV DNA 1 to 6 has also been identified and defined as the major common region (CR-M) and is only present in BBTV (6).

In previous studies, two groups of BBTV, the South Pacific (including Indian and Egyptian isolates) and the Asian groups, have been identified from sequence analysis of coding and noncoding sequences of DNA 1, 3, and 6 in isolates obtained from different geographic regions (37, 38, 84). The past few years have seen the accumulation of extensive sequence information about BBTV components, including several full sets of BBTV genomes. These sequences provide the basis for our analysis of the evolutionary history of BBTV.

Although BBTV is an ssDNA virus, the basic mechanisms of molecular evolution, including mutation, recombination, and reassortment (for multipartite viruses), are qualitatively similar to those of most other plant viruses with a plus-sense RNA

* Corresponding author. Mailing address for Jer-Ming Hu: Institute of Ecology and Evolutionary Biology, National Taiwan University, 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan. Phone: 886 2 3366 2472. Fax: 886 2 2368 6750. E-mail: jmhu@ntu.edu.tw. Mailing address for Hsin-Hung Yeh: Department of Plant Pathology and Microbiology, National Taiwan University, 1, Sec. 4, Roosevelt Road Taipei 106, Taiwan. Phone: 886 2 3366 4601. Fax: 886 2 2363 6490. E-mail: hyeh@ntu.edu.tw.

▽ Published ahead of print on 29 November 2006.

genome (for a review, see reference 19). However, studies of Rep and coat protein genes of cotton leaf curl geminivirus, a bipartite plant DNA virus, revealed a higher degree of natural variation than several plant and animal RNA viruses (64). Whether plant DNA and RNA viruses show distinct properties remains to be examined since the information on plant DNA viruses is mostly limited to *Geminiviridae* and *Caulimoviridae*. BBTV is an important and unusual multipartite ssDNA virus, unique among all known DNA viruses and nanoviruses, and serves as a great example for studying plant DNA virus evolution.

In this study, we first performed phylogenetic network analyses based on coding regions of the master Rep (DNA 1) and the coat protein (DNA 3) genes of BBTV. Phylogenetic network methods were developed to visualize a non-tree-like network of target sequences and are particularly useful in studying organisms with reticulate evolutionary history (34). Such methods have been incorporated in phylogenetic analysis of viruses as evidence of recombination or conflicting signals in the genomes of, for example, Dengue virus (28), primate lentivirus (61), and hepatitis E virus (81). To examine the congruence of the phylogenetic trees obtained from study of DNA 1 and 3 sequences, we conducted a preliminary screening of conflict phylogenetic signals by an incongruence length difference (ILD) test (14). An ILD test is usually used to assess whether the data partitions are combinable, but it can be applied to detect hybridization (2) or horizontal gene transfer (42). Since the test can give false-positive results (9, 11, 36), we conducted nonparametric and parametric analysis for comparing obtained trees by the Kishino-Hasegawa (KH) (40) and Shimodaira-Hasegawa (SH) (69) analysis. In addition, we performed a phylogenetic analysis using CR-M regions from all six components of BBTV.

We included two BBTV isolates from Taiwan, one from a typical BBTV Taiwanese strain (Taiwan), and one from a newly identified mild strain of BBTV, the Taiwan type V (TW4) strain (72). The TW4 genome has only five components, corresponding to DNA 1 to DNA 5 of other BBTV genomes.

Finally, to obtain more evolutionary information of BBTV genomes, we conducted multivariate analysis based on codon and amino acid usages for different components of BBTV and other nanoviruses. Such analysis has been used for examining disparity in codon usages among genes in the same genome and/or patterns of biochemical composition among gene products of orthologous genes (56). An unusual pattern showing a specific gene positioned away from the gene clusters of the same genome on the two-dimensional (2-D) plot will indicate a different codon usage (codon bias) of the particular gene (7, 52). Although the virus genome is small, the coding sequence lengths of the BBTV genome are all longer than the minimal requirement for multivariate analysis of codon usage (i.e., 150 nt) (56). Clustering of codon usage according to specific genomes, as seen on the 2-D plot, is expected, and any outliers of the clusters are possible indicators of horizontal transferred genes, as has been demonstrated in *Escherichia coli* (49).

MATERIALS AND METHODS

Sequences. The sequences were all directly obtained from GenBank or determined by this study and are listed in Table 1. The genetic diversity among different components was calculated for full-length sequences as well as the separated regions of ORF, CR-SL, and CR-M. We constructed a data matrix comprising 102 CR-M sequences of six BBTV components in the GenBank, including five additional Rep gene sequences and 14 new sequences from Taiwan. All sequences were aligned by CLUSTAL X, version 1.83 (76), with default settings (gap opening penalty, 10.0; gap extension penalty, 0.1).

Cloning and sequencing of BBTV genomic DNA of Taiwanese isolates. BBTV components were amplified by PCR with primer pairs designed at conserved regions of BBTV SL (DNA 1, 3, and 5) or a specific region of corresponding components (DNA 2, 4, and 6). The nucleic acids were extracted from a specific infected banana plant at Taiwan. The mild strain BBTV TW4 was collected from a field-infected banana and has been characterized previously (72). The primer pairs were designed in an immediately adjacent outward extending direction, and the sequences are available upon request. The PCR conditions were as in previous reports (71, 72). The amplified products were cloned into pGEM-T vector (Promega, Madison, WI) by incubating with DNA ligase overnight at 4°C, followed by transformation into *E. coli* DH5α or Top 10. DNA 1, 3, and 5 of BBTV were differentiated by restriction enzyme AluI digestion patterns. Nucleotide sequences of the cloned cDNAs were determined in both directions with use of an ABI Prism DNA sequencer 310 or 377 (Applied Biosystems, Foster city, CA). At least three individual clones of each component were sequenced, and the consensus sequence derived from the alignment of sequences of each component was used in this study.

Phylogenetic analysis. To obtain detailed phylogenetic information, we compiled data sets of coding regions of 43 nucleotide sequences of DNA 1 isolates and 29 sequences of DNA 3 isolates of BBTV. Corresponding sequences according to Hughes (32) from three other nanoviruses, *Faba bean necrotic yellow virus* (FBNYV), *Milk vetch dwarf virus* (MDV), and *Subterranean clover stunt virus* (SCSV), were used as outgroups. Phylogenetic network analysis involved use of Neighbor-Net (5), a distance method for constructing phylogenetic networks, incorporated in the program SplitsTree4 (34). Nucleotide substitution of the HKY85 model was used in the analysis, and branch support was estimated by bootstrapping with 1,000 replicates.

Phylogenetic analysis of DNA 4 and DNA 5 ORFs involved Bayesian inference (BI), neighbor-joining (NJ), and maximum parsimony (MP) methods. BI analysis was by use of MrBayes, version 3.1 (31), and NJ and MP by PAUP*, version 4.0b10 (73). In BI analysis, nucleotide substitution of the HKY85 model (23) was used with gamma distribution (alpha = 0.5) set as the rate distribution among sites in the analyses. We performed BI analyses with four chains of Markov chain Monte Carlo, sampling 1 tree per 5,000 generations for 1,500,000 generations, and the first 500 trees were excluded for calculation of posterior probability on each node. An HKY85 model for the nucleotide sequences was incorporated in NJ analysis by PAUP*, version 4.0b10 (73). In MP analysis, we conducted heuristic searches with 1,000 random addition replicates and tree bisection-reconnection branch swapping, and 10 trees were saved from each replicate. Because of the limited availability of data, only 10 and 11 sequences were included in the data sets for DNA 4 and DNA 5, respectively. The branch support was estimated by bootstrapping with 1,000 replicates for both NJ and MP analyses.

Phylogenetic analysis of 102 CR-M sequences involved NJ with PAUP*, version 4.0b10 (73), and the HKY85 nucleotide substitution model, and support for the branches was evaluated by bootstrap analyses with 1,000 bootstrap replicates. For all analyses, gaps were treated as missing data, and no sites containing insertions or deletions were excluded.

To compare the phylogenies of DNA 1 and DNA 3, two smaller data matrices were constructed for strains with both DNA 1 and DNA 3 sequences, since the comparison among components should be based on the same source of data. We selected 17 isolates with both DNA 1 and DNA 3 sequences, including 14 BBTV and three other nanoviruses, FBNYV, MDV, and SCSV, as outgroups. Phylogenetic analysis was conducted by BI methods as described above.

The incongruence of phylogenetic information among data partitions of BBTV DNA 1 and DNA 3 was first evaluated by the ILD test (14), followed by KH and SH tests. Three partitions, CR-SL, ORF, and CR-M, were categorized for each BBTV component and subjected to the ILD test, which is implemented in PAUP*, version 4.0b10 (73), and the branch-and-bound option was chosen for tree searching with 1,000 replicates. The tree topological test was conducted by use of KH tests (40) under parsimony criteria for comparing tree lengths of the obtained trees from the BI analysis above and SH tests (69) for comparing the likelihood scores between trees. SH tests involved a resampling estimated log-likelihood method (41). Both analyses were conducted by use of PAUP*, version 4.0b10 (73), for the DNA 1 or DNA 3 data matrix.

Measurement of codon and amino acid usage. A total of 132 coding sequences from all components of BBTV and other nanoviruses, including five identified

TABLE 1. Voucher information of BBTV sequences used in this study[a]

Protein family	Accession no.[b]	Region of origin	Reference or source[c]	Strain[d]
DNA 1	AusNC_003479	Australia	22	Australia
	ChiAF110266	China: Zhangzhou	P. La et al. 2001*	
	ChiAF238874	China: Gaozhou	24	NS
	ChiAF246123	China: Guangdong	H. Xiao et al.*	Guangdong-1
	ChiAF238875	China: Guangzhou	24	NSP
	ChiU97525	China	H. Xiao et al.*	C4
	ChiAY450396	China: Hainan	77	Hainan
	EgyAF102780	Egypt: Kalubia	A. A. Rezk et al.*	Kalubia
	EgyAF416465	Egypt	38	
	FijAF416466	Fiji	38	Fiji
	HawU18077	United States Hawaii	89	
	IndAF416470	India	38	
	IndAY222303	India	S. B. Ghosh et al.*	
	IndAM055641	India: Karnataka	S. Rahman et al.*	
	IndAY845437	India: Tamil Nadu	S. Harish et al.*	TN
	InaAB186924	Indonesia: Central Java	17	IG33
	InaAB186925	Indonesia: Central Java	17	IG64
	InaAB186926	Indonesia: Central Java	17	IJs11
	JapAB108452	Japan: Okinawa	16	JN4
	JapAB108453	Japan: Kume	16	JK3
	JapAB108454	Japan: Miyako	16	JM5
	JapAB108455	Japan: Kurima	16	JM6
	JapAB108456	Japan: Ishigaki	16	JY1
	JapAB108457	Japan: Taketomi	16	JY3
	JapAB108458	Japan: Iriomote	16	JY7
	PhiAB189067	Philippines: Luzon	16	bP5
	PhiAF416469	Philippines	38	
	TaiAF416468	Taiwan	38	
	TaiDQ817617	Taiwan	This study	Taiwan
	853027	Taiwan	This study	TaiV1
	857902	Taiwan	This study	TaiV1
	TonAF416467	Tonga	38	
	VieAF416464	Vietnam	38	
	VieAB113659	Vietnam: Hanoi	16	V6
	VieAB113660	Vietnam: north of Hanoi	16	V14
	VieAF416472	Vietnam: Son La	4	
	VieAF416473	Vietnam: Dien Bien Phu	4	
	VieAF416474	Vietnam: Bac Ninh	4	
	VieAF416475	Vietnam: Hue	4	
	VieAF416476	Vietnam: Buon Ma Thout	4	
	VieAF416477	Vietnam: Da Nang	4	
	VieAF416478	Vietnam: Ho Chi Minh City	4	
	VieAF416479	Vietnam: Ven Bai	4	
DNA 2	AusNC_003475	Australia	6	Australia
	AusL41576	Australia: Nambour of southern Queensland	6	
	ChiAY606084	China: Hainan	77	Hainan
	HawU18078	United States: Hawaii	89	
	EgyAF102781	Egypt: Kalubia	A. A. Rezk et al.*	
	IndAY884172	India	J. Ananthi et al.*	
	IndAY884173	India	J. Ananthi et al.*	
	TaiDQ817867	Taiwan	This study	Taiwan
	TaiDQ825708	Taiwan	This study	Taiwan
	853087	Taiwan	This study	TaiV1
DNA 3	AusNC_003473	Australia	6	Australia
	AusL41574	Australia	6	
	BurAF148943	Burundi	84	
	ChiAF238876	China: Gaozhou	25	NS
	ChiAF238877	China: Guangzhou	25	NSP
	ChiAF246122	China	H. Xiao et al.*	Guangdong-3
	ChiAF330706	China: Zhangzhou	W. Q. Cai and G. J. Mao*	
	ChiAY337715	China: Hainan	E. Tian and Z. X. Liu*	
	ChiAY494786	China: Hainan	77	Hainan
	ChiU97526	China	H. Xiao et al.*	C4
	EgyAF102782	Egypt: Kalubia	A. A.Rezk et al.*	Kalubia
	FijAF148944	Fiji	84	Fiji
	IndAY272038	India	J. Ananthi et al.*	

Continued on facing page

TABLE 1—*Continued*

Protein family	Accession no.[b]	Region of origin	Reference or source[c]	Strain[d]
	IndAY534140	India: Tamil Nadu	R. Selvarajan et al.*	TN
	InaAB186927	Indonesia: Central Java	17	IG33
	InaAB186928	Indonesia: Central Java	17	IG64
	InaAB186929	Indonesia: Central Java	17	IJs11
	JapAB078023	Japan: Okinawa	K. Kawabe and M. Onuki*	
	JapAB108449	Japan: Okinawa	16	JN4
	JapAB108450	Japan: Kume	16	JK3
	JapAB108451	Japan: Ishigaki	16	JY1
	PhiAB189068	Philippines: Luzon	16	bP5
	PhiAF148068	Philippines	84	
	TaiAF148942	Taiwan	84	
	TaiDQ817893	Taiwan	This study	Taiwan
	857918	Taiwan	This study	TaiV1
	VieAF148945	Vietnam	84	
	VieAB113661	Vietnam: Hanoi	16	V6
	VieAB113662	Vietnam: north of Hanoi	16	V14
DNA 4	AusNC_003474	Australia	6	Australia
	AusL41575	Australia	6	
	ChiAF246124	China: Guangdong	90	NS
	ChiAF349568	China: Zhangzhou	J. D. Sun et al.*	
	ChiAY494788	China: Hainan	77	Hainan
	ChiU97527	China	H. Xiao et al.*	C4
	EgyAF102783	Egypt: Kalubia	A. A. Rezk et al.*	Kalubia
	IndAY845635	India: Tamil Nadu	S. Harish et al.*	TN
	IndAY953429	India	M. Kavino et al.*	
	TaiDQ825714	Taiwan	This study	Taiwan
	853095	Taiwan	This study	TaiV1
DNA 5	AusNC_003477	Australia	6	Australia
	AusL41578	Australia	6	
	ChiAY264347	China: Guangzhou	92	NSP
	ChiAY266417	China: Gaozhou	Y. Zheng et al.*	NS
	ChiAY606085	China: Hainan	77	Hainan
	EgyAF102784	Egypt: Kalubia	A. A. Rezk et al.*	Kalubia
	HawU18079	United States: Hawaii	89	
	IndAY267898	India	J. Ananthi et al.*	
	IndAY845636	India: Tamil Nadu	S. Harish et al.*	TN
	TaiDQ817921	Taiwan	This study	Taiwan
	857944	Taiwan	This study	TaiV1
DNA 6	AusNC_003476	Australia	6	Australia
	AusL41577	Australia	6	
	ChiAF238878	China: Gaozhou	26	NS
	ChiAF238879	China: Guangzhou	26	NSP
	EgyAF148139	Egypt: Kalubia	A. A. Rezk et al.*	Kalubia
	HaiAY494787	China: Hainan	77	Hainan
	IndAY273170	India	J. Ananthi et al.*	
	IndAY845637	India: Tamil Nadu	M. Kavino et al.*	TN
	TaiDQ825730	Taiwan	This study	Taiwan
Additional Rep				
	TaiAF216221	Taiwan	30	BBTV-S1
	TaiAF216222	Taiwan	30	BBTV-S2
	VieAF416471	Vietnam: Son La	4	BBTV-S3
	U12586	Taiwan	This study	BBTV-W3
	U12587	Taiwan	This study	BBTV-W4
	L32166	Taiwan	88	BBTV-W1
	L32167	Taiwan	88	BBTV-W2

[a] Other nanovirus and circovirus information was obtained from Hughes (32).
[b] Accession numbers from NCBI GenBank database are given after the three-letter abbreviation of country or region names. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Fij, Fiji; Haw, Hawaii; Phi, Philippines; Ton, Tonga; Bur, Burundi.
[c] Asterisks mark sequences found in the GenBank database without references.
[d] Isolates with strain names are given for those in which multiple components were identified, as in Fig. 4.

additional Rep genes (as described in reference 32), were used in codon and amino acid usage analysis. Relative synonymous codon usage (RSCU) values (67) were used for multivariate analysis to measure codon bias. The RSCU value is the observed frequency of a codon divided by the expected frequency under equal usage of all codons for a given amino acid. Amino acid usage was calculated from the direct counts of amino acids in a particular sequence.

Multivariate analyses involved principal component analysis (PCA) for both RSCU and amino acid usage data for all 101 coding sequences of BBTV (PCA/RSCU and PCA/AA, respectively, hereafter), and 31 sequences of nanoviruses, including *Coconut foliar decay virus* (CFDV), FBNYV, MDV, and SCSV. PCA was chosen for multivariate analyses of codon and amino acid usages as suggested by Perriere (56), because other methods such as corresponding analysis are very sensitive to codons of rarely used amino acids in the sequences. The multivariate analyses involved use of ADE-4 (75), in which the indices of codon

TABLE 2. Genetic diversity of all used BBTV genomic DNA, calculated as the number of substitutions per site

Protein family and subgroup (no. of isolates)[a]	No. of substitutions			
	Full-length	ORF	CR-SL	CR-M
DNA 1 (42)	0.057 ± 0.033	0.051 ± 0.027	0.006 ± 0.014	0.119 ± 0.095
Pacific (10)	0.016 ± 0.006	0.014 ± 0.006	0.000 ± 0.000	0.055 ± 0.049
Asia s.s. (21)	0.029 ± 0.021	0.024 ± 0.016	0.007 ± 0.017	0.040 ± 0.036
Vietnam-N (8)	0.024 ± 0.009	0.017 ± 0.003	0.005 ± 0.007	0.038 ± 0.021
Vietnam-S (3)	0.024 ± 0.005	0.022 ± 0.006	0.019 ± 0.017	0.032 ± 0.011
DNA 2 (9)	0.152 ± 0.081	0.087 ± 0.058[b]	0.196 ± 0.195	0.228 ± 0.137
DNA3 (28)	0.060 ± 0.044	0.042 ± 0.026	0.016 ± 0.016	0.111 ± 0.131
Pacific (7)	0.023 ± 0.020	0.035 ± 0.030	0.029 ± 0.024	0.026 ± 0.014
Asia (21)	0.033 ± 0.212	0.025 ± 0.017	0.009 ± 0.013	0.018 ± 0.018
DNA 4 (10)	0.134 ± 0.066	0.111 ± 0.051	0.054 ± 0.061	0.170 ± 0.123
DNA 5 (10)	0.078 ± 0.050	0.074 ± 0.048	0.017 ± 0.010	0.154 ± 0.113
DNA 6 (8)	0.094 ± 0.052	0.064 ± 0.044	0.014 ± 0.012	0.150 ± 0.125

[a] The subgroups correspond to the phylogenies of Fig. 1. The sequences from Taiwanese type V strain are not included.

[b] The location of the DNA 2-encoded ORF was proposed by Beetham et al. (3). Although the ORF could not be detected from isolates reported in Hainan (ChiAY606084) or from Taiwan, the corresponding position of the predicted ORF in these isolates was used to calculate the sequence divergence.

usage were calculated by CodonW (55) on a web server hosted by the Institut Pasteur, Paris, France (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html). Stop codons were excluded from the analyses because only one stop codon exists for each gene.

Nucleotide sequence accession numbers. The obtained sequences of BBTV Taiwan strain isolated by this study were deposited in GenBank under the following accession numbers: DNA 1, DQ817617; DNA 2a, DQ817867; DNA 2b, DQ825708; DNA 3, DQ817893; DNA 4, DQ825714; DNA 5, DQ817921; and DNA 6, DQ825730. Accession numbers for the TW4 strain are as follows: DNA 1a, 853027; DNA 1b, 857902; DNA 2, 853087; DNA 3, 857918; DNA 4, 853095; and DNA 5, 857944.

RESULTS

Genetic diversity of the BBTV isolates. The sequence divergence of BBTV isolates varies among components and among regions for each component (Table 2). Subgroups of DNA 1 and 3 are designated according to well-supported clades based on further phylogenetic analyses (see following discussion). In general, sequences of ORFs show less than 10% variation among isolates, except for DNA 4 (11.09%). The CR-SL among all components is highly conserved, except for DNA 2. In contrast, the CR-M shows the greatest divergence among all components, ranging from 11.12% (DNA 3) to 22.76% (DNA 2) (Table 2). However, the divergence of the CR-M within subgroups of DNA 1 and 3 is quite low (i.e., less than 3% variation within the two subclades of DNA 3 isolates but more than 11% divergence between the subclades) (Table 2). It is worth noting that the ORF of DNA 2 proposed by Beetham et al. (3) could not be detected from isolates reported from Hainan (ChiAY606084) or Taiwan (this study). In these isolates the proposed DNA 2 translation start codon is TTG instead of AUG. To avoid sequencing errors or mutation from a single clone of DNA 2, we sequenced other DNA 2 clones from the same isolate as well as from four individual isolates collected in Taiwan, but none contained a typical ATG start codon in the proposed position (data not shown).

Phylogenetic relationship among BBTV isolates. Results of phylogenetic network analysis based on nucleotide sequences of the DNA 1 ORF are shown in Fig. 1. Two major groups can be identified: isolates from Australia (Aus), India (Ind), Egypt (Egy), and Pacific islands, named the "Pacific group," and all other sequences, named the "Asian group." Three subgroups can be identified in the Asian group: three isolates from southern regions of Vietnam, sensu Bell et al. (4) (Vietnam-S); eight isolates from northern Vietnam (Vietnam-N); and all other Asian isolates (Asian s.s.). A DNA 1 isolate from Hainan (ChiAY450396) did not group with other Asian DNA 1 on the tree and was close to the Vietnam-N group (Fig. 1).

Phylogenetic analysis for the DNA 3 ORF also revealed a non-tree-like pattern (Fig. 2). The Pacific isolates (Australia, India, Egypt, and Pacific islands) are still clustered together, but the Egyptian isolate (EgyAF102782) seems to have reticulate evolution. Sequences from Vietnam-N isolates are quite close, but several Chinese isolates (ChiAY494786, ChiAY337715, and ChiAF238877) show a web-like network, although most of the bootstrap supports for reticulate relationships are low (Fig. 2).

The phylogenies of DNA 4 and DNA 5 show a distinct incongruence for the TW4 isolate (Fig. 3). Taiwan V1DNA4 (TaiV1DNA4) is grouped with other Asian DNA 4 isolates, whereas TaiV1DNA5 is grouped with the Pacific DNA 5 isolates. Both relationships received very high posterior probability (1.0) and bootstrap supports (100 or over 90).

The unrooted phylogram based on 102 CR-M regions of BBTV sequences used with NJ is shown in Fig. 4A. Two major clades can be seen in the phylogeny, the Pacific group and the Asian group. The Pacific group contains the CR-M sequences of all six components from Australia, India, Egypt, and the Pacific islands (Fig. 4A), whereas all other sequences form another clade, the Asian group (Fig. 4A). Note that within the two major clades, sequences from the same components tend to cluster together with only some exceptions, which all received low to moderate bootstrap support. CR-M sequences of four additional Rep genes of BBTV all clustered with the CR-M of DNA 1 and 2 of the Asian group (Fig. 4A). The result is different from the null hypothesis that the CR-M of BBTV has a single origin, which will produce a phylogenetic tree similar to that in Fig. 4B. In contrast, the obtained tree is more like the hypothetical tree of Fig. 4C, which suggests a concerted evolution in the CR-M of the Pacific or the Asian group. Aphylogenetic tree constraining the monophyly of the CR-M of each DNA component was manually reconstructed and compared with the Fig. 4 tree in an SH test under maximum likelihood criteria. The two phylogenies were significantly different ($P < 0.001$); the CR-M of each DNA component does not result in monophyletic groups in the BBTV tree, which rejects the hypothesis of the tree in Fig. 4B.

The alignment of the 102 CR-M sequences is shown in Fig. 5. Conserved sequences can be easily visualized within the Asian and Pacific groups but are quite distinct between these two groups. This result is consistent with the data of Table 2 that the overall divergence of CR-M is high but low within the Asian and the Pacific groups for all BBTV components.

Phylogenetic analysis of ORFs of DNA 1 and DNA 3 from 17 nanoviruses resulted in similar BI tree topologies (Fig. 6).

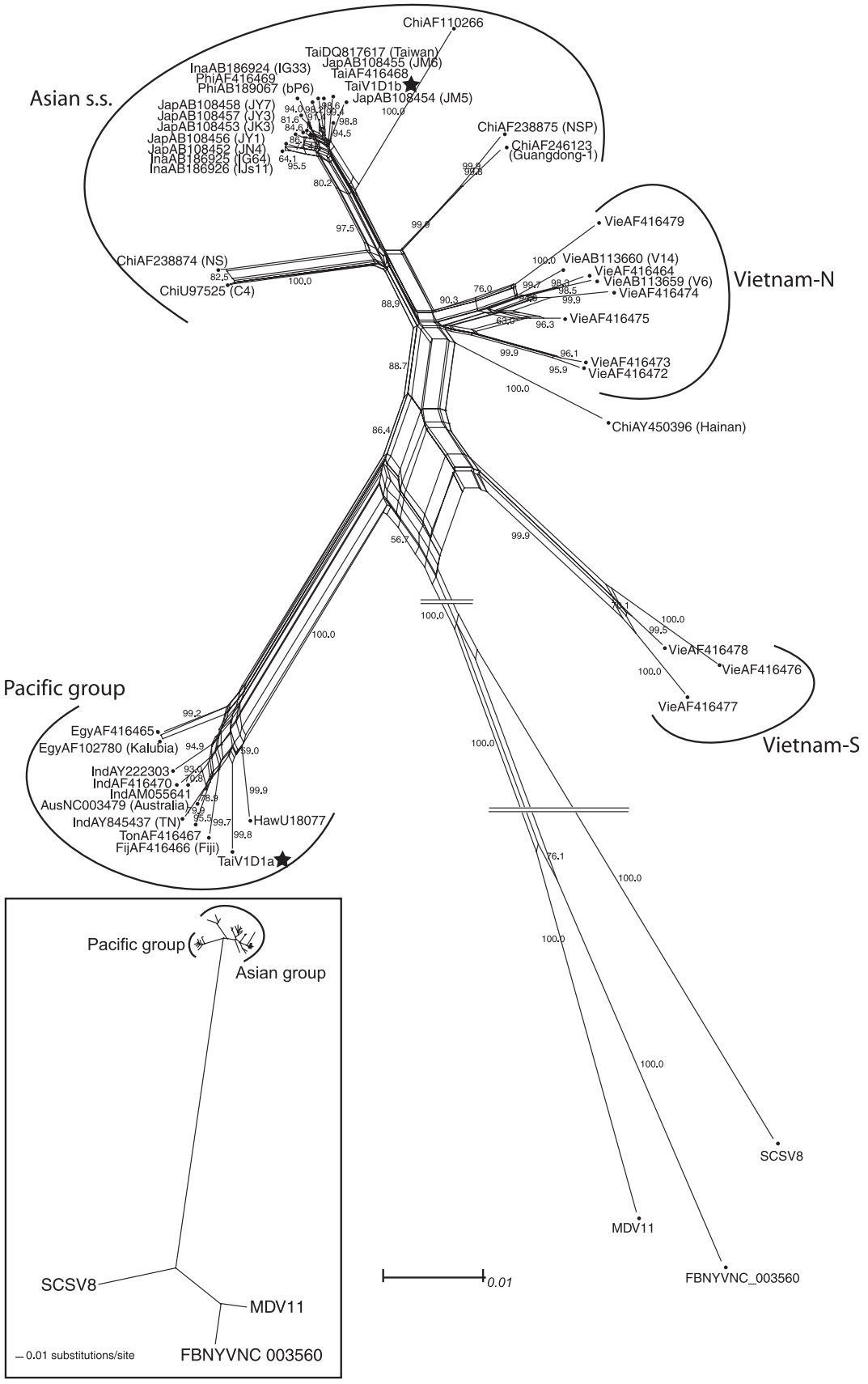

FIG. 1. Neighbor-Net trees based on 46 ORF nucleotide sequences of BBTV DNA 1 and three other nanoviruses. The Pacific group, and three Asian groups—Vietnam-N, Vietnam-S, and Asian s.s. groups—are marked accordingly. Bootstrap values over 50 are marked on the branches. The branches leading to FBNYV, MDV11, and SCSV8 sequences are shortened and marked by broken lines to save space. The inset is a phylogram based on neighbor-joining criteria, showing the relative branch length of sequences of BBTV and other nanoviruses. Accession numbers are preceded by abbreviations for the places of origin. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Fij, Fiji; Haw, Hawaii; Phi, Philippines; Ton, Tonga.

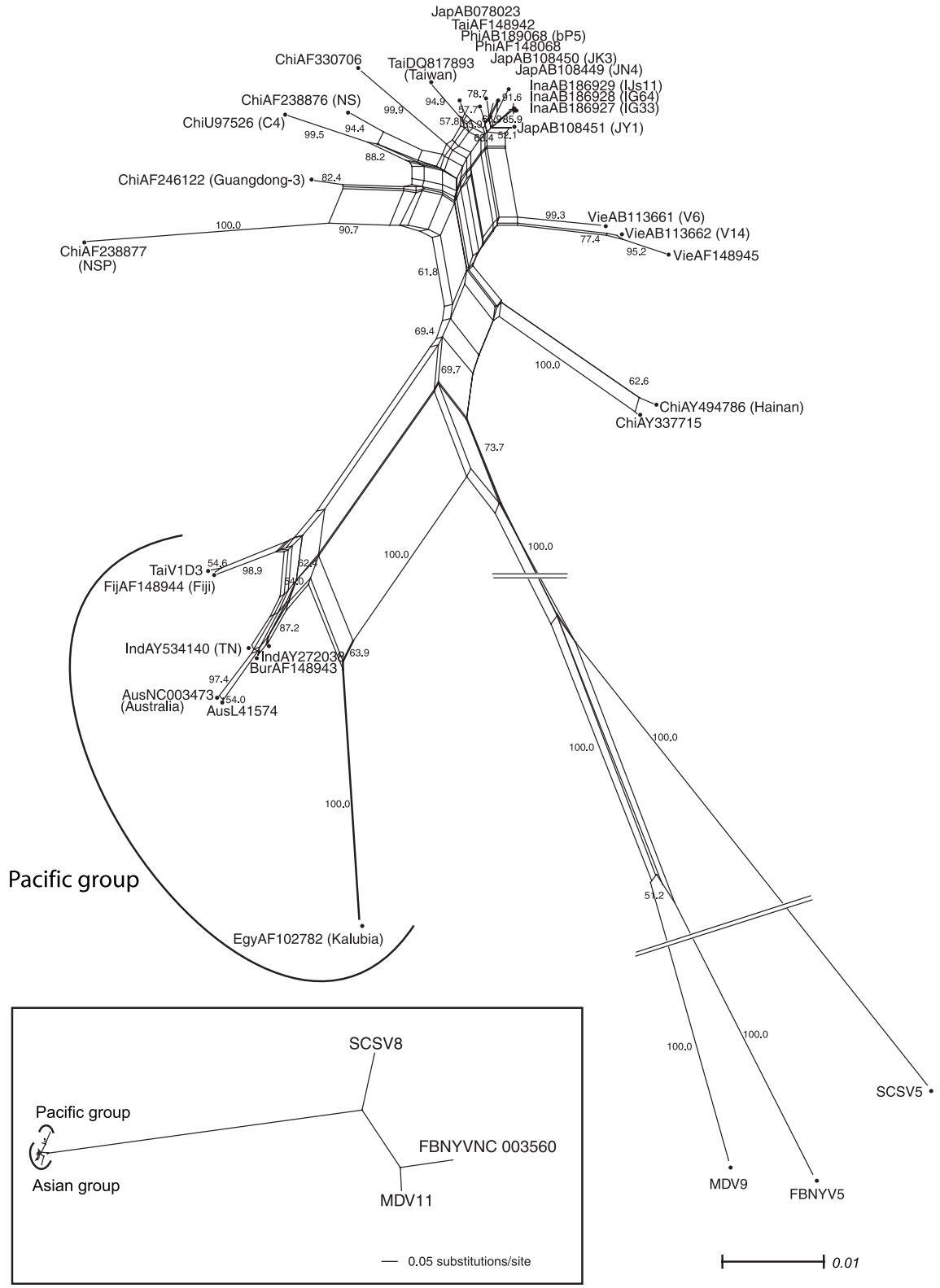

FIG. 2. Neighbor-Net trees based on 32 ORF nucleotide sequences of BBTV DNA 3 and three other nanoviruses. The Pacific group is marked on the figure, and the rest of the BBTV sequences are Asian group. Bootstrap values over 50 are marked on the branches. The branches leading to FBNYV, MDV11, and SCSV8 sequences are shortened and marked by broken lines to save space. The inset is a phylogram based on neighbor-joining criteria, showing the relative branch length of sequences of BBTV and other nanoviruses. Accession numbers are preceded by abbreviations for the places of origin. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Fij, Fiji; Phi, Philippines; Bur, Burundi.

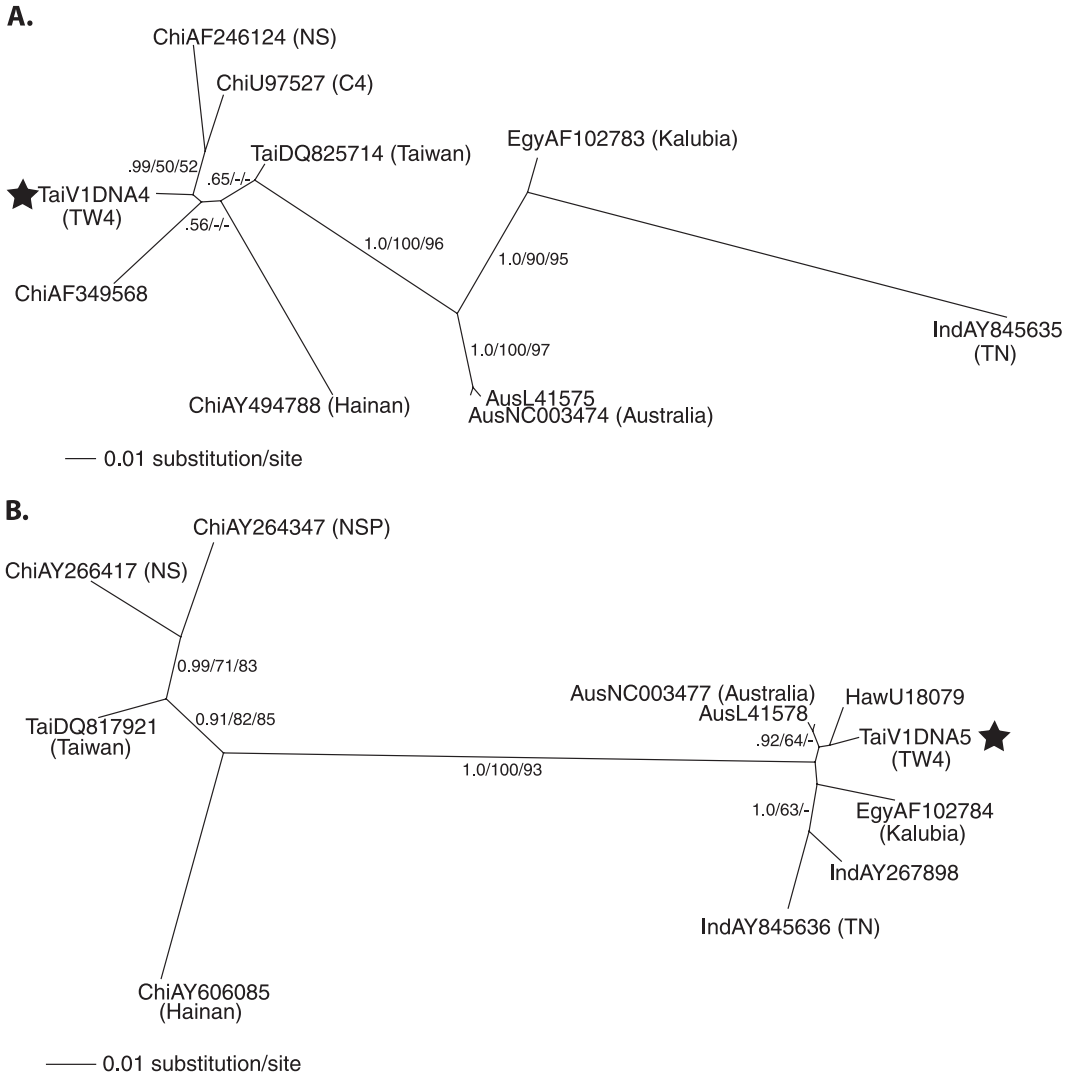

FIG. 3. Phylogenies obtained by the BI method based on ORF nucleotide sequences of DNA 4 and DNA 5 BBTV isolates. (A) DNA 4 tree. (B) DNA 5 tree. Along the branches are the supports of posterior probabilities from BI, followed by bootstrap supports of neighbor-joining and most parsimonious methods; only values that are >50% are shown. Accession numbers are preceded by abbreviations for the places of origin. Haw, Hawaii.

The phylogenies differ at the positions of the three Pacific isolates (Australia, Kalubia1, and Fiji) and three Chinese isolates (Hainan, NSP, and Guangdong). Partition homogeneity analysis based on ILD tests show conflicted signals in the pairwise comparison of DNA 1 and DNA 3 (Table 3). The ORF of DNA 1 shows significant incongruence with the CR-M of DNA 1 as well as the CR-SL and ORF of DNA 3. Tree topological incongruence analysis by use of SH and KH tests showed significant differences between the BI trees of DNA 1 and DNA 3 (consensus at 0.5 level for posterior probability) (Table 4). The *P* value is 0.000 with use of DNA 1 as a data matrix and 0.06 to 0.1 with use of DNA 3 as a data matrix.

Amino acid and codon usage patterns. PCA results of amino acid and codon usages of 132 nanoviruses are shown in Fig. 7. PCA/AA results reveal that the BBTV genes can be separated into four groups. The first three axes represent 23.7%, 19.8%, and 13.8%, respectively. DNA 1, 3, and 4 are separated, whereas DNA 5 and 6 are clustered together (Fig. 7A). All other nanovirus genes are scattered, corresponding to the homologues of different components of BBTV. DNA 1 genes are tightly clustered and not far from all other Rep genes, including the 14 additional Rep genes currently identified. DNA 3 genes are clustered with coat protein genes of CFDV2, FBNYV5, MDV9, and SCSV5. DNA 4 and MDV8 are at the lower right of Fig. 7A but are not close to the other putative homologues SCSV1 and FBNYV4. The position of SCSV1 is in fact close to two master Rep genes, MDV10 and FBNYV7. DNA 5 and 6 are clustered with the clink protein genes and the nonstructural proteins of other nanoviruses at the left side of Fig. 7A.

PCA/RSCU results are shown in Fig. 7B. The first three axes represent 17.8%, 15.6%, and 9.6%, respectively. DNA 1 and 3 form two distinct clusters in the plot, whereas all other genes are loosely grouped together, which suggests that the codon usages of DNA 1 and 3 are very different from each other and also from other components of BBTV and all other nanovirus

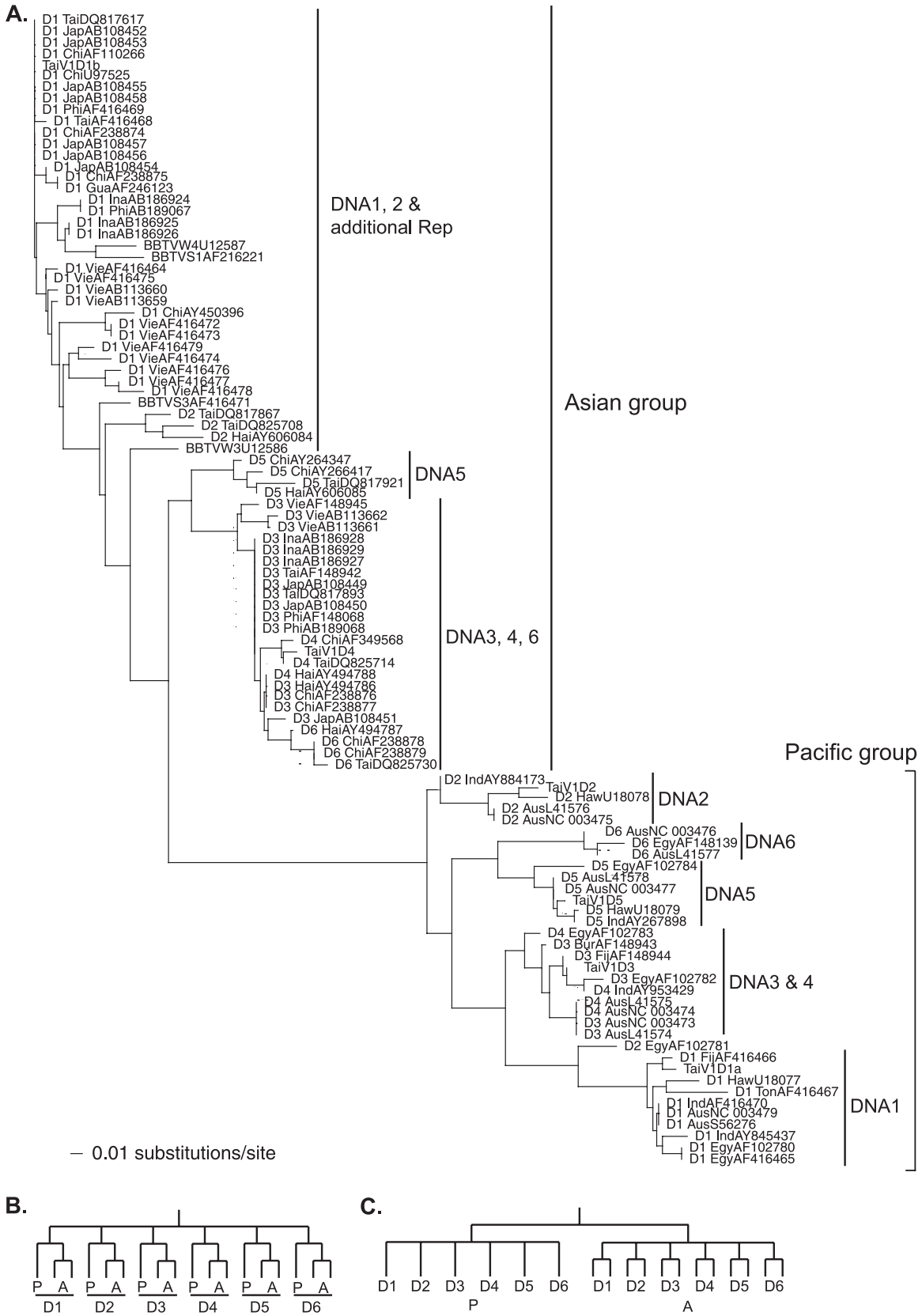

FIG. 4. (A) Unrooted phylogram of the neighbor-joining tree based on nucleotide sequences of the CR-Ms from 102 BBTV isolates under an HKY85 model. Bootstrap values for internal support of the branches are given along the branches. Two major clades are marked: the Asian and Pacific groups. The component names are given before the sequence names (i.e., DNA 1 to DNA 6 are named D1 to D6, respectively). CR-M

genes. Importantly, all five BBTV additional Rep genes are within the large cluster of various genes but not with BBTV DNA 1 or DNA 3. This codon usage pattern suggests that the DNA 1 and 3 may have a different codon bias or even origins compared with other components of BBTV.

DISCUSSION

Our detailed sequence analysis of 14 isolates in different geographical regions reveals several interesting findings. Phylogenetic analysis confirmed two major groups of BBTV, a Pacific and an Asian group, but show evidence of web-like phylogenies for some genes. The analysis of 102 CR-Ms from all six components showed a possible concerted evolution within the Pacific group, which is likely due to recombination in the sequences. We also demonstrated a case of genome reassortment in the BBTV TW4 strain from Taiwan.

Use of Rep and coat protein gene sequences to reconstruct phylogenies has been the prime approach for elucidating the evolutionary history of BBTV and other nanoviruses. Our results indicate that although such an approach could reflect the genealogy of individual genes of BBTV, the phylogenies might not be applicable to other components since different components can have different phylogenies. We demonstrated a clear case of a chimeric BBTV isolate, TW4, as having a mixture of the Asian and the Pacific components in a single isolate; TW4 likely contains DNA 2, 3, and 5 of the Pacific group and DNA 4 of the Asian group by genome reassortment, as supported by the CR-M phylogeny (Fig. 4) and the ORF phylogenies (Fig. 1, 2, and 3). Interestingly, TW4 has both the Pacific and the Asian type of DNA 1. The reason for maintaining two copies of DNA 1 could be due to a necessity for the two types of components, since the CR-M regions are quite distinct between the Pacific and the Asian group.

Phylogenetic network analysis is designed for analyzing non-tree-like phylogenies in the case of reticulate evolution such as hybridization, horizontal gene transfer, and recombination (34). It has also been proven very useful in studying virus evolution (28, 81). Our results indicate that, in general, DNA 1 of BBTV is more tree-like than DNA 3. The inclusion of FBNYV, MDV, and SCSV sequences did not help much in the rooting problems in BBTV, because they showed a network-like pattern at the base of BBTV (Fig. 1 and 2). However, a network-like pattern simply reflects conflicting phylogenetic signals, but the actual cause for the pattern could be reassortment, recombination, or concerted evolution. Other methods, such as use of genetic distances or substitution distribution as estimates (58), are needed for further evaluation of the possibility of recombination.

Our phylogenetic results for DNA 1 and DNA 3 mostly agree with the two-group hypothesis of DNA 1 and 3 in BBTV as suggested by previous studies (37, 38, 84). However, results of BI-based phylogenies of DNA 1 and 3 may be problematic in the rooting. One of the error sources for rooting in phylogenetic reconstruction is the inclusion of distantly related outgroup sequences. Use of a very divergent outgroup can cause precarious rooting in phylogenetic analysis (74, 85), likely due to the long-branch attraction effect (15), a well-known source of error in phylogenetic analysis. Although FBNYV, MDV, and SCSV represent the closest known viruses to BBTV, they are still quite different from BBTV, averaging more than 40% variation compared to DNA 1 and more than 57% compared to DNA 3 at the nucleotide level. In contrast, the three nanoviruses show only 8% to 20% variation on Rep genes and 21% to 38% variation on coat protein genes among themselves. Unfortunately, no other nanoviruses closer to BBTV are known; thus, no better outgroup is currently available.

The phylogeny of the CR-M complicates the story of BBTV. The rationale is that, if the ancestor genome of BBTV possessed all six components, the CR-M of each component should follow the evolution of BBTV, in accordance with the coding regions; therefore, the CR-M sequences should form separate monophyletic groups on the resulting tree according to each component. If not, then the evolution of CR-M will be shown to be independent of the evolution of components as a whole (i.e., some genomic reassortment or recombination has occurred). Our results strongly suggest two origins of the CR-M in BBTV, one that gave rise to the Pacific group and another that gave rise to the Asian group of BBTV. However, when or where these two CR-M types originated is not clear because no other comparable homologous sequence has been found, even in other nanoviruses. Since the additional Rep genes have been considered as outgroups of all nanovirus DNA 1 genes (32) and their CR-M grouped with the Asian group, the CR-M of the Asian group is likely the ancient type. A possible scenario is that the original BBTV of the Pacific group acquired a new CR-M sequence and then a homogenization occurred among the six components in this region. This "new" set of the BBTV genome became the origin of the Pacific group BBTV and then gave rise to strains in Australia, the Pacific, India, and Egypt. The phylogeny shows that the CR-Ms of additional Rep genes are similar to the CR-M of DNA 1, which suggests that they may be regulated by the master Rep (80). The phylogeny also shows that the CR-Ms of DNA 1 and 2 form a monophyletic group distinct from that of DNA 3 to 6 (Fig. 4A). In sum, this evidence suggests a concerted evolution in the CR-M of BBTV components, and the most likely cause is recombination in these regions. Our further analysis with use of the Recombination Analysis Tool (13) to rapidly screen possible recombinations showed a likely recombination breaking point between the ORF and CR-M of DNA 1 and 3 (J.-M. Hu, unpublished data). Even though the trees we obtained for individual components are identical—e.g., two groups, Asian and Pacific, for three components,

sequences of additional Reps have full names (i.e., BBTV S1, S3, W3, and W4) before the accession numbers. (B and C) Hypothetical trees showing two scenarios of CR-M evolution: a phylogeny showing a single origin of CR-M regions, in which each component forms monophyletic groups (B) and a phylogeny showing two origins of CR-M, grouped in two major clades, accordingly (C). See text for more explanation. Accession numbers are preceded by abbreviations for the places of origin. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Fij, Fiji; Haw, Hawaii; Phi, Philippines; Ton, Tonga; Bur, Burundi.

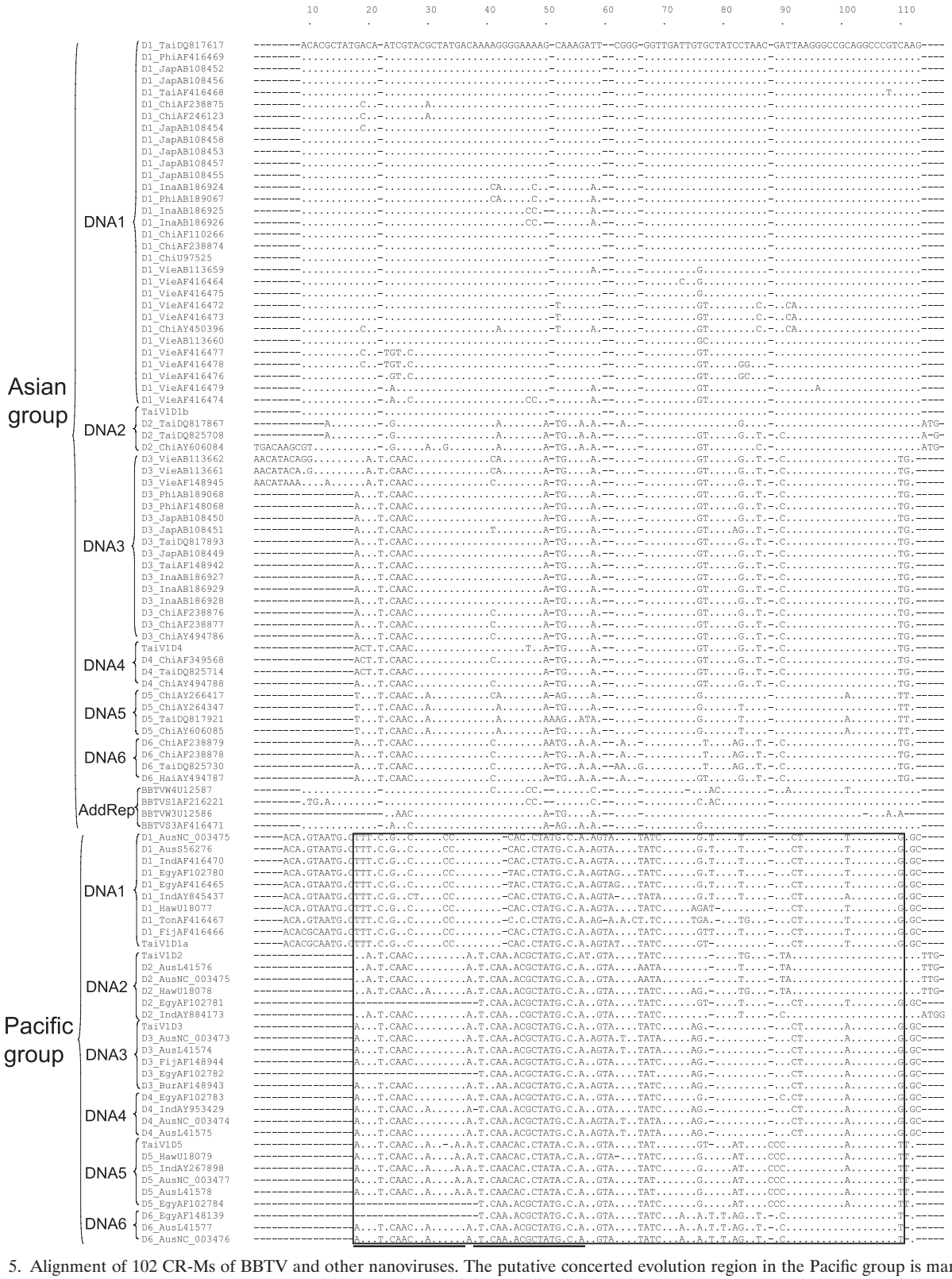

FIG. 5. Alignment of 102 CR-Ms of BBTV and other nanoviruses. The putative concerted evolution region in the Pacific group is marked by a square. The tandem repeated sequence proposed by Burns et al. (6) is underlined. Accession numbers are preceded by abbreviations for the places of origin. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Fij, Fiji; Haw, Hawaii; Phi, Philippines; Ton, Tonga; Bur, Burundi.

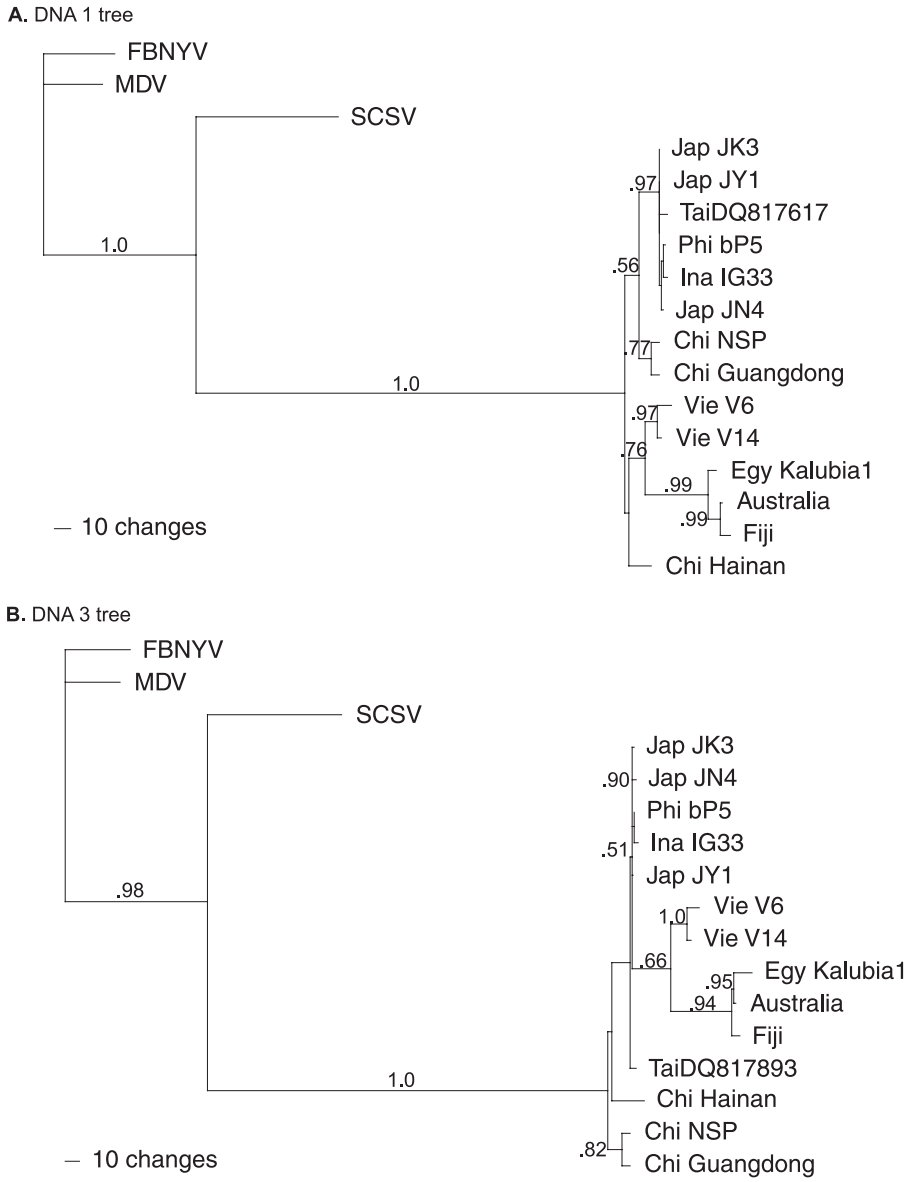

FIG. 6. Phylogenies obtained by BI based on ORF nucleotide sequences of DNA 1 and DNA 3 of 14 BBTV and three other nanovirus isolates. (A) DNA 1 tree. (B) DNA 3 tree. Posterior probabilities for each branch are marked. Ina, Indonesia; Jap, Japan; Vie, Vietnam; Phi, Philippines.

DNA 1, 3, and 6—the true evolutionary history of BBTV may be more complicated than demonstrated in this study. Since the function of the CR-M has yet to be identified in BBTV, whether the difference indicates distinct regulation patterns in replication or signals of virion encapsidation for the two groups of components remains to be examined.

The use of incongruence tests such as the ILD test in phylogenies has been questioned for the feasibility in detecting incongruent partitions (8, 11, 36). We consider the ILD test a first screen for identifying potential incongruence, since it is generally more susceptible to type I error, as suggested by several authors (27, 47, 57). Our ILD test result showing significance in the CR-SL of DNA 3 with other regions could have been due to random effects and type I error, since the noncoding regions in our analysis are quite short, with CR-SL and CR-M approximately 100 bp or shorter. Further analysis is certainly needed to validate this incongruence result. Therefore, we focused on the incongruence in coding regions between DNA 1 and 3. The tree topologies based on DNA 1 and DNA 3 are significantly different as calculated from the DNA 1 data matrix on SH and KH testing (Table 4), although the difference was less significant when the DNA 3 data matrix was used. The results suggest that the evolutionary histories of DNA 1 and 3 are very likely different from each other. The results of both the ILD test and the phylogenies indicate that the DNA 1 and 3 coding sequences contain incongruence signals, likely because of a birth-and-death evolution and/or genome reassortment among components, which is not uncommon in viruses (32). Our PCA/AA and PCA/RSCU results also indicate that the Rep and coat protein genes, including the

TABLE 3. ILD test result based on 14 BBTV and three other nanovirus strains with full sets of DNA 1 and DNA3

Component	Region	DNA 1			DNA 3		
		CR-SL	ORF	CR-M	CR-SL	ORF	CR-M
DNA 1	CR-SL						
	ORF	1.000					
	CR-M	1.000	0.004[a]				
DNA 3	CR-SL	1.000	0.074[b]	0.001[b]			
	ORF	1.000	0.005[a]	0.648	0.160		
	CR-M	1.000	0.728	0.295	0.001[a]	0.931	

[a] $P < 0.01$.
[b] $P < 0.1$.

TABLE 4. KH and SH tests comparing the trees resulting from analysis of DNA 1 and DNA 3 data matrices from Fig. 6

Tree[a]	DNA 1				DNA 3			
	MP[b]		ML[c]		MP[b]		ML[c]	
	Length[e]	*P*	-ln *L*	*P*	Length	*P*	-ln *L*	*P*
$Tree_{DNA1}$	687		3,925		628	0.10	2,983	0.06
$Tree_{DNA3}$	709	<0.01[d]	3,983	<0.01[d]	620		2,959	

[a] The $Tree_{DNA1}$ and $Tree_{DNA3}$ trees used are 50% majority-rule consensus trees. The tests were conducted by using data matrices of DNA 1 and 3, respectively.
[b] KH (40) test using parsimony scores.
[c] SH test using likelihood scores under an HKY85 (gamma distribution) model of substitution (69). ML, maximum likelihood; *L*, likelihood.
[d] 0.05 significance level.
[e] Evolutionary steps.

additional Rep genes, have a distinct codon bias compared with other nanovirus genomes.

One other problem we encountered is the identification of DNA 2 ORF in BBTV. Although the corresponding mRNA of DNA 2 has been detected by Beetham et al. (3), the identification of an ORF in DNA 2 is problematic in other isolates. Two possible explanations may account for the ambiguity. First, BBTV may utilize a nontypical translation start site for protein initiation as reported for *Rice tungro bacilliform virus* (18) and *Tobacco mosaic virus* (66). Alternatively, the function of the DNA 2 gene product could act at the RNA rather the protein level, because increasing evidence has indicated that noncoding RNA plays an important role in eukaryotic cell gene regulation (48). However, the function of DNA 2 of BBTV remains to be resolved. In addition, based on the inoculation assay with cloned *Faba bean necrotic yellow virus*, DNA revealed a possible functional redundancy or complementation between distinctive nanovirus genomic components. It is possible that other components may have compensated for the function of DNA 2 and allowed DNA 2 to accumulate mutations (79).

Multivariate analysis of codon usage analysis shows an unusual bias in BBTV DNA 1 and 3 that differs from other components of BBTV, whereas the codon usage of DNA 4 to 6 of BBTV is similar to that of other nanoviruses. Although the Eigen values were generally low from PCA/AA and PCA/CA and the significance of the clustering was not accessed by randomized data sets, we still think the analyses reflect certain patterns that bear evolutionary information. We have examined four axes of the data, and most of the axis pairs show the clustering of DNA 1 and 3, respectively (data not shown). The unusual codon usages of DNA 1 and 3 might be due to their having origins different from the origins of other components, as explained below.

The main causes of codon bias among genes in a specific genome are limitations in translational efficiency reflected by expression level and/or the need to maintain a certain genomic composition (i.e., the GC content) (43, 68). Although quantitative expression data for different components of BBTV are largely unavailable, we assume that expression level is not a major force in shaping codon usage in BBTV, since corresponding genes in FBNYV, MDV, and SCSV all clustered together with other components of BBTV. In addition, since virus gene expression depends on the host translation apparatus, codon bias may be influenced by host background. However, no such effect was detected in our analysis as no clustering patterns according to host species were seen in the PCA plot (data not shown).

The second source of codon bias is maintenance of genomic nucleotide composition. The base compositions are assumed to be similar among components of BBTV since they represent a single and complete genome. The average nucleotide content among all ORFs of nanoviruses is similar. In addition, the third-position nucleotide composition of nanoviruses does not influence any corresponding patterning in the 2-D PCA/RSCU plot (Fig. 6B). For example, although thymine content in the synonymous third codon position (T3s) of BBTV ORFs range from 27% to 48%, the values are not as extreme as in other nanoviruses (e.g., 23% in CFDV or 63% in SCSV4, where they are clustered in the middle of the PCA/RSCU plot) (Fig. 6B). Therefore, codon bias due to differences in genome base composition is also not a major factor in nanoviruses.

Gene length has been shown to shape codon usage in both prokaryotes and eukaryotes (12, 51), although the factor is usually significant only among genes with a 10-fold or greater difference in length. The length differences of the genes used in this study are within 100 to 300 codons. Therefore, the codon bias cannot be due to gene length differences; furthermore, FBNYV and MDV genes, with different lengths, were clustered together, which suggests that gene length is not a major factor in this case. Another possible source of codon bias is the nature of overlapping genes in many viruses, but the effect has only been extensively evaluated by a few studies (33, 54, 70). The possible influence of overlapping genes in codon usage should be limited in nanoviruses since most are monopartite for each component; thus, the core genes do not overlap.

We favor a reassortment event as the most plausible cause of codon bias among components of BBTV. This explanation has been demonstrated in prokaryotes, for which multivariate analyses revealed horizontally acquired genes showing distinct codon usage patterns (20, 49). Our results also showed the five identified additional Rep genes in BBTV with a codon usage similar to that of DNA 4 to 6 of BBTV and other nanoviruses, which suggests that they all derived from the same origin.

The combination and reassortment of virus components from different sources, known as a "symbiogenesis" process, have been recently reviewed by Roossinck (60). Mixed infection of viruses in a single host plant is common in the field. In such mix-infected plants, genome reassortment events are fre-

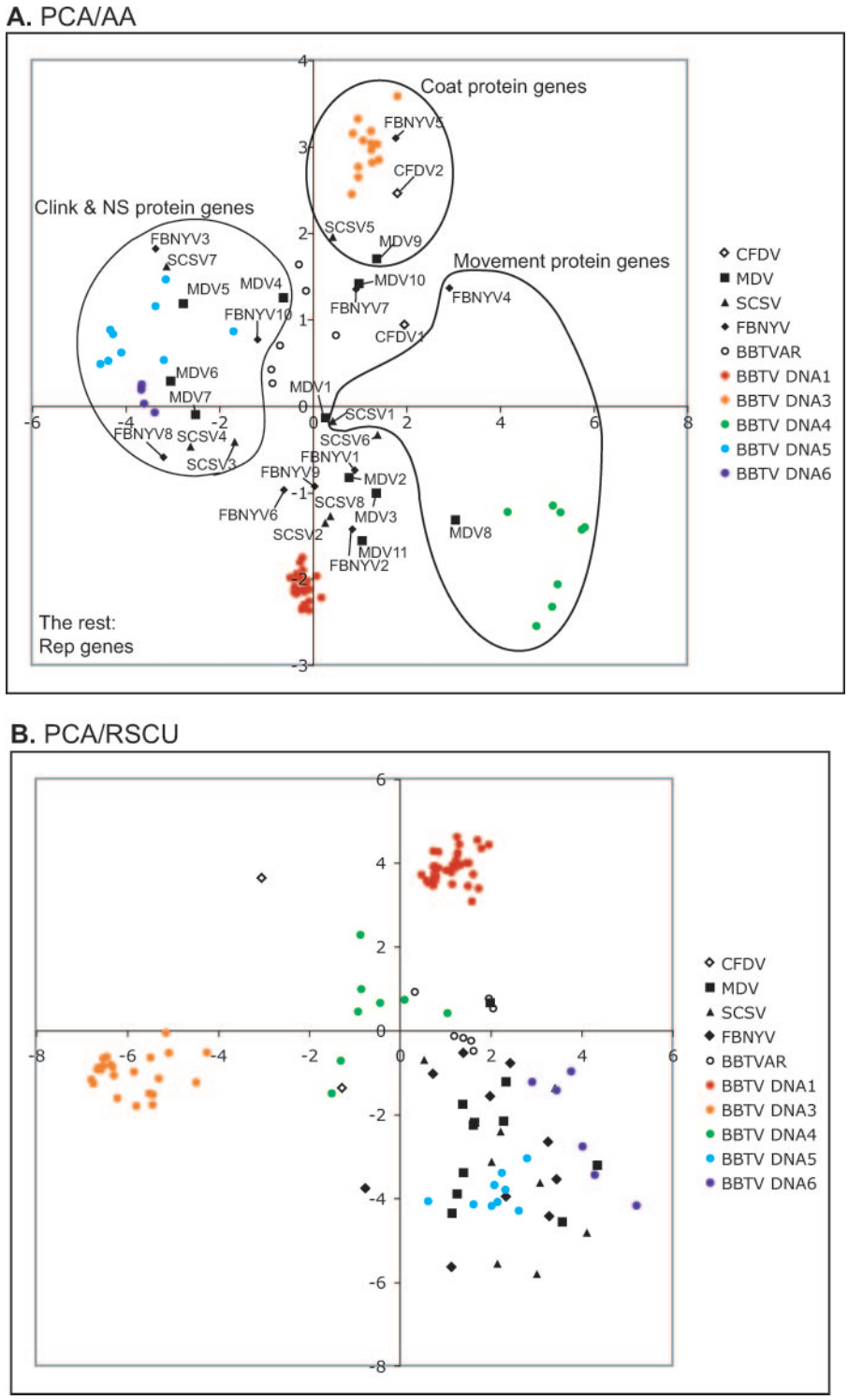

FIG. 7. Result of combined analysis of PCA on amino acid and codon usage values of 132 nanovirus genes. (A) PCA map of the two first factors realized by amino acid usage of gene products. (B) PCA map plotted by the two first factors realized by RSCU of all genes. Different components of BBTV are marked in different colors.

quently observed in viruses with segmented genomes, and these events have a strong impact on virus speciation (44, 60). Examples can be seen in tobravirus strains I6 and N5, containing RNA 1 sequences from *Tobacco rattle virus* and RNA 2 sequences derived from *Pea early browning virus* (59), and *Bean distortion mosaic virus*, which contains RNAs 1 and 2 of *Cucumber mosaic virus* and RNA 3 of *Peanut stunt virus* (86). The potential for genetic reassortment among nanoviruses has

been demonstrated, as the Rep genes of FBNYV, MDV, and SCSV are able to trigger replication of heterologous nanovirus DNAs (80). In addition, several Rep-encoded components associated with nanovirus infection have been discovered. Meanwhile, a replication-competent, nanovirus-like DNA component was found to be associated with *Cotton leaf curl begomovirus* (46, 65). It is believed that this particular nanovirus-like DNA component requires *Cotton leaf curl begomovirus* for encapsidation and transmission. Collectively, this information supports a possible scenario in which a nanovirus Rep gene was introduced into the ancestors of BBTV populations and took over the function of replication. The original Rep genes were able to remain in the BBTV genome only in some isolates and eventually lost the ability to replicate themselves. The case of a mild strain, TW4, demonstrated here is a strong proof that genome reassortment did occur in BBTV.

The results of this study suggest that the traditional approach of using a single component to reconstruct phylogenies in BBTV allows only limited inferences about the evolutionary history of the virus even though it is still informative in comparing isolates from different regions. We demonstrated a special case of genome reassortment in one Taiwanese isolate (TW4) and a concerted evolution in the CR-M of the Pacific group of BBTV. In addition, we demonstrated the possible uses of multivariate analyses of codon and amino usages for nanovirus genes for detecting underlying evolutionary patterns that are not obvious from other analyses; these methods are applicable to other virus genomes with similar situations. Finally, since little information is available for BBTV with full sets of components, we urge that such data be collected for other components of BBTV to better elucidate evolution in nanoviruses.

ACKNOWLEDGMENTS

We thank Nicole Maturen and Laura Heraty for helping edit the manuscript and Yu-Tza Liu for excellent technical assistance in cloning and sequencing.

H.-H.Y. is supported by grants from National Science Council, Taiwan (NSC 92-2313-B-002-066, NSC 93-2313-B-002-112, and NSC 94-2313-B-002-105).

REFERENCES

1. **Allen, R. N.** 1987. Further studies on epidemiological factors influencing control of banana bunchy top disease, and evaluation of control measures by computer simulation. Aus. J. Agric. Res. **38:**373–382.
2. **Avise, J. C.** 2000. Phylogeography. Harvard University Press, Cambridge, MA.
3. **Beetham, P. R., R. M. Harding, and J. L. Dale.** 1999. Banana bunchy top virus DNA-2 to 6 are monocistronic. Arch. Virol. **144:**89–105.
4. **Bell, K. E., J. L. Dale, C. V. Ha, M. T. Vu, and P. A. Revill.** 2002. Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. Arch. Virol. **147:**695–707.
5. **Bryant, D., and V. Moulton.** 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. **21:**255–265.
6. **Burns, T. M., R. M. Harding, and J. L. Dale.** 1995. The genome organization of *Banana bunchy top virus*: analysis of six ssDNA components. J. Gen. Virol. **76:**1471–1482.
7. **Carbone, A., F. Kepes, and A. Zinovyev.** 2005. Codon bias signatures, organization of microorganisms in codon space, and lifestyle. Mol. Biol. Evol. **22:**547–561.
8. **Cunningham, C. W.** 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14:**733–740.
9. **Cunningham, C. W.** 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. **46:**464–478.
10. **Dale, J. L.** 1987. Banana bunchy top: an economically important tropical plant virus disease. Adv. Virus Res. **33:**301–325.
11. **Dolphin, K., R. Belshaw, C. D. L. Orme, and D. L. J. Quicke.** 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. **17:**401–406.
12. **Duret, L., and D. Mouchiroud.** 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc. Natl. Acad. Sci. USA **96:**4482–4487.
13. **Etherington, G. J., J. Dicks, and I. N. Roberts.** 2005. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. Bioinformatics **21:**278–281.
14. **Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult.** 1995. Testing significance of incongruence. Cladistics **10:**315–319.
15. **Felsenstein, J.** 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27:**401–410.
16. **Furuya, N., S. Kawano, and K. T. Natsuaki.** 2005. Characterization and genetic status of banana bunchy top virus isolated from Okinawa, Japan. J. Gen. Plant Pathol. **71:**68–73.
17. **Furuya, N., S. Somowiyarjo, and K. T. Natsuaki.** 2004. Virus detection from local banana cultivars and the first molecular characterization of *Banana bunchy top virus* in Indonesia. J. Agric. Sci. Tokyo Univ. Agric. **49:**75–81.
18. **Futterer, J., I. Potrykus, Y. M. Bao, L. Li, T. M. Burns, R. Hull, and T. Hohn.** 1996. Position-dependent ATT initiation during plant pararetrovirus rice tungro bacilliform virus translation. J. Virol. **70:**2999–3010.
19. **Garcia-Arenal, F., A. Fraile, and J. M. Malpica.** 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. **39:**157–186.
20. **Garcia-Vallve, S., A. Romeu, and J. Palau.** 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res. **10:**1719–1725.
21. **Harding, R. M., T. M. Burns, and J. L. Dale.** 1991. Virus-like particles associated with banana bunchy top disease contain small single-stranded-DNA. J. Gen. Virol. **72:**225–230.
22. **Harding, R. M., T. M. Burns, G. Hafner, R. G. Dietzgen, and J. L. Dale.** 1993. Nucleotide-sequence of one-component of the *Banana bunchy top virus* genome contains a putative replicase gene. J. Gen. Virol. **74:**323–328.
23. **Hasegawa, M., H. Kishino, and T. Yano.** 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22:**160–174.
24. **He, Z. F., H. P. Li, H. G. Xiao, and H. Z. Fan.** 2000. Cloning and sequencing of DNA component 1 of two BBTV strains. Chih Wu Ping Li Hsueh Pao **30:**364–369. [In Chinese.]
25. **He, Z. F., H. P. Li, H. G. Xiao, and H. Z. Fan.** 2001. Cloning and sequencing of DNA component 3 of two BBTV strains. Nongye Shengwu Jisu Xuebao **9:**145–148. [In Chinese.]
26. **He, Z. F., H. P. Li, H. G. Xiao, and H. Z. Fan.** 2001. Cloning and sequencing of DNA component 6 of two BBTV strains. Hua Nan Nung Yeh Ta Hsueh Hsueh Pao **22:**33–36. [In Chinese.]
27. **Hipp, A. L., J. C. Hall, and K. J. Sytsma.** 2004. Congruence versus phylogenetic accuracy: Revisiting the incongruence length difference test. Syst. Biol. **53:**81–89.
28. **Holmes, E. C., M. Worobey, and A. Rambaut.** 1999. Phylogenetic evidence for recombination in Dengue virus. Mol. Biol. Evol. **16:**405–409.
29. **Horser, C. L., R. M. Harding, and J. L. Dale.** 2001. Banana bunchy top nanovirus DNA-1 encodes the "master" replication initiation protein. J. Gen. Virol. **82:**459–464.
30. **Horser, C. L., M. Karan, R. M. Harding, and J. L. Dale.** 2001. Additional Rep-encoding DNAs associated with *Banana bunchy top virus*. Arch. Virol. **146:**71–86.
31. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
32. **Hughes, A. L.** 2004. Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. Mol. Phylogenet. Evol. **30:**287–294.
33. **Hughes, A. L., and M. A. K. Hughes.** 2005. Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes. Virus Res. **113:**81–88.
34. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23:**254–267.
35. **INIBAP: International Network for the Improvement of Banana and Plantain.** 2000. Banana: food for the poor. INIBAP, Montpellier, France.
36. **Jakobsen, I. B., S. R. Wilson, and S. Easteal.** 1997. The partition matrix: Exploring variable phylogenetic signals along nucleotide sequence alignments. Mol. Biol. Evol. **14:**474–484.
37. **Karan, M., R. M. Harding, and J. L. Dale.** 1997. Association of *Banana bunchy top virus* DNA components 2 to 6 with bunchy top disease. Mol. Plant Pathol. http://194.247.68.33/mppol/1997/0624karan/paper.htm.
38. **Karan, M., R. M. Harding, and J. L. Dale.** 1994. Evidence for two groups of *Banana bunchy top virus* isolates. J. Gen. Virol. **75:**3541–3546.
39. **Katul, L., T. Timchenko, B. Gronenborn, and H. J. Vetten.** 1998. Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the *Faba bean necrotic yellows virus* genome. J. Gen. Virol. **79:**3101–3109.
40. **Kishino, H., and M. Hasegawa.** 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidea*. J. Mol. Evol. **29:**170–179.
41. **Kishino, H., T. Miyata, and M. Hasegawa.** 1990. Maximum likelihood infer-

ence of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31:**151–160.
42. **Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur.** 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Mol. Biol. Evol. **15:**1685–1695.
43. **Li, W. H.** 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:**337–344.
44. **MacFarlane, S. A.** 1997. Natural recombination among plant virus genomes: evidence from tobraviruses. Semin. Virol. **8:**25–31.
45. **Magee, C. J. P.** 1940. Transmission studies on the banana bunchy top virus. J. Austr. Inst. Agric. Sci. **6:**109–110.
46. **Mansoor, S., S. H. Khan, A. Bashir, M. Saeed, Y. Zafar, K. A. Malik, R. Briddon, J. Stanley, and P. G. Markham.** 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. Virology **259:**190–199.
47. **Mason-Gamer, R. J., and E. A. Kellogg.** 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. **45:**524–545.
48. **Mattick, J. S.** 2005. The functional genomics of noncoding RNA. Science **309:**1527–1528.
49. **Medigue, C., T. Rouxel, P. Vigier, A. Henaut, and A. Danchin.** 1991. Evidence for horizontal gene-transfer in *Escherichia coli* speciation. J. Mol. Biol. **222:**851–856.
50. **Moffat, A. S.** 2001. Finding new ways to fight plant diseases. Science **292:** 2270–2273.
51. **Moriyama, E. N., and J. R. Powell.** 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. **26:**3188–3193.
52. **Naya, H., A. Zavala, H. Romero, H. Rodriguez-Maseda, and H. Musto.** 2004. Correspondence analysis of amino acid usage within the family *Bacillaceae*. Biochem. Biophys. Res. Commun. **325:**1252–1257.
53. **Noueiry, A. O., W. J. Lucas, and R. L. Gilbertson.** 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell **76:** 925–932.
54. **Pavesi, A., B. DeIaco, M. I. Granero, and A. Porati.** 1997. On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. J. Mol. Evol. **44:**625–631.
55. **Peden, J. F.** 1997. Codon W: codon usage analysis package, 1.3 ed. Oxford University Molecular Biology Data Centre, Oxford, United Kingdom.
56. **Perriere, G., and J. Thioulouse.** 2002. Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. **30:**4548–4555.
57. **Planet, P. J.** 2006. Tree disagreement: measuring and testing incongruence in phylogenies. J. Biomed. Inform. **39:**86–102.
58. **Posada, D., K. A. Crandall, and E. C. Holmes.** 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. **36:**75–97.
59. **Robinson, D. J., W. D. O. Hamilton, B. D. Harrison, and D. C. Baulcombe.** 1987. Two anomalous tobravirus isolates: evidence for RNA recombination in nature. J. Gen. Virol. **68:**2551–2561.
60. **Roossinck, M. J.** 2005. Symbiosis versus competition in plant virus evolution. Nat. Rev. Microbiol. **3:**917–924.
61. **Salemi, M., T. De Oliveira, V. Courgnaud, V. Moulton, B. Holland, S. Cassol, W. M. Switzer, and A. M. Vandamme.** 2003. Mosaic genomes of the six major primate lentivirus lineages revealed by phylogenetic analyses. J. Virol. **77:**7202–7213.
62. **Sanderfoot, A. A., and S. G. Lazarowitz.** 1995. Cooperation in viral movement: the geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. Plant Cell **7:**1185–1194.
63. **Sano, Y., M. Wada, Y. Hashimoto, T. Matsumoto, and M. Kojima.** 1998. Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. J. Gen. Virol. **79:**3111–3118.
64. **Sanz, A. I., A. Fraile, J. M. Gallego, J. M. Malpica, and F. Garcia-Arenal.** 1999. Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. J. Mol. Evol. **49:**672–681.
65. **Saunders, K., and J. Stanley.** 1999. A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: evidence for interfamilial recombination between plant DNA viruses. Virology **264:**142–152.
66. **Schmitz, J., D. Prufer, W. Rohde, and E. Tacke.** 1996. Non-canonical translation mechanisms in plants: efficient in vitro and in planta initiation at AUU codons of the tobacco mosaic virus enhancer sequence. Nucleic Acids Res. **24:**257–263.
67. **Sharp, P. M., and W.-H. Li.** 1986. An evolutionary perspective on codon usage in unicellular organisms. J. Mol. Evol. **24:**28–38.
68. **Sharp, P. M., M. Stenico, J. F. Peden, and A. T. Lloyd.** 1993. Codon usage: mutational bias, translational selection, or both? Biochem. Soc. Trans. **21:** 835–841.
69. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16:**1114–1116.
70. **Smith, T. F., and M. S. Waterman.** 1980. Protein constraints induced by multiframe encoding. Math. Biosci. **49:**17–26.
71. **Su, H. J., L. Y. Tsao, and T. H. Hung.** 1998. Pathological and molecular characterization of *Banana bunchy top virus* (BBTV) strains in Asia, p. 164. *In* A. B. Molina, V. A. Roa, J. Bay-Petersen, A. T. Carpio, and J. E. A. Joven (ed.), Managing banana and citrus diseases, INIBAP, Montpellier, France.
72. **Su, H. J., L. Y. Tsao, M. L. Wu, and T. H. Hung.** 2003. Biological and molecular categorization of strains of *Banana bunchy top virus*. J. Phytopathol. **151:**290–296.
73. **Swofford, D. L.** 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0beta10. Sinauer Associates, Inc., Sunderland, MA.
74. **Swofford, D. L., G. K. Olsen, P. J. Waddell, and D. M. Hillis.** 1996. Phylogeny reconstruction, p. 407–514. *In* D. M. Hillis, C. Moritz, and B. K. Mable (ed.), Molecular systematics, 2nd ed. Sinauer Associates, Sunderland, MA.
75. **Thioulouse, J., D. Chessel, S. Dolédec, and J.-M. Olivier.** 1997. ADE-4: a multivariate analysis and graphical display software. Stat. Computing **7:**75–83.
76. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL-X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
77. **Tien, E., J. Zhuang, and Z. X. Liu.** 2004. Cloning and sequencing of DNA components of *Banana bunchy top virus* Hainan isolate. J. Agric. Biotechnol. **12:**680–684.
78. **Timchenko, T., F. de Kouchkovsky, L. Katul, C. David, H. J. Vetten, and B. Gronenborn.** 1999. A single Rep protein initiates replication of multiple genome components of *Faba bean necrotic yellows virus*, a single-stranded DNA virus of plants. J. Virol. **73:**10173–10182.
79. **Timchenko, T., L. Katul, M. Aronson, J. C. Vega-Arreguin, B. C. Ramirez, H. J. Vetten, and B. Gronenborn.** 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of *Faba bean necrotic yellow virus*. J. Gen. Virol. **87:**1735–1743.
80. **Timchenko, T., L. Katul, Y. Sano, F. de Kouchkovsky, H. J. Vetten, and B. Gronenborn.** 2000. The master Rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology **274:**189–195.
81. **van Cuyck, H., J. Fan, D. L. Robertson, and P. Roques.** 2005. Evidence of recombination between divergent hepatitis E viruses. J. Virol. **79:**9306–9314.
82. **Wanitchakorn, R., G. J. Hafner, R. M. Harding, and J. L. Dale.** 2000. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. J. Gen. Virol. **81:**299–306.
83. **Wanitchakorn, R., R. M. Harding, and J. L. Dale.** 1997. *Banana bunchy top virus* DNA-3 encodes the viral coat protein. Arch. Virol. **142:**1673–1680.
84. **Wanitchakorn, R., R. M. Harding, and J. L. Dale.** 2000. Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. Arch. Virol. **145:**593–602.
85. **Wheeler, W. C.** 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics **6:**363–368.
86. **White, P. S., F. Morales, and M. J. Roossinck.** 1995. Interspecific reassortment of genomic segments in the evolution of cucumoviruses. Virology **207:**334–337.
87. **Wu, R. Y., and H. J. Su.** 1990. Purification and characterization of banana bunchy top virus. J. Phytopathol. **128:**153–160.
88. **Wu, R. Y., L. R. You, and T. S. Soong.** 1994. Nucleotide-sequences of 2 circular single-stranded DNAs associated with *Banana bunchy top virus*. Phytopathology **84:**952–958.
89. **Xie, W. S., and J. S. Hu.** 1995. Molecular-cloning, sequence-analysis, and detection of *Banana bunchy top virus* in Hawaii. Phytopathology **85:**339–347.
90. **Xu, C. X., H. P. Li, and H. G. Xiao.** 2001. Cloning and sequencing of DNA component 4 of BBTV NS strain. Chin. J. Virol. **17:**288.
91. **Yeh, H. H., H. J. Su, and Y. C. Chao.** 1994. Genome characterization and identification of viral-associated dsDNA component of *Banana bunchy top virus*. Virology **198:**645–652.
92. **Zheng, Y., H. P. Li, H. G. Xiao, and H. Z. Fan.** 2005. Cloning and nucleotide sequence analysis of DNA-5 of *Banana bunchy top virus* NSP strain. Acta Phytopathol. Sinica **35:**289–292.

JOURNAL OF VIROLOGY, June 2008, p. 6088 Vol. 82, No. 12
0022-538X/08/$08.00+0 doi:10.1128/JVI.00758-08

AUTHORS' CORRECTION

Reassortment and Concerted Evolution in *Banana Bunchy Top Virus* Genomes

Jer-Ming Hu, Hui-Chuan Fu, Chia-Hua Lin, Hong-Ji Su, and Hsin-Hung Yeh

Institute of Ecology and Evolutionary Biology, National Taiwan University, and Department of Plant Pathology and Microbiology, National Taiwan University, Taipei 106, Taiwan

Volume 81, no. 4, p. 1746–1761, 2007. Pages 1748 and 1749, Table 1: Some accession numbers listed in Table 1 should be corrected as shown below.

Accession no.	Should read:
TaiDQ817617	TaiDQ826390
853027	TaiEF095162
857902	TaiEF095161
TaiDQ817867	TaiDQ826391
TaiDQ825708	TaiDQ826392
853087	TaiEF095163
TaiDQ817893	TaiDQ826393
857918	TaiEF095164
TaiDQ825714	TaiDQ826394
853095	TaiEF095165
TaiDQ817921	TaiDQ826395
857944	TaiEF095166
TaiDQ825730	TaiDQ826396