Rice black-streaked dwarf virus Genome in China: Diversification, Phylogeny, and Selection

Yu Zhou, College of Agronomy, Northeast Agricultural University, Harbin, Heilongjiang Province, China; and Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China; **Lin Zhang**, **Xiaoming Zhang**, **Hongyue Zu**, **Hong Di**, **Ling Dong**, **Xianjun Liu**, and **Xing Zeng**, College of Agronomy, Northeast Agricultural University, Harbin, Heilongjiang Province, China; **Jianfeng Weng**, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China; **Zhenhua Wang**,[†] College of Agronomy, Northeast Agricultural University, Harbin, Heilongjiang Province, China; and **Xinhai Li**,[†] Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China

Abstract

Rice black-streaked dwarf virus (RBSDV), a *Fijivirus*, causes maize rough dwarf disease and rice black-streaked dwarf disease in the summer maize-growing regions of the Yellow and Huai rivers, respectively, in China. Nevertheless, the diversification and selection of the entire genome from S1 to S10 have not been illuminated. Molecular variation, evolution, conserved regions, and other genomic properties were analyzed in 21 RBSDV isolates from maize (*Zea mays* L.) and rice (*Oryza sativa*) hosts sampled from nine geographic locations in China. Low codon adaptation index values ranging from 0.1878 to 0.2918 indicated a low degree of codon-usage bias and low potential expression for all 13 RBSDV open reading frames (ORFs). ORF9-2 showed a stronger effect of codon usage bias than did other ORFs, as the majority of points for this ORF lay close to the standard curve in the *Nc* plot (the effective number of codons [*Nc*] versus the frequency of G+C at synonymous third-base positions [GC3]). A 9-bp deletion mutation was detected in the RBSDV genome in the 3' UTR of S8. Nucleotide diversity analysis indicated that the structural proteins of RBSDV, such as S2 and S4, were all more conserved than nonstructural proteins such as S9. Nucleotide diversity (π) was highest among S9 sequences (0.0656), and was significantly higher than among S4 sequences

(0.0225, $P < 0.01$). The number of conserved regions among the 10 segments varied substantially. The highest number of conserved regions (5) was found in S5, whereas no conserved regions were identified in S9. Nucleotide diversity and the number of conserved regions were independent of the lengths of segments. Nucleotide diversity was also not correlated with the number of conserved regions in segments. Ten recombination events in 21 isolates were found in seven segments with breakpoint positions in UTRs, intergenic spacer regions, and gene coding regions. The number of recombination events was also independent of the lengths of segments. RBSDV isolates from China could be phylogenetically classified into two groups using either 10 segment sequences or the concatenated sequence of S1 through S10, regardless of host or geographical location. The phylogenetic tree generated from pairwise nucleotide identities of individual RBSDV segments such as S9 and S3, with nucleotide identity values of 93.74% and 95.86%, respectively, is similar to the tree constructed from the concatenated sequences of the entire RBSDV genome. The 13 RBSDV ORFs were under negative and purifying selection ($Ka/Ks < 1$). ORF5-2 was under the greatest selection pressure; however, ORF2, which encodes the core protein of RBSDV, was under the lowest selection pressure.

Maize rough dwarf disease (MRDD), which is caused by viruses in the genus *Fijivirus* in the family *Reoviridae*, is a disease of maize (*Zea mays* L.) that results in severe yield losses worldwide, particularly in the Yellow and Huai river summer maize-growing regions of China (Dovas et al. 2004; Harpaz 1959; Meng et al. 2008; Shi et al. 2013). The main causative agents of MRDD are Southern rice black-streaked dwarf virus (SRBSDV) in the southern region, such as Hainan and Guangdong provinces, and Rice black-streaked dwarf virus (RBSDV) in the Yellow and Huai river summer maize-growing regions of China, which can also cause rice black-streaked dwarf disease (RBSDD) (Bai et al. 2002; Fang et al. 2001; Li et al. 2012).

The entire genome of RBSDV is composed of 10 linear double-stranded RNAs (dsRNAs), which have been designated segment 1 (S1) through segment 10 (S10) and range in size from 1.8 to 4.5 kb (Li et al. 2012, 2013; Milne et al. 1973; Z. H. Wang et al. 2003; Zhang et al. 2001). A total of 13 proteins ranging in length from 203 to 1,464 amino acids are encoded by these 10 dsRNAs. Three of the RBSDV segments, S5, S7, and S9, each have two open reading frames (ORFs).

ORF5-1 and ORF5-2 are the first and second ORFs of S5, ORF7-1 and ORF7-2 are the first and second ORFs of S7, and ORF9-1 and ORF9-2 are the first and second ORFs of S9, respectively. The other seven segments each contain only one ORF designated ORF1, ORF2, ORF3, ORF4, ORF6, ORF8, and ORF10 (Azuhata et al. 1993; Firth and Atkins 2009; Isogai et al. 1998). Molecular variation, recombination, genetic structure, and evolution have been analyzed in S1 using 121 RBSDV isolates from Jinling and other regions of China (Y. Zhou et al. 2016), in S7 using 111 RBSDV isolates from eight regions in China (Y. Zhou et al. 2015b), in S8 and S10 using 120 RBSDV isolates from maize and rice from Shandong and the neighboring provinces of Anhui and Jiangsu in China (Yin et al. 2013), and in S9 using 49 RBSDV isolates from maize and rice hosts from various regions in China (Y. Zhou et al. 2015a). Nevertheless, molecular variation, conserved regions, and phylogenies have not yet been reported for the entire RBSDV genome, which can be used to detect the diversification and phylogeny of the concatenated S1 through S10 sequences and differences among the 10 segments of RBSDV.

The appearance of synonymous codons at unequal frequencies in protein sequences is known as codon usage bias (CUB). Knowledge of the CUB of viruses can help to understand the evolution of RBSDV, particularly to detect the difference CUBs among these 10 segments and the immune responses of maize or rice hosts. CUB has been studied in eukaryotes (Diamant et al. 2014), prokaryotes, and human and animal viruses (Pintó et al. 2007). CUB has also been analyzed in some plant viruses, such as in sobemovirus using the genome sequences of 14 isolates (H. Zhou et al. 2005), in *Soybean dwarf virus* (SDV) (Kyrychenko et al. 2012), and in *Citrus tristeza virus* (CTV) using the full coding region (Cheng et al. 2012). So far, codon usage bias has been detected only in RBSDV S7 in an analysis of 111

[†]Corresponding authors: Xinhai Li, lixinhai@caas.cn; Zhenhua Wang, zhenhuawang_2006@163.com

Yu Zhou and Lin Zhang contributed equally to this research.

*The e-Xtra logo stands for “electronic extra” and indicates that two supplementary tables are published online.

Accepted for publication 22 May 2017.

isolates. CUB analysis in RBSDV S7 showed that all CUB parameters for S7-2 were significantly lower than those for S7-1 ($P < 0.01$) (Y. Zhou et al. 2015b). However, CUB analyses have not been reported for the remaining RBSDV ORFs.

Previous studies indicated that RBSDV populations can be classified into two groups based on the sequences of S1 (Y. Zhou et al. 2016), S7 (Y. Zhou et al. 2015b), S9 (Y. Zhou et al. 2015a), and S10 (Li et al. 2012; Yin et al. 2011), or into three groups based on S8 sequences (Yin et al. 2011), independent of host or geographical origin (Yin et al. 2011; Y. Zhou et al. 2015a, b). But the year in which these isolates were collected influences the grouping of RBSDV isolates to some degree (Y. Zhou et al. 2015b). Phylogenetic analyses have been conducted for each separate 10 RBSDV segment. But phylogenetic analysis of the entire RBSDV genome has not been performed for all 10 segments of RBSDV nor for the concatenated sequences from S1 to S10. So the phylogeny of the entire RBSDV genome was examined to gain insights into the evolution of this virus and the functions of its encoded proteins.

The evolutionary mechanism of genome dynamics and the genetic stability of virus populations can be understood by the analysis of complete genomic sequence data of viral populations (Seo et al. 2009). The molecular evolutionary histories related to emergence, dispersion, and virulence of epidemics will be explored by analyzing diversification, selection, and phylogeny (Kim et al. 2014; Y. Zhou et al. 2017) in order to explain the hypothesis that the variation and evolutionary mechanisms of each of the 10 segments of RBSDV genome are different. In the present study, CUB, molecular variation and nucleotide diversity, conserved sequences, and phylogenies were constructed for all 10 segments of the RBSDV genome using 21 isolates of maize and rice hosts collected from nine geographic locations in China, which only have been used in the evolutionary analysis on S1 (Y. Zhou et al. 2016), S7 (Y. Zhou et al. 2015b), and S9 (Y. Zhou et al. 2015a) as contrast or partial isolates. These results provide further insights into the variation and evolution of RBSDV based on the molecular characterization of the entire RBSDV genome.

Materials and Methods

Sampling of virus isolates. A total of 21 maize or rice plants with symptoms of MRDD or RBSD were collected from nine locations in which these diseases prevailed from 2012 to 2013 (Supplementary Table S1). Virus-infected samples included two maize plants collected from Beijing (I) in 2012 and 2013; one maize plant from Tangshan (II), Hebei Province in 2013; one maize and one rice plant from Baoding, Hebei Province (III); one maize plant from Jinan, Shandong Province (IV); three maize and three rice plants from Jining, Shandong Province (V); two maize plants and one rice plant from Zhengzhou, Henan Province (VI); three maize plants from Yancheng, Jiangsu Province (VII); one maize and one rice plant from Nanjing, Jiangsu Province (VIII); and one maize plant from Xinyang, Henan Province (IX). All leaf samples were frozen in liquid nitrogen and stored at -80°C . A total of 21 isolates from nine geographic locations (from I through IX) were processed and used for analyses of RBSDV sequences S1 through S10.

RNA extractions, RT-PCR, and sequencing. RBSDV dsRNA was extracted from the 21 individual maize and rice isolates following the methods described previously (Dodds et al. 1984; Morris and Dodds 1979; Z. H. Wang et al. 2003). The quality and integrity of dsRNAs were assessed using 1.2% native agarose gels, and a NanoDrop 2000 spectrophotometer (Thermo Scientific, U.S.A.) was used to estimate dsRNA concentrations. First-strand cDNA was synthesized using a Fast Quant RT Kit (TIANGEN, China), and PCR products were amplified in reactions using KOD-Plus-Neo enzyme (TOYOBO, Japan) with 26 pairs of primers specific to S1 through S10 (Supplementary Table S2). These PCR products were then sequenced using the dideoxy chain-termination method at the AuGCT DNA-SYN Biotechnology Company (Beijing, China). To confirm sequence quality, three independent PCR reactions were sequenced for each segment. DNAMAN and Jembooss v. 1.5 (EMBOSS, Cambridge, U.K.) software (Carver and Mullan

2005) were used to assemble and analyze sequence data for these 10 RBSDV segments. The raw data have been uploaded into the NCBI GenBank database. The GenBank accession numbers of these sequences were from KX557466 to KX557486 and from KY662060 to KY662246.

Codon usage bias in the RBSDV genome. CUB was assessed using the program CodonW 1.4.4 (<http://sourceforge.net/projects/codonw/>). Base compositions, including GC, GC3, A3, U3, G3, and C3, of all sequenced RBSDV ORFs were analyzed. GC indicates the frequency of GC contents in all ORFs. GC3, A3, U3, G3, and C3 denote the frequencies of G+C, A, U, G, or C at synonymous third-base positions, respectively.

The codon adaptation index (*CAI*), the codon bias index (*CBI*), and the effective number of codons (*N_c* value) were calculated for the 13 RBSDV ORFs. *CAI*, with values from zero to one, represents the extent of codon bias in expressed genes. A *CAI* value of one indicates potential expression level and a high degree of CUB (Sharp and Li 1987). Proportions of preferred codons were estimated using *CBI* values. A nonrandom process in which only preferred codons are used for all triplets is indicated by a *CBI* value of one (Bennetzen and Hall 1982). *N_c* values range from 20 (when only one codon is used per amino acid) to 61 (when all possible codons are used equally). Low *N_c* values indicate highly expressed genes that tend to have a high degree of codon bias (Sharp and Cowe 1991).

Sequence variants and diversity in RBSDV. The nucleotide and amino acid sequences of S1 through S10 from these 21 RBSDV isolates were aligned using MEGA 6.06 software (Ma et al. 2013). Sequence variants, variable sites, haplotype diversity (a measure of the uniqueness of a particular haplotype in a given population), and conserved regions of these 10 RBSDV sequences were estimated using the software DnaSP version 5.0 (Librado and Rozas 2009). The conserved regions were detected by the procedure of “conserved DNA regions” which selected “dynamic defined” and “given the observed S” options. Sliding-window analysis of nucleotide diversities (π) (a concept in molecular genetics which is used to measure the degree of polymorphism within a population) in each segment of the RBSDV genome was performed with 200-bp windows in 100-bp steps in TASSEL 3.0 (Weng et al. 2013).

Genetic recombination and phylogenetic analyses in 10 RBSDV segments. The nucleotide sequences of these 10 RBSDV segments were aligned using CLUSTAL W in MEGA 6.06 with default settings (Tamura et al. 2013). Possible recombination sites in each of the 10 sequences were examined using the RDP, GENECONV, BOOTSCAN, Maximum Chi SQUARE (MAXCHI), CHIMAERA, Sister Scanning (SISCAN), and 3Seq algorithms in RDP version 4.22 (Martin et al. 2010) set to the default configurations, except that the “linear sequence” and “disentangling overlapping signals” options were selected. Recombination events were validated only if they were detected by more than four of these algorithms with a *P*-value cutoff of 0.05 and the default number (100) of simulated datasets.

Phylogenetic trees for all 10 RBSDV segment sequences from the 21 isolates were constructed separately using the maximum likelihood method in MEGA 6.06 software (Tamura et al. 2013). A single phylogenetic tree was also constructed using the concatenated S1 through S10 sequences for all 21 isolates. The number of bootstrap replicates was set to 1,000 and only replicates with bootstrap values greater than 70% are shown.

Detection of selection pressure on RBSDV. The levels of selection pressure on each segment of RBSDV were estimated as the *K_a/K_s* ratio, where the average number of nonsynonymous substitutions per nonsynonymous site is represented by *K_a* and the average number of synonymous substitutions per synonymous site is represented by *K_s*. MEGA 6.06 (Tamura et al. 2013) was used to calculate the average values of *K_a* and *K_s* as was done in our previous studies (Y. Zhou et al. 2015a, b). Genes have been under negative or purifying selection if their *K_a/K_s* ratio is less than one. However, genes have likely been under positive or diversifying selection if their *K_a/K_s* ratio is greater than one.

Results

Genome-wide codon usage in RBSDV. The GC contents of the RBSDV segment sequences S1 through S10 varied from 30.24 to 37.84%, among which there were significant differences ($P < 0.01$) (Fig. 1a). There were also significant differences in T3, C3, A3, G3, and GC3 percentages, and estimates of *Nc*, *CAI*, and *CBI* among RBSDV segment sequences S1 through S10 ($P < 0.05$). T3 percentages were the highest in all segments except ORF5-2 (T3, 29.49%), and ranged from 34.89 to 47.42%. ORF5-2 had the highest percentage of A3 at 31.15%. The GC3 percentages in all 13 ORFs ranged from 20.83% in ORF1 to 38.20% in ORF5-2. The *CAI* estimates for all 13 ORFs were closer to zero than to one, with values ranging from 0.1878 to 0.2918. This indicates that the 13 RBSDV ORFs all had relatively low CUB and potential expression. Analysis of these ORFs, excluding ORF6, ORF9-1, and ORF10, showed that nonpreferred codons appeared more often than expected, as indicated by the negative *CBI* values (Fig. 1a).

Average *Nc* values for ORF1 through ORF10 ranged from 39.89 to 54.73, with significant differences among ORFs ($P < 0.01$). About half of the points lie on or above the standard curve in the *Nc* plot (a plot of *Nc* versus GC3s) for ORF9-2, but the remaining points lie below the standard curve in the *Nc* plots of the remaining ORFs (Fig. 1b). These results suggest the occurrence of CUB in most RBSDV ORFs. Most points for ORF9-2 lie close to the standard curve, which indicates that mutational bias had a stronger effect on codon usage variation in ORF9-2 than in the other ORFs. Points for ORF7-2 and ORF6 were far away from the standard curve, which suggests weak codon usage bias in these two ORFs (Fig. 1b).

Molecular diversification in the entire RBSDV genome. In the present study, 15 maize isolates with typical MRDD symptoms and six rice isolates with typical RBSDD symptoms were collected from nine locations in 2012 and 2013. A 9-bp deletion mutation was detected in S8, but no other deletions or insertions were found in other segments. A total of either 3,402 or 3,404 nucleotide mutation sites, including 1,183 singleton variable sites and either 2,306 or 2,308 parsimony informative sites, were detected in the entire RBSDV genome sequence across these 21 viral isolates. There are two values for mutation sites due to the 9-bp deletion mutation in S8. The average number of nucleotide variants (10) was greatest in S2, but lowest (6) in S9. The number of parsimony informative sites was greater than the number of singleton variable sites among the 10 RBSDV segments. In S9, the number of parsimony informative sites was 3.09-fold greater than the number of singleton variable sites. However, the number of parsimony informative sites was 1.16-fold higher than the number of singleton variable sites in S4.

Point mutations represented the majority of mutation types in nucleotide sequences, from 93.73% in S9 to 96.11% in S7. Transition mutations occurred in higher proportions (from 76.87% in S1 to 85.92% in S3) than did transversions (from 8.69% in S3 to 17.43% in S1). Transitions between pyrimidines occurred at frequencies from 47.72% in S1 to 59.60% in S2 (Fig. 2). A total of 379 amino acid changes were detected in S1 through S10. P5-2 showed the highest degree of variation in amino acid sequence among isolates, with an average of one mutation site per 10 amino acids, whereas P2 showed the lowest level of amino acid variation, with an average of one mutation site per 45 amino acids. The levels of amino acid variation in different ORFs even within the same segment could differ (Table 1).

Analysis of sequence conservation in RBSDV. Nucleotide diversity and conserved regions were analyzed in the 10 segments of these 21 RBSDV isolates. The relative nucleotide diversity values of these 10 RBSDV segments ranked from highest to lowest were S9 > S3 > S6 > S10 > S5 > S1 > S8 > S7 > S2 > S4. The nucleotide diversity of S9 ($\pi = 0.0656$) was significantly higher ($P < 0.01$) than that of S4 ($\pi = 0.0225$). The greatest nucleotide diversity ($\pi = 0.0977$) was detected from bp position 300 to 500 in the S9 sequence (Fig. 3). The haplotype diversity values for S1 through S9 were all 1.0, and that for S10 was 0.9950.

Nucleotide diversity and the number of conserved regions were independent of the lengths of segments. And the number of conserved

regions was not correlated with the nucleotide diversity of segments. In S9, no conserved region was detected. However, there were five conserved regions in S5. The most conserved region was detected in a region from 2,400 to 2,700 bp in S5 that included the overlapping regions of ORF5-1 and ORF5-2. Three conserved regions were detected in each of S1, S6, and S8. Otherwise, two conserved regions were detected in each of S3, S4, and S10. Only one conserved region was detected in each of S2 and S7 (Fig. 3).

Recombination across the RBSDV genome. A total of 10 recombination events were detected in segments S1, S2, S3, S5, S6, S8, and S9 in these 21 isolates. Two recombination events each occurred in S2 (S2_13VR-1 and S2_13IIM-1), S6 (S6_13VM-2 and S6_13VM-3), and S9 (S9_12IM-1 and S9_13VIIM-2). Only one recombination event each was detected in S1 (S1_13VR-2), S3 (S3_13VR-2), S5 (S5_13VIIM-2), and S8 (S8_13VIIM-1) (Table 2). No recombination events were detected in S4, S7, or S10. The number of recombination events was independent of the lengths of segments. About 59% of recombination events were detected in isolates collected from Jinan in Shandong provinces. Among the 10 observed recombination events, four (S1_13VR-2, S2_13VR-1, S6_13VM-2, and S6_13VM-3) begin at breakpoint positions within 5' UTRs. One breakpoint position was detected in the 3' UTR (S5_13VIIM-2) and one was detected in an intergenic spacer region (IGS) (S9_13VIIM-2). All other observed breakpoint positions were detected within coding regions.

Phylogenetic analyses of different RBSDV segments. Ten separate phylogenetic trees were constructed from the RBSDV nucleotide sequences of S1 through S10 to determine the evolutionary relationships among these 21 RBSDV isolates. Based on the sequences of these 10 segments, these 21 isolates could be classified into two main groups designated A and B, independent of host or geographical origin. However, different numbers of isolates clustered into group A or B for each segment. Isolate 13VM-3 was classified into group B based on the sequences of eight segments (S1, S2, S3, S4, S6, S7, S9, and S10). Three isolates, 13IVM-1, 13VM-1, and 13VM-2, were classified into group B based on the sequences of seven segments (S2, S3, S5, S7, S8, S9, and S10 of 13IVM-1; S1, S2, S3, S6, S7, S9, and S10 of 13VM-1; and S1, S2, S3, S4, S7, S9, and S10 of 13VM-2). Two isolates, 13IIM-1 and 13VIR-1, were classified into group B based on the sequences of six segments (S2, S3, S6, S7, S9, and S10 of 13IIM-1; and S1, S2, S3, S6, S7, and S9 of 13VIR-1). Two isolates, 13VR-1 and 13VR-2, were classified into group B based on the sequences of five segments (S3, S4, S7, S9, and S10 of 13VR-1; and S1, S2, S3, S8, and S9 of 13VR-2). One isolate, 13IXM-1, was classified into group B based on the sequences of four segments (S2, S3, S7, and S9 of 13VR-2) (Fig. 4a–j).

To understand the evolutionary relationships among RBSDV isolates based on the entire genome of the virus, a single phylogenetic tree was constructed from the concatenated segment sequences of each of these 21 isolates (Fig. 4k). These 21 isolates also could be classified into two main groups, designated A and B. Group A included nine maize isolates and three rice isolates sampled from six locations including Beijing, Baoding, Jinan, Zhengzhou, Yancheng, and Nanjing. Group B included six maize isolates and three rice isolates from five geographic locations including Tangshan, Jinan, Jinan, Zhengzhou, and Xinyang.

Detection of selection pressure in S1 through S10. To detect selection pressure on RBSDV, the ratios of non-synonymous to synonymous sites (*Ka/Ks*) were calculated for the entire genome from ORF1 through ORF10. The *Ka/Ks* ratios suggest that the 13 ORFs of RBSDV were under negative and purifying selection (Table 3). The *Ka/Ks* ratios for ORF5-1 and ORF5-2 on the same segment even differed significantly ($P < 0.01$), with *Ka/Ks* ratios of 0.0671 and 0.1295, respectively. That was caused primarily by the lower *Ks* value (0.0865) for ORF-2, which was not in the same level of *Ks* value of other 12 ORFs. ORF5-2 had the highest *Ka/Ks* ratio (0.1295), significantly higher ($P < 0.01$) than that of ORF2 with the lowest *Ka/Ks* ratio (0.0126). This result suggests that these 13 ORFs each experienced different levels of selection pressure, and that ORF5-2 was under the least purifying selection pressure

Fig. 1. Codon usage characteristics in the entire RBSDV genome from S1 through S10. **A**, Values for N_c , CAI, CBI, GC3s, and GC are shown for S1 to S10 in 21 RBSDV isolates. **B**, Distribution of effective number of codons (N_c) and GC3 for 10 segments. N_c = the effective number of codons, CAI = the codon adaptation index, CBI = the codon bias index, GC3 = the frequency of G+C at synonymous third-base positions, and GC = the proportion of GC contents.

than other ORFs. ORF2 experienced the most purifying selection pressure compared with the other 12 ORFs.

Discussion

Analyses of codon usage have been reported in the *Flaviviridae* (Lobo et al. 2009), SDV (Kyrychenko et al. 2012), and sobemovirus (H. Zhou et al. 2005). So far in the *Reoviridae*, CUB has only been reported for RBSDV S7 (Y. Zhou et al. 2015b). Previous studies revealed a low degree of CUB in RBSDV S7; thus the major force driving codon usage bias in this virus might be mutational bias. Analysis of CUB in the 10 segments of the RBSDV genome showed that codon usage bias other than GC3 occurred in ORF9-2. However, mutational bias appears to be the major force driving codon usage in the other 12 RBSDV ORFs, consistent with results of previous studies on human RNA viruses (Jenkins and Holmes 2003), CTV (Cheng et al. 2012), and S7 of RBSDV (Y. Zhou et al. 2015b). The major forces driving codon usage in particular individual RBSDV ORFs

might differ, which may be because of the different secondary structure or function of these ORFs.

Variation in nucleotide and amino acid sequences did not depend on the lengths of segments. A single 9-bp deletion that did not affect the amino acid coding sequence was detected in the 3' UTR of RBSDV S8. This result indicated that point mutations are the main source of variation in RBSDV. Previous studies have analyzed the nucleotide diversity of S9 (Y. Zhou et al. 2015a), S7 (Y. Zhou et al. 2015b), and S1 (Y. Zhou et al. 2016); the result is consistent with this result. However, analysis of nucleotide diversity had not previously been reported for the entire genome of RBSDV with the differential analysis among these 10 segments. Here, nucleotide diversity and the number of conserved regions were independent of RBSDV segment lengths. The most conserved region was detected from 2,400 bp to 2,700 bp in RBSDV S5. This result is the first example of a conserved region detected by analysis of nucleotide diversity in RBSDV. No conserved regions were detected in S9, but conserved regions

Fig. 2. Analysis of nucleotide variation in S1 through S10 sequences of 21 RBSDV isolates collected in nine locations in China.

Table 1. Nucleotide and amino acid variations in RBSDV S1 through S10

Name of segment	Total length of segment (bp)	Number of variations in nucleotide	Avg. number variations in nucleotide	Singleton variable sites	Parsimony informative site	Number of amino acids	Number of variations in amino acid	Avg. number of variations in amino acid	Nucleotide diversity	Haplotype diversity
S1	4,501	614	7	164	450	1,464	84	17	0.0385 ± 0.0082	1.000 ± 0.0150
S2	3,812	396	10	130	266	1,226	27	45	0.0342 ± 0.0023	1.000 ± 0.0150
S3	3,572	426	8	117	309	1,146	42	27	0.0414 ± 0.0019	1.000 ± 0.0150
S4	3,617	402	9	186	216	1,169	87	13	0.0225 ± 0.0047	1.000 ± 0.0150
S5	3,164	383	8	120	263	937/203	64/20	15/10	0.0379 ± 0.0028	1.000 ± 0.0150
S6	2,645	339	8	90	249	792	63	12	0.0402 ± 0.0036	1.000 ± 0.0150
S7	2,192	283	8	128	155	362/309	10/26	36/12	0.0356 ± 0.0034	1.000 ± 0.0150
S8	1,936/1,927	286/284	7/7	87	199/197	591	46	12	0.0361 ± 0.0030	1.000 ± 0.0150
S9	1,900	319	6	78	241	347/209	31/12	11/17	0.0626 ± 0.0029	1.000 ± 0.0150
S10	1,801	242	7	83	159	558	30	19	0.0396 ± 0.0025	0.9950 ± 0.0160

Fig. 3. Nucleotide diversity and conserved sequences in entire genomes of 21 RBSDV isolates collected in nine locations in China. Sliding-window analysis of nucleotide diversity in RBSDV S1 through S10 with 200-bp sliding windows in 100-bp steps. **A**, Nucleotide diversity and conserved sequences in RBSDV S1. **B**, Nucleotide diversity and conserved sequences in RBSDV S2. **C**, Nucleotide diversity and conserved sequences in RBSDV S3. **D**, Nucleotide diversity and conserved sequences in RBSDV S4. **E**, Nucleotide diversity and conserved sequences in RBSDV S5. **F**, Nucleotide diversity and conserved sequences in RBSDV S6. **G**, Nucleotide diversity and conserved sequences in RBSDV S7. **H**, Nucleotide diversity and conserved sequences in RBSDV S8. **I**, Nucleotide diversity and conserved sequences in RBSDV S9. **J**, Nucleotide diversity and conserved sequences in RBSDV S10. The red boxes represent conserved regions in these segments.

Table 2. Analysis of possible recombination events in RBSDV sequences S1 through S10

Recombinant sequence	Major parental sequence	Minor parental sequence	P-value							Beginning breakpoint positions	Ending breakpoint positions
			RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiScan	3Seq		
S1_13VR-2	S1_13VM-1	S1_13VIR-1	1.22×10^{-21}	9.79×10^{-16}	1.73×10^{-19}	1.12×10^{-16}	5.88×10^{-17}	9.98×10^{-41}	1.82×10^{-61}	31	3,281
S2_13VR-1	S2_13VM-2	S2_13VIR-1	4.58×10^{-17}	2.08×10^{-15}	5.26×10^{-23}	1.28×10^{-19}	1.86×10^{-19}	2.41×10^{-26}	2.23×10^{-46}	14	2,415
S2_13IIM-1	S2_13VR-2	Unknown (S2_13IIM-1)	1.88×10^{-10}	NS	6.89×10^{-11}	7.73×10^{-14}	6.23×10^{-13}	3.19×10^{-12}	1.66×10^{-25}	1,438	2,415
S3_13VR-2	S3_13VM-1	S3_13VIR-1	6.67×10^{-24}	3.31×10^{-21}	3.74×10^{-23}	4.81×10^{-13}	2.49×10^{-12}	3.63×10^{-13}	5.28×10^{-35}	1,340	2,075
S5_13VIR-2	S5_13VM-1	S5_13VIR-1	2.23×10^{-12}	7.01×10^{-13}	7.02×10^{-20}	1.90×10^{-18}	1.31×10^{-18}	7.00×10^{-27}	2.63×10^{-49}	1,157	3,164
S6_13VM-2	S6_13VIR-1	S6_13VIR-1	3.17×10^{-14}	2.53×10^{-14}	1.13×10^{-17}	6.72×10^{-17}	1.02×10^{-16}	3.82×10^{-23}	3.07×10^{-39}	11	1,401
S6_13VM-3	S6_13VR-3	S6_13VIR-1	1.03×10^{-13}	3.22×10^{-11}	1.04×10^{-13}	1.00×10^{-15}	2.26×10^{-15}	5.27×10^{-19}	7.26×10^{-36}	22	1,360
S8_13VIR-1	S8_13VM-1	S8_13VIR-1	NS	4.58×10^{-3}	1.15×10^{-4}	1.50×10^{-07}	9.74×10^{-08}	2.10×10^{-11}	2.04×10^{-11}	44	1,086
S9_12IM-1	S9_13IIM-1	S9_13VIR-1	2.29×10^{-23}	4.21×10^{-20}	2.29×10^{-23}	8.63×10^{-18}	2.32×10^{-18}	2.26×10^{-21}	8.93×10^{-46}	1,030	1,785
S9_13VIR-2	S9_13IIM-1	S9_13IIM-1	NS	7.78×10^{-5}	NS	1.29×10^{-2}	6.47×10^{-3}	NS	9.21×10^{-10}	1,030	1,123

Fig. 4. Maximum likelihood phylogenetic trees and pairwise identities of nucleotide sequences from S1 to S10 from 21 RBSDV isolates from China. The number of bootstrap replicates was set to 1,000. Only bootstrap values >70% are shown. Green lines represent the isolates that clustered into group A; purple lines represent the isolates that clustered into group B. **A.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S1. **B.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S2. **C.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S3. **D.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S4. **E.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S5. **F.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S6. **G.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S7. **H.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S8. **I.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S9. **J.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in RBSDV S10. **K.** Maximum likelihood phylogenetic trees and pairwise nucleotide identities in the concatenated RBSDV sequences from S1 through S10.

were detected in the coding regions of the other nine segments. S9 also has the highest level of nucleotide diversity among the 10 RBSDV segments, which might be related to protein secondary structure, genomic architecture, or protein function.

The nucleotide diversity of S4, which encodes the B-spike protein of the outer shell of RBSDV (Li et al. 2012; Zhang et al. 2001), is the lowest among the 10 RBSDV segments. RBSDV S2, which encodes core protein P2, had the lowest average variation in nucleotide and amino acid sequences and the second lowest level of nucleotide diversity. ORF2 was also under the lowest selection pressure among the 13 RBSDV ORFs. S8, with the fourth lowest nucleotide diversity, encodes another major capsid protein of RBSDV (Li et al. 2012). These results indicated that the structural proteins of RBSDV were all more conserved than its nonstructural proteins. However, the nucleotide diversity of S10, which encodes an RBSDV coat protein (Liu et al. 2007), was the sixth lowest among the 10 segments. The nucleotide diversity of S10 could be higher than that of segments encoding other structural proteins due to other possible functions of P10, such as self-interaction (Liu et al. 2007). The nucleotide diversity of S9 was significantly higher than that of the other nine segments ($P < 0.01$), and the highest nucleotide diversity in the entire RBSDV genome was detected in ORF9-1. P9-1, which is encoded by the first ORF of S9, is recruited by P6 into viroplasm-like structures (VLSs) by direct protein-protein interaction (Q. Wang et al. 2011). S6 showed the third highest nucleotide diversity among the 10 RBSDV segments, which might be related to the interaction of P6 with P9-1.

Previous studies have shown that RBSDV populations in China can be organized into two groups based on the sequences of S9 (Y. Zhou et al. 2015a) and S10 (Li et al. 2012; Yin et al. 2013), independent of host or geographic origin. The 21 isolates here also clustered into two groups based on either these 10 genome segments or the concatenated sequences for S1 through S10. This study further confirmed that the classification of RBSDV based on the sequences of these 10 segments and the entire genome is also independent of host or geographical origin. These results also indicated that an isolate could be clustered into group B based on the sequences of at least four segments. The classifications of S3, S7, and S9 were more similar to the classifications obtained from the concatenated S1 through S10 sequences than were those of the other seven segments. Clustering based on S3 and S9 sequences were consistent. The pairwise nucleotide identities of S9 (93.74%) and S3 (95.86%) were low relative to those of the other nine segments (Fig. 4). This suggests that phylogenies based on sequences with low levels of pairwise nucleotide identity were similar to those based on concatenated sequences of the entire RBSDV genome. Previous results showed that synonymous codon usage related to the phylogeny in S7 (Y. Zhou et al. 2015b). In this study, only two isolates were classified in the same group based on all the 10 segments, which may be because of the different synonymous codon usage among these 10 segments from S1 to S10.

In conclusion, analyses of codon usage bias, conserved regions, and phylogenies were used to characterize molecular diversification

and selection in the entire genome of 21 RBSDV isolates from the Yellow and Huai river summer maize-growing regions of China. These results reveal variation in the genetic structure of RBSDV and should accelerate the exploration of the pathogenic mechanisms of this virus.

Acknowledgments

This study was funded by the Postdoctoral Science Foundation of Heilongjiang Province of China (LBH-Z16019), Postdoctoral Science Foundation of China (2017M611344), Young Talent of Northeast Agricultural University, the Agricultural Science and Technology Innovation Program at CAAS, and the National Key Research and Development Program of China (2016YFD0101201). Plant samples were provided by Jianhua Yuan, Zhaodong Meng, Shuanggui Tie, Qingchang Meng, Yanping Chen, Jirong Wu, Jie Shi, Xiaohua Han, Wenyue Tong, Qi Sun, Tao Guo, and Zhaowen Sun, whom we would like to thank for their kind assistance.

Literature Cited

- Azuhata, F., Uyeda, I., Kimura, I., and Shikata, E. 1993. Close similarity between genome structures of rice black-streaked dwarf and maize rough dwarf viruses. *J. Gen. Virol.* 74:1227-1232.
- Bai, F. W., Yan, J., Qu, Z. C., Zhang, H. W., Xu, J., Ye, M. M., and Shen, D. L. 2002. Phylogenetic analysis reveals that a dwarfing disease on different cereal crops in China is due to rice black streaked dwarf virus (RBSDV). *Virus Genes* 25:201-206.
- Bennetzen, J. L., and Hall, B. D. 1982. Codon selection in yeast. *J. Biol. Chem.* 257:3026-3031.
- Carver, T. J., and Mullan, L. J. 2005. JAE: Jemboss Alignment Editor. *Appl. Bioinformatics* 4:151-154.
- Cheng, X. F., Wu, X. Y., Wang, H. Z., Sun, Y. Q., Qian, Y. S., and Luo, L. 2012. High codon adaptation in citrus tristeza virus to its citrus host. *Virol. J.* 9: 113-121.
- Diamant, A., Pinter, R. Y., and Tuller, T. 2014. Three-dimensional eukaryotic genomic organization is strongly correlated with codon usage expression and function. *Nat. Commun.* 5:5876-5887.
- Dodds, J. A., Morris, T. J., and Jordan, R. L. 1984. Plant viral double-stranded RNA. *Annu. Rev. Phytopathol.* 22:151-168.
- Dovas, C. I., Eythymiou, K., and Katis, N. I. 2004. First report of maize rough dwarf virus (MRDV) on maize crops in Greece. *Plant Pathol.* 53:238.
- Fang, S., Yu, J., Feng, J., Han, C., Li, D., and Liu, Y. 2001. Identification of rice black-streaked dwarf fivirus in maize with rough dwarf disease in China. *Arch. Virol.* 146:167-170.
- Firth, A. E., and Atkins, J. F. 2009. Analysis of the coding potential of the partially overlapping 3' ORF in segment 5 of the plant fiviruses. *Virol. J.* 6:32-36.
- Harpaz, I. 1959. Needle transmission of a new maize virus. *Nature* 184:B.A.77-B.A.78.
- Isogai, M., Uyeda, I., and Lee, B. 1998. Detection and assignment of proteins encoded by rice black streaked dwarf fivirus S7, S8, S9 and S10. *J. Gen. Virol.* 79:1487-1494.
- Jenkins, G. M., and Holmes, E. C. 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92:1-7.
- Kim, M. K., Seo, J. K., Kwak, H. R., Kim, J. K., Kim, K. H., Cha, B. J., and Choi, H. S. 2014. Molecular genetic analysis of Cucumber mosaic virus populations infecting pepper suggests unique patterns of evolution in Korea. *Virology* 104: 1-8.
- Kyrychenko, A. M., Hordeichik, O. I., and Shcherbatenko, I. S. 2012. Codon bias and nucleotide substitutions in soybean dwarf virus. *Mikrobiol. Z.* 74:90-97.
- Li, Y. Q., Jia, M. G., Jiang, Z. D., Zhou, T., and Fan, Z. F. 2012. Molecular variation and recombination in RNA segment 10 of rice black-streaked dwarf virus isolated from China during 2007-2010. *Arch. Virol.* 157:1351-1356.
- Li, Y. Q., Xia, Z. H., Peng, J., Zhou, T., and Fan, Z. F. 2013. Evidence of recombination and genetic diversity in southern rice black-streaked dwarf virus. *Arch. Virol.* 158:2147-2151.
- Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.
- Liu, H. J., Wei, C. H., Zhong, Y. W., and Li, Y. 2007. Rice black-streaked dwarf virus outer capsid protein P10 has self-interactions and forms oligomeric complexes in solution. *Virus Res.* 127:34-42.
- Lobo, F. P., Mota, B. E., Pena, S. D., Azevedo, V., Macedo, A. M., Tauch, A., Machado, C. R., and Franco, G. R. 2009. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One* 4:e6282.
- Ma, H. X., Pan, J. J., Kang, K., Xie, Z. Q., Ru, W. P., Chen, H. M., Huang, X. Y., and Xu, B. L. 2013. Genome sequences of coxsackievirus B5 isolated from viral encephalitis patients in Henan Province, 2011. *Chinese J. Epidemiol.* 34: 1213-1215.
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefevre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462-2463.

Table 3. Ratios of non-synonymous to synonymous substitutions in 13 RBSDV ORFs

	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>
ORF1	0.0076	0.2058	0.0369
ORF2	0.0023	0.1824	0.0126
ORF3	0.0051	0.2121	0.0240
ORF4	0.0052	0.1041	0.0500
ORF5-1	0.0106	0.1580	0.0671
ORF5-2	0.0112	0.0865	0.1295
ORF6	0.0104	0.1974	0.0527
ORF7-1	0.0037	0.1765	0.0210
ORF7-2	0.0087	0.1753	0.0496
ORF8	0.0080	0.1623	0.0493
ORF9-1	0.0191	0.3435	0.0556
ORF9-2	0.0083	0.3693	0.0225
ORF10	0.0061	0.1929	0.0316

- Meng, Y., Meng, F., Han, T., and Liu, K. 2008. Causes and prevention measures of summer maize rough dwarf disease in Yellow and Huai River valleys of China. *China Acad. J.* 7:29-31.
- Milne, R. G., Conti, M., and Lisa, V. 1973. Partial purification, structure and infectivity of complete maize rough dwarf virus particles. *Virology* 53:130-141.
- Morris, T. J., and Dodds, J. A. 1979. Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* 69:854-858.
- Pintó, R. M., Aragones, L., Costafreda, M. I., Ribes, E., and Bosch, A. 2007. Codon usage and replicative strategies of hepatitis A virus. *Virus Res.* 127:158-163.
- Seo, J. K., Ohshima, K., Lee, H. G., Son, M., Choi, H. S., Lee, S. H., Sohn, H. S., and Kim, K. H. 2009. Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. *Virology* 393:91-103.
- Sharp, P. M., and Cowe, E. 1991. Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast* 7:657-678.
- Sharp, P. M., and Li, W. H. 1987. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:1281-1295.
- Shi, L. Y., Weng, J. F., Liu, C. L., Song, X. Y., Miao, H. Q., Hao, Z. F., Xie, C. X., Li, M. S., Zhang, D. G., Bai, L., Pan, G. T., Li, X. H., and Zhang, S. H. 2013. Identification of promoter motifs regulating *ZmELF4E* expression level involved in maize rough dwarf disease resistance in maize (*Zea mays* L.). *Mol. Genet. Genomics* 288:89-99.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725-2729.
- Wang, Q., Tao, T., Zhang, Y. J., Wu, W. Q., Li, D. W., Yu, J. L., and Han, C. G. 2011. Rice black-streaked dwarf virus P6 self-interacts to form punctate, viroplasm-like structures in the cytoplasm and recruits viroplasm-associated protein P9-1. *Virology* 418:24-38.
- Wang, Z. H., Fang, S. G., Xu, J. L., Sun, L. Y., Li, D. W., and Yu, J. L. 2003. Sequence analysis of the complete genome of rice black-streaked dwarf virus isolated from maize with rough dwarf disease. *Virus Genes* 27:163-168.
- Weng, J. F., Li, B., Liu, C. L., Yang, X. Y., Wang, H. W., Hao, Z. F., Li, M. S., Zhang, D. G., Ci, X. K., Li, X. H., and Zhang, S. H. 2013. A non-synonymous SNP within the *isopentenyl transferase2* locus is associated with kernel weight in Chinese maize inbreds (*Zea mays* L.). *BMC Plant Biol.* 13:98-108.
- Yin, X., Xu, F. F., Zheng, F. Q., Li, X. D., Liu, B. S., and Zhang, C. Q. 2011. Molecular characterization of segments S7 to S10 of a southern rice black-streaked dwarf virus isolate from maize in northern China. *Virology* 426:47-53.
- Yin, X., Zheng, F. Q., Tang, W., Zhu, Q. Q., Li, X. D., Zhang, G. M., Liu, H. T., and Liu, B. S. 2013. Genetic structure of rice black-streaked dwarf virus populations in China. *Arch. Virol.* 158:2505-2515.
- Zhang, H. M., Chen, J. P., and Adams, M. J. 2001. Molecular characterisation of segments 1 to 6 of rice black-streaked dwarf virus from China provides the complete genome. *Arch. Virol.* 146:2331-2339.
- Zhou, H., Wang, H., Huang, L. F., Naylor, M., and Clifford, P. 2005. Heterogeneity in codon usages of sobemovirus genes. *Arch. Virol.* 150:1591-1605.
- Zhou, Y., Chen, H. M., Cao, M. J., Wang, X. F., Jin, X., Liu, K. H., and Zhou, C. Y. 2017. Occurrence, distribution, and molecular characterization of Citrus yellow vein clearing virus in China. *Plant Dis.* 101:137-143.
- Zhou, Y., Meng, Q. C., Chen, Y. P., Wu, J. R., Hao, Z. F., Wang, Z. H., Zhang, D. G., Li, M. S., Yong, H. J., Zhang, S. H., Li, X. H., and Weng, J. F. 2016. Molecular variation and expansion of a rice black-streaked dwarf virus population based on analysis of segment 1 in Jining, China. *Arch. Virol.* 161:3435-3443.
- Zhou, Y., Weng, J. F., Chen, Y. P., Liu, C. L., Han, X. H., Hao, Z. F., Li, M. S., Yong, H. J., Zhang, D. G., Zhang, S. H., and Li, X. H. 2015a. Phylogenetic and recombination analysis of rice black-streaked dwarf virus segment 9 in China. *Arch. Virol.* 160:1119-1123.
- Zhou, Y., Weng, J. F., Chen, Y. P., Wu, J. R., Meng, Q. C., Han, X. X., Hao, Z. F., Li, M. S., Yong, H. J., Zhang, D. G., Zhang, S. H., and Li, X. H. 2015b. Molecular genetic analysis and evolution of segment 7 in rice black-streaked dwarf virus in China. *PLoS One* 10:e0131410.