

RESEARCH

Open Access



Close evolutionary relationship between rice black-streaked dwarf virus and southern rice black-streaked dwarf virus based on analysis of their bicistronic RNAs

Zenghui Wang^{1,2†}, Chengming Yu^{2†}, Yuanhao Peng^{1†}, Chengshi Ding¹, Qingliang Li¹, Deya Wang^{1*}  and Xuefeng Yuan^{2*}

Abstract

Background: *Rice black-streaked dwarf virus* (RBSDV) and *Southern rice black-streaked dwarf virus* (SRBSDV) seriously interfered in the production of rice and maize in China. These two viruses are members of the genus *Fijivirus* in the family *Reoviridae* and can cause similar dwarf symptoms in rice. Although some studies have reported the phylogenetic analysis on RBSDV or SRBSDV, the evolutionary relationship between these viruses is scarce.

Methods: In this study, we analyzed the evolutionary relationships between RBSDV and SRBSDV based on the data from the analysis of codon usage, RNA recombination and phylogenetic relationship, selection pressure and genetic characteristics of the bicistronic RNAs (S5, S7 and S9).

Results: RBSDV and SRBSDV showed similar patterns of codon preference: open reading frames (ORFs) in S7 and S5 had with higher and lower codon usage bias, respectively. Some isolates from RBSDV and SRBSDV formed a clade in the phylogenetic tree of S7 and S9. In addition, some recombination events in S9 occurred between RBSDV and SRBSDV.

Conclusions: Our results suggest close evolutionary relationships between RBSDV and SRBSDV. Selection pressure, gene flow, and neutrality tests also supported the evolutionary relationships.

Keywords: RBSDV, SRBSDV, Phylogenetic analysis, Recombination, Selection pressure

Introduction

In major rice-growing regions, virus diseases occurred frequently and caused severe damages to rice, among which Rice black-streaked dwarf virus (RBSDV) and Southern rice black-streaked dwarf virus (SRBSDV) are major viral pathogens in China. RBSDV is a member of the genus *Fijivirus* in the family *Reoviridae*, which infects rice, maize and wheat [1–3]. RBSDV-infected rice plants have typical dwarf symptoms, such as dark leaves and small white waxy galls on the underside of the leaves [2, 3]. Another disease with analogous symptoms

was first reported in Guangdong Province, China in 2001 [4], and later was identified as a disease caused by *Southern rice black-streaked dwarf virus* (SRBSDV), a new member of the genus *Fijivirus* [5]. SRBSDV was subsequently reported to be found throughout southern China and Vietnam [6]. In recent years, both viruses caused a significant reduction in rice yield. However, the identification of these diseases simply based on the symptoms on rice leaves is problematic.

RBSDV and SRBSDV are effectively transmitted by the small brown planthopper (SBPH, *Laodelphax striatellus* Fallen) and the white-backed planthopper (WBPH, *Sogatella furcifera* Horvath), respectively in a persistent propagative manner [5–7]. RBSDV and SRBSDV have similar genome organizations with 10 dsRNA segments that totally encode 13 proteins [1, 5, 8]. Among the 10 dsRNA

* Correspondence: wangdeyasdy@163.com; snowpeak77@163.com

†Zenghui Wang, Chengming Yu and Yuanhao Peng are co-first author
¹College of Life Sciences, Zaozhuang University, Zaozhuang 277160, People's Republic of China

²College of Plant Protection, Shandong Agricultural University, Tai'an 271018, People's Republic of China



segments, the dsRNA S5, S7 and S9 are comprised of two open reading frames (ORFs) that encode for two proteins. It is suggested that the expression of the second protein encoded by the bicistronic RNAs is strictly regulated. P7-1 can cause male sterility due to non-dehiscent anthers in *Arabidopsis* [9]. P7-2 can interact with SKP1, a core subunit of SCF ubiquitin ligase [10]. P9-1 may participate in the processes of viroplasma nucleation and virus morphogenesis through the recruitment of P6 [11]. Although some studies have reported the phylogenetic analysis on RBSDV or SRBSDV [12–14], the evolutionary relationship between RBSDV and SRBSDV is scarce. In this study, we accessed the evolutionary relationship between RBSDV and SRBSDV using on the data the analysis of codon usage, RNA recombination and phylogenetic relationship, selection pressure and genetic characteristic of bicistronic RNAs (S5, S7 and S9) (Fig. 1). By analyzing the bicistronic RNAs (S5, S7 and S9) in RBSDV and SRBSDV, we identified the evolutionary relationships between the viruses, and potentially discovered some information of gene expression and regulation. (For the abbreviations, refer to Table 1).

Results

Analysis of codon usage bias based on ORFs of S5, S7 and S9 in RBSDV and SRBSDV

Using DnaSP 5.0, codon usage bias analysis was performed on ORFs of S5, S7 and S9 in both RBSDV and SRBSDV. When the gene was highly expressed, the gene had high codon bias with the low effective number of codons (ENC) value. Data from RBSDV and SRBSDV showed that ORFs in S5 had higher ENC, and ORFs in S7 had lower ENC (Table 2). For RBSDV, ORF7-2 had the lowest ENC (39.958), and ORF5-2 had the highest ENC (55.334). For SRBSDV, ORF7-1 has lowest ENC (41.394) and ORF5-1 had highest ENC (51.455) (Table 1). Codon bias index (CBI) showed the pattern of negative

Table 1 List of abbreviations

| Abbreviations | Unabbreviated Form |
|-----------------|--|
| RBSDV | Rice black-streaked dwarf virus |
| SRBSDV | Southern rice black-streaked dwarf virus |
| ORFs | Open reading frames |
| ENC | Effective number of codons |
| CBI | Codon bias index |
| GC ₃ | G + C content at third position |
| GC _C | G + C content at coding positions |
| SBPH | Small brown planthopper |
| WBPH | White-backed planthopper |
| RSCU | Relative synonymous codon usage |
| RDP4 | Recombination Detection Program version |

correlation with ENC. Both ENC and CBI suggested that ORFs in S7 of RBSDV and SRBSDV has relatively higher codon usage bias, while ORFs in S5 of RBSDV and SRBSDV had relatively lower codon usage bias (Table 2). In addition, data for GC₃ and GC_C had shown that G + C content was low in all analyzed ORFs.

For determining the optimal codons used in all ORFs, the average relative synonymous codon usage (RSCU) value was determined (Additional file 1: Table S1). GUA (V) in S7-1, GAA (E) /GAG (E) in S7-2 of SRBSDV as well as GGC (G) /GGG (G) in S5-2 of RBSDV, and UGG in ORFs of S5/S7 in RBSDV and SRBSDV showed no codon usage bias, except AUG (M) (Additional file 1: Table S1). In addition, the majority of optimal codons with RSCU values approximately equal to 1 end with U or A, indicating that the codon usage in RBSDV and SRBSDV was biased towards synonymous codons. In addition, we produced a graphical heat map for showing the codon usage in the coding region (Fig. 2).

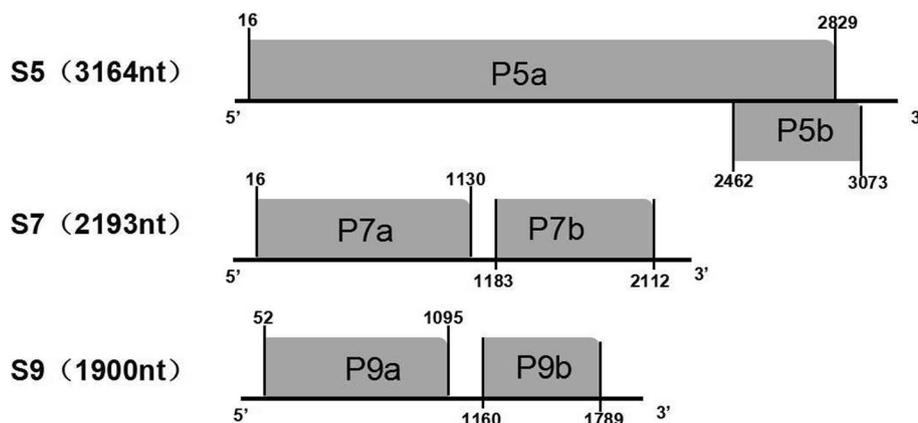


Fig. 1 Genome organization of S5, S7 and S9. The dark area is ORF region

Recombination analysis

Recombination analysis, used to identify potential evolutionary relationships between RBSDV and SRBSDV, detected no recombination for S5. Only one recombination was detected for S7 (Table 3). All of the recombination sequences, major and minor parents were the sequences from RBSDV S7. For S9, we detected higher recombination frequency with a total of 9 recombination events (Table 3). We also detected recombination events between RBSDV S9 and SRBSDV S9. For event 9–7, recombination of position 82–1184 in RBSDV S9 (HQ731500) was derived from major parent SRBSDV S9(HM998852) and minor parent SRBSDV S9 (HF955003). For other events, drivers were RBSDV S9 and minor parents were SRBSDV S9. These results suggest that S9 of RBSDV and SRBSDV underwent frequent RNA recombination during natural evolution. It also suggests that genome information of RBSDV and SRBSDV may be exchanged when infecting the same host plant.

Phylogenetic analysis of S5, S7 and S9 in RBSDV and SRBSDV

For further identifying the evolutionary relationship between RBSDV and SRBSDV, phylogenetic trees were constructed based on S5, S7 and S9 via MEGA 5.0, respectively (Fig. 3). In a phylogenetic tree of S5, RBSDV S5 formed two clades (I and II) and SRBSDV S5 formed clade III (Fig. 3a).

Table 2 Codon usage for ORFs of S5, S7 and S9 in Rice black-streaked dwarf virus and Southern rice black-streaked dwarf virus

| Virus | ORFs | ENC | CBI | GC _{3S} | GC _C |
|--------|--------|--------|-------|------------------|-----------------|
| RBSDV | ORF5-1 | 49.982 | 0.323 | 0.329 | 0.372 |
| | ORF5-2 | 55.334 | 0.312 | 0.384 | 0.378 |
| | ORF7-1 | 45.780 | 0.498 | 0.231 | 0.354 |
| | ORF7-2 | 39.958 | 0.598 | 0.227 | 0.324 |
| | ORF9-1 | 47.382 | 0.422 | 0.298 | 0.356 |
| SRBSDV | ORF9-2 | 43.354 | 0.541 | 0.189 | 0.296 |
| | ORF5-1 | 51.455 | 0.323 | 0.365 | 0.379 |
| | ORF5-2 | 46.595 | 0.342 | 0.334 | 0.360 |
| | ORF7-1 | 41.394 | 0.569 | 0.210 | 0.343 |
| | ORF7-2 | 43.882 | 0.539 | 0.217 | 0.304 |
| | ORF9-1 | 45.992 | 0.445 | 0.291 | 0.356 |
| | ORF9-2 | 47.494 | 0.516 | 0.256 | 0.322 |

It implies S5 independently evolved in RBSDV and SRBSDV. In phylogenetic tree of S7, there are three clades. Clade I and II consist of RBSDV S7. Clade III includes RBSDV S7 (KC134295) and SRBSDV S7 (Fig. 3b). These results highlight a close relationship between RBSDV S7 and SRBSDV S7. Phylogenetic tree of S9 also produced three clades with clade I consists of three SRBSDV S9 s and

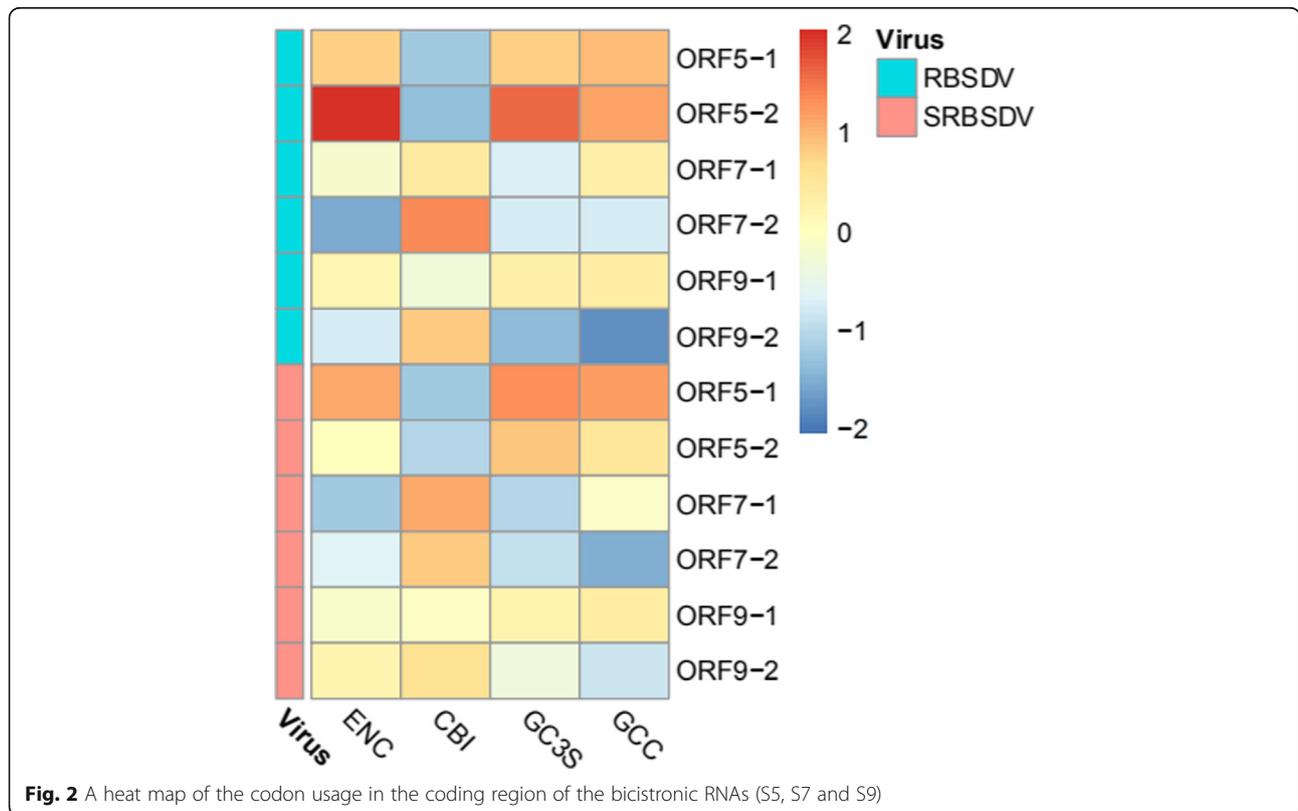


Fig. 2 A heat map of the codon usage in the coding region of the bicistronic RNAs (S5, S7 and S9)

Table 3 Recombination events of S7 and S9 in RBSDV and SRBSDV

| Event number | Recombination sequence | Breakpoint | | Supporting software | Major Parent | Minor Parent | P-value |
|--------------|------------------------|------------|------|---------------------|--------------------|--------------------|-------------------------|
| | | Begin | End | | | | |
| 7-1 | HF955011(S7) | 1262 | 1313 | RGMCS | HF954991 | AY147039 | 1.456×10^{-9} |
| 9-1 | AF459812(S9) | 16 | 1472 | RGMS | HQ7315000 | HF955003 | 2.390×10^{-10} |
| 9-2 | AF536564(S9) | 16 | 1472 | RGMS | HQ7315000 | Unknown (HF955003) | 2.390×10^{-10} |
| 9-3 | AJ291706(S9) | 4 | 1472 | RGMS | HQ7315000 | Unknown (HF955003) | 1.081×10^{-62} |
| 9-4 | AY039705(S9) | 16 | 1469 | RGMS | HQ731500 | Unknown (HF955003) | 4.518×10^{-21} |
| 9-5 | AY050487(S9) | 16 | 1469 | RGMS | HQ731500 | Unknown (HF955003) | 1.081×10^{-62} |
| 9-6 | HF955013(S9) | 16 | 1469 | RGMCS | HQ731500 | Unknown (HF955003) | 2.673×10^{-9} |
| 9-7 | HQ731500(S9) | 82 | 1184 | RGBMS | Unknown (HM998852) | HF955003 | 1.011×10^{-14} |
| 9-8 | KC134297(S9) | 4 | 1472 | RGMS | HQ731500 | Unknown (HF955003) | 7.524×10^{-14} |
| 9-9 | KM921681(S9) | 4 | 1469 | RGBMS | HQ731500 | Unknown (HF955003) | 4.625×10^{-19} |

Note: R, RDP; G, GENECONV; M, MAXCHI; C, CHIMAERA; S, SISCAN; B, BOOTSCAN

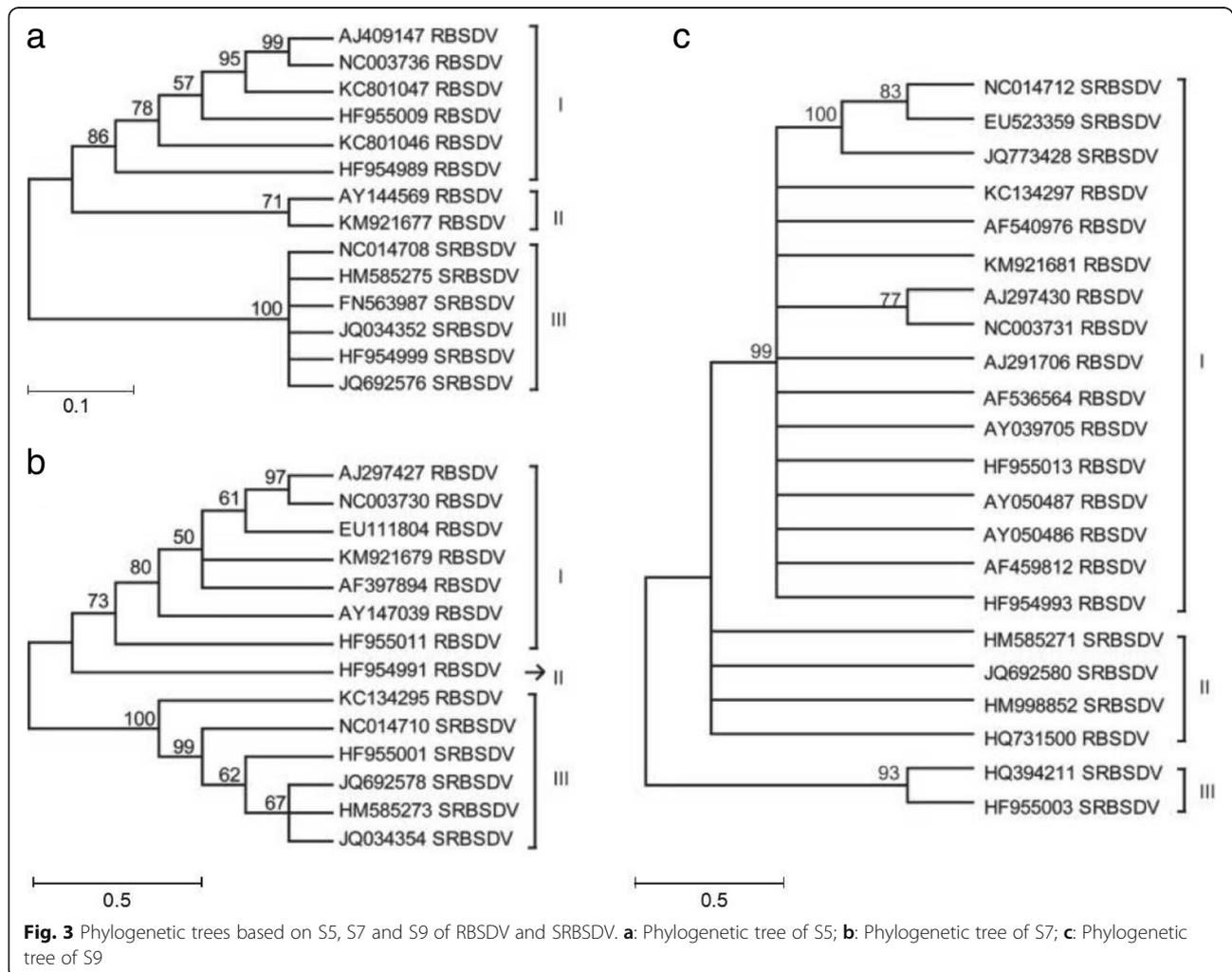


Fig. 3 Phylogenetic trees based on S5, S7 and S9 of RBSDV and SRBSDV. **a:** Phylogenetic tree of S5; **b:** Phylogenetic tree of S7; **c:** Phylogenetic tree of S9

thirteen RBSDV S9s, clade II includes two SRBSDV S9s and two RBSDV S9s and clade III contains two SRBSDV S9s (Fig. 3c). These results a potential genetic exchange between RBSDV S9 and SRBSDV S9 during natural evolution.

Selection pressure on ORFs of S5, S7 and S9

The selection pressure on S5, S7 and S9 in RBSDV and SRBSDV were measured by calculating the ratios of non-synonymous (dN) and synonymous sites (dS). When the value of ratio is more than 1, the gene is considered to be under positive or diversifying selection; when the value of ratio is less than 1, the selection is negative or purifying selection; and when the ratio is equal to 1, the selection is neutral. The statistics were performed on different subpopulations based on host and virus types (Table 4). Based on dN/dS value, ORF7-2 in viruses infecting maize underwent neutral selection with dN/dS ratio equal to 1, while other ORFs in different subpopulations underwent negative selection with dN/dS ratio less than 1 (Table 4). Different ratios indicate that ORFs experienced different negative selection pressure.

Based on dN and dS value, non-synonymous and synonymous substitution rates in ORFs were also identified. ORF5-1 and ORF5-2 in viruses infecting rice have relatively high dN value (0.605 and 1.289, respectively), indicating the high rates of non-synonymous substitutions in these ORFs (Table 3). dS values of ORF9-1 (0.259) and ORF9-2 (0.348) in RBSDV and ORF9-2 (0.193) in SRBSDV were relatively high, indicating frequent synonymous nucleotide substitutions in these ORFs (Table 4). For maize infecting viruses, ORF9-1 and ORF9-2 had relatively high dS value (0.763 and 1.081, respectively). However, for rice infecting viruses, all ORFs, especially ORF5-1 (dS : 1.039) and ORF5-2 (dS : 1.363), have relatively high dS values with ORF5-1

(dS : 1.039) and ORF5-2 (dS : 1.363) had the highest values. It suggests that RBSDV and SRBSDV have a higher synonymous substitution rate when infecting rice, rather than maize.

Genetic differentiation and gene flow in S5, S7 and S9

Genetic differentiation and gene flow were analyzed using DnaSP 5.0. Test statistics Ks^* , Z , and Snn (see methods) were used to access genetic differentiation in S5, S7 and S9; and Fst and Nm were used to measure gene flow. The statistics were performed on different subpopulations according to virus type, host type and region (Table 5). The value of Ks^* ranged from 0 to 6.06298 (Table 5), which indicated the variations between the sequences (Table 5). Compared to several kb genomes, the genetic difference was relatively low. The highest genetic difference (Ks^* : 6.06298) was in S9 of SRBSDV, but there was no genetic difference in S5, S7 or S9 (Ks^* : 0) between Zhejiang and Anhui (Table 5).

Different subpopulations showed different gene flow. For S5, gene flow within RBSDV or SRBSDV was frequent with the absolute value of Fst less than 0.33 (Table 5). This pattern also applied to S7 and S9 within RBSDV and SRBSDV. For S5, S7 and S9 between RBSDV and SRBSDV, there was an infrequent gene flow with Fst more than 0.33. We also observed a substantial local differentiation with Nm less than 1. For S5 and S7 between maize and rice, there was an infrequent gene flow with Fst more than 0.33. However, we found a low frequency of genetic differentiation with Nm more than 1. For S9 between maize and rice, there is a frequent gene flow with Fst less than 0.33 and low frequency of genetic differentiation with Nm more than 1. Different regions had different gene flow patterns. For S5 and S7 between Jiangsu and Zhejiang, or between Zhejiang and

Table 4 Non-synonymous and synonymous substitution in ORFs of S5, S7 and S9

| | ORFs | dN | dS | dN/dS | | ORFs | dN | dS | dN/dS |
|-------|--------|---------------|---------------|---------|--------|--------|---------------|---------------|---------|
| RBSDV | ORF5-1 | 0.013 ± 0.002 | 0.105 ± 0.008 | 0.124 | SRBSDV | ORF5-1 | 0.006 ± 0.001 | 0.021 ± 0.003 | 0.286 |
| | ORF5-2 | 0.015 ± 0.004 | 0.053 ± 0.012 | 0.283 | | ORF5-2 | 0.005 ± 0.002 | 0.012 ± 0.006 | 0.417 |
| | ORF7-1 | 0.002 ± 0.001 | 0.071 ± 0.009 | 0.028 | | ORF7-1 | 0.001 ± 0.001 | 0.009 ± 0.004 | 0.111 |
| | ORF7-2 | 0.010 ± 0.002 | 0.089 ± 0.013 | 0.112 | | ORF7-2 | 0.001 ± 0.001 | 0.012 ± 0.006 | 0.083 |
| | ORF9-1 | 0.033 ± 0.004 | 0.259 ± 0.028 | 0.127 | | ORF9-1 | 0.008 ± 0.002 | 0.014 ± 0.003 | 0.571 |
| | ORF9-2 | 0.032 ± 0.012 | 0.348 ± 0.101 | 0.092 | | ORF9-2 | 0.090 ± 0.011 | 0.193 ± 0.049 | 0.466 |
| Maize | ORF5-1 | 0.002 ± 0.001 | 0.025 ± 0.006 | 0.08 | Rice | ORF5-1 | 0.605 ± 0.067 | 1.039 ± 0.135 | 0.582 |
| | ORF5-2 | 0.004 ± 0.003 | 0.028 ± 0.015 | 0.143 | | ORF5-2 | 1.289 ± 0.391 | 1.363 ± 0.362 | 0.946 |
| | ORF7-1 | 0.001 ± 0.001 | 0.023 ± 0.007 | 0.043 | | ORF7-1 | 0.085 ± 0.012 | 0.772 ± 0.140 | 0.110 |
| | ORF7-2 | 0.003 ± 0.002 | 0.003 ± 0.011 | 1.0 | | ORF7-2 | 0.168 ± 0.055 | 0.994 ± 0.316 | 0.169 |
| | ORF9-1 | 0.101 ± 0.010 | 0.763 ± 0.113 | 0.132 | | ORF9-1 | 0.091 ± 0.010 | 0.683 ± 0.093 | 0.133 |
| | ORF9-2 | 0.102 ± 0.046 | 1.081 ± 0.394 | 0.094 | | ORF9-2 | 0.122 ± 0.030 | 0.980 ± 0.280 | 0.124 |

Table 5 Genetic differentiation and gene flow between subpopulations

| sequences | subpopulations | Ks* | Z | Snn | Nm | Fst |
|-----------|------------------|---------|---------|---------|------|----------|
| S5 | RBSDV | 4.68574 | 2.63348 | 0.06250 | 5.08 | -0.21377 |
| | SRBSDV | 3.52466 | 1.99848 | 0.16667 | 1.40 | -0.10526 |
| | RBSDV-SRBSDV | 4.03210 | 2.77618 | 1.0000 | 0.02 | 0.96173 |
| | Maize-Rice | 3.82262 | 7.93333 | 1.0000 | 4.56 | 0.78107 |
| | Jiangsu-Zhejiang | 4.75212 | 7.83333 | 0.83333 | 0.71 | 0.40672 |
| | Zhejiang-Anhui | 0.0000 | 0.00000 | 0.75000 | 0.50 | 0.42143 |
| | Jiangsu-Anhui | 4.75212 | 1.83501 | 0.33333 | 12.0 | 0.13397 |
| S7 | RBSDV | 4.52336 | 2.77535 | 0.16667 | 1.52 | -0.01867 |
| | SRBSDV | 1.90126 | 1.45543 | 0.2000 | 0.99 | -0.18605 |
| | RBSDV-SRBSDV | 3.70347 | 2.79015 | 0.92857 | 0.09 | 0.86866 |
| | Maize-Rice | 3.54044 | 1.95144 | 0.85714 | 5.20 | 0.14586 |
| | Jiangsu-Zhejiang | 3.04377 | 1.13373 | 1.0000 | 0.61 | 0.57143 |
| | Zhejiang-Anhui | 0.0000 | 0.0000 | 0.75000 | 0.50 | 0.61463 |
| | Jiangsu-Anhui | 3.04377 | 0.59725 | 0.80000 | 30.0 | 0.48168 |
| S9 | RBSDV | 3.89158 | 3.54343 | 0.35714 | 1.75 | 0.04670 |
| | SRBSDV | 6.06298 | 2.36085 | 0.62500 | 2.80 | -0.13159 |
| | RBSDV-SRBSDV | 4.85276 | 4.08636 | 0.73810 | 0.62 | 0.33722 |
| | Maize-Rice | 3.79931 | 2.70675 | 0.9000 | 8.51 | 0.02095 |
| | Jiangsu-Zhejiang | 2.55985 | 1.66296 | 0.71429 | 4.62 | 0.00696 |
| | Zhejiang-Anhui | 0.00000 | 0.00000 | 0.71429 | 0.61 | -0.00378 |
| | Jiangsu- Anhui | 2.55985 | 1.85037 | 0.25000 | 4.62 | -0.10000 |

Anhui, there is an infrequent gene flow with substantial local differences. For S5, S7 and S9 between Jiangsu and Anhui, there was a low frequency of genetic differentiation with *Nm* more than 1.

Neutrality tests on S5, S7 and S9 in RBSDV and SRBSDV

The neutrality test was based on the differences between the number of segregating sites and the average number of nucleotide differences. To testing the neutrality hypothesis and perform population demography, values for Tajima's D, Fu and Li's D, and Fu and Li's F were calculated using DnaSP version 5.0 (Table 6). Except for the Tajima's D value of ORF7-2 in RBSDV, all Tajima's D

values of other ORFs in RBSDV and SRBSDV are significantly far away from 0, which indicates these ORFs were under natural selection.

In terms of Fu & Li's D and F statistical tests, we found different patterns exist in RBSDV and SRBSDV. For SRBSDV, all values of Fu & Li's D and F are negative, indicating that S5, S7 and S9 of SRBSDV have a tendency to expand with a low frequency of polymorphism. RBSDV, ORF7-1, ORF9-1 and ORF9-2 had a low frequency of polymorphism with a negative value of Fu & Li's D and F. In contrast ORF5-1, ORF5-2 and ORF7-2 may have a high frequency of polymorphism with a positive value of Fu & Li's D and F. For SRBSDV, all values of Fu and Li's D and F were negative, indicating that S5, S7

Table 6 Neutrality test of S5, S7 and S9 in RBSDV and SRBSDV

| RBSDV | Tajima's D | Fu & Li's D | Fu & Li's F | SRBSDV | Tajima's D | Fu & Li's D | Fu & Li's F |
|--------|------------|-------------|-------------|--------|------------|-------------|-------------|
| ORF5-1 | -0.00508 | 0.40941 | 0.34886 | ORF5-1 | -0.66188 | -0.61629 | -0.68738 |
| ORF5-2 | -0.06959 | 0.36868 | 0.29669 | ORF5-2 | -0.79480 | -0.85062 | -0.90575 |
| ORF7-1 | -1.53102 | -1.60532 | -1.78365 | ORF7-1 | -0.07339 | -0.07339 | -0.07686 |
| ORF7-2 | -0.24789 | 0.69109 | 0.52040 | ORF7-2 | -0.80734 | -0.80734 | -0.84548 |
| ORF9-1 | -2.01232 | -2.30495 | -2.55773 | ORF9-1 | -1.26116 | -1.05424 | -1.23976 |
| ORF9-2 | -2.04066 | -2.35516 | -2.60827 | ORF9-2 | -2.03684 | -2.31023 | -2.52287 |

and S9 of SRBSDV may have a tendency to expand with a low frequency of polymorphism.

Discussion

Consistent with increasing trends in global export of agriculture commodities, RBSDV and SRBSDV have been spreading rapidly worldwide. During a rapid spread phase, these viruses may have maintained different evolutionary trajectories based on specific host and the environment. Although some previous studies have reported the phylogenetic analyses on RBSDV or SRBSDV separately [12–14], but the studies on the evolutionary relationship between RBSDV and SRBSDV was rare. In this study, we examined a potential evolutionary relationship between these viruses by performing an evolutionary analysis of S5, S7 and S9 in RBSDV and SRBSDV.

Codon usage bias has been reported for viruses (CITATION), including those infect humans (CITATION). Such biases provide information about virus-host coevolution [15, 16]. A detailed comparative analysis was performed to evaluate the degrees of codon usage bias about ORFs of S5, S7 and S9. In general, RBSDV and SRBSDV showed similar patterns of codon usage bias for ORFs of S5, S7 and S9 (Table 1). Compared with other ORFs, ORFs in S7 of RBSDV and SRBSDV had relatively higher codon usage bias with lowest ENC value. Low ENC value was usually correlated to high expression [17]. The correlation implies the evolutionary requirements for high expression of the ORFs, particularly ORF7–2, in S7. This potential high expression of ORF7–2 was supported by a study of that evaluated a intergenic region between ORF7–1 and ORF7–2 that contains a high activity IRES (Yuan et al., in preparation). In addition, 5'UTR in S3 or S10 of RBSDV enhanced the translation of FLuc reporter gene and possess IRES activity in the absence or presence of the 5'cap structure. To the best of our knowledge, this is the first report on the effect of untranslated regions of ds RNA viruses on translation [18].

Recombination is a major evolutionary mechanism that commonly found in many plant RNA viruses, including RBSDV and SRBSDV [12, 13, 19, 20]. Such recombination may prevent accumulation of mutations, help adapt to new hosts or environmental changes and overcome host resistance [21–23]. In this study, recombination events were detected between RBSDV and SRBSDV (e.g., recombination event 9–7) (Table 3). These results highlight the potential evolutionary relationship between RBSDV and SRBSDV. Moreover, phylogenetic analysis supported a evolutionary relationship. Clade III in the phylogenetic tree of S7 contained both RBSDV S7 and SRBSDV S7. Similar pattern was also found in clade I and II in the phylogenetic tree of S7. The close evolutionary

relationship highlights the frequent genetic exchange between RBSDV and SRBSDV when they infected the same host. It also suggests that these viruses may evolved from a single ancestor during evolution. Although the high nucleotide identity of these RBSDV and SRBSDV isolates, they may take different evolutionary path, and showed different host range or pathogenicity in the field. Finally, the selection pressure, gene flow, and recombination together promoted the evolution of the RBSDV and SRBSDV. The evolutionary trends under natural conditions may have a directing significance to the prevention and control of viral diseases.

Conclusions

RBSDV and SRBSDV presented similar patterns of codon usage bias: ORFs in S7 with higher codon usage bias and ORFs in S5 with lower codon usage bias. Some isolates from RBSDV and SRBSDV formed a clade in the phylogenetic tree of S7 and S9. In addition, some recombination events in S9 occurred between RBSDV and SRBSDV. Our results implied a close evolutionary relationship between RBSDV and SRBSDV. Analysis of selection pressure, gene flow, and neutrality test also supports this potential relationship.

Materials and methods

Sources of virus sequences

The sequences of bicistronic RNAs (S5, S7 and S9) in RBSDV and SRBSDV were referenced from <http://www.ncbi.nlm.nih.gov/nucleotide/>.

Codon usage bias analysis

Codon usage bias analysis was performed using the DnaSP 5.0 [24–26]. An effective number of codons (ENC) indicates a bias for synonymous codons rather than codon number or amino acid composition [26, 27]. The value of ENC ranges from 20 to 61 [17]. ENC value 20 indicates that only one type of codon is used for each amino acid and the codon bias is maximum; When the value is 61, all synonymous codons of each amino acid are equally used and there was no codon bias [26]. When the gene is highly expressed, it has high codon bias with low ENC value [17]. The codon bias index (CBI) is a measure of the deviation from the equal use of synonymous codons. The value of CBI ranges from 0 to 1 [28]. If the value of CBI is higher, it indicates that the codon bias is higher. $G + C_3s$ is the $G + C$ content at the third position. $G + C_c$ is $G + C$ content at coding positions [26]. Relative Synonymous Codon Usage (*RSCU*) values represent the number of times a particular codon is observed, relative to the number of times that the codon would be observed for a uniform synonymous codon usage. In the absence of any codon usage bias,

the *RSCU* values would be 1.00. A value less than 1 (or more than 1) indicates that the codons are used less frequently (or more frequently) than expected [24].

Recombination and phylogenetic analysis

Recombination analysis was performed using RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, 3SEQ and SISCAN programs in RDP 4 software package with a default setting [28]. When the *p*-values was less than 10^{-6} and the value of *Z* was more than 3 at the same time, the events supported by at least four programs were considered to be recombination [19, 29, 30]. Phylogenetic analysis was performed using MAGE5. Phylogenetic trees were constructed using the neighbor-joining (NJ) method as described previously [31]. The number of bootstrap replicates was 1000. Branches with less than 50% bootstrap value were collapsed.

Detection of selection pressure

Selection pressure was performed using the software MAGE5.0 as described in previous studies [31–33]. The ratios of *dN/dS* was used to describe the selection pressure. Here, *dN* is the average number of non-synonymous substitutions per site. *dS* is the average number of synonymous substitutions per site [31]. When a value of *dN/dS* is more than 1, the gene is considered to be under positive or diversifying selection; when a value of *dN/dS* is less than 1, the selection is negative or purifying; Finally, when *dN/dS* is equal to 1, the selection is neutral [31, 32].

Analysis of genetic differentiation and gene flow

Genetic differentiation and gene flow analyses were performed using the software DnaSP 5.0 [34]. Genetic differentiation was evaluated using sequence based statistics, *Ks**, *Z*, and *Snn* (the nearest-neighbor statistic). where *Ks** is a weighted average of differences between the sequences. *Z* is a rank statistic, and *Snn* represents how often the nearest neighbors' in the sequences are from the same location [35, 36]. Gene flow was estimated by measuring *Fst* and *Nm*, where *Fst* represents the component of genetic variation between populations and *Nm* represents the female effective size of female population (*N*) and migration rate among populations (*m*) [34]. The *Fst* values ranges from 0 to 1. When a value is more than 0.33, it implies that there is an infrequent gene flow. A value of < 0.33 implies a frequent gene flow [20, 34]. A *Nm* value of less than 1, it implies a genetic drift with a substantial local differentiation. A value of *Nm* is more than 1 implies a frequent gene flow with a low frequency of genetic differentiation [34].

Neutrality tests and population demography

Software DnaSP5.0 was used to detect the value of Tajima's *D*, Fu & Li's *D* and *F* statistics [34]. The Tajima's *D* test statistic was proposed for testing the hypothesis that states all the mutations are selectively neutral [37]. The test compares the differences between the number of segregating sites and the average number of nucleotide differences. A Tajima's *D* value away from 0 (i.e. $<$ or $>$ 0) implies population under natural selection. The Fu & Li's *D* test statistics measure the differences between the number of singletons and the total number of mutations [38]. The *F* test statistics measures the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences [38, 39]. A negative value for Fu & Li's *D* and *F* implies a low population diversity but still tends to expand. A negative value further implies a population with a low frequency of polymorphism.

Additional file

Additional file 1: Table S1. Relative synonymous codon usage (*RSCU*) values for each codon in *S5*, *S7* and *S9* of *RBSDV* and *SRBSDV*. **Table S2.** The sequence of *RBSDV* and *SRBSDV* information used in this paper. (DOCX 29 kb)

Acknowledgements

We are grateful to Prof. Hongkai Wang, University of Zaozhuang for valuable suggestions and critical modification on the manuscript, and acknowledge TopEdit LLC for the linguistic editing and proofreading during the preparation of this manuscript.

Funding

This work was supported by grants from The National Natural Science Foundation of China (31872638 and 31670147), The Shandong Province Natural Sciences Foundation of China (ZR2013CM015), Scientific Research Foundation for Ph.D. Programs of Zaozhuang University (2018BS040, 2018BS045), and Science and technology Program of Zaozhuang.

Availability of data and materials

The sequences of bicistronic RNAs (*S5*, *S7* and *S9*) in *RBSDV* and *SRBSDV* were referenced from <http://www.ncbi.nlm.nih.gov/nucleotide/>. (Additional file 1: Table S2); And the data is analyzed by bioinformatics using different software packages and methods.

Authors' contributions

ZH contributed to the design of the study. CM contributed to the statistical analysis, and drafting the manuscript. YH contributed to the data analysis. CS and QL contributed to drafting the manuscript. XF contributed to data analysis and drafting the manuscript. DY contributed to the design of the study, data analysis and drafting the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors declare that they agreed to publish this paper with the permission of the publishing houses.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 January 2019 Accepted: 15 April 2019

Published online: 27 April 2019

References

- Bai FW, Qu ZC, Yan J, Zhang HW, Xu J, Ye MM, Wu HL, Liao XG, Shen DL. Identification of rice black streaked dwarf virus in different cereal crops with dwarfing symptoms in China. *Acta Virol.* 2001;45:335–9.
- Shikata E, Kitaqawa Y. Rice black-streaked dwarf virus: its properties, morphology and intracellular localization. *Virology.* 1977;77:826–42.
- Zhang HM, Chen JP, Adams MJ. Molecular characterisation of segments 1 to 6 of Rice black-streaked dwarf virus from China provides the complete genome. *Arch Virol.* 2001;146:2331–9.
- Zhou GH, Xu DL, Li HP. Identification of rice black streaked dwarf virus infecting rice in Guangdong. In: Peng YL (ed) Proceedings of the conference on Chinese plant pathology, 4–7 august 2004. Beijing, China, Agricultural Sciencetech Press, 2004, 210–212.
- Zhou GH, Wen JJ, Cai DJ, Li P, Xu DL, Zhang SG. Southern rice black-streaked dwarf virus: a new proposed Fijivirus species in the family Reoviridae. *Chinese Sic Bull.* 2008;53:3677–85.
- Zhang HM, Yang J, Chen JP, Adam MJ. A black-streaked dwarf disease on rice in China is caused by a novel Fijivirus. *Arch Virol.* 1893–1898;2008:153.
- Wang ZH, Fang SG, Xu JL, Sun LY, Li DW, Yu JL. Sequence analysis of the complete genome of rice black-streaked dwarf virus isolated from maize with rough dwarf disease. *Virus Genes.* 2003;27:163–8.
- Wang Q, Yang J, Zhou G, Zhang HM, Chen JP, Adam MJ. The complete genome sequence of two isolates of southern rice black-streaked dwarf virus, a new member of the genus Fijivirus. *J Phytopathol.* 2010;158:733–7.
- Sun F, Yuan X, Xu Q, Zhou T, Fan Y, Zhou Y. Over expression of rice black-streaked dwarf virus p7-1 in *Arabidopsis* results in male sterility due to non-dehiscent anthers. *PLoS One.* 2013;8(11):e79514.
- Wang Q, Tao T, Han Y, Chen X, Fan Z, Li D, Yu J, Han C. Nonstructural protein P7-2 encoded by Rice black-streaked dwarf virus interacts with SKP1, a core subunit of SCF ubiquitin ligase. *Virol J.* 2013;10(1):325.
- Isogai M, Uyeda I, Lee BC. Detection and assignment of protein encoded by rice black streaked dwarf fivirus S7, S8, S9 and S10. *J Gen Virol.* 1998;79:1487–94.
- Yin X, Zheng FQ, Tang W, Zhu QQ, Li XD, Zhang GM, Liu HT, Liu BS. Genetic structure of rice black-streaked dwarf virus populations in China. *Arch Virol.* 2013;158:2505–15.
- Zhou Y, Weng JF, Chen YP, Liu CL, Han XH, Hao ZF, Li MS, Yong HJ, Zhang SH, Li XH. Phylogenetic and recombination analysis of rice black-streaked dwarf virus segment 9 in China. *Arch Virol.* 2015;160:1119–23.
- Li Y, Jia M, Jiang Z, Zhou T, Fan Z. Molecular variation and recombination in RNA segment 10 of rice black-streaked dwarf virus isolated from China during 2007–2010. *Arch Virol.* 2012;157:1351–6.
- Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol.* 2009;5:311.
- Lobo FP, Mota BE, Pena SD, Azevedo V, Macedo AM, Tauch A, Machado CR, Frabco GR. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One.* 2009;4:e6282.
- Sharp PM, Cowe E. Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast.* 1991;7:657–78.
- Yuan T, Wang Z, Yu C, Geng G, Su C, Yuan X. Effect of untranslated regions of S3 and S10 from Rice black-streaked dwarf virus on translation in the absence or presence of 5'cap. *Acta Pharmacologica Sinica* 2018, doi:<https://doi.org/10.13926/j.cnki.apps.000383>. (In Chinese).
- Ohshima K, Tomitaka Y, Wood JT, Minematsu Y, Kajiyama H, Tomimura K, Gibbs AJ. Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *J Gen Virol.* 2007;88:298–315.
- Wei TY, Yang JG, Liao FR, Gao FL, Lu LM, Zhang XT, Li F, Wu ZJ, Lin QY, Xie LH, Lin HX. Genetic diversity and population structure of rice stripe virus in China. *J Gen Virol.* 2009;90:1025–34.
- Martin DP. Rapid host adaptation by extensive recombination. *J Gen Virol.* 2009;90:734–46.
- Valli A, Lopez-Moya JJ, Garcia JA. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J Gen Virol.* 2007;88:1016–28.
- vander Walt E, Rybicki EP, Varsani A, Polston JE, Billharz R, Donaldson L, Monjane AL, Martin DP. Rapid host adaptation by extensive recombination. *J Gen Virol.* 2009;90:734–46.
- Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 1986;14(13):5125–43.
- Shields DC, Sharp PM, Higgins DG, Wright F. "silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Bio Evol.* 1988;5:704–16.
- Wright F. The 'effective number of codons' used in a gene. *Gene.* 1990; 87(1):23–9.
- Sau K, Gupta SK, Sau S, Mandal SC, Ghosh TC. Factors influencing synonymous codon and amino acid usage biases in Mimivirus. *Biosystems.* 2006;85:107–13.
- Morton BR. Chloroplast DNA codon use: evidence for selection at the psb a locus based on tRNA availability. *J Mol Evol.* 1993;37:273–80.
- Tomitaka Y, Ohshima K. A phylogeographical study of the turnip mosaic virus population in East Asia reveals an 'emergent' lineage in Japan. *Mol Ecol.* 2006;15:4437–57.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 2010;6:2462–3.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007;24:1596–9.
- Li WH. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol.* 1993;36:96–9.
- Pamilo P, Bianchi NO. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol Biol Evol.* 1993;10:271–81.
- Rozas J. DNA sequence polymorphism analysis using DnaSP. *Methods Mol Biol.* 2009;537:337–50.
- Hudson RR. A new statistic for detecting genetic differentiation. *Genetics.* 2000;155:2011–4.
- Hudson RR, Boss DD, Kaplan NL. A statistical test for detecting geographic subdivision. *Mol Biol Evol.* 1992;9:138–51.
- Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics.* 1983;105:437–60.
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics.* 1993;133: 693–709.
- Simonsen KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics.* 1995;141:413–29.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

