Molecular characterization of a novel wheat-infecting virus of the family Betaflexiviridae

Shuai Fu1 · Tianze Zhang1 · Mengzhu He1 · Bingjian Sun2 · Xueping Zhou1,3 · Jianxiang Wu1

Received: 7 April 2021 / Accepted: 27 May 2021 / Published online: 23 July 2021 © The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract
Wheat plants showing yellowing and mosaic in leaves and stunting were collected from wheat fields in Henan Province, China. Analysis of these plants by transmission electron microscopy showed that they contained two types of filamentous virus-like particles with a length of 200-500 nm and 1000-1300 nm, respectively. RNA-seq revealed a coinfection with wheat yellow mosaic virus (WYMV) and an unknown wheat-infecting virus. The genome of the unknown virus is 8,410 nucleotides long, excluding its 3’ poly(A) tail. It has six open reading frames (ORFs). ORF1 encodes a putative viral replication-associated protein (Rep), and ORFs 2, 3, and 4 encode the triple gene block (TGB) proteins. ORFs 5 and 6 encode the capsid protein (CP) and a protein with unknown function, respectively. Phylogenetic analysis showed that this novel virus is evolutionarily related to members of the subfamily Quinvirinae, family Betaflexiviridae. It is, however, distinct from the viruses in the currently established genera. Based on the species and genus demarcation criteria set by the International Committee on Taxonomy of Viruses (ICTV), we tentatively name this novel virus "wheat yellow stunt-associated betaflexivirus" (WYSaBV), and we propose it to be a member of a new genus in the family Betaflexiviridae.

Virus diseases often cause serious yield losses in wheat fields in China and in many other countries [6]. Based on the virus taxonomy released by the ICTV in 2019 (http://ictv.global/report), the order Tymovirales has five families: Alphaflexiviridae, Betaflexiviridae, Gammaplexiviridae, Deltaflexiviridae, and Tymoviridae. The family Betaflexiviridae is the largest of these and is divided into two subfamilies (Quinvirinae and Trivirinae), containing 13 genera and 108 species. The subfamilies Quinvirinae and Trivirinae are distinguished by the types of their viral movement proteins. Viruses in the subfamily Quinvirinae encode triple gene block (TGB) movement proteins, while the viruses in the subfamily Trivirinae encode 30K-like movement proteins [1]. The genera Foveavirus, Carlavirus, and Robigovirus are in the subfamily Quinvirinae, while the genera Capillovirus, Chordovirus, Citrivirus, Divavirus, Prunevirus, Ravavirus, Tepovirus, Trichovirus, Vitivirus, and Wamavirus are in the subfamily Trivirinae [5]. Viruses in the family Betaflexiviridae contain a monopartite positive-sense single-stranded RNA genome with 6.5-9.0 kilobases (kb) and a 3’-terminal polyadenylate tail. The viruses in some genera have a genome that is capped at the 5’ end. Virions of viruses in the family Betaflexiviridae are non-enveloped flexuous filamentous particles that are 600-1000 nm long, or even longer, and 12-13 nm in diameter [1].

Recently, RNA-seq technology has been widely used to identify viruses in field-collected plant samples, and this has
resulted in the identification of many novel plant-infecting viruses \[9, 10\]. In 2017, we sampled some wheat plants showing yellowing and mosaic in leaves and stunting from wheat fields in the city of Kaifeng in Henan Province, China (Fig. 1a and b). Examination of these samples under a transmission electron microscope revealed the presence of two different types of flexuous filamentous virus-like particles with dimensions of 200-500 × 13 nm (Fig. 1c and d) and 1000-1300 × 13 nm, respectively (Fig. 1e and f).

To characterize these virus-like particles, total RNA was extracted from the sampled leaf tissues using an EASYspin Plus Complex Plant RNA Kit (Aidlab Biotech, Beijing, China) followed by RNA-seq. An RNA library was constructed after removing ribosomal RNAs using a TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA) and then sequenced on an Illumina HiSeq X-ten platform (Biomarker Technologies, Beijing, China). Adapters were trimmed from the paired-end raw reads, and low-quality reads were filtered using CLC Genomics Workbench 9.5 software (QIAGEN, Germantown, MD, USA). The clean reads were first mapped to the wheat genome to identify and remove host sequences, and the remaining 9,561,825 clean reads were then assembled de novo into 13,481 larger contigs using the Trinity program \[2, 4, 8\]. The assembled contigs were then subjected to BLASTn and BLASTx searches against nucleotide (nt) and amino acid (aa) sequences of the NCBI databases. Sequences representing viral RNAs were obtained, and the 5’ and 3’ end sequences of the assembled contigs were then determined by 5’ and 3’ RACE PCR as instructed (Roche Diagnostic, Milan, Italy). The amplified products were inserted individually into the pLB cloning vector (Tiangen, Beijing, China), and the resulting constructs were introduced by transformation into Escherichia coli DH5α cells. At least five positive clones were sequenced, and the resulting nucleotide sequences and predicted protein sequences were analyzed using Snap V4.1.9 software (GSL Biotech, San Diego, CA). Multiple alignments using nucleotide and amino acid sequences were done using MEGA X software as described, with default settings \[3\].

A BLASTx search identified a contig of 7621 nt that covered 99.8% (7621 nt out of 7635 nt) and shared 99% nt sequence identity with genomic RNA1 of wheat yellow mosaic virus (WYMV) (FJ361766.1). The BLASTx search also identified a contig of 3634 nt that covered 99.6% (3634 nt out of 3650 nt) and shared 99% nt sequence identity with genomic RNA2 of WYMV (FJ361769.1). This result supported the result obtained by transmission electron microscopy (Fig. 1c and d) and indicated that WYMV was one of the viruses found in the diseased wheat plants. In addition, another contig of 8403 nt was found to share the highest amino acid sequence identity (37.0%) with the replication-associated particles of approximately 200-500 × 13 nm in size. (e, f) Transmission electron micrographs showing negatively stained filamentous virus-like particles of approximately 1000-1300 × 13 nm in size. Bar ≈ 200 nm.
protein of peach chlorotic mottle virus (PCMV). Identification of this novel contig indicated the presence of an unknown virus in the infected wheat plants that might be a member of the family Betaflexiviridae. To confirm this, we amplified the full-length novel virus sequence by RT-PCR, followed by 5’ and 3’ RACE. The sequencing results showed that the genomic RNA of the novel virus contained 8140 nt, excluding its poly(A) tail (GenBank accession no. MW771279). ORFs of the viral genomic RNA were then predicted using ORF Finder software (http://www.ncbi.nlm.nih.gov/orffinder), and the proteins were predicted using CDD software [7]. The results showed that this novel virus has six ORFs (Supplementary Fig. S1), a 59-nt 5’ untranslated region (UTR), and an 11-nt 3’ UTR. ORF1 starts from an AUG start codon at nt position 60 and ends at a TAA stop codon at nt position 6213. This ORF1 was predicted to encode a putative Rep protein with 2070 aa residues and an estimated molecular mass (M_r) of 233.8 kDa. This putative Rep contains four conserved domains: a methyltransferase domain (aa position 44-357, pfam01660), a peptidase domain (aa position 1080-1167, cl05111), a helicase domain (aa position 1254-1506, cl03049), and an RNA-dependent RNA polymerase domain (aa position 1678-1976, cl03049). Sequence alignment analysis showed that the Rep aa sequence of this novel virus shared the highest aa sequence identity (37.0%) with the Rep protein of PCMV (Supplementary Table S1). Further analysis of ORF2, ORF3, and ORF4, which encode triple gene block (TGB1, TGB2, and TGB3) proteins, showed that ORF2-4 of the novel virus were similar to those of other viruses in the subfamily Quinvirinae [1]. For example, ORF2 of the novel virus was predicted to start from an AUG codon at nt position 6286 and end at a UAG stop codon at nt position 6957 (Supplementary Fig. S1). It was predicted to encode a 24.5-kDa polypeptide containing a conserved helicase domain (aa position 24-218, pfam01443) and shared the highest aa sequence identity (42.5%) with TGB1 of PCMV (Supplementary Fig. S1, Supplementary Table S1). ORF3 of the novel virus was predicted to start at nt position 6947 and end at nt position 7309. It was predicted to encode a 13.4-kDa protein containing a conserved viral movement protein domain (aa position 5-104, cl03157) and shared the highest aa sequence identity (43.2%) with TGB3 of elderberry carlavirus A (EBCVA) (Supplementary Fig. S1, Supplementary Table S1). ORF4 of the novel virus was predicted to start at nt position 7218 and end at nt position 7475. It was predicted to encode a 9.2-kDa protein that shared the highest aa sequence identity (33.8%) with TGB4 of butterbur mosaic virus (BMV) (Supplementary Table S1). ORF5 of the novel virus was predicted to start at nt position 7487 and end at nt position 8278 (Supplementary Fig. S1). It was predicted to encode a 28.5-kDa protein with a conserved flexi CP domain (aa 83-220, cl02836). This protein shared the highest aa sequence identity (33.5%) with the CP of PCMV (33.5%) (Supplementary Fig. S1, Supplementary Table S1). ORF6 of the novel virus partially overlaps with ORF5 and was predicted to start at nt position 7941 and end at nt position 8399. This ORF was predicted to encode a 17.4-kDa protein with unknown function, and it was not found to have any significant sequence similarity to other known proteins. An alignment of the genome sequences of this novel virus and eight representative members of the subfamily Quinvirinae showed that the genome of this novel virus shared the highest nt sequence similarity (45.6% identity) with the genome of PCMV (Supplementary Table S1).

Phylogenetic analysis using whole genome sequences of the novel virus and representative viruses of five different families in the order Tymovirales indicated that the novel virus was most closely related to members of the family Betaflexiviridae (Fig. 2a). Phylogenetic trees constructed using whole genome sequences as well as Rep or CP sequences of the novel virus and the representative viruses of the family Betaflexiviridae showed that the novel virus should be placed in the subfamily Quinvirinae but is distinct from the members of existing genera (Fig. 2b–d). Alignment analysis showed that the viral genome sequences and protein sequences of the novel virus and the other members of the subfamily Quinvirinae were 41.8% to 45.6% identical and 14.7% to 43.2% identical, respectively (Supplementary Table S1). For Rep and CP, the aa sequence identity values in the comparison between the novel virus and the other members ranged from 22.9% to 37%. These genome, Rep, and CP sequence differences not only exceed the cutoff values for species discrimination (i.e., 72% for the nt sequence and 80% for the aa sequences) but also exceed the cutoff values for genus discrimination (i.e., 45% for the nt and aa sequences) [1]. Consequently, we consider the novel virus a member of a new species in the subfamily Quinvirinae. Due to differences in virus genome size, genome structure, and phylogenetic relationships to other members in the family Betaflexiviridae, we were unable to place this novel virus in any existing genus in the family Betaflexiviridae. Therefore, we tentatively named it "wheat yellow stunt-associated betaflexivirus" (WYSaBV) and propose that it be placed in a new genus in the family Betaflexiviridae.

To investigate the infectivity of WYSaBV, leaves of the plants showing typical yellow mosaic and stunting were ground in 0.1 M phosphate buffer and then mechanically inoculated to several assay plants, including wheat, Nicotiana benthamiana, N. occidentalis, N. tabacum cv. Samsun, Chenopodium quinoa, C. amaranticolor, and Gomphrena globosa grown in a greenhouse at 25°C. All of the inoculated plants remained symptomless during the study, and the
novel virus was not detected in any of them by RT-PCR at 30 days post-inoculation. It therefore remains to be investigated how WYSaBV is transmitted in the field. Nonetheless, identification of WYSaBV on wheat plants in this study indicates the need for caution in developing an effective

Clusters with bootstrap values less than 50 are not shown. The trees are drawn to scale, and the lengths of branches in the same cluster represent the evolutionary distance computed using the Poisson correction method. Evolutionary analysis was conducted using MEGA X software.

Fig. 2 Phylogenetic analysis using the genomic sequences or the Rep and CP sequences of WYSaBV and representative members in several genera of the order Tymovirales or family Betaflexiviridae. Evolutionary relationships among these viruses were determined using the neighbor-joining method with 1000 bootstrap replications.
A novel wheat-infecting virus of the family *Betaflexiviridae*

management strategy for virus disease control in wheat fields in China.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00705-021-05175-y.

**Acknowledgements** This research was supported by grants from the National Natural Science Foundation of China for Youth Scholars (31801701) and the China Postdoctoral Science Foundation (2019T120521 and 2018M642449). We thank Dr. Xinshun Ding (retired, New York, USA) for his help on the language of this manuscript.

**Author contributions** JW designed the experiments; SF, MH, and BS performed the experiments; TZ performed the bioinformatics analysis; JW, XZ, and SF wrote the manuscript.

**Declarations**

**Conflict of interest** All authors declare they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**References**

1. Adams MJ, Antoniw JF, Bar-Joseph M, Brunt AA, Candresse T, Foster GD, Martelli GP, Milne RG, Zavriev SK, Fauquet CM (2004) The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. Adv Virol 149:1045–1060
2. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacochen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652
3. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
4. Lai K, Berkman PJ, Lorenc MT, Duran C, Smits L, Manoli S, Stiller J, Edwards D (2012) WheatGenome.info: an integrated database and portal for wheat genome information. Plant Cell Physiol 53:e2
5. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB (2018) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res 46:D708-d717
6. Liu Y, Khine MO, Zhang P, Fu Y, Wang X (2020) Incidence and distribution of insect-transmitted cereal viruses in wheat in China from 2007 to 2019. Plant Dis 104:1407–1414
7. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Feng JH, Ger L, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thakker N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res 39:D225-D229
8. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2014) The pangenome of hexaploid bread wheat. Plant J 90:1007–1013
9. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ (2016) Redefining the invertebrate RNA virosphere. Nature 540:539–543
10. Wu Q, Ding SW, Zhang Y, Zhu S (2015) Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. Annu Rev Phytopathol 53:425–444

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.