

Complete genome sequence of a novel iflavivirus from wheat sawfly (*Dolerus tritici*)

Jiashu Guo

Chinese Academy of Agricultural Sciences Institute of Plant Protection

Wenwen Liu

liuwenwen01@caas.cn

Chinese Academy of Agricultural Sciences Institute of Plant Protection <https://orcid.org/0000-0001-7239-0851>

Chen Chen

Chinese Academy of Agricultural Sciences Institute of Plant Protection

Frederic Francis

Universite de Liege Gembloux Agro-Bio Tech

Xifeng Wang

Chinese Academy of Agricultural Sciences Institute of Plant Protection

Research Article

Keywords:

Posted Date: November 4th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-5294652/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Archives of Virology on December 17th, 2024. See the published version at <https://doi.org/10.1007/s00705-024-06206-0>.

Abstract

Little is known about the commensal viruses in wheat sawfly, *Dolerus tritici*, which is an important agricultural insect feeding on wheat leaves. Here, we identified a novel single positive-strand RNA virus from the larvae of wheat sawfly collected in northern China through RNA sequencing, and then determined its complete genome sequence by rapid amplification of cDNA ends. The complete genome consists of 9,594 bp in size including a poly A tail at 3' terminus, which is predicted encoding a 326.3 kD polyprotein. The phylogenetic tree analysis of deduced polyprotein revealed that this RNA virus was clustered in the clade with deformed wing virus in genus *Iflavirus*, family *Iflaviridae*. The full genome of this RNA virus shows 38.9-50.0% sequence identities with other iflaviruses. Sequence identity at the amino acid level of coat proteins shows this RNA virus has highest identity with slow bee paralysis virus at 33.6%, which no more than 90% suggesting that it is a new member in the genus *Iflavirus*. Thus, we tentatively designated as "Dolerus tritici iflavirus 1" (DtIV1). According to our knowledge, this is the first report of the commensal virus in wheat sawfly.

Full Text

Wheat sawfly (*Dolerus tritici* Chu) is an important global pest of common wheat (*Triticum aestivum* L.). The larvae of wheat sawfly damage wheat leaves, causing a severe decline in wheat production [1]. With the development of high-throughput sequencing technology, more commensal viruses have been found in agricultural insects [2]. Insects from order Hymenoptera account for the highest proportion with nearly 300 viruses, and most of them come from beneficial insect bees [2]. As pests, the sawflies have rarely been found any commensal virus until now. In the present study, we isolated and identified a novel putative iflavirus from wheat sawfly, tentatively named Dolerus tritici iflavirus 1 (DtIV1).

Viruses in the genus *Iflavirus*, the only genus of family *Iflaviridae*, possess a positive-strand RNA genome with 9-11 kilobases (kb), with non-enveloped virions [3]. The genome of iflavirus has an open reading frame (ORF) that encodes a polyprotein followed by proteolytic cleavage into functional structural proteins and non-structural proteins [4]. A genome-linked viral protein (VPg) participating in viral life cycle is covalently attached to 5'-end of the genome, and the 3'-end of the genome is polyadenylated [5]. According to the International Committee on Taxonomy of Viruses (ICTV), the genus *Iflavirus* includes 16 approved species [3]. All iflaviruses are identified from arthropods insects [6-8]. Some of them are associated with symptomless infections [9,10]. While some can be harmful to host insects, such as deformed wing virus, which causes characteristic wing deformity and premature mortality of honeybees [11].

During a field investigation in March 2024, wheat sawflies were found feeding on wheat plants in Yuanyang, Henan province, China. We randomly pooled three larvae for RNA sequencing (RNA-seq). Total RNA was extracted from the sample using TRIzol reagent according to the manufacturer's instructions (Invitrogen, Carlsbad, CA) and then a library was constructed for RNA-seq by the Illumina HiSeq X Ten platform. After removal of adaptor sequences, The CLC Genomics Workbench 9.5 was used

for de novo assembly of RNA-seq data. The assembled contigs were subsequently screened against the NCBI databases using BLASTn and BLASTx searches with default options (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Three viral contigs were derived from an unidentified virus that was related to members of the family *Iflaviridae*.

To obtain the full genome sequence of DtIV1, we used SMARTer RACE 5'/3' Kit (Takara Bio, USA) for performing both 5'- and 3'-rapid amplification of cDNA ends (RACE) to get the terminal sequences. Subsequently, RT-PCR was carried out using the total RNA extracted from the single larva and primers which was designed using the software Oligo 7.60 (OLIGO, Colorado Springs, CO) (Table S1). Three fragments of DtIV1 with overlap sequences were amplified using RT-PCR. (Fig. S1). We amplified the fragments of 3071 bp, 2893 bp and 3752 bp as well as well as 562 bp at 5' terminal and 1433 bp at 3' terminal with overlap sequences. Then, sequences were assembled using the DNAMAN (v6) program (Lynnon Biosoft, San Ramon, CA). The whole viral genome was 9,594 nucleotides (nt) in length including the polyA tail (GenBank accession PQ323359), with the 5' untranslated region (UTR) of 807 nt and 3' UTR of 180 nt. Moreover, it was polyadenylated as viruses in family *Iflaviridae* [3], suggested that the virus sequence was complete (Fig.1). The genome was predicted to have a single ORF (808 nt- 9414 nt) using Open Reading Frame Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) that encoded a 326.3 kDa polyprotein comprising 2868 aa. Compared to the deduced amino acid (aa) sequences obtained from RNA-seq, it revealed seven aa were varied. Totally 18 ATG triplets existed in 5' UTR and complicated stem-loop structures were predicted in the 5' UTR by a secondary structure prediction RNAfold Webserver (<http://rna.tbi.univie.ac.at/>). Sequence identities were computed using LALIGN program of EMBL's EBI European Bioinformatics Institute (EMBL-EBI) with the default settings (EMBOSS Needle < EMBL-EBI). The identities of deduced amino acid sequence of coat protein (CP) with those of species in family *Picornavirales* were 6.6% to 33.6%, which was much lower than the species demarcation thresholds in the genus *Iflavirus* [3](Table S2).

The deduced polyprotein of DtIV1 was predicted to cleaved into mature and functional proteins by the putative autocatalytic and 3C cysteine protease cleavage sites according to the characteristics of iflaviruses [10-12] (Fig. 1A). The Conserved Domain Database (CDD) at NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd>) and InterProScan (<https://www.ebi.ac.uk/interpro/>) were used to analysis of the conserved domains of DtIV1 and showed a typical characteristic of genus *iflavirus* [3]. At the N-terminus of polyprotein, two rhinovirus-like (Rhv) structural protein domains: Rhv1 domain (accession no. IPR033703) and Rhv2 domain (accession no. IPR033703), and a cricket paralysis virus coat domain (CRPV) (accession no. IPR014872) were identified. The three conserved domains at residues 358-524, 632-793, and 995-1207 corresponded to VP3, VP1 and VP2 of DtIV1, respectively. The predicted coat proteins of DtIV1 contained four conserved sequences in structural protein region, as described in picorna-like viruses [13,14] (Fig. 1B). The conserved VP4 cleavage site (Nx/DxP) was also found at aa 535 position, at the end of the C-terminus of the VP3, suggesting that a small protein VP4 was positioned right between VP3 and VP1 [15]. At the C-terminus of polyprotein, non-structural proteins appeared in the following order: RNA helicase protein (domain accession no. IPR014759) at residues 1509-1676, 3C cysteine protease protein (domain accession no. IPR009003) at residues 2125-2327, and

RNA-dependent RNA polymerase (domain accession no. IPR001205) at residues 2386-2839. Three conserved helicase superfamily domains were identified in DtIV1 that were usually associated with NTP-binding, which were Hel-A (Gx2GxGKS), Hel-B (Qx5DD), and Hel-C (KKx4Px5NSN). While the motif Hel-C of DtIV1 was slightly different from the consensus sequence KGx4Sx5STN [16] (Fig. 1C). The 3C cysteine protease domains of picornaviruses, including a cysteine-protease motif (GxCG) and a substrate binding motif (GxHx2G) were found in DtIV1 [16,17] (Fig. 1C). All eight recognized RNA-dependent RNA polymerase (RdRp) domains featured in picorna-like family were also found in DtIV1 (Fig. 1D). The core domains (IV, V, and VI) thought to be involved in catalysis and NTP binding of virus life [16,18,19]. Besides that, a highly conserved TSxGxP domain, similar to that found in some picornaviruses, was located immediately prior to the RdRp domain I [11,20]. Overall, DtIV1 possessed characteristic domains of structural proteins, RNA helicase, 3C cysteine protease, and RdRp of iflaviruses.

The identity values for the genomic nucleotide sequence and amino acid sequence of the polyprotein and CP of DtIV1 with the viruses in order *Picornavirales* were calculated using EMBL. The identities were only 38.9-50.0% for total genome, 9.7-32.6% for polyprotein, and 6.6-33.6% for CP (Table S2). Phylogenetic trees based on the deduced amino acid sequences of polyprotein (Fig. 2A) and RdRp (Fig. 2B) were constructed using the neighbor-joining (NJ) method with 1,000 bootstrap in MEGA 11. foot-and-mouth disease virus A and human poliovirus 1 of picorna-like viruses belonging to *Picornaviridae*, and heterosigma akashiwo RNA virus of *Marnaviridae* were used as outgroups. In both trees, the sequences of DtIV1 that we identified were clustered in a clade of genus *iflavirus*, with the closest relationship with deformed wing virus.

In conclusion, we report the identification and genomic characterization of a new virus naturally infecting wheat sawfly (*D. tritici*). Based on its overall genome sequence, structure, and phylogenetic relationships, DtIV1 can be regarded as a member of the genus *iflavirus*, family *Iflaviridae*. To our knowledge, DtIV1 is the first novel commensal virus in wheat sawfly. Further research is needed to investigate the origin and host range of the virus and to assess the possible impact of DtIV1 on its host insect in the wheat field ecosystem.

Declarations

Acknowledgements The authors thank Prof. Yingdang Ren (Henan Academy of Agricultural Sciences) and Dr. Zhongtian Xu (Ningbo University) for assistance with collection of insect samples and data analysis.

Funding This research was supported by National Natural Science Foundation of China (32270173).

Data availability The complete nucleotide sequence of DtIV1 has been deposited in the NCBI GenBank database under the accession number PQ323359.

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

Conflict of interest The authors declare that there are no conflicts of interest.

References

1. Chu HF (1949) The wheat sawfly, *Dolerus tritici* new species. Contributions from the Institute of Zoology, National Academy of Peiping 5(3): 79-92
2. Qi YH, Ye ZX, Zhang CX, Chen JP, Li JM (2023) Diversity of RNA viruses in agricultural insects. Comput Struct Biotech 21: 4312-4321
3. Valles SM, Chen Y, Firth AE, Dma G, Hashimoto Y, Herrero S, de Miranda JR, Ryabov E, Consortium IR (2017) ICTV Virus Taxonomy Profile: *Iflaviridae*. J Gen Virol 98(4):527–528
4. van Oers MM (2010) Genomics and biology of Iflaviruses. Insect Virol 231–250
5. Hashimoto Y, Watanabe A, Kawase S (1986) Evidence for the presence of a genome-linked protein in infectious flacherie virus. Arch Virol 90 (3-4):301-312
6. Wu N, Zhang PP, Liu WW, Cao MJ, Massart S, Wang XF (2019) Complete genome sequence and characterization of a new iflavirus from the small brown planthopper (*Laodelphax striatellus*). Virus Res 272: 197651
7. Wang H, Liu Y, Liu W, Cao M, Wang X (2019) Full genome sequence of a novel iflavirus from the leafhopper *Psammotettix alienus*. Arch Virol 164(1): 309-311
8. Wu N, Zhang P, Liu W, Cao M, Wang X (2018) Sequence analysis and genomic organization of a new insect iflavirus, Sogatella furcifera honeydew virus 1. Arch Virol 163(7): 2001-2003
9. Murakami R, Suetsugu Y, Kobayashi T, Nakashima N (2013) The genome sequence and transmission of an iflavirus from the brown planthopper, *Nilaparvata lugens*. Virus Res 176:179–187
10. Parry R, Naccache F, Ndiaye EH, Fall G, Castelli I, Lühken R, Medlock J, Cull B, Hesson JC, Montarsi F, Failloux AB, Kohl A, Schnettler E, Diallo M, Asgari S, Dietrich I, Becker SC (2020) Identification and RNAi profile of a novel iflavirus infecting senegalese *Aedes vexans arabiensis* mosquitoes. Viruses 12(4):440
11. Lanzi G, de Miranda JR, Boniotti MB, Cameron CE, Lavazza A, Capucci L, Camazine SM, Rossi C (2006) Molecular and biological characterization of deformed wing virus of honeybees (*Apis mellifera* L.). J Virol 80(10):4998-5009
12. Geng P, Li W, Lin L, de Miranda JR, Emrich S, An L (2014) Genetic Characterization of a novel iflavirus associated with vomiting disease in the chinese oak silkworm *Antheraea pernyi*. Plos One 9(3): e92107
13. Liljas L, Tate J, Lin T, Christian P, Johnson JE (2002) Evolutionary and taxonomic implications of conserved structural motifs between picornaviruses and insect picorna-like viruses. Arch Virol 147: 59-84
14. Chen B, Chen YL, Chen HZ, Liang ZY, Chen JH, Wu RF, Zhang T, Zhou GH, Yang X (2023) Identification, characterization and prevalence in southern China of a new iflavirus in the leafhopper *Recilia dorsalis* (Hemiptera: Cicadellidae). Virus Res 323:199005

15. Jia WX, Wang F, Li JJ, Chang XF, Yang Y, Yao HW, Bao YY, Song QS and Ye GY (2021) A novel iflavirus was discovered in green rice leafhopper *Nephotettix cincticeps* and its proliferation was inhibited by infection of rice dwarf virus. *Front Microbiol* 11:621141
16. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol* 28:375–430
17. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV (1989) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Lett* 243(2):103–114
18. Baker AC, Schroeder DC (2008) The use of RNA-dependent RNA polymerase for the taxonomic assignment of Picorna-like viruses (order *Picornavirales*) infecting *Apis mellifera* L. populations. *Virol J* 5(1):10
19. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72(9): 2197-2206
20. de Miranda JR, Dainat B, Locke B, Cordonni G, Berthoud H, Gauthier L, Neumann P, Budge GE, Ball BV, Stoltz DB (2010) Genetic characterization of slow bee paralysis virus of the honeybee (*Apis mellifera* L.). *J Gen Virol* 91:2524-2530.

Figures

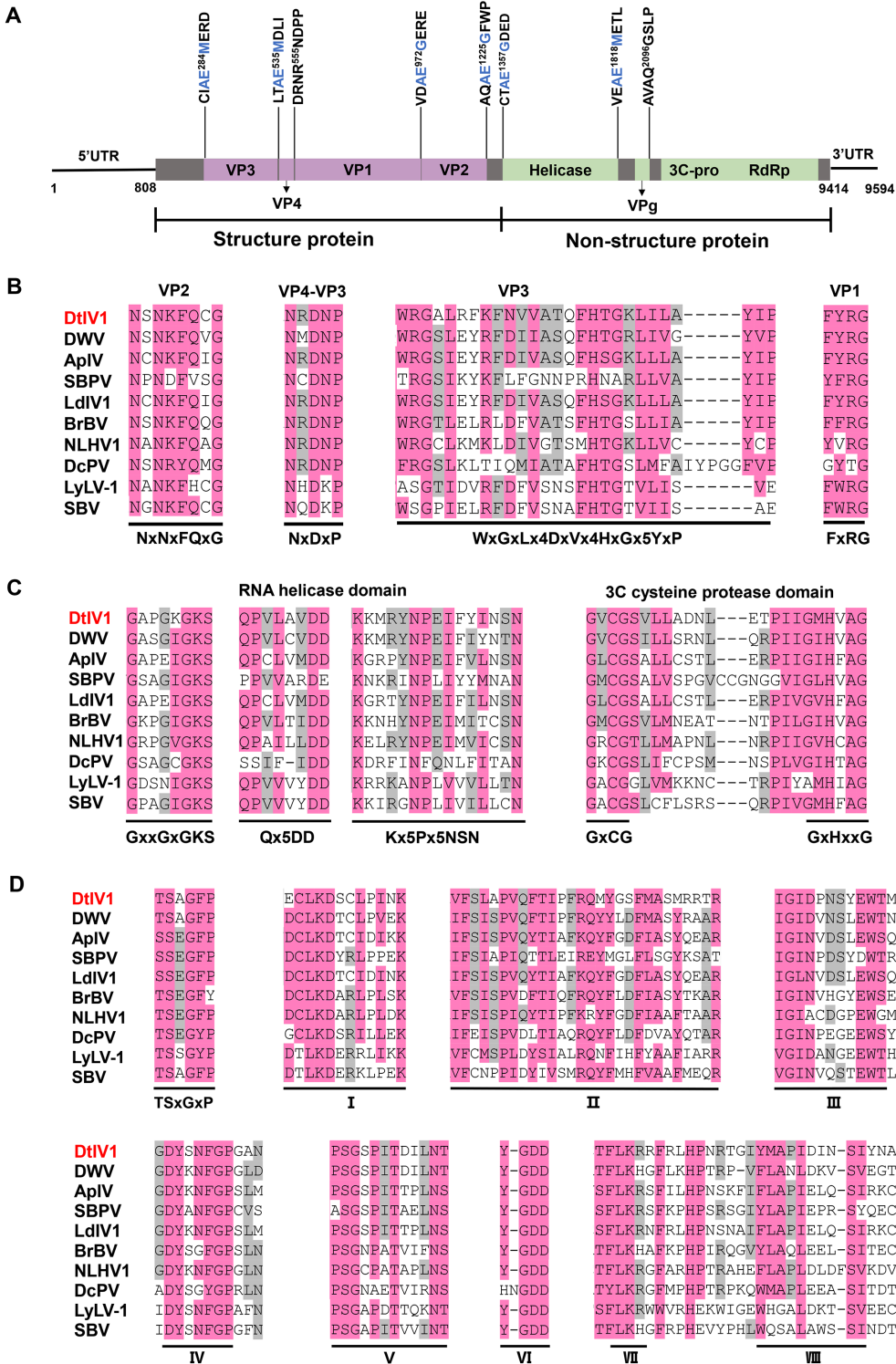


Figure 1

Analysis of the genomic characteristics of *Dolerus tritici* iflavirus 1 (DtIV1). (A) Genome organization schematic of DtIV1 with vertical lines indicating conserved 3C-protease cleavage sites inferred by amino acid alignment and conserved protease consensus sites (conserved amino acids in blue font; variable amino acids in black font). Alignment of VP1-VP4 proteins (B), RNA helicase and 3C-protease proteins (C) and RdRp (D) of DtIV1 among the iflaviruses using the deduced amino acids. The amino

acids shaded in pink indicate the conservative and similar residues. Viral information is given in Table S2.

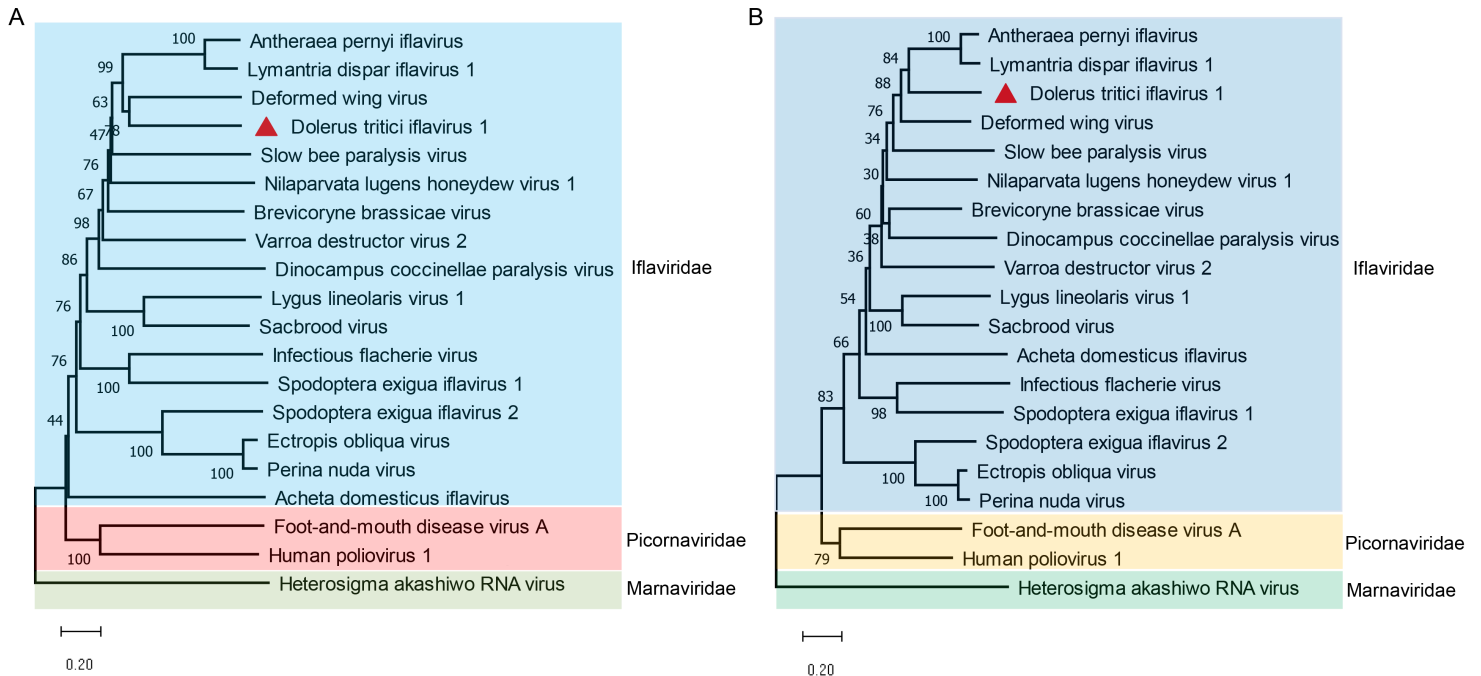


Figure 2

Phylogenetic analysis among the species in genus *Iflavirus* with other picorna-like viruses using deduced amino acids of polyprotein(A) and RdRp (B).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [1019Supplementaryinformation.docx](#)