Archives of Virology (2025) 170:21
https://doi.org/10.1007/s00705-024-06206-0

ANNOTATED SEQUENCE RECORD

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

Jiashu Guo[1,2] · Wenwen Liu[1] · Chen Chen[1] · Zhongtian Xu[3] · Frederic Francis[2] · Xifeng Wang[1]

Received: 19 October 2024 / Accepted: 13 November 2024

Abstract

Little is known about the insect viruses in wheat sawfly, *Dolerus tritici*, which is an important agricultural insect feeding on wheat leaves. Here, we used RNA sequencing to identify a novel single positive-strand RNA virus from the larvae of wheat sawfly collected in northern China and then determined its complete genome sequence by rapid amplification of cDNA ends. The complete genome is 9,594 nt in length, including a polyA tail at its 3′ terminus, and it is predicted to encode a 326.3-kDa polyprotein. Phylogenetic analysis based on deduced amino acid sequences of the polyprotein revealed that this RNA virus clustered in a clade with deformed wing virus of the genus *Iflavirus*, family *Iflaviridae.* The full genome of this RNA virus shows 42.0–50.0% sequence identity with other iflaviruses. Comparisons of amino acid sequences showed that the coat protein of this RNA virus is most similar to that of slow bee paralysis virus, with 33.6% identity, suggesting that this virus is a new member in the genus *Iflavirus*. Thus, we have tentatively designated it as "Dolerus tritici iflavirus 1" (DtIV1). To our knowledge, this is the first report of an insect virus in wheat sawfly.

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, many viruses have been found in agricultural insects [2–4]. Insects of the order Hymenoptera account for the highest proportion, with nearly 300 viruses, most of which were found in beneficial insects, such as bees [2, 3]. In contrast, little is known about viruses that infect wheat sawflies. In the present study, we isolated and identified a novel putative iflavirus from wheat sawfly, which we have tentatively named "Dolerus tritici iflavirus 1" (DtIV1).

The family *Iflaviridae* contains a single genus, *Iflavirus*. According to the International Committee on Taxonomy of Viruses (ICTV), the genus *Iflavirus* includes 16 approved species [5], and all of the iflaviruses described so far were identified in arthropod species [6–8]. Members of the genus *Iflavirus* possess a positive-strand RNA genome, 9–11 kilobases (kb) in length, and form non-enveloped virions [5]. Iflavirus genomes usually have a single open reading frame (ORF) that encodes a polyprotein that is processed by proteolytic cleavage to generate functional structural proteins and non-structural proteins [9]. A genome-linked viral protein (VPg) that is important in the viral life cycle is covalently attached to 5′ end of the genome, and the 3′ end of the genome is polyadenylated [10]. While some iflaviruses are associated with symptomless infections [11, 12], some can be harmful to the host insect, such as deformed wing virus, which causes characteristic wing deformity and premature mortality of honeybees [13].

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 a library was constructed and used for RNA-seq on an Illumina HiSeq X Ten platform. After removal of adaptor sequences,

Handling Editor Simona Abba'

✉ Wenwen Liu
liuwenwen01@caas.cn

[1] State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, 100193 Beijing, China

[2] Functional & Evolutionary Entomology, Gembloux Agro-BioTech, University of Liège, 5030 Gembloux, Belgium

[3] Institute of Plant Virology, Ningbo University, Ningbo, China

Published online: 17 December 2024

CLC Genomics Workbench 9.5 was used for *de novo* assembly of the RNA-seq data. The assembled contigs were used to search the NCBI databases for related sequences, using BLASTn and BLASTx with default options (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Three viral contigs were found to be derived from an unidentified virus that was related to members of the family *Iflaviridae*.

To obtain the full genome sequence of DtIV1, we used a SMARTer RACE 5′/3′ Kit (Takara Bio, USA) to perform both 5′ and 3′ rapid amplification of cDNA ends (RACE) to determine the terminal sequences. Subsequently, RT-PCR was carried out using total RNA extracted from a single larva and specific primers that were designed using the software Oligo 7.60 (OLIGO, Colorado Springs, CO) (Supplementary Table S1). Three fragments of the DtIV1 genome with overlapping sequences were amplified using RT-PCR (Supplementary Fig. S1). Fragments of 3071, 2893, and 3752 nt from the coding region as well as 562 nt from the 5′ terminus and 1433 nt from the 3′ terminus were amplified and sequenced, and these sequences were assembled using the DNAMAN (v6) program (Lynnon Biosoft, San Ramon, CA). The whole viral genome was found to be 9,594 nucleotides (nt) in length, including the polyA tail (GenBank accession no. PQ323359), with a 5′ untranslated region (UTR) of 807 nt and a 3′ UTR of 180 nt (Fig. 1A). When the clean reads from RNA-seq were aligned with the DtIV1 genome, almost complete coverage of the DtIV1 genome was observed. In total, 985,214 reads were mapped to the DtIV1 genome, with an average sequencing depth of 14,831x. (Fig. 1A). Using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/), the genome was predicted to have a single ORF (nt 808–9414) that encoded a 326.3-kDa polyprotein comprising 2868 amino acids (aa). The deduced amino acid (aa) sequences obtained by RNA-seq differed at seven positions from those obtained by RT-PCR and RACE. A total of 18 ATG triplets were found in the 5′ UTR, which was predicted using the RNAfold Webserver (http://rna.tbi.univie.ac.at/) to form complex stem-loop structures. Sequence identity values were computed using the LALIGN program of the EMBL European Bioinformatics Institute (EMBL-EBI) with default settings (EMBOSS Needle < EMBL-EBI). The deduced amino acid sequence of the coat protein (CP) of DtIV1 was 6.6–33.6% identical to those of members of the order *Picornavirales*, which is far below the species demarcation threshold for the genus *Iflavirus* [5] (Supplementary Table S2).

The putative polyprotein of DtIV1 was predicted to be cleaved into mature and functional proteins at autocatalytic and 3C cysteine protease cleavage sites corresponding to those of other iflaviruses [12–14] (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 identify conserved domains, and the results showed DtIV1 to have typical characteristics of members of the genus *Iflavirus* [5]. At the N-terminus of polyprotein, two rhinovirus-like (Rhv) structural protein domains – an Rhv1 domain (accession no. IPR033703) and an Rhv2 domain (accession no. IPR033703) – and a cricket paralysis virus coat domain (CRPV) (accession no. IPR014872) were identified. Three conserved domains at residues 358–524, 632–793, and 995–1207 corresponded to VP3, VP1, and VP2, respectively. The predicted coat proteins of DtIV1 contained four conserved sequences that have been described in picorna-like viruses [15, 16] (Fig. 1B). The conserved VP4 cleavage site (Nx/DxP) was also found at aa position 535, at the C-terminal end of VP3, suggesting that a small protein, VP4, was positioned between VP3 and VP1 [17]. In the C-terminal portion of the polyprotein, non-structural proteins are encoded 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 are usually associated with NTP binding: Hel-A (Gx2GxGKS), Hel-B (Qx5DD), and Hel-C (KKx4Px5NSN), with the Hel-C motif of DtIV1 differing slightly from the consensus sequence KGx4Sx5STN [18] (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 [18, 19] (Fig. 1C). All eight recognized RNA-dependent RNA polymerase (RdRp) domains that are found in picorna-like viruses were also found in DtIV1 (Fig. 1D). The core domains (IV, V, and VI) are thought to be involved in catalysis and NTP binding [18, 20, 21]. A highly conserved TSxGxP domain, similar to those found in some picornaviruses, was located immediately prior to the RdRp domain I [13, 22]. Overall, DtIV1 possessed characteristic domains of structural proteins, RNA helicase, 3C cysteine protease, and RdRp of iflaviruses.

The identity values from a comparison of the genomic nucleotide sequence and the amino acid sequences of the polyprotein and CP of DtIV1 to those of selected members of the order *Picornavirales* were calculated using EMBOSS Needle. These values were 38.9–50.0% for the whole genome, 9.7–32.6% for the polyprotein, and 6.6–33.6% for the CP (Supplementary Table S2). Phylogenetic trees based on the deduced amino acid sequences of the polyprotein (Fig. 2A) and the RdRp (Fig. 2B) were constructed using the neighbor-joining (NJ) method with 1,000 bootstrap replicates in MEGA 11. Foot-and-mouth disease virus A and human poliovirus 1 of the family *Picornaviridae* and

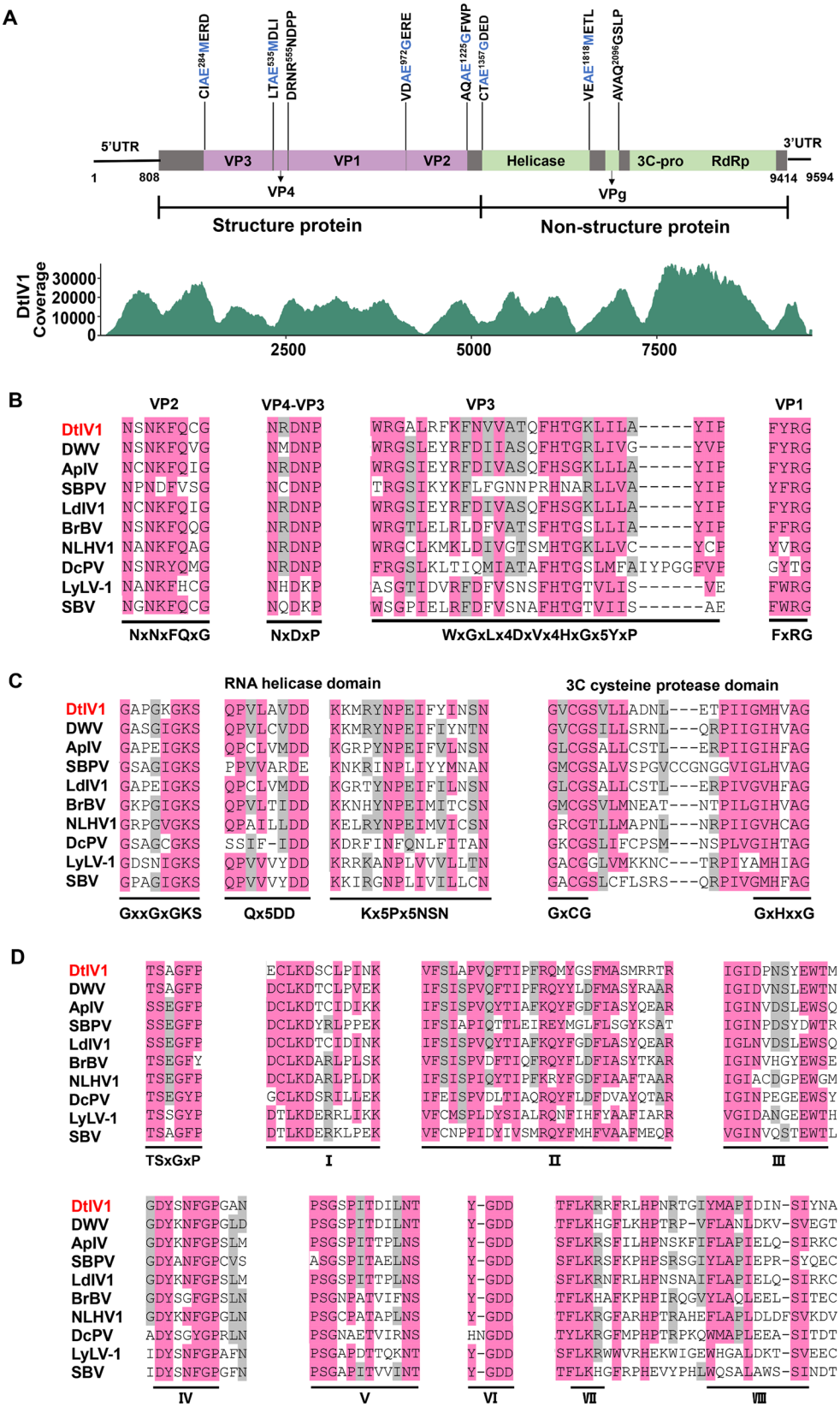

Fig. 1 Analysis of the genomic characteristics of Dolerus tritici iflavirus 1 (DtIV1). (**A**) Genome organization of DtIV1 and the coverage of transcriptome reads. Vertical lines indicate conserved 3C protease cleavage sites, which were inferred by amino acid alignment and conserved protease consensus sites are shown with conserved amino acids in blue and variable amino acids in black. (**B-D**) Amino acid sequence alignment of VP1-VP4 proteins (B), RNA helicase and 3C-protease proteins (**C**), and RdRp (**D**) of DtIV1 and other iflaviruses. The amino acids shaded in pink indicate the chemically similar residues. Information about the viruses used in the comparison is given in Supplementary Table S2.

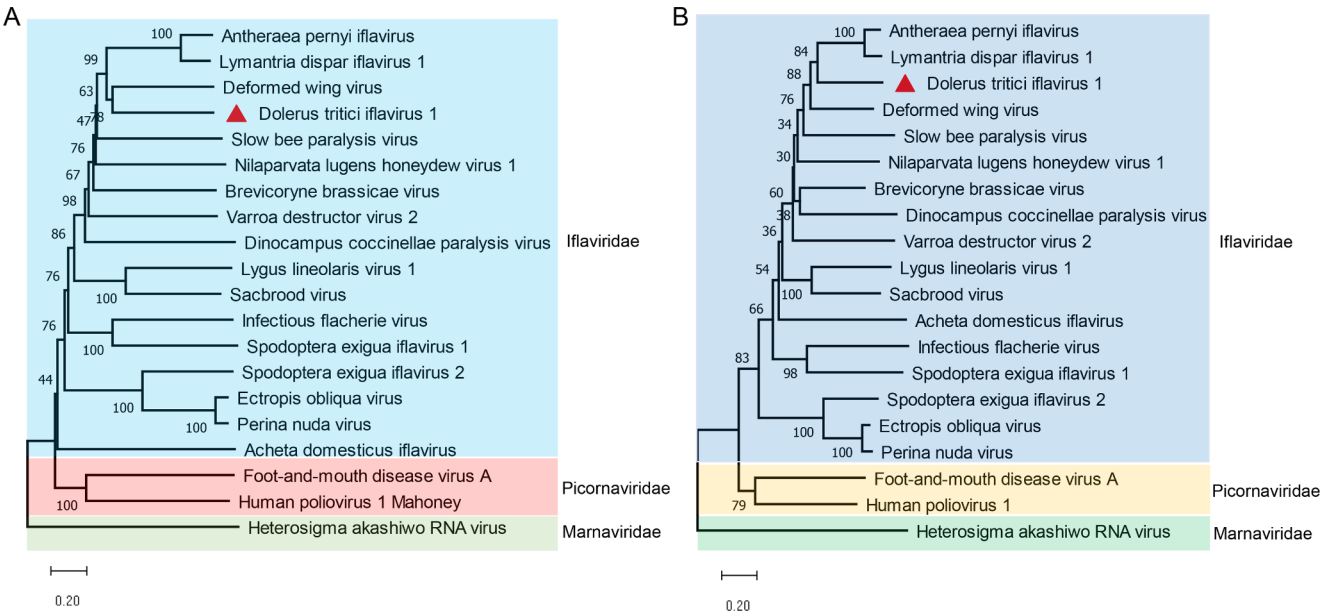

Fig. 2 Phylogenetic analysis of members of the genus *Iflavirus* and other picorna-like viruses based on deduced amino acid sequences of the polyprotein (**A**) and the RdRp (**B**)

Heterosigma akashiwo RNA virus of the family *Marnaviridae* were used as outgroups. In both trees, DtIV1 clustered with members of the genus *Iflavirus*, with deformed wing virus being the closest relative.

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 insect virus identified in a wheat sawfly. Further research is needed to investigate the origin and host range of the virus and to assess its possible impact on its host in the wheat field ecosystem.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-024-06206-0.

Acknowledgements The authors thank Prof. Yingdang Ren (Henan Academy of Agricultural Sciences) for assistance with collection of insect samples.

Funding This research was supported by the 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. The raw data from RNA sequencing have been deposited in the NCBI Sequence Read Archive database under the accession number PRJNA1176038.

Declarations

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

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

References

1. Chu H (1949) The wheat sawfly, *Dolerus tritici* new species. Contrib Inst Zool Natl Acad Peiping 5(3):79–92
2. Qi Y, Ye Z, Zhang C, Chen J, Li J (2023) Diversity of RNA viruses in agricultural insects. Comput Struct Biotec 21:4312–4321
3. Shi M, Lin X, Tian J, Chen L, Chen X, Li C, Qin X, Li J, Cao J, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang Y (2016) Redefining the invertebrate RNA virosphere. Nature 540(7634):539–543
4. Wu H, Pang R, Cheng T, Xue L, Zeng H, Lei T, Chen M, Wu S, Ding Y, Zhang J, Shi M, Wu Q (2020) Abundant and diverse RNA viruses in insects revealed by RNA-seq analysis: ecological and evolutionary implications. MSystems 5(4):e00039–e00020
5. 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
6. 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
7. 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
8. Wu N, Zhang P, Liu W, Cao M, Massart S, Wang X (2019) Complete genome sequence and characterization of a new iflavirus from the small brown planthopper (*Laodelphax striatellus*). Virus Res 272:197651
9. van Oers MM (2010) Genomics and biology of iflaviruses. Insect Virol 231–250

10. 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
11. 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
12. 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
13. 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
14. 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 silkmoth *Antheraea pernyi*. PLoS ONE 9(3):e92107
15. 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
16. Chen B, Chen Y, Chen H, Liang Z, Chen J, Wu R, Zhang T, Zhou G, 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
17. Jia W, Wang F, Li J, Chang X, Yang Y, Yao H, Bao Y, Song Q, Ye G (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
18. 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
19. 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
20. 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
21. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J Gen Virol 72(9):2197–2206
22. de Miranda JR, Dainat B, Locke B, Cordoni 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