Acta virologica **65**: 320 – 323, 2021 doi: 10.4149/av_2021_310

Sequencing and phylogenetic characterization of a novel RNA virus in *Arma chinensis*

Yonghao Dong[1, 2], Pengjun Xu[2*], Guangwei Ren[2], Changchun Feng[4], Dongyang Liu[4], Lianqiang Jiang[4], Fangzhao Jia[5], Changhua Zhang[5], Qinggang Gao[3], Yongjie Liu[1*]

[1]College of Plant Protection, Shandong Agricultural University, Taian 271018, P. R. China; [2]Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, P. R. China; [3]Heze Bureau of Agriculture and Rural Affairs, Heze 274000, P. R. China; [4]Liangshan Branch of Sichuan Provincial Tobacco Company, Technical Center, Xichang 615000, P. R. China; [5]Zunyi Branch of Guizhou Tobacco Company, Zunyi 563000, P. R. China

Received June 25, 2020; revised December 18, 2020; accepted July 14, 2021

Summary. – The complete genome of a novel virus from *Arma chinensis* was determined by RNA sequencing and rapid amplification of cDNA ends. This virus has a single-stranded RNA genome of 10,540 nucleotides (nt) excluding the poly(A) tail. Two non-overlapping open reading frames (ORFs) in the sense direction were predicted: one long ORF at the 5' end of the genome (6,219 nt) that encodes a polypeptide of 2,072 amino acids (aa), and one short ORF at the 3' end of the genome (3,033 nt) that encodes a polypeptide of 1,010 aa. Phylogenetic analysis indicated that the virus clusters within a large cluster of currently unidentified picorna-like viruses with a high bootstrap value. We named the virus isolate *Arma chinensis* picorna-like virus 1 (AcPV-1). The prevalence of AcPV-1 infection in samples of *Arma chinensis* from the wild was at a low level (5.48%, 8 positives in 146 samples).

Keywords: Arma chinensis; genomic characterization; phylogenetic analysis; Arma chinensis picorna-like virus 1; prevalence

Arma chinensis (Fallou) (*Hemiptera*: *Pentatomidae*) is a predaceous insect species in its native Asian range that can effectively suppress a wide range of agricultural and forest insect pests, including lepidopteran, coleopteran, hymenopteran and hemipteran species (Zou *et al*., 2015, 2019; Pan *et al*., 2019). Recently, the development of next-generation sequencing facilitated the discovery of many novel covert viruses, wherein several recently reported viruses belong to the order *Picornavirales* (picorna-like viruses) (Shi *et al*., 2016, 2018; Yang *et al*., 2019). Insect picorna-like viruses have a nonenveloped, single-stranded, positive-sense RNA genome (+ssRNA) of approximately 9 kbp and consisting of five viral families, namely *Dicistroviridae*, *Iflaviridae*, *Marnaviridae*, *Picornaviridae*, and *Secoviridae*, and an unassigned group (Cholleti *et al*., 2018; Valles *et al*., 2017). Moreover, Shi (2016) reported that the Kelp Fly Virus Related Clade contained numerous unclassified insect picorna-like viruses. Herein we report the identification of a novel RNA virus by RNA sequencing, isolated from *A. chinensis,* tentatively named *A. chinensis* picorna-like virus 1 (AcPV-1).

The samples of *A. chinensis* were collected at an experimental station of natural enemy breeding in Zunyi, Guizhou Province, China (27.45° N, 106.53° E) in 2019. To detect viruses, we used the whole bodies of *A. chinensis*, including nymphs (n = 100) and adults (n = 100). Per 10 individuals were ground up under liquid nitrogen separately and put into a 1.5 ml tube with 1 ml TRIzol (Invitrogen, Grand Island, USA). After vortexing and centrifugation for 5 min at 12,000 × g, 50 µl of each sample were combined together for extracting total RNA. RNA sequencing

*Corresponding authors: E-mails: lyj@sdau.edu.cn (Y. Liu), xupeng-jun@163.com (P. Xu); phone: +86-0532-88701012.

Abbreviations: aa = amino acids; AcPV-1 = *Arma chinensis* picorna-like virus 1; ML = maximum-likelihood; NJ = neighbor-joining; nt = nucleotide; ORFs = open reading frames; RdRp = RNA-dependent RNA polymerase

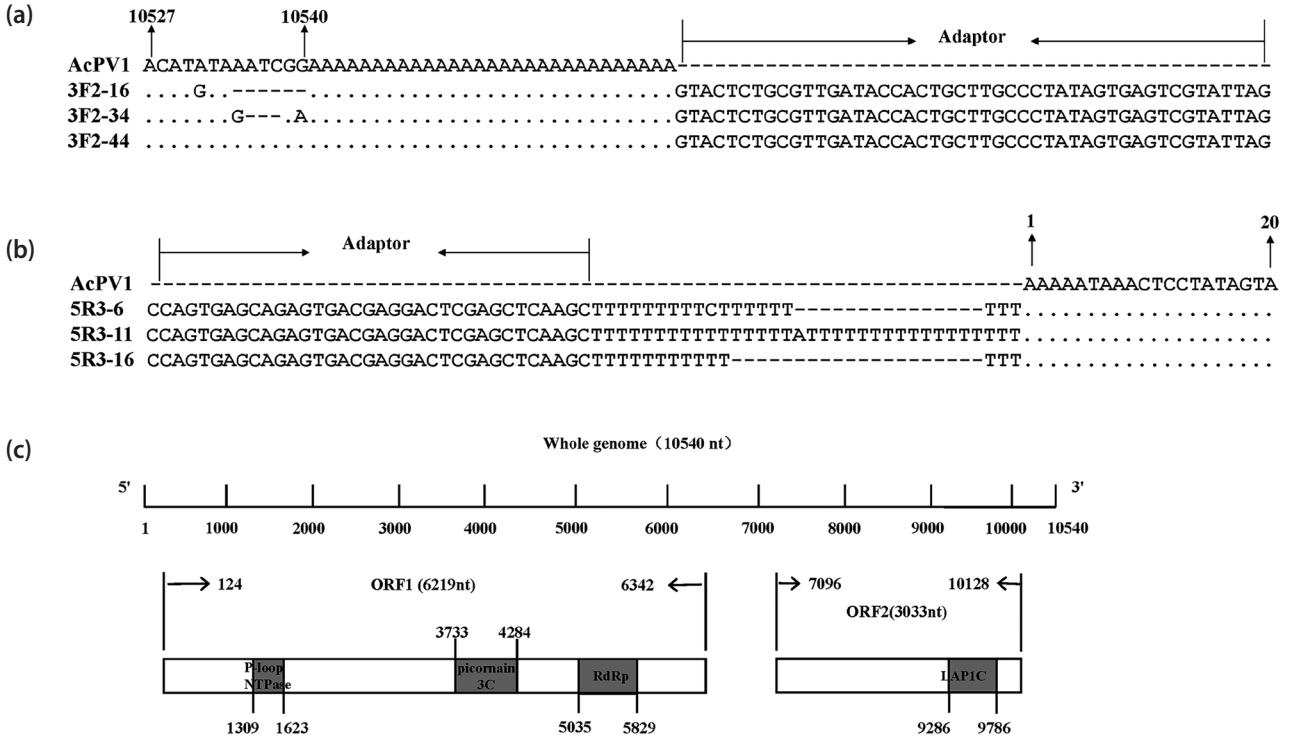
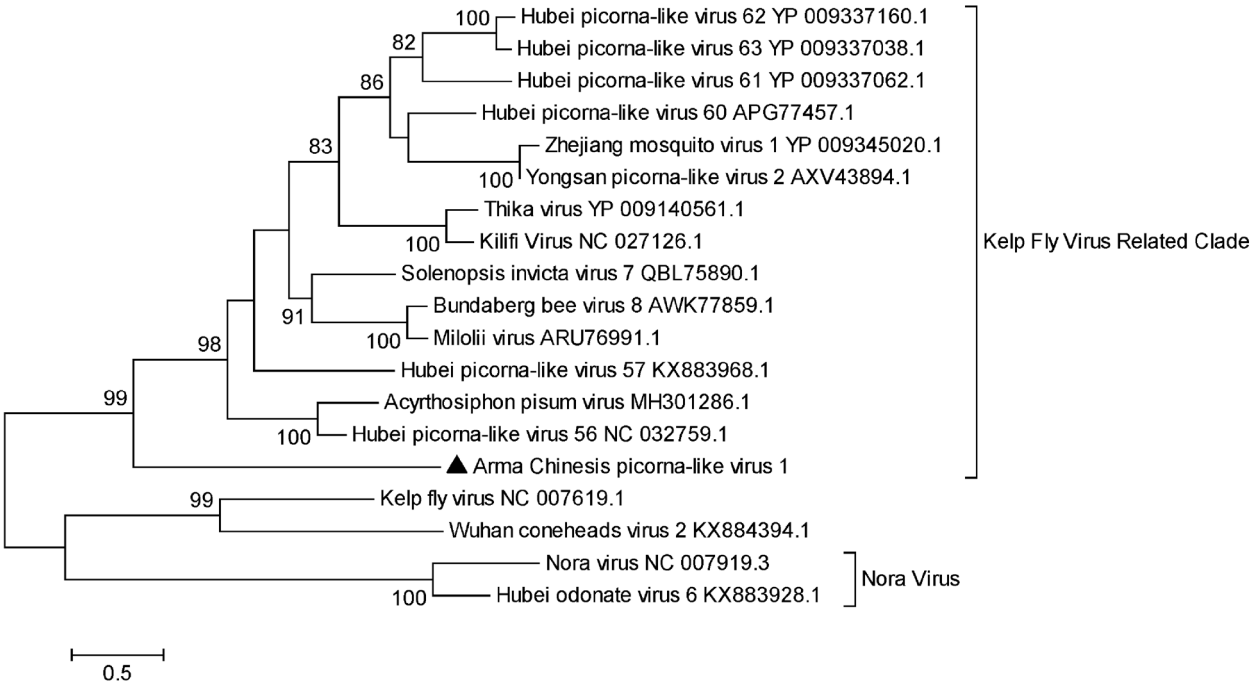


Fig. 1

Amplifcation of the 3'/5' ends and the genome structure of *Arma Chinensis* picorna-like virus 1 (AcPV-1)

(a) Amplifcation of 3' end of AcPV-1 genome. **(b)** Amplifcation of 5' end of AcPV-1 genome. AcPV-1 indicates the genome sequence. The adaptor sequences are shown and the numbers on the sequence indicate the nucleotide location based on the genomic sequences. "-" = absence of nucleotides. "." = identical nucleotides. Clone names shown on the left indicate their amplifcation primers and clone numbers (e.g., clone 3F2-16 stands for one clone amplifed by primer 3F2 and universal primer). **(c)** Genome structure of AcPV-1. The line represents the entire genome, and the two ORFs are shown below. The conserved domains are indicated by the shaded regions: P-loop NTPase = P-loop NTPase super family domain, picornain 3C = 3C cysteine protease, LAP1C = Lamina-associated polypeptide 1C, RdRP = RNA dependent RNA polymerase. The start and stop locations of the ORFs and the conserved domains are shown.

analysis on the samples was performed using an Illumina HiSeq™ instrument (Illumina, San Diego, USA). First, the samples were used for total RNA extraction, mRNA isolation, and fragmentation as described previously (Xu *et al*., 2017). Subsequently, sequencing was performed using the paired-end and 150-nt read length options with no strand specificity, and *de novo* assembly was performed using Trinity (v. 2.0.6) (Grabherr *et al*., 2011). Then, the assembled contigs with an average length of 850.75 bp were aligned to multiple databases such as NR, String, Swissprot, and KEGG using BLASTx (E-value ≤ 10-5) for functional annotation. The RNA-Seq data were submitted to the NCBI Sequence Read Archive (SRA) database (accession number: SRR10098905). Lastly, a total of 6 gigabases (Gb) of clean data were obtained by RNA-seq, producing a total of 27,058 contigs including 21,972 unigenes (Table S1). Using functional annotation with a nonredundant protein database, one assembled contig (10,467 nt in length) showed a 45% aa sequence similarity with the Milolii virus isolated from the ghost ants *Tapinoma melanocephalum* (GenBank Acc. No. ARU76991.1) (Laura *et al*., 2017). The cDNA templates were prepared as described previously and specific primers were designed to confirm the sequence of the contig (Table S2) by sequencing the PCR products (Xu *et al*., 2017). The PCR was performed as follows: 30 s at 94°C, 30 s at 60°C, and 90 s at 72°C for 35 cycles. The 3' and 5' ends were amplified by using primers listed in Table S2 according to the method described by Wang *et al*. (2005). The PCR products were cloned into the pEASY-T cloning vector and then sequenced (TransGen, Beijing, China). The ORFs within the virus genomic sequence were determined using DNAMAN software version 6 (Lynnon Corporation, Quebec, Canada). Conserved domains were predicted using the NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Percentage identities and alignments of the nt and aa sequences were calculated using the CLUSTAL W software (Thompson *et al*., 1994). Maximum-likelihood (ML) and neighbor-joining



Maximum-likelihood (ML) tree constructed using the amino acid sequences of the conserved RdRp domain from AcPV-1 and 18 other viruses

"▲" indicates the AcPV-1 reported in this study. Bootstrap values (1000 pseudoreplicates) > 50% are indicated on the nodes.

(NJ) trees with 1000 bootstrap replicates were constructed using MEGA 7.0 software packages with conserved aa sequences (Hartley *et al*., 2016).

We successfully amplified the 3' and 5' ends of the AcPV-1 genome (Fig. 1a,b). The whole genome of AcPV-1 was 10,540 nt in length excluding the poly(A) tail, had the nucleotide base composition of 28.6% A, 29.3% T, 22.5% C, and 19.6% G, and contained 123 nt 5' and 414 nt 3' untranslated regions (Fig. 1c). Analysis of AcPV-1 using the DNAMAN software showed that the AcPV-1 genome contained two deduced ORFs on the same strand (Fig. 1c). ORF1 (nt 124–6342) encoded a polypeptide of 2072 aa that had a predicted molecular mass of 232.71 kDa, a theoretical isoelectric point (pI) of 7.26, and contained three conserved domains, including a P-loop_NTPase superfamily domain (aa 396–500), a Peptidase_C3 superfamily domain (aa 1204–1387), and a RNA-dependent RNA polymerase (RdRp) domain (aa 1638–1902). ORF2 (nt 7096–10128) encoded a polypeptide with 1010 aa that possessed a predicted molecular mass of 111.44 kDa, a pI of 5.79, and contained one conserved domain of the LAP1C superfamily domain (aa 731–897). All the conserved domains are shown in Fig. 1c according to the nucleotide location.

Blast searches using the whole genome nucleotide and deduced aa sequences of RdRp indicated that AcPV-1 exhibited low identity with the known viruses, of which Hubei picorna-like virus 60 exhibited the highest similarity (34% in RdRp). However, Hubei picorna-like virus 60 has only a single large ORF (Shi *et al*., 2016). Thereafter, we focused on the phylogenetic analysis using the virus whole genome nucleotide sequences as well as deduced aa sequences of RdRp of AcPV-1 and 18 other virus species. Model tests were performed using the ML tree with a LG+Gamma distributed with Invariant sites (G+I) model, which possessed the lowest Bayesian information criterion scores. The ML tree based on RdRp showed that AcPV-1 clustered within Kelp Fly Virus Related Clade (Fig. 2). Moreover, the NJ tree with the whole genome nucleotide sequences showed similar topological structures as the ML trees (Fig. S1). All the ML and NJ trees revealed that AcPV-1 clustered within the Kelp Fly Virus Related Clade. These results suggested that AcPV-1 could be considered a new prototype for a *Arma chinensis* picorna-like virus taxon.

The prevalence of microbes in their hosts can reveal potential interactions between them and their hosts. Pathogenic viruses always have a low field infection rate, e.g., *Helicoverpa armigera* nucleopolyhedrovirus. However, mutualistic viruses also have a high field infection rate such as *Helicoverpa armigera* densovirus (Xu *et al*.

2014). We designed specific primers, ZCF4/ZCR4 (Table S2), to amplify a 737-bp fragment based on the genomic sequence of AcPV-1 to determine the infection rate. We detected 146 individuals collected from Zunyi (27°45'N, 106°53'E) and found the *Arma chinensis* infection rate of 5.48% (8 positives in 146 samples). Determination of the virulence of AcPV-1 will require further investigation by bioassays. We investigated the genetic diversity of AcPV-1 in the 8 positive samples using the fragments amplified with the primers ZCF4/ZCR4 (Supplementary material 3). The identities among these fragments were more than 99.5% and there were only 2 variable sites, suggesting a low level variability in the viruses from different host individuals.

Nucleotide sequence Acc. No. The complete genome sequences of AcPV-1 was submitted to the GenBank and the Acc. No. is MN966682.

Acknowledgment. This research was supported by the Agricultural Science and Technology Innovation Program (Grant No. ASTIP-TRIC04), the Key Laboratory Project of China Tobacco Corporation (Grant No. 110201603011), and the Major Project of Guizhou Branch of China Tobacco Corporation (Grant No. 201938).

Supplementary information is available in the online version of the paper.

References

Cholleti H, Hayer J, Fafetine J, Berg M, Anne-Lie Blomström (2018): Genetic characterization of a novel picorna-like virus in culex spp. mosquitoes from mozambique. Virol. J. 15(1), 71. https://doi.org/10.1186/s12985-018-0981-z

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen 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. 9, 644–652. https://doi.org/10.1038/nbt.1883

Hartley C J, Greenwood D R, Gilbert R J, Masoumi A, Gordon K H, Hanzlik T N, Fry E E, Stuart D I, Kumar S, Stecher G, Tamura K (2016): MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Laura EB, Gideon JM, Purnima P, Stephen JM (2017): Novel RNA Virus Genome Discovered in Ghost Ants (Tapinoma melanocephalum) from Hawaii. Genome Announc. 5, e00669-17. https://doi.org/10.1128/genomeA.00669-17

Pan M, Zhang H, Zhang L, Chen H (2019): Effects of Starvation and Prey Availability on Predation and Dispersal of an Omnivorous Predator Arma chinensis Fallou. J. Insect Behav. 32, 134–144. https://doi.org/10.1007/s10905-019-09718-9

Shi M, Lin XD, Chen X, Tian JH, Zhang YZ et al. (2018): The evolutionary history of vertebrate RNA viruses. Nature 556, 197–202. https://doi.org/10.1038/s41586-018-0012-7

Shi M, Lin XD, Tian J, Chen LJ, Chen X et al. (2016): Redefining the invertebrate RNA virosphere. Nature 540, 539–543. https://doi.org/10.1038/nature20167

Thompson JD, Higgins DG, Gibson TJ (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Valles SM, Chen Y, Firth AE, Guerin DMA, Hashimoto Y et al (2017): ICTV virus taxonomy profile: Iflaviridae. J. Gen. Virol. 98, 527–528. https://doi.org/10.1099/jgv.0.000757

Wang J, Zhang J, Jiang H, Liu C, Yi F, Hu Y (2005): Nucleotide sequence and genomic organization of a newly isolated densovirus infecting Dendrolimus punctatus. J. Gen. Virol. 86, 2169–2173. https://doi.org/10.1099/vir.0.80898-0

Xu P, Liu Y, Graham RI, Wilson K, Wu K (2014): Densovirus is a mutualistic symbiont of a gGlobal crop pest (Helicoverpa armigera) and protects against a baculovirus and Bt Biopesticide. PLoS Pathog. 10, e1004490.10. https://doi.org/10.1371/journal.ppat.1004490

Xu P, Song X, Yang X, Tang Z, Ren G, Lu Y (2017): A novel single-stranded RNA virus in Nesidiocoris tenuis. Arch. Virol. 162, 1125–1128. https://doi.org/10.1007/s00705-016-3195-y

Yang X, Xu P, Yuan H, Graham RI, Wilson K, Wu K (2019): Discovery and characterization of a novel picorna-like RNA virus in the cotton bollworm Helicoverpa armigera. J. Invertebr. Pathol. 160, 1–7. https://doi.org/10.1016/j.jip.2018.11.003

Zou DY, Coudron TA, Wu HH, Gu XS, Xu WH, Zhang LS, Chen HY (2015): Performance and Cost Comparisons for Continuous Rearing of Arma chinensis (Hemiptera: Pentatomidae: Asopinae) on a Zoophytogenous Artificial Diet and a Secondary Prey. J. Econ. Entomol. 108, 454–461. https://doi.org/10.1093/jee/tov024

Zou DY, Coudron TA, Zhang LS, Gu XS, Hu WH, Liu XL, Wu HH (2019): Performance of Arma chinensis reared on an artificial diet formulated using transcriptomic methods. Bull. Entomol. Res. 109, 1–10. https://doi.org/10.1017/S0007485318000111