Complete genome analysis of a novel chuvirus from a southern green stink bug, *Nezara viridula*

Peng-Peng Ren
Ningbo University
Zhuang-Xin Ye
Ningbo University
Sai-Nan Wang
Ningbo University
Jian-Ping Chen
Ningbo University
Jun-Min Li
Ningbo University
Chuan-Xi Zhang
Ningbo University
Jia-Bao Lu (lujiabao@nbu.edu.cn)
Ningbo University https://orcid.org/0000-0002-8257-3384

Research Article

Keywords: Nbu southern green stink bug chuvirus-1, Chuviridae, Virus discovery, Nezara viridula

Posted Date: May 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1557313/v1

Abstract

A novel chuvirus from the southern green stink bug (*Nezara viridula*) was identified by RNA sequencing in this study, which was named "Nbu southern green stink bug chuvirus-1" (NbuSGSBV-1). The complete genome sequence of NbuSGSBV-1 consisted of 11,375 nucleotides, which was confirmed to be a circular form by 'around-the-genome' reverse transcription polymerase chain reaction (RT-PCR) and sanger sequencing. Furthermore, three open reading frames (ORFs) were predicted in the NbuSGSBV-1 genome, including a large polymerase protein (L protein), a glycoprotein (G protein) and a nucleocapsid protein (N protein). Thereafter, a phylogenetic tree was constructed based on the RNA-dependent RNA polymerase amino acid sequences of all the currently available viruses in the family *Chuviridae*. As a result, NbuSGSBV-1 was clustered together with Sanya chuvirus 2 and Hubei odonate virus 11, indicating that NbuSGSBV-1 might belong to the genus *Odonatavirus*. Meanwhile, motif prediction results revealed five conserved sites among the L proteins of NbuSGSBV-1 and its homologous chiviruses. Moreover, the high abundance and typical characteristics of the NbuSGSBV-1-derived small interfering RNAs suggested the active replication of NbuSGSBV-1 in the host insect. To the best of our knowledge, this is the first report of a chuvirus identified from the insect family Pentatomidae. The discovery and characterization of NbuSGSBV-1 will help to understand the diversity of chuviruses in insects.

Full Text

With the advancement of next-generation sequencing (NGS) technology and metagenomic analysis in recent years, a growing number of novel viruses have been discovered and identified [1, 2]. Most of them are discovered from arthropods such as insects, arachnids, and chilopoda [2, 3]. As a result, arthropods are assumed as the primary store of viral genetic variety, which may play an important role in viral evolution [2, 4]. As the number of novel viruses increases dramatically, new families of viruses are gradually divided. For instance, the family *Chuviridae* was defined in 2015, which belongs to the order Jingchuvirales, class Monjiviricetes [2]. Most of the viruses in the family *Chuviridae* were discovered in the ancient Chinese region called Chu, which referred to the middle and lower reaches of the Yangtze River at that time [2, 5]. The phylogenetic diversity of chuviruses is between those of segmented and non-segmented viruses, and their genome structures are also diverse, including unsegmented, bi-segmented, and a circular form [2, 4, 5]. Typically, the circular structure of *Chuviridae* is distinct from the pseudo-circular structure of some other negative-sense RNA viruses like the families *Buniaviridae* and *Orthomyxoviridae*. The genome sequences of linear chuviruses are composed of a glycoprotein (G), a nucleoprotein (N) and a polymere (L) genes, while those of circular chuviruses are generally considered in the order L-(G)- N (displayed in a linear form), such as Bole Tick Virus 3 (GenBank NO. NC_028259.1). It is also reported that G gene is probably lost during the long-term evolution of some chuviruses [2, 4, 6].

The southern green stink bug (*Nezara viridula*), belonging to family Pentatomidae, order Hemiptera, is a highly polyphagous cosmopolitan pest. *N. viridula* feeds on a variety of important economic crops, and is widely distributed in Americas, Asia, Australia and Europe [7, 8]. The damage of *N. viridula* is mainly induced by its piercing-sucking mouthparts, resulting in plant damage, reduced seed germination and

survival, as well as the spread of plant pathogens [9, 10]. Previous studies have revealed several viruses in *N. viridula*. In 1992, two pathogenic viruses, namely, Nezara viridula virus-1 and Nezara viridula virus-2, were isolated and identified in the *N. viridula* [11]. A honeybee virus called Israel acute paralysis virus (IAPV) has recently been discovered in the *N. viridula*, indicating that some viruses can spread among species [12]. In this study, a novel chuvirus in *N. viridula* was identified by RNA sequencing (RNA-seq), which represented the first chuvirus identified in the family Pentatomidae.

In August 2019, a single adult stink bug was collected from a rice field in Hangzhou, Zhejiang, China. The total RNA of the *N.viridula* was later extracted using TRIzol Reagent (Invitrogen, MA, USA) according to the manufacturer's instructions. The quality of the extracted RNA was then evaluated using a Nano Drop spectrophotometer (Termo Scientifc, MA, USA). Afterwards, paired-end (150 bp) sequencing of the RNA library was performed using the Illumina HiSeq 4000 sequencer (Novogene, Tianjin, China). After trimming adaptor sequences [13], a total of 45,044,688 clean reads were obtained. First of all, the assembled contigs were compared with Barcode of Life Data Systems (https://www.boldsystems.org/), and one contig representing the potential cytochrome oxidase subunit 1 (COI) sequence of the stink bug was extracted. Thereafter, the extracted contig was further searched on the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/). According to our results, it was almost identical (99.70% similarity) to the deposited COI sequence of *N. viridula*, confirming that the stink bug used in this study is *N. viridula*. The identified contig (COI sequence of *N. viridula*) was further confirmed by sanger sequencing before it was submitted to the GenBank (Accession number: ON171205, File S1).

Then, the assembled contigs were searched against with the local virus database downloaded from the NCBI viral reference database (https://www.ncbi.nlm.nih.gov/genome/viruses). As the result, a chu-like viral sequence of about 11,500 nt, which represented almost the complete virus genome, was identified. To avoid false-positives and identify homology viruses, the obtained viral sequence was compared with the NCBI nucleotide (NT) and non-redundant (NR) protein databases (Table S1). At the same time, both Bowtie2 and Samtools were used to map the adaptor- and quality-trimmed reads of the transcriptome back to the obtained viral contig. Based on the results, a high coverage (≈ 14,000X coverage) was approved for this chu-like viral contig. To confirm the presence and validity of this contig, the viral contig was divided into six regions and specific primers were designed accordingly (Table S2). Then RT-PCR and sanger sequencing were performed for each of the region. As a result, the sequences of all the six regions were perfectly matched with the original contig retrieved from transcriptome. Moreover, ‘around-the-genome RT-PCR’ was smoothly performed from the 3' end to the 5' end to verify whether this virus was circular. Results of sanger sequencing clearly confirmed that this chu-like viral contig is in a circular form (Figure S1).

The obtained chu-like virus with verified full genome sequence of 11,375 nucleotides was submitted to the NCBI GenBank database (accession number: ON191814, File S2), and named “Nbu Southern green stink bug chuvirus – 1” (NbuSGSBV-1). Using the NCBI Open Reading Frame Finder (ORF Finder, https://www.ncbi.nlm.nih.gov/orffinder/), three non-overlapping ORFs (ORF1: 48-6578 nt; ORF2: 6619–

8628 nt; ORF3: 8698–10179 nt) were found (Fig. 1A, B). Typically, ORF1-3 were predicted to encode a 261.12-kDa large polymerase protein (L protein), a 80.28-kDa glycoprotein (G protein) and a 59.16-kDa nucleoprotein (N protein), respectively, which is consistent with the typical structure of a previously reported circular chuvirus (L-(G)-N) [2]. To identify the homologous chiviruses of NbuSGSBV-1, BlastP was applied to predict protein sequence against the NCBI reference viral sequence database. The results indicated that NbuSGSBV-1 L protein and G protein sequences exhibited the highest homology to Chuviridae sp. (GenBank NO. QXV86379.1), with the amino acid homologies of 41.58% and 48.91%, respectively. And the N protein sequence displayed highest homology with Changping Tick Virus 2 (GenBank NO. YP_009177706.1), with the amino acid homology of 27.69% (Table S1). Thereafter, InterProScan (https://www.ebi.ac.uk/interpro) was utilized to predict the conserved domains of the three ORFs. Consequently, three conserved domains in L protein were successfully predicted, including a Mononeg_mRNAcap, a Mononeg_RNA_pol_cat and a Mononega_L_MeTrfase, whereas no conserved structures were predicted in G or N protein (Fig. 1A). To further understand the abundance and coverage of the NbuSGSBV-1-derived sequenced reads, the RNA-seq reads were realigned back to the confirmed full genome of NbuSGSBV-1. Noteworthily, viral reads were apparently accumulated in the 3' region of the genome, especially in ORF3 (N) (Fig. 1A).

Moreover, to investigate the taxonomic status of NbuSGSBV-1, the RNA-dependent RNA polymerase (RdRp) sequences of all the available viruses in the family *Chuviridae* were retrieved from NCBI to generate a phylogenetic tree, and Atrato Chu-like virus 5 (GenBank NO. QHA33675.1) in the family *Aliusviridae* was used as the outgroup. The substitution model was firstly evaluated by ModelTest-NG, and then a maximum likelihood tree was constructed using RAxMLNG (version 0.9.0) with 1000 bootstrap replicates[14–16]. As indicated in Fig. 2A, NbuSGSBV-1 was clustered with the Sanya chuvirus 2 (GenBank NO. UHK03098.1) in the tree, as evidenced by a high bootstrap value (100). In addition, another virus on the same branch of NbuSGSBV-1 was Hubei odonate virus 11 (GenBank YP 009336946.1), which belongs to the genus *Odonatavirus* and has the same structure with NbuSGSBV-1. As indicated by the International Committee on Taxonomy of Viruses (ICTV, https://talk.ictvonline.org/), only three viruses have been found currently in the genus *Odonatavirus*. Therefore, it is speculated that NbuSGSBV-1 may be a new member in the genus *Odonatavirus*, family *Chuviridae*. Although NbuSGSBV-1 has a circular form of genome, the genomes of Sanya chuvirus 2 and Hubei odonate virus 11 are both linear, which may imply the unique evolutionary characteristics of chuvirus genomes. In addition, MEME (https://meme-suite.org/meme/tools/meme) was adopted to identify the conserved motifs in L genes of NbuSGSBV-1 with default parameters, together with another five closely related chiviruses (clustered into the same branch of the tree). As a result, totally five conserved motifs were identified in the six chuviruses (Fig.S2). Such conserved regions identified in these chiviruses may provide important information for the classification of novel chuviruses in this clade in the future. In conclusion, the above results indicate that the newly discovered NbuSGSBV-1 may belong to the genus *Odonatavirus* in the family *Chuviridae.*

To better understand the small interfering RNA (siRNA)-based host antiviral immunity in response to NbuSGSBV-1 infection, small RNAs (sRNAs) from *N. viridula* were sequenced and comprehensively characterized. Firstly, a library of sRNA from *N. viridula* was prepared using the Illumina TruSeq sRNA

Sample Preparation Kit (Illumina, USA) and later sequenced with the Illumina HiSeq 2500 platform. Secondly, the sRNAs were processed to obtain clean reads, and sRNAs with a length of 18–30 nt were extracted. Finally, Bowtie was employed to map the processed sRNA reads back to the whole viral genome sequence of NbuSGSBV-1 (allowing zero mismatch) for the identification of vsiRNAs. Thereafter, the obtained vsiRNAs were further analyzed using custom perl scripts and Linux bash scripts [17]. According to our results, a total of 12,907 vsiRNA reads (including 3,187 unique ones) were perfectly mapped to the assembled genome of NbuSGSBV-1. The length of vsiRNAs was mainly concentrated in 22 nt, accounting for 31.9% and 23.7% of the total and unique vsiRNAs, respectively. Moreover, they were almost equally derived from the sense and antisense strands of the viral genome (Fig. 2B-a,b). In addition, vsiRNAs displayed a strong A/U preference in their 5' terminal nucleotides and were uniformly distributed throughout the viral genome, as observed from Fig. 2B-C. The typical characteristics of vsiRNAs suggested that the host antiviral RNAi pathway was actively involved in the response to NbuSGSBV-1 infection.

In conclusion, a novel circular chuvirus, NbuSGSBV-1, is discovered and identified in *N.viridula*. This is the first report of a chuvirus in the insect family Pentatomidae. The discovery and identification of NbuSGSBV-1 will contribute to a better understanding of the diversity of chuviruses in insects.

Declarations

Acknowledgements

This work was supported by the National Natural Science Foundation of China (U20A2036) and Ningbo Science and Technology Innovation 2025 Major Project (2019B10004).

Compliance with ethical standards

Conflict of interest

The authors declare no conflict of interest.

Ethical statement

No experimental work with humans was done in this study.

References

1. Edgar RC, Taylor J, Lin V, Altman T, Barbera P, Meleshko D, Lohr D, Novakovsky G, Buchfink B, Al-Shayeb B, Banfield JF, de la Pena M, Korobeynikov A, Chikhi R, Babaian A (2022) Petabase-scale sequence alignment catalyses viral discovery. Nature 602:142–147
2. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ (2015) Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 4

3. Niu J, Li XL, Wu YL, Sun QZ, Zhang W, Cao M, Wang JJ (2020) RNA virome screening in diverse but ecologically related citrus pests reveals potential virus-host interactions. J Invertebr Pathol 170:107329
4. Qin XC, Shi M, Tian JH, Lin XD, Gao DY, He JR, Wang JB, Li CX, Kang YJ, Yu B, Zhou DJ, Xu J, Plyusnin A, Holmes EC, Zhang YZ (2014) A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. Proc Natl Acad Sci U S A 111:6744–6749
5. 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
6. Cook S, Chung BY, Bass D, Moureau G, Tang S, McAlister E, Culverwell CL, Glucksman E, Wang H, Brown TD, Gould EA, Harbach RE, de Lamballerie X, Firth AE (2013) Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. PLoS ONE 8:e80720
7. Todd JJAroe (1989) Ecology and behavior of *Nezara viridula*. Annu Rev Entomol 34:273–292
8. Panizzi A, Lucini T (2016) What happened to *Nezara viridula* (L.) in the Americas? Possible reasons to explain populations decline. Neotrop Entomol 45:619–628
9. Miller L, Rose H, McDonald F (1978) The effects of damage by the green vegetable bug, Nezara viridula (L.) on yield and quality of soybeans. Australian J Entomol 16:421–426
10. Leskey TC, Hamilton GC, Nielsen AL, Polk DF, Rodriguez-Saona C, Bergh JC, Herbert DA, Kuhar TP, Pfeiffer D, Dively Gp (2012) Pest status of the brown marmorated stink bug, Halyomorpha halys in the USA. Outlooks Pest Manage 23:218–226
11. Williamson C, Von Wechmar M (1992) Two novel viruses associated with severe disease symptoms of the green stinkbug *Nezara viridula*. J Gen Virol 73:2467–2471
12. Susevich ML, Reynaldi FJ, Marti GA, Echeverría MG (2019) Primer hallazgo del Virus de la Parálisis Aguda Israelí (IAPV) en *Nezara viridula* (Hemiptera: Pentatomidae). Revista de la Sociedad Entomológica Argentina 78:1–5
13. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652
14. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274
15. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455
16. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2020) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol 37:291–294
17. Li J, Andika IB, Shen J, Lv Y, Ji Y, Sun L, Chen J (2013) Characterization of rice black-streaked dwarf virus-and rice stripe virus-derived siRNAs in singly and doubly infected insect vector *Laodelphax striatellus*. PLoS ONE 8:e66007

Figures

Figure 1

Characterization of Nbu southern green stink bug chuvirus 1 (NbuSGSBV-1). (A) Genome structure and transcripts coverage of NbuSGSBV-1. (B) Schematic diagram of the circular genome structure of NbuSGSBV-1. RNA-dependent RNA polymerase, glycoprotein and nucleoprotein genes are shaded blue, orange and gray, respectively. L, G and N represent large polymerase protein, glycoprotein and nucleoprotein, separately.

Figure 2

Phylogenetic tree and virus derived small interfering RNAs (vsiRNAs) analysis of NbuSGSBV-1. (A) Maximum-likelihood phylogenetic tree based on the amino acid sequences of the conserved RdRp domain of NbuSGSBV-1 and the other chuviruses. Bootstrap values > 50 are shown in each node of the tree. Scale bars represent percent divergence. NbuSGSBV-1 is indicated with a red star. (B) The size distribution (a, b) and 5' terminal nucleotide percentage (c) of NbuSGSBV-1 siRNAs. (C) Distribution of NbuSGSBV-1 siRNAs alongside the viral genome. Red and black represent siRNAs derived from the sense (plus strand) and antisense (minus strand) genomic strands of NbuSGSBV-1, respectively.

Supplementary Files

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

- fig.s1.tif
- fig.s2.tif
- Supp.file.s1.COIsequenceoftheNezaraviridula.fasta
- Supp.file.s2.FullgenomesequenceofNbuGSBV1.fasta
- SupplementaryTableS1.BLASTresultsofNbuGSBV1.docx
- SupplementaryTableS2.Primersusedinthisstudy.docx