

RESEARCH ARTICLE

Open Access



Full-length genome sequence of segmented RNA virus from ticks was obtained using small RNA sequencing data

Xiaofeng Xu^{1†}, Jinlong Bei^{2,3†}, Yibo Xuan¹, Jiayuan Chen⁴, Defu Chen⁴, Stephen C. Barker⁵, Samuel Kelava⁵, Xiaoi Zhang^{2,3}, Shan Gao^{4*}  and Ze Chen^{1*}

Abstract

Background: In 2014, a novel tick-borne virus of the *Flaviviridae* family was first reported in the Mogiana region of Brazil and named the Mogiana tick virus (MGTV). Thereafter, the Jingmen tick virus (JMTV), Kindia tick virus (KITV), and Guangxi tick virus (GXTV)—evolutionarily related to MGTV—were reported.

Results: In the present study, we used small RNA sequencing (sRNA-seq) to detect viruses in ticks and discovered a new MGTV strain in *Amblyomma testudinarium* ticks collected in China's Yunnan Province in 2016. We obtained the full-length genome sequence of this MGTV strain Yunnan2016 (GenBank: MT080097, MT080098, MT080099 and MT080100) and recommended it for its inclusion in the NCBI RefSeq database for future studies on MGTV, JMTV, KITV and GXTV. Phylogenetic analysis showed that MGTV, JMTV, KITV and GXTV are monophyletic and belong to a MGTV group. Furthermore, this MGTV group of viruses may be phylogenetically related to geographical regions that were formerly part of the supercontinents Gondwana and Laurasia.

Conclusions: To the best of our knowledge, this is the first study in which 5' and 3' sRNAs were used to generate full-length genome sequences of, but not limited to, RNA viruses. We also demonstrated the feasibility of using the sRNA-seq based method for the detection of viruses in pooled two and even possible one small ticks. MGTV may preserve the characteristic of ancient RNA viruses, which can be used to study the origin and evolution of RNA viruses. In addition, MGTV can be used as novel species for studies in phylogeography.

Keywords: MGTV, JMTV, Full-length genome, 5' sRNA, 3' sRNA

Background

Next generation sequencing (NGS) technologies have been widely applied to virus and viroid discovery in plants and animals. Compared to other NGS based methods, the small RNA sequencing (sRNA-seq) based

method simplifies virus detection and has several other advantages [1]. The sRNA-seq based method was originally used for viral detection and identification in plants [2] and in invertebrates [3]. Although the sRNA-seq based method does not perform as well in the detection of mammalian viruses as in the detection of plant or invertebrate viruses, we still detected eight mammalian viruses: human papillomavirus type 18 (HPV-18) [4], hepatitis B virus (HBV) [4], hepatitis C virus (HCV) [4], human immunodeficiency virus type 1 (HIV-1) [4], squirrel monkey retrovirus (SMRV) [4], Epstein-Barr virus (EBV) [4], severe acute respiratory syndrome

* Correspondence: gao_shan@mail.nankai.edu.cn; chenze@yeah.net

[†]Xiaofeng Xu and Jinlong Bei contributed equally to this work.

⁴College of Life Sciences, Nankai University, Tianjin, Tianjin 300071, People's Republic of China

¹Hebei Key Laboratory of Animal Physiology, Biochemistry and Molecular Biology, College of Life Sciences, Hebei Normal University, Shijiazhuang, Hebei 050024, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

coronavirus (SARS-CoV) [5] and a DNA segment of African swine fever virus (ASFV) [6]. The discovery of featured RNA fragments including small interfering RNA (siRNA) duplexes [7], 5' and 3' end sRNAs [8, 9] palindromic sRNAs (psRNAs) and complemented psRNAs (cpsRNAs) [5] increased our capacity to detect viruses in mammals. Moreover, we found that 5' and 3' sRNAs can be used to annotate nuclear non-coding and mitochondrial genes at 1 bp resolution [10, 11]. In the present study, we report that 5' and 3' sRNAs can be used to complete 5' and 3' ends of genome sequences of RNA viruses at 1-bp resolution and generate full-length genome sequences.

Ticks transmit a multitude of infectious agents to humans and other animal species, including viruses of the *Flaviviridae* family, which are among the most common tick-borne viruses [12]. With the widespread use of NGS, a number of studies have applied metagenomic methods to detect tick-associated pathogens [13]. However, metagenomic methods using DNA cannot be used to detect RNA viruses without DNA stages. Consequently, transcriptomic approaches using total RNA have also been used to detect tick viruses [14]. The sRNA-seq based method has been successfully used in the detection of *Rickettsia* in ticks [15]. To the best of our knowledge, there are no previous studies of virus detection in ticks using the sRNA-seq based method until the detection of the DNA segment of ASFV [6]. Although high-depth sRNA-seq data was used to detect a DNA segment of ASFV, the coverage of the ASFV reference genome was still very low. This suggested that the sRNA-seq based method does not perform as well in the detection of DNA viruses as it does in the detection of RNA viruses.

In a previous study [6], we used sRNA-seq to detect viruses in ticks. Subsequent analysis of these viruses led to the discovery of a new strain of RNA virus. In 2014, the virus was first reported as a novel tick-borne virus of the *Flaviviridae* family in the Mogiana Region of Brazil [12] and was named Mogiana tick virus (MGTV). Subsequently, Jingmen tick virus (JMTV), Kindia tick virus (KITV) and Guangxi tick virus (GXTV)—evolutionarily related to MGTV (Results and Discussion)—were detected in ticks. In 2018, viruses closely related to JMTV were detected in the sera samples of three Crimean-Congo hemorrhagic fever (CCHF) patients collected from 2013 and 2015 in Kosovo and two of these patients had exposed to tick bites [16]. In the present study, we identified a new MGTV strain Yunnan2016 detected in *Amblyomma testudinarium* ticks [17] and aimed to achieve the following research goals: (1) establish a method to generate the full-length genome sequence of an RNA virus using sRNA-seq data; (2) determine the feasibility of using the sRNA-seq based method in the

detection of viruses in a small tick; and (3) provide a high-quality and well-curated reference genome for future studies on MGTV, JMTV, KITV and GXTV.

Results and discussion

Detection of viruses in ticks using sRNA-seq data

Amblyomma testudinarium, *Dermacentor nuttalli*, *D. niveus* and *D. silvarum* ticks were collected for our previous studies [8, 18]. The sRNA-seq data from these ticks were generated and deposited in the NCBI SRA database under the accession numbers SRP084097 and SRP178347 (Table 1). Using VirusDetect (Refer to Methods), MGTV was detected in *A. testudinarium* ticks (SRA: SRP084097) but not in *D. nuttalli*, *D. niveus* or *D. silvarum* (SRA: SRP178347) ticks that were used as negative controls. Since the *A. testudinarium* ticks were collected in China's Yunnan Province in 2016, the new MGTV strain was named Yunnan2016. As a segmented RNA virus, MGTV is composed of four RNAs in its genome, two of which (RNA1 and RNA3) are related to the nonstructural protein genes of the genus *Flavivirus* (family *Flaviviridae*), while the other two segments (RNA2 and RNA4) are unique to MGTV. VirusDetect (Refer to Methods) uses the closest reference sequence to report the detected virus. Used as reference to report Yunnan2016, the genome of the JMTV strain Xinjiang2016 (GenBank: MK174251, MK174244, MK174230 and MK174237) was sequenced from wild rodents collected in China's Xinjiang Province. The sRNA-seq data from *A. testudinarium* (SRA: SRR4116826) covered 86.71% of the Xinjiang2016 genome with an average depth of 46.66 (Table 1). The sRNA-seq data from the *A. testudinarium* contained significantly more reads aligned to the Yunnan2016 genome (Fig. 1a) and the Xinjiang2016 genome (Fig. 1b) than the sRNA-seq data from the three other species (Fig. 1c). RNA1, RNA2, RNA3 and RNA4 of the MGTV strain Yunnan2016 were assembled at the contig level. Subsequently, PCR amplification coupled with Sanger sequencing (Additional file 1) was used to fill the gaps between contigs and confirm the genome assembly: 93.7% (2879/3073) of RNA1, 90.6% (2528/2790) of RNA2, 88.3% (2468/2795) of RNA3 and 95.2% (2619/2752) of RNA4 were confirmed by Sanger sequencing [the polyA tails of 3' untranslated regions (UTRs) were not part of these calculations].

Full-length genome sequence of the MGTV strain

Yunnan2016

We used 5' and 3' sRNAs (Fig. 2a) to generate the full-length genome sequence of the new MGTV strain Yunnan2016 at 1 bp resolution (Additional file 1). The 5' ends of all RNAs in the Yunnan2016 genome have the sequence motif AG [T]₂₋₃[A]₄₋₆[C/G]_nAAGTGC

Table 1 Genome coverage and average depth of the MGTV strain Yunnan2016

Data	Tick	Reference	Coverage	Depth
SRR4116826	<i>Amblyomma testudinarium</i>	Xinjiang2016	86.47%	46.52
		Yunnan2016	97.37%	91.11
SRR8439389	<i>Dermacentor nuttalli</i>	Yunnan2016	18.40%	2.58
SRR8439390				
SRR8432408	<i>D. niveus</i>	Yunnan2016	18.62%	2.91
SRR8432409		Yunnan2016	20.72%	2.94
SRR811197093	<i>D. silvarum</i>	Yunnan2016	24.38%	10.31
SRR811197094		Yunnan2016	8.11%	4.22

"Data" (the first column) was aligned to the reference genome (the 3rd column) to obtain the information on the fourth and fifth columns. "Data" was unique for each run of sRNA-seq data in the NCBI SRA database. The virus strain Yunnan2016 (GenBank: MT080097, MT080098, MT080099 and MT080100) was detected in the present study. The reference genome of the MGTV strain Xinjiang2016 (GenBank: MK174251, MK174244, MK174230 and MK174237) was used to report Yunnan2016. "Coverage" and "Depth" indicated the genome coverage and average depth, respectively (Refer to [Methods](#))

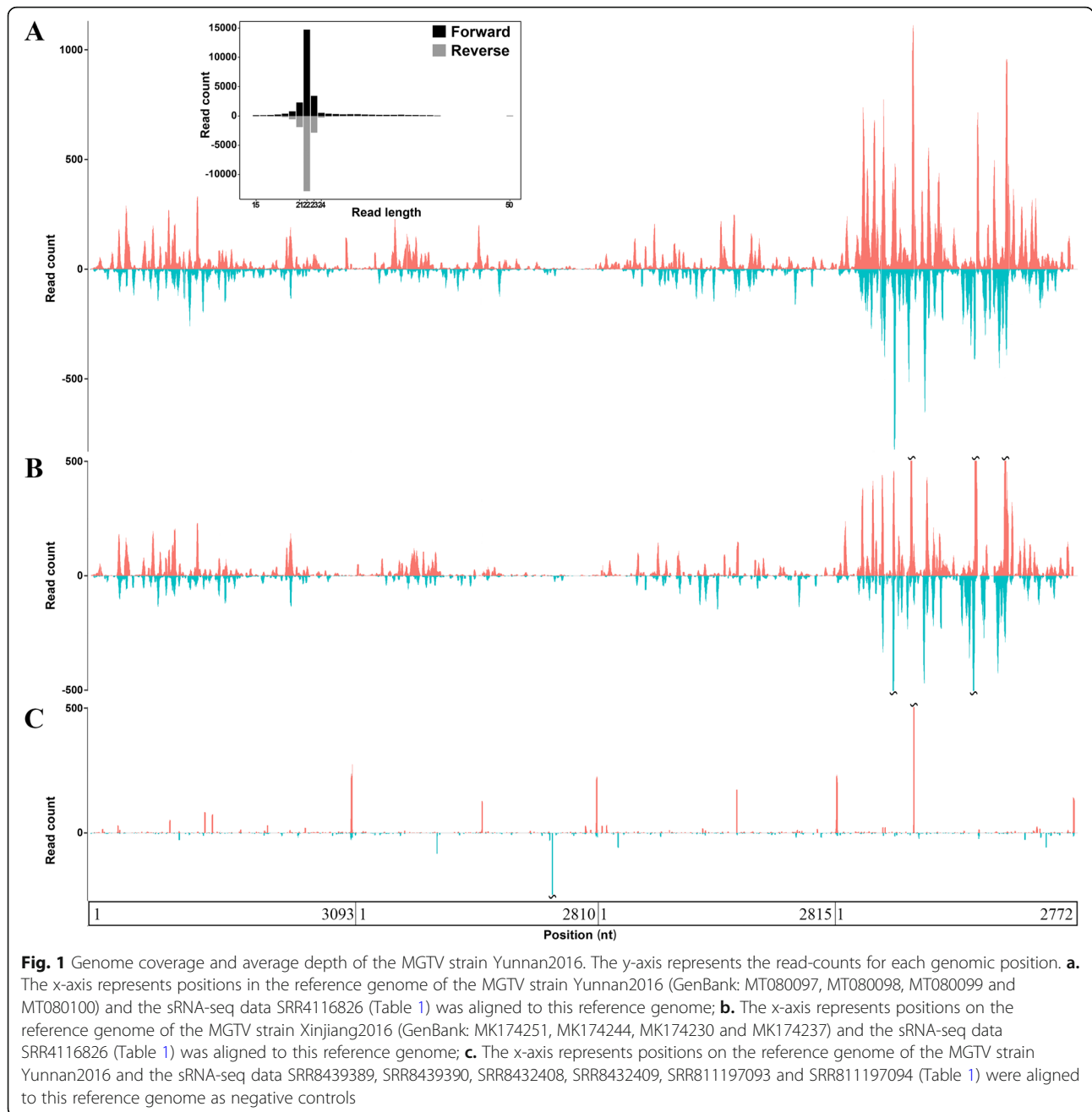
(Fig. 2b), where $[C/G]_n$ represents a GC-enriched region. The 3' ends of all RNAs in the Yunnan2016 genome have an AC-enriched region (Fig. 2b). RNA1, RNA2, RNA3 and RNA4 of Yunnan2016 with respective lengths of 3093, 2810, 2815 and 2772 nt were submitted to the NCBI GenBank database under the accession numbers MT080097, MT080098, MT080099 and MT080100, respectively. The length of the polyA tail in each 3'-UTR of these RNAs was set as 20 nt. The sRNA-seq data from *A. testudinarium* (SRA: SRR039620) covered 97.37% of the full-length genome sequence of Yunnan2016 with an average depth of 91.11 (Table 1); 58.5% (26,668/45,563) of the virus reads were aligned to RNA4 (Fig. 1b). Although MGTV is a positive-sense single-stranded RNA (+ssRNA) virus, the proportion of sRNA-seq reads aligned to the viral positive- and negative-strands was 1.42 (26,767/18,796).

Compared to the genome coverage of 86.47% and average depth of 46.52 (Refer to [Methods](#)) when we used the Xinjiang2016 genome as a reference (Fig. 1a), genome coverage increased to 97.37% and average depth to 91.11 when we used the Yunnan2016 genome as a reference (Fig. 1b). A high-quality virus genome should contain the full-length 5' UTRs, as these regions contain useful information for the analysis of such genomes. In a previous study, we analyzed 5' UTRs in Betacoronaviruses and developed 5'-UTR barcoding to be used in detection, identification, classification and phylogenetic analysis of, but not limited to, coronaviruses [19]. Comparing the full-length genome sequence of Yunnan2016 with those of 16 other MGTV, JMTV, KITV and GXTV complete genomes (Refer to [Methods](#)), we found that none of these other genomes had the correct full-length 5' and 3' ends. Particularly, RNA1 (GenBank: MN025516) and RNA4 (GenBank: MN025515) of the JMTV strain TT2017-2 had 56 and 48 nt additional sequences at their 5' ends, respectively. Further analysis showed the additional sequences were identical to internal regions of the genomes. Obviously, these

additional sequences had been assembled incorrectly in the previous studies. Therefore, this high-quality and full-length genome sequence of the MGTV strain Yunnan2016 should be included in the NCBI RefSeq database for future studies on MGTV, JMTV, KITV and GXTV.

Phylogenetic analysis of MGTV genomes

In total, 17 MGTV, JMTV, KITV and GXTV genomes were used for the further analysis (Refer to [Methods](#)). Five protein-coding genes were annotated for each of 17 genomes. The putative proteins encoded by RNA1, RNA2 and RNA3 are the RNA-dependent RNA polymerase, glycoprotein and protease, respectively, whereas the putative proteins encoded by RNA4 are the capsid protein and the membrane protein. The RNA-dependent RNA polymerase from RNA1, the protease from RNA3, and the capsid protein from RNA4 had lengths of 915, 809 and 255 aa (amino acid residues), respectively. In all 17 virus genomes, the lengths of these three proteins were constant, whereas those of the other two proteins (the glycoprotein from RNA2 and the membrane protein from RNA4) varied. The lengths of the glycoprotein from RNA2 varied because of two mutations (Fig. 2c): T/C mutations in the start codons shortened the coding sequences (CDSs) of RNA2 by 21 nt and a small insertion/deletion (InDels) shortened the CDSs by 3 nt. Theoretically, four types of glycoproteins with lengths of 745, 746, 754 or 755 aa would be translated from RNA2; however, a glycoprotein with 746 aa was not observed in the 17 virus genomes. Since the lengths of the membrane protein from RNA4 varied because of one mutation—T/C (Fig. 2d)—two types of membrane proteins with lengths of 522 or 539 aa can be translated from RNA4. The multiple-aligned RNA1, RNA3 and RNA4 had 2745, 2427 and 2351 nt CDSs, whereas the multiple-aligned RNA2 had a 2265 nt CDS with a 15 or 18 nt variable region removed (Fig. 2c). CDS 1, 2, 3 and 4 of RNA 1, 2, 3 and 4 could then be connected into a



combined 9788 nt CDS. Using paired Pearson correlations between the CDSs of 17 viruses, the degrees of evolutionary conservation are ranked as CDS 2, 1, 3 and 4 (Fig. 3a).

Five phylogenetic trees from the CDS 1, 2, 3 and 4, as well as the combined CDSs, were built using nine non-redundant genome sequences (Refer to Methods). Although CDS 1, 2, 3 and 4 exhibited substantial differences in their degrees of evolutionary conservation, the tree topologies from them remained congruent using the unweighted pair group method with arithmetic mean (UPGMA), maximum parsimony (MP) and neighbour

joining (NJ) methods (Fig. 3b-f). MGTV, JMTV, KITV and GXTV belong to an MGTV group with two major clades. The two branches of Clade I contain the virus strains isolated from Brazil (South America) and Guinea (West Africa), in addition to the virus strains Yunnan2016 and Xinjiang2016 (Fig. 3f). Clade II contains the virus strains isolated from Martinique of France (Central America), Trinidad and Tobago (Central America) and Kosovo (Central Europe). Phylogenetic analysis of these viruses in relation to their geographic information showed that Clade I and Clade II approximately correspond with

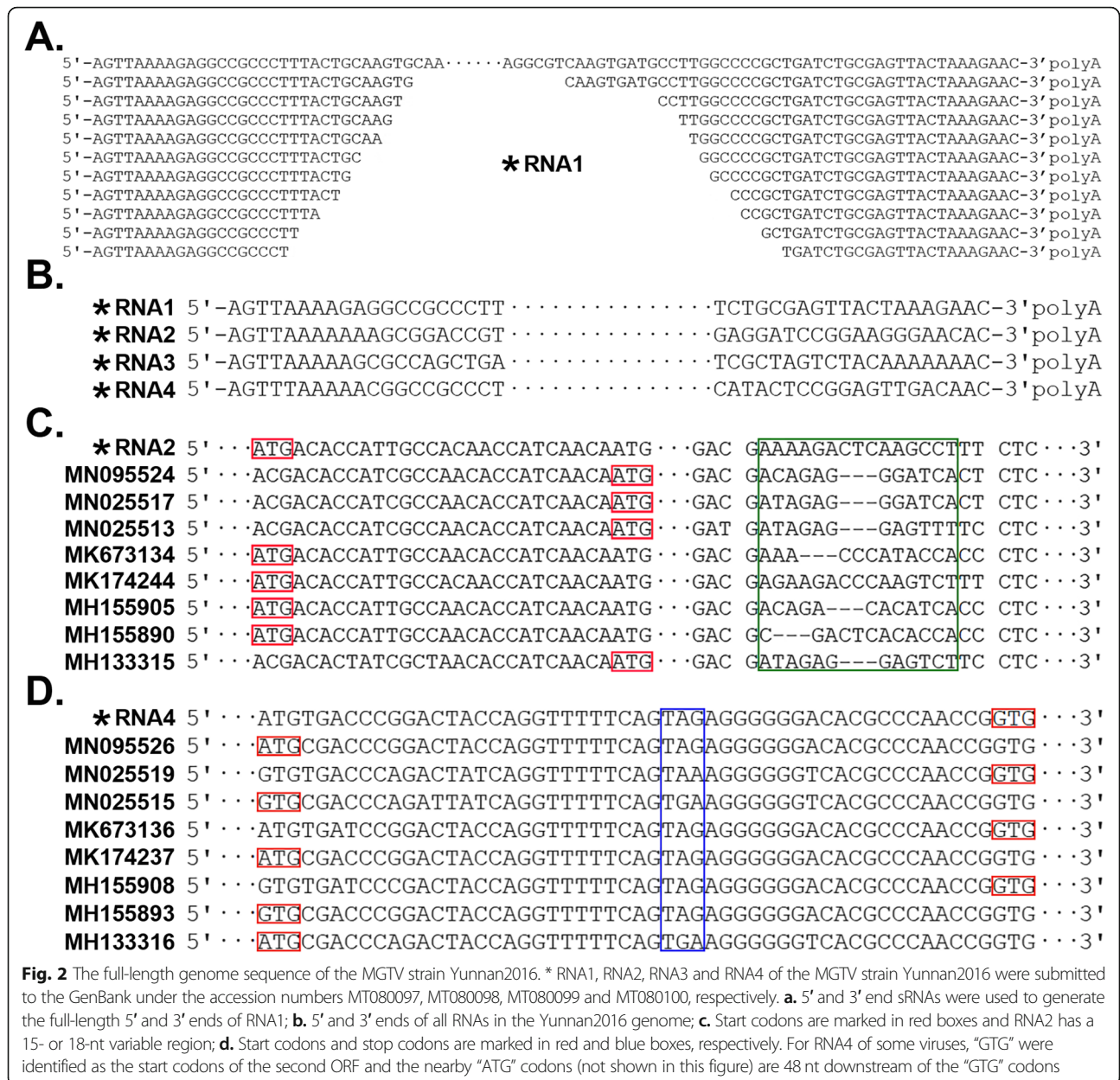
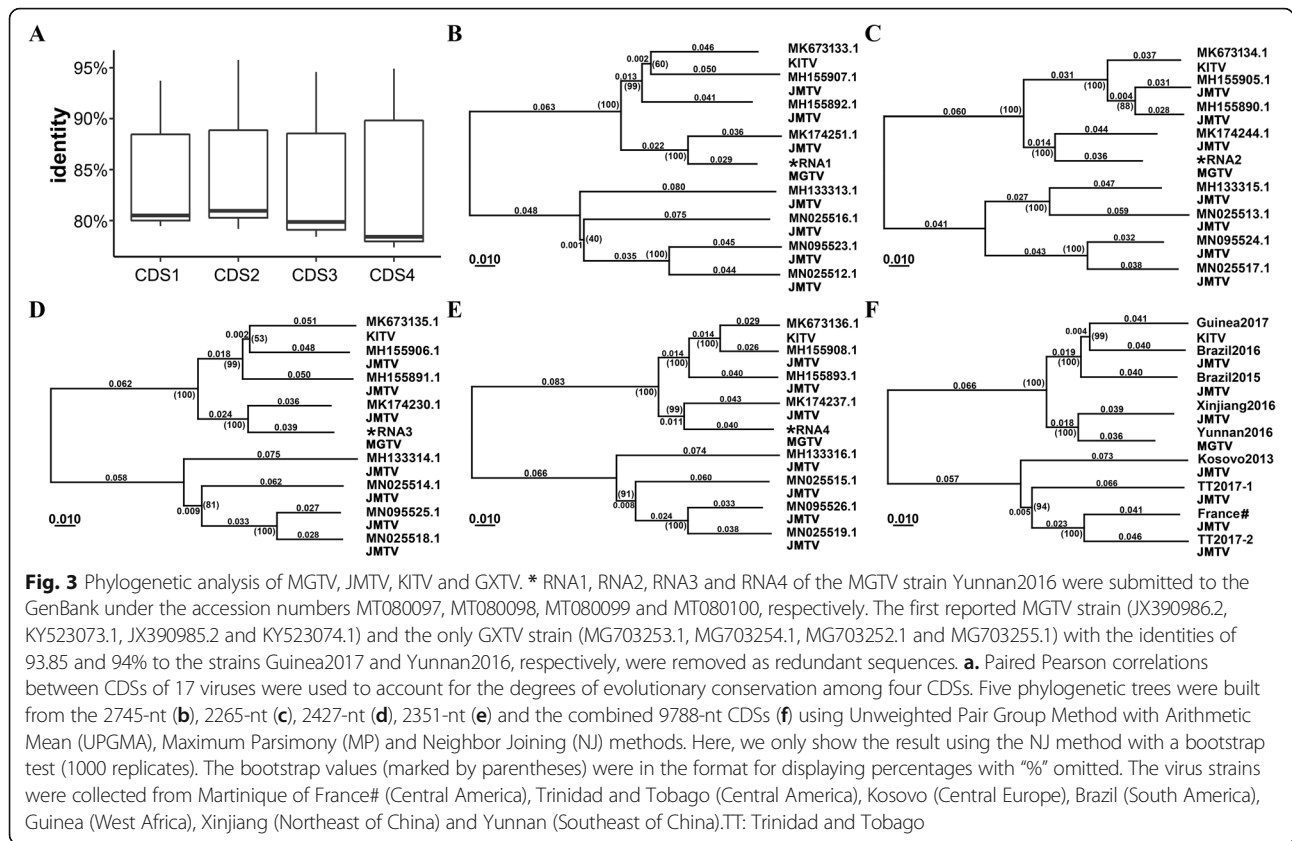


Fig. 2 The full-length genome sequence of the MGTV strain Yunnan2016. * RNA1, RNA2, RNA3 and RNA4 of the MGTV strain Yunnan2016 were submitted to the GenBank under the accession numbers MT080097, MT080098, MT080099 and MT080100, respectively. **a.** 5' and 3' end sRNAs were used to generate the full-length 5' and 3' ends of RNA1; **b.** 5' and 3' ends of all RNAs in the Yunnan2016 genome; **c.** Start codons are marked in red boxes and RNA2 has a 15- or 18-nt variable region; **d.** Start codons and stop codons are marked in red and blue boxes, respectively. For RNA4 of some viruses, "GTG" were identified as the start codons of the second ORF and the nearby "ATG" codons (not shown in this figure) are 48 nt downstream of the "GTG" codons

the supercontinents Gondwanaland and Laurasia, respectively. Brazil and Guinea were very close in Gondwanaland, while Martinique of France, Trinidad and Tobago (TT) and Kosovo were close in Laurasia. Based on Wegener's concept, Pangea is considered to have formed from the assembly of Earth's continents in the time range of 300–250 Ma (mega-annum: one million years) and consisted of Gondwana (Australia, India, Sri Lanka, Madagascar, East Antarctica, South America and Africa) as its southern half and Laurasia (North America, Greenland, and Eurasia) as its northern half [20]. Our results suggest that the MGTV group of viruses are phylogenetically related to geographical regions that were formerly part of Gondwana and Laurasia.

Conclusions

In the present study, we conclude: (1) the high-quality, well-curated and full-length Yunnan2016 genome (MT080097, MT080098, MT080099 and MT080100) should be included in the NCBI RefSeq database for future studies on MGTV, JMTV, KITV and GXTV; (2) To the best of our knowledge, this is the first study in which 5' and 3' sRNAs were used to generate full-length genome sequences of, but not limited to, RNA viruses; (3) it is feasible to use the sRNA-seq based method for the detection of viruses in pooled two and even possible one small ticks; (4) MGTV may preserve the characteristic of ancient RNA viruses, which can be used to study the



origin and evolution of RNA viruses; and (5) MGTV can be used as novel species for studies in phylogeography. Future studies should be conducted to confirm the viability of MGTV in ticks and the hosts of these ticks.

Methods

The workflow to generate a full-length genome using 5’ and 3’ sRNAs can be seen in the Additional file 1. The full-length genome sequence of the MGTV strain Yunnan2016 has been deposited into NCBI GenBank database under the accession numbers MT080097, MT080098, MT080099 and MT080100. The sRNA-seq data was deposited in the NCBI SRA database under the accession numbers SRP084097 and SRP178347 (Table 1). In the present study, 17 MGTV, JMTV, KITV and GXTV complete genome sequences (including the Yunnan2016) were downloaded from the NCBI GenBank database (Additional file 1) and analyzed together. One genome sequence (GenBank: MN095531, MN095532, MN095533 and MN095534) was removed because it had too many ambiguous nucleotides. The online tool CD-HIT [21] (Date: 20191212) was then used to remove redundant sequences with the sequence identity cut-off 0.93 and default settings for other parameters, resulting in 9 complete genome sequences for the phylogenetic analysis. The multi-alignment of sequences and the

phylogenetic analysis were performed using the online tool ClustalW2 [22] and the software MEGA v7.0.26 [23], respectively.

The software Fastq_clean [24] was used for sRNA data cleaning and quality control. The virus detection was performed using the pipeline VirusDetect [25]. For each detected virus, VirusDetect assigned a closest reference genome from the NCBI GenBank database to help characterize that virus. VirusDetect used reference genome coverage and average depth to quantify the detected viruses for validation. Genome coverage was defined as the proportion of read-covered positions divided by genome length whereas average depth was the total number of base pairs of aligned reads divided by the read-covered positions of the reference genome. Statistical computation and plotting were performed using the software R v2.15.3 with the Bioconductor packages [26].

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-07060-5>.

Additional file 1 : Figure S1. A workflow to generate full-length genome sequence of an RNA virus. Table S1. Collection of ticks. **Table S2.** Primers for PCR amplification coupled with Sanger sequencing. **Table S3.** PCR reagent for each sample. **Table S4.** 17 complete genomes for further analysis.

Abbreviations

NGS: Next generation sequencing; MGTV: Mogiana tick virus; JMTV: Jingmen tick virus; KITV: Kindia tick virus; GXTV: Guangxi tick virus; sRNA-seq: small RNA sequencing; HPV-18: Human papillomavirus type 18; HBV: Hepatitis B Virus; HCV: Hepatitis C Virus; HIV-1: Human immunodeficiency virus type 1; SMRV: Squirrel monkey retrovirus; EBV: Epstein-barr virus; SARS-CoV: Severe Acute Respiratory Syndrome Coronavirus; ASFV: African Swine Fever Virus; cpsRNAs: complemented palindromic small RNAs; CDSs: Coding sequences; InDels: Insertions/Deletions

Acknowledgments

We appreciate the help equally from the people listed below. They are Wenjun Bu, Guoqing Liu, Dawei Huang, Yanqiang Liu, Bingjun He, Qiang Zhao, Zhen Ye and Xiufeng Jin from College of Life Sciences, Nankai University.

Authors' contributions

ZC and SG conceived the project and supervised this study. SG drafted the manuscript. SB and SK contributed to subsequent drafts of the manuscript. XX and JB executed the experiments. SG, JC and DC downloaded, managed and analyzed the data. YX prepared the figures and Tables. SG and XZ revised the manuscript. All authors have read and approved the manuscript.

Funding

This work was supported by the Natural Science Foundation of Guangdong Province, China (No. 2018A030310195) and Guangzhou Municipal Science and Technology Bureau, China (No. 201804010338) to Xiaoli Zhang, Tianjin Key Research and Development Program of China (19YFZCZY00500) to Shan Gao, and Science Foundation of Hebei Normal University (L2020B17) and Hebei Provincial Higher Education Science and Technology Research Foundation (QN2020162) to Ze Chen. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The complete genome sequence of Yunnan2016 is available at the NCBI GenBank database under the accession numbers MT080097, MT080098, MT080099 and MT080100. The sRNA-seq data used for virus detection is available at the NCBI SRA database under the accession numbers SRP084097 and SRP178347.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Hebei Key Laboratory of Animal Physiology, Biochemistry and Molecular Biology, College of Life Sciences, Hebei Normal University, Shijiazhuang, Hebei 050024, People's Republic of China. ²Guangdong Provincial Key Laboratory for Crop Germplasm Resources Preservation and Utilization, Agro-Biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong 510640, People's Republic of China. ³Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, Guangdong 510642, People's Republic of China. ⁴College of Life Sciences, Nankai University, Tianjin, Tianjin 300071, People's Republic of China. ⁵School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.

Received: 6 April 2020 Accepted: 10 September 2020

Published online: 16 September 2020

References

- Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, et al. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and

- sequencing of viruses. *Virology*. 2009;388(1):1–7. <https://doi.org/10.1016/j.virol.2009.03.024>.
- Li RG, Gao S, Hernandez AG, Wechter WP, Fei ZJ, Ling KS. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One*. 2012;7(5):1–10. <https://doi.org/10.1371/journal.pone.0037127>.
- Nayak A, Tassetto M, Kunitomi M, Andino R. RNA interference-mediated intrinsic antiviral immunity in invertebrates, in intrinsic immunity. Springer. 2013:183–200. https://doi.org/10.1007/978-3-642-37765-5_7.
- Wang F, Sun Y, Ruan J, Chen R, Chen X, Chen C, et al. Using small RNA deep sequencing data to detect human viruses. *Biomed Res Int*. 2016;2016(2016):1–9. <https://doi.org/10.1155/2016/2596782>.
- Liu C, Chen Z, Hu Y, Ji H, Yu D, Shen W, et al. Complemented palindromic small RNAs first discovered from SARS coronavirus. *Genes*. 2018;9(9):1–11. <https://doi.org/10.3390/genes9090442>.
- Chen Z, Xu X, Wang Y, Bei J, Jin X, Dou W, et al. DNA segments of African swine fever virus detected for the first time in hard ticks from sheep and bovines. *Syst Appl Acarol*. 2019;24(1):180–4. <https://doi.org/10.11158/saa.24.1.13>.
- Niu X, Sun Y, Chen Z, Li R, Padmanabhan C, Ruan J, et al. Using small RNA-seq data to detect siRNA duplexes induced by plant viruses. *Genes*. 2017;8(6):163. <https://doi.org/10.3390/genes8060163>.
- Chen Z, Sun Y, Yang X, Wu Z, Guo K, Niu X, et al. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS One*. 2017;12(4):1–9. <https://doi.org/10.1371/journal.pone.0176458>.
- Xu X, Ji H, Jin X, Cheng Z, Yao X, Liu Y, et al. Using pan RNA-seq analysis to reveal the ubiquitous existence of 5' and 3' end small RNAs. *Front Genet*. 2019;10:1–11. <https://doi.org/10.3389/fgene.2019.00105>.
- Ji H, Xu X, Jin X, Yin H, Luo J, Liu G, et al. Using high-resolution annotation of insect mitochondrial DNA to decipher tandem repeats in the control region. *RNA Biol*. 2019;16(6):830–7. <https://doi.org/10.1080/1547286.2019.1591035>.
- Jin X, Cheng Z, Wang B, Yau TO, Chen Z, Barker SC, et al. Precise annotation of human, chimpanzee, rhesus macaque and mouse mitochondrial genomes leads to insight into mitochondrial transcription in mammals. *RNA Biol*. 2020;17(1):1–8.
- Maruyama SR, Castro-Jorge LA, Ribeiro JMC, Gardinassi LG, Garcia GR, Brandão LG, et al. Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in Rhipicephalus microplus ticks from Brazil. *Mem Inst Oswaldo Cruz*. 2014;109(1):38–50. <https://doi.org/10.1590/0074-0276130166>.
- Waits K, Edwards MJ, Cobb IN, Fontenele RS, Varsani A. Identification of an anellovirus and genomoviruses in ixodid ticks. *Virus Genes*. 2017;54(1):155–9. <https://doi.org/10.1007/s11262-017-1520-5>.
- Harvey E, Rose K, Eden JS, Lo N, Abeyasuriya T, Shi M, et al. Extensive diversity of RNA viruses in Australian ticks. *J Virol*. 2018;93(3):1–33. <https://doi.org/10.1128/JVI.01358-18>.
- Zhuang L, Zhang Z, An X, Fan H, Ma M, Anderson BD, et al. An efficient strategy of screening for pathogens in wild-caught ticks and mosquitoes by reusing small RNA deep sequencing data. *PLoS One*. 2014;9(3):1–7. <https://doi.org/10.1371/journal.pone.0090831>.
- Emmerich P, Jakupi X, Von Possel R, Berisha L, Halilij B, Gunther S, et al. Viral metagenomics, genetic and evolutionary characteristics of Crimean-Congo hemorrhagic fever orthonaviruses in humans, Kosovo. *Infect Genet Evol*. 2018;65:6–11. <https://doi.org/10.1016/j.meegid.2018.07.010>.
- Chen Z, Yang X. Systematics and taxonomy of Ixodida (Chinese edition). Beijing: Science Press; 2020.
- Chen Z, Xuan Y, Liang G, Yang X, Yu Z, Barker SC, et al. Precise annotation of tick mitochondrial genomes reveals multiple copy number variation of short tandem repeats and one transposon-like element. *BMC Genomics*. 2020;21(488):1–11. <https://doi.org/10.1186/s12864-020-06906-2>.
- Duan G, Shi J, Xuan Y, Chen J, Liu C, Ruan J, et al. 5' UTR barcode of the 2019 novel coronavirus leads to insights into its virulence. *Chinese J Virol*. 2020;36(3):365–9. <https://doi.org/10.13242/j.cnki.bingduxuebao.003681>.
- Zhao G, Sun M, Wilde SA, Li S. A Paleo-Mesoproterozoic supercontinent: assembly, growth and breakup. *Earth-Sci Rev*. 2004;67(1–2):91–123. <https://doi.org/10.1016/j.earscirev.2004.02.003>.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–9. https://doi.org/10.1007/978-1-4899-7478-5_221.
- Larkin M, Blackshields G, Brown N, Chenna R, Higgins D. ClustalW and ClustalX version 2. *Bioinformatics*. 2007;23(21):2947–8. <https://doi.org/10.1093/bioinformatics/btm404>.

23. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 2008; 9(4):299–306. <https://doi.org/10.1093/bib/bbn017>.
24. Zhang M, Zhan F, Sun H, Gong X, Fei Z, Gao S. Fastq_clean: An optimized pipeline to clean the Illumina sequencing data with quality control. In: *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on: 2014*; Belfast: IEEE; 2014:44–48. <https://doi.org/10.1109/BIBM.2014.6999309>.
25. Zheng Y, Gao S, Padmanabhan C, Li R, Galvez M, Gutierrez D, et al. VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology.* 2016;500(2017):130–8. <https://doi.org/10.1016/j.virol.2016.10.017>.
26. Gao S, Ou J, Xiao K. *R language and bioconductor in bioinformatics applications* (Chinese edition). 2014. Tianjin: Tianjin Science and Technology Translation Publishing Ltd; 2014.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

