Gemykibivirus in acute encephalitis patients

Title Page

Article summary line:

Viruses in the genus Gemykibivirus can be potential causal agents of encephalitis.

Running Title:

Gemykibivirus in acute encephalitis patients

Key words:

acute encephalitis syndrome, Gemykibivirus, metagenomic analysis, emerging infection

Title:

Gemykibivirus detection in acute encephalitis patients from Nepal

Authors:

Eans Tara Tuladhar¹, Smita Shrestha¹, Susan Vernon, Lindsay Droit, Kathie A. Mihindukulasuriya, Mamta Tamang, Lata Karki, Annie Elong Ngono, Bimlesh Jha, Bal Krishna Awal, Bimal Sharma Chalise, Runa Jha, Sujan Shresta, David Wang², Krishna Das Manandhar²

Author affiliations:

Tribhuvan University Central Department of Biotechnology, Kathmandu, Nepal (E.T. Tuladhar, S. Shrestha, L. Karki, M. Tamang, K.D. Manandhar); National Public Health Laboratory, Kathmandu, Nepal (B. Jha, B.K Awal, R. Jha); Sukraraj Tropical and Infectious Disease Hospital, Kathmandu, Nepal (B.S Chalise); La Jolla Institute for Immunology, California, United States (A. Ngono, S. Shrestha); Washington University in St. Louis, Missouri, United States (S. Vernon, L. Droit, K. Mihindukulasuriya, D. Wang)

Footnotes

¹These authors contributed equally to this article; ²Corresponding authors
Gemykibivirus in acute encephalitis patients

Abstract

Acute Encephalitis Syndrome (AES) causes significant morbidity and mortality worldwide. In Nepal, Japanese encephalitis virus (JEV) accounts for ~5-20% of AES cases, but ~75% of AES cases are of unknown etiology. We identified a gemykibivirus in CSF collected in 2020 from a male child with AES using metagenomic next-generation sequencing. Gemykibiviruses are single stranded, circular DNA viruses in the family Genomoviridae. The complete genome of 2211 nucleotides was sequenced which shared 98.69% nucleotide identity to its closest relative, Human associated gemykibivirus 2 isolate SAfia-449D. Two real-time PCR assays were designed, and screening of 337 CSF and 164 serum samples from AES patients in Nepal collected in 2020 and 2022 yielded 11 CSF and 1 serum sample that were positive in both PCR assays. Complete genomes of 7 of the positives were sequenced. These results identify a candidate etiologic agent of encephalitis in Nepal.
Introduction

Encephalitis is a neurological disorder associated with a high mortality rate on a global scale (1). It is the inflammation of the brain parenchyma with clinical features of fever, altered mental state, and/or new onset of seizures. At present, low sociodemographic index regions in Asia and Africa carry the highest burden of encephalitis (2). In the year 2021, a total of 512 cases of AES were reported in Nepal (3). More than 100 different infectious agents that cause encephalitis are known including bacteria, viruses, fungi, and parasites (4, 5). The major known etiologic agents of encephalitis as reported in Nepal and internationally are Japanese encephalitis virus (JEV), enteroviruses, herpes simplex, and varicella zoster viruses (6, 7). In Nepal, there has been a national surveillance program for JE since 2004 wherein CSF and serum samples of suspected viral cases of encephalitis are collected from sentinel sites throughout Nepal with technical support from WHO and analyzed at the National Public Health Laboratory (NPHL) by serology for anti-JEV IgM. In numerous studies in Nepal assessing cases since 2000, ~70-95% of the AES cases per year have no diagnosis (3, 8-11). A significant fraction of encephalitis in other countries similarly lack diagnosis, despite extensive testing (5, 12, 13). In recent years, the application of metagenomic analysis to patients with encephalitis has begun to identify a range of emerging viruses linked to encephalitis (14-19).

The Genomoviridae family of viruses has single stranded DNA genomes of ~2.1-2.2 kb (20) that encode a capsid protein (CP) and a replication associated protein (REP). They have been identified from a wide range of hosts including plants, insects, animals, and humans (20-22). There are ten genera in the Genomiviridae family (20). Viruses in the genus Gemykibivirus have been identified in multiple human cases and in multiple specimen types including: blood of febrile Tanzanian children (23); the...
Gemykibivirus in acute encephalitis patients

respiratory tract of an elderly woman with respiratory distress in China (24); feces from diarrhea patients in Brazil (25); blood of healthy blood donors of Brazil (21); blood of HIV positive Cameroonian males (26); and in cervical swab of HIV/HPV infected pregnant females (27). Specific to encephalitis, there are reports of gemykibiviruses in CSF from an encephalitic child from China (15) and CSF from three patients with encephalitis from Sri Lanka (25). Furthermore, analysis of Nepalese sewage yielded the complete genome of a gemykibivirus (25). Here, we used metagenomic next generation sequencing (NGS) to identify the presence of a gemykibivirus in CSF from a patient with encephalitis in Nepal. Further PCR screening identified 12 additional positive cases.

Materials and Methods

Ethical clearance, study population, and collection of biospecimens:

This study was approved by the Nepal Health Research Council of Nepal [approval# 274-2020] and the Human Research Protection Office of Washington University in Saint Louis [202004087]. The study population focused on AES patients who were negative for JE IgM. Residual samples from national JE surveillance sentinel sites in Nepal were utilized. The index case CSF sample in the NPHL repository was collected in 2020 from Rupandehi district. For prevalence studies, 122 repository specimens (82 CSF, 40 sera) collected in 2020 and 379 repository specimens (255 CSF, 124 sera) from 2022 were tested by PCR for gemykibivirus.

Total nucleic acid extraction

Total nucleic acid extraction was performed using the Invitrogen Pure Link™ Viral RNA/DNA mini kit [Thermo Fisher Scientific] and eluted in 50 µL volume following the instruction manual of the kit. Samples were stored at -80°C.
Gemykibivirus in acute encephalitis patients

Metagenomic NGS analysis

Extracted total nucleic acid were randomly amplified as described previously (28) and used for library construction with NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). The sample library was sequenced on the Illumina MiSeq instrument using the 2 × 250 bp paired-end protocol. NGS data was analyzed for presence of viruses using CZID (29). NGS data is available at ENA: PRJEB72279.

Genome sequencing of index case and additional cases

The NGS contigs were confirmed using PCR, cloning, and sanger sequencing with primers (Appendix Table 1). Using a pair of primers that amplified the entire circular genome, Gemy1xgenomeF (5’TTAATCGATCTAGAGGATCCTTGTTAGATATCCATATGGCGG-3’) and Gemy1xgenomeR (5’-TTAGTAATGGGCCCGGATCCACGAGAGGAACACG-3’), three independent PCR reactions were performed, and the resulting fragments were cloned into pCR4.0 and sequenced using the Oxford Nanopore technologies (Plasmidsaurus). Additional positive cases were similarly amplified and whole genomes sequenced to 3X coverage. Complete genomes sequences are available at Genbank (Accession# PP270194-PP270201).

Gemykibivirus qPCR development

Two sets of Taqman real time PCR primers and probes were designed using express software (Applied Biosystems), one targeting the CP gene and the other targeting the REP gene. The primers and the probes were supplied by IDT (Integrated DNA Technologies, USA). The first assay, Gemy_1, targets the CP gene using primers GemykibiTM_8917F (5’-ACCTCTTATCCGTTTGGA-3’) and GemykibiTM_8917R (5’-AGCGCGAAATTCCTCTTGAC-3’) and the probe
Gemykibivirus in acute encephalitis patients

The first assay, Gemy_1, targets the 6' gene with GemykibiTM_8917F (5'-CGACCTGA[ZEN]CCGGATGCCC[3|ABkFQ]-3') that uses FAM and the dual quencher Zen and Iowa Black. The second assay, Gemy_2, targets the REP gene with GemykibiTM_9967F (5'-GGTCAGAGCCTAGTGTTGTATG-3')

For both assays, standard curves were generated using serial 10-fold dilution ranging from 2x10^8 to 2x10^1 copies of positive control plasmid [plasmid PCR4 containing the respective target sequence]. A 20 µL PCR mixture was made comprising 2 µL of extracted nucleic acid sample, 10 µL of 2x TaqMan Fast Advanced Master Mix [Thermo Fisher Scientific], and 5 pmol of each primer and probe. The PCR reactions were performed in 96-well plates on a CFX Opus 96 thermocycler [Bio-Rad] with one negative control nuclease free water in each row and one positive control of 2 x10^3 copies per plate. The cycling conditions were 50°C for 2 mins, 95°C for 30 secs and 40 cycles of 95°C for 5 secs followed by 60°C for 30 secs. The threshold of all plates was set at standard value and data was analyzed using Bio-Rad CFX Maestro 2.3 software. Samples were counted as positive if their threshold cycle (Ct) value was less than 33.

Phylogenetic analysis

Representative protein sequences of the REP gene of prototypes of each genus in the Genomoviridae were downloaded from GenBank. Alignments were generated using Clustal Omega (30). The alignment converted to fasta via http://sequenceconversion.bugaco.com/converter/biology/sequences/clustal_to_fasta.php. Maximum likelihood trees were generated with bootstrapping, using W-IQ-TREE.
Gemykibivirus in acute encephalitis patients

Trees were visualized using iTOL (31). All available complete genomes in the species *Gemykibivirus humas* were downloaded from Genbank along with representative genomes from the species *Gemykibivirus humas* 1, 3, 4, and 5, and the top 10 additional complete genomes with highest BLASTn scores. Alignments were generated using Clustal Omega (30). Maximum likelihood trees were generated with bootstrapping, using W-IQ-TREE (31). Trees were visualized using iTOL (32).

**Results**

**Detection of a gemykibivirus by metagenomic NGS**

NGS of nucleic acids extracted from the CSF of a male child yielded reads that could be assembled into two contigs that shared 97 and 99% nucleotide similarity with human associated Gemykibivirus2 SAfia-449D (accession# MN765187.1). Using PCR, gaps between the two contigs were spanned to generate a complete circular genome of 2211 nt. To formally assess the taxonomic relationship of this virus to viruses in the family *Genomoviridae*, we generated a maximum likelihood phylogenetic tree of the REP protein with the type species of each genus, in accordance with the ICTV guidelines (20, 22) (Figure 1A). The virus was most similar to the prototype virus from the genus *Gemykibivirus*. To further assess its relationship within the genus *Gemykibivirus*, we generated a maximum likelihood tree using the whole genome sequence (Figure 1B), which demonstrated it is most closely related to Human gemykibivirus 2 SAfia-449D, a virus detected in blood of Tanzanian children (23), with 98.69% nucleotide identity. Based on these criteria, the virus genome from the index case was designated Human gemykibivirus 2 Nepal/N0000051/2020. Human gemykibivirus 2 Nepal/N0000051/2020 was also closely related to Gemycircularvirus-SL1 (accession# KP133075), a previously
reported gemykibivirus detected in CSF from an encephalitis patient in Sri Lanka (25).

sharing 97.96% nucleotide identity (33 SNPs across the genome and a 12 bp insertion located in a region of tandem repeat hexamers) and it shared 98.64% identity to another gemykibivirus, Gemycircularvirus NP (accession# KP133080), detected in sewage from Nepal (25).

Gene predictions identified three ORFs, characteristic of Gemykibiviruses (Figure 2). The CP ORF is 969 bp, the REP ORF, generated by splicing is 1114 bp, and the unknown ORF3, which overlaps with the REP ORF, is 702 bp. A large intergenic region (LIR) of 127bp is present, and the putative viral origin of replication nona-nucleotide motif 5′-TAAAATTTA-3′ described in Gemycircularvirus NP (accession# KP133080) is conserved. A predicted stem loop in the LIR is observed from nucleotides 23 to 57. The stem loop structure is present in genomoviruses and geminiviruses where it is necessary for rolling circle replication (33, 34).

Identification of additional positive specimens from encephalitis patients

To define the prevalence of this virus, two Taqman real-time PCR assays were designed and validated, one targeting the REP gene and one targeting the CP gene (Appendix Figure 1). A total of 337 CSF samples and 164 serum samples were tested using both assays. Sample with Ct. values <33 for both assays were considered positive. There were three gemykibivirus positive CSF specimens from 2020 and 9 positive specimens (8 CSF and 1 serum) from 2022 (Table). The prevalence rate in CSF was 3.3% (11/337) and in serum was 0.6% (1/164). Positive patients ranged from 4 months to 72 years of age. Including the index case, there were 7 male and 6 female patients. Geographically, most positive patients were from districts in south-central Nepal (Figure 3). Unfortunately, no additional clinical details are available for the patients in this study besides meeting the acute case definition of encephalitis.
Whole genome sequencing and phylogenetic analysis of additional positive cases

From the 10 positive specimens with highest Gemykibivirus copy number, we tried to amplify the whole genome using PCR. For 7 of the samples, we obtained amplicons that corresponded to the whole genome. These were cloned and sequenced to 3X coverage to determine their consensus sequences. These 7 genomes varied from the index genome by 4 to 21 SNPs. The intron of the Rep gene contains a region with seven tandemly repeated hexamers, and deletions of one or more of the hexamers are observed in some of the genomes. Phylogenetic analysis of the complete genomes demonstrated that they formed a clade that included known Gemykibiviruses previously detected in human CSF, stool, and blood.

Discussion

We used metagenomic NGS to detect Human gemykibivirus2 Nepal/N0000051/2020 in CSF of a male encephalitic child from Rupandehi district (Figure 3) of Nepal which is alongside the border with India. Further screening through qPCR identified an additional 12 positive specimens from patient samples collected in 2020 and 2022. We found that these cases were mostly concentrated in and around the south-central region of Nepal (Figure 3). The identified virus genomes were very closely related (97.96% nt. identity) to a gemykibivirus previously detected in CSF from three encephalitis patients from Sri Lanka (25). In addition, a distinct gemykibivirus (Human gemykibivirus 4) has also been reported in CSF of an encephalitis patient from China (15). Together with our study, these data implicate viruses in the genus Gemykibivirus as potential causal agents of encephalitis. There is one report of Gemykibivirus in sewage from Nepal from 2012, which is also highly similar to the sequences we detected in Nepalese patients from 2020-2022. This suggests that gemykibivirus has been circulating in
Gemykibivirus in acute encephalitis patients

Nepal for at least the past decade. Furthermore, detection in sewage, raises the possibility that gemykibivirus may be transmitted fecal-orally, similarly to some neurotropic viruses such as polio and enteroviruses. Detection of highly similar viruses in patients in Tanzania (23) and Brazil (21) suggests that gemykibiviruses are globally widespread.

One limitation of this study is that the samples analyzed were residual specimens from a surveillance repository without additional available clinical metadata, thus limiting our knowledge of the precise symptoms, disease severity, and outcomes of these patients. While detection of gemykibiviruses in presumptively sterile CSF supports the hypothesis that they could be causal agents of encephalitis, additional research to culture the virus and establish animal models to fulfill Koch’s postulates are needed to definitively establish causality. In addition, more prevalence studies in encephalitis and other diseases are also needed as are serological studies to define the extent of human infection by gemykibiviruses. Finally, additional, unbiased approaches are needed to define the etiologies of encephalitis in Nepal, and worldwide.

Acknowledgment.

This project was supported by National Institutes of Health grant U01AI151810.

Biographical sketch of first author: Dr. Tuladhar, MD is currently pursuing a PhD at Tribhuvan University Central Department of Biotechnology, Nepal studying emerging viral infections. [ORCiD:0000-0001-5705-3953]
Gemykivivirus in acute encephalitis patients

Corresponding author’s mailing address

David Wang
ORCiD: 0000-0002-0827-196X
314-286-1123 (phone)
314-286-1232 (FAX)
davewang@wustl.edu
Washington University in St. Louis School of Medicine
660 S. Euclid Ave
St. Louis, MO 63110
USA
Krishna Das Manandhar
ORCiD: 0000-0002-6798-4935
+977 9860 960 577 (Phone)
krishna.manandhar@gmail.com
Central Department of Biotechnology, Tribhuvan University
TU Road, Kirtipur, Nepal
References


Table: Patient demographics and qPCR Ct. values of gemykibivirus positive cases

<table>
<thead>
<tr>
<th>Sample ID†</th>
<th>Specimen</th>
<th>Age group [yrs]</th>
<th>Gender</th>
<th>Ct. Gemy_1‡</th>
<th>Ct. Gemy_2§</th>
<th>Specimen collection date</th>
<th>Districts¶</th>
<th>Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>N0000218</td>
<td>CSF</td>
<td>11-15</td>
<td>M</td>
<td>32.42</td>
<td>30.68</td>
<td>2020 Feb</td>
<td>Kaski</td>
<td>Gandaki</td>
</tr>
<tr>
<td>N0000256</td>
<td>CSF</td>
<td>0-5</td>
<td>F</td>
<td>29.41</td>
<td>29.34</td>
<td>2020 July</td>
<td>Myagdi</td>
<td>Gandaki</td>
</tr>
<tr>
<td>N0000300</td>
<td>CSF</td>
<td>6-10</td>
<td>F</td>
<td>32.95</td>
<td>31.46</td>
<td>2020 Sep</td>
<td>Palpa</td>
<td>Lumbini</td>
</tr>
<tr>
<td>N0000358</td>
<td>Serum</td>
<td>0-5</td>
<td>M</td>
<td>29.27</td>
<td>29.94</td>
<td>2022 Feb</td>
<td>Sarlahi</td>
<td>Madhesh</td>
</tr>
<tr>
<td>N0000359</td>
<td>CSF</td>
<td>0-5</td>
<td>F</td>
<td>27.4</td>
<td>27.08</td>
<td>2022 Feb</td>
<td>Kapiivastu</td>
<td>Lumbini</td>
</tr>
<tr>
<td>N0000395</td>
<td>CSF</td>
<td>0-5</td>
<td>M</td>
<td>30.74</td>
<td>30.73</td>
<td>2022 April</td>
<td>Kaski</td>
<td>Gandaki</td>
</tr>
<tr>
<td>N0000434</td>
<td>CSF</td>
<td>56-60</td>
<td>M</td>
<td>29.02</td>
<td>30.06</td>
<td>2022 June</td>
<td>Lalitpur</td>
<td>Bagmati</td>
</tr>
<tr>
<td>N0000468</td>
<td>CSF</td>
<td>30-35</td>
<td>F</td>
<td>30.84</td>
<td>31.47</td>
<td>2022 July</td>
<td>Chitwan</td>
<td>Bagmati</td>
</tr>
<tr>
<td>N0000490</td>
<td>CSF</td>
<td>unknown</td>
<td>M</td>
<td>29.86</td>
<td>32.08</td>
<td>2022 July</td>
<td>Syangja</td>
<td>Gandaki</td>
</tr>
<tr>
<td>N0000545</td>
<td>CSF</td>
<td>70-75</td>
<td>F</td>
<td>26.84</td>
<td>25.35</td>
<td>2022 Aug</td>
<td>Kathmandu</td>
<td>Bagmati</td>
</tr>
<tr>
<td>N0000546</td>
<td>CSF</td>
<td>20-25</td>
<td>M</td>
<td>24.65</td>
<td>23.53</td>
<td>2022 Aug</td>
<td>Kathmandu</td>
<td>Bagmati</td>
</tr>
<tr>
<td>N0000722</td>
<td>CSF</td>
<td>20-25</td>
<td>F</td>
<td>28.74</td>
<td>29.72</td>
<td>2022 Nov</td>
<td>Chitwan</td>
<td>Bagmati</td>
</tr>
</tbody>
</table>

*Ct., cycle threshold; yrs, years.
†The sample IDs are coded and are not known outside the research group.
‡Gemy_1: qPCR targeting capsid protein gene.
§Gemy_2: qPCR targeting replication associated protein (REP) gene.
¶Districts involved are highlighted in Figure 3 as shade.
### Appendix Table: Primers used for PCR

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’ to 3’</th>
<th>Binding Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gemy8917-9967F</td>
<td>TCAACGACCTCTGATACATAACC</td>
<td>429 &lt;- 451</td>
</tr>
<tr>
<td>Gemy8917-9967R</td>
<td>TAGATGAGTCCACCATCAGC</td>
<td>2034 -&gt; 2055</td>
</tr>
<tr>
<td>Gemi1625confirmR</td>
<td>GCCATCGCAATTATGGCTTATGTT</td>
<td>88 -&gt; 113</td>
</tr>
<tr>
<td>Gemi1625confirmF</td>
<td>GGCAGAATCTCGACACTGGG</td>
<td>1194 &lt;- 1215</td>
</tr>
<tr>
<td>Gemy9967-8917R</td>
<td>TCATAATCTGCTCCGTGTTCT</td>
<td>1133 -&gt; 1154</td>
</tr>
<tr>
<td>GemykibiTM_9967F</td>
<td>GGTAGACCTGTGTTGTTATG</td>
<td>1419 &lt;- 1440</td>
</tr>
<tr>
<td>Gemy1xgenomeF</td>
<td>TTAATCGATCTAGAGTCTCTGTGGATATGACATGAGAAGCCCG</td>
<td>1160 -&gt; 1187</td>
</tr>
<tr>
<td>Gemy1xgenomeR</td>
<td>TTAGTATGGCCCGGATCCACGAGAAGAAGCCCG</td>
<td>1146 &lt;- 1165</td>
</tr>
</tbody>
</table>

### Figure Legends.

**Figure 1.** Phylogenetic trees of (A) amino acid sequences of the index case REP protein and representatives of each genus in the family *Genomoviridae*; (B) whole genome nucleotide sequences of the positive samples from Nepal compared to other genomes within the genus *Gemykibivirus*.

**Figure 2.** Schematic of the genome of Human gemykibivirus 2 Nepal/N0000051/2020.

**Figure 3.** Map of Nepal and locations of the gemykibivirus positive cases [shaded].
Human Gemykibivirus 2
Nepal/N0000051/2020
2211 bp

Replication associated protein
Capsid

ORF3
Figure 3. Map of Nepal and locations of the genykhiviruses positive cases [shaded].