Investigation of acute encephalitis syndrome with implementation of metagenomic next generation sequencing in a low-middle-income setting

Shrestha R*1,2, Katuwal N*1, Tamrakar D1,3, Tato CM4, Vanaerschot M5, Ahyong V4, Gil J4, Madhup SK6, Gupta BP7

1 Center for Infectious Diseases, Dhulikhel Hospital Kathmandu University Hospital
2 Department of Pharmacology, Kathmandu University School of Medical Sciences
3 Department of Community Medicine, Kathmandu University School of Medical Sciences
4 Rapid Response Team, Chan Zuckerberg Biohub
5 Roche Diagnostics
6 Department of Microbiology, Kathmandu University School of Medical Sciences
7 World Health Organisation-Nepal

* Contributed equally

Word Count (Intro, Methods, Results, Discussion and Conclusion): 3264

Abstract Word Count: 371

Figures: 4

Tables: 1

Correspondence:
Dr. Rajeev Shrestha
Chief, Center for Infectious Diseases, Dhulikhel Hospital Kathmandu University Hospital
Professor, Department of Pharmacology, Kathmandu University School of Medical Sciences
Dhulikhel, Kavre, Nepal
Email: rajeev.shrestha@kusms.edu.np
Ph: +977-9801002489
Abstract

Background

Acute Encephalitis Syndrome is characterised by acute onset of fever and altered sensorium followed by a rapidly worsening clinical condition and even death. The causative agents vary with season and geographic location, while the etiologies remain unknown in 68-75% of the cases. In Nepal, the cases of acute encephalitis syndrome are tested only for Japanese encephalitis, which constitutes about 15% of the cases globally. However, there could be several causative organisms, including vaccine-preventable etiologies that cause acute encephalitis, which upon identification could direct public health efforts for prevention, including expanded use of vaccines or address gaps in vaccine coverage.

Objective

This study employs metagenomic next generation sequencing in the exploration of infectious etiologies contributing to acute encephalitis syndrome in a low-and-middle-income setting.

Methods

In this study, we have investigated 90, Japanese encephalitis-negative, banked retrospective cerebrospinal fluid samples that were collected as part of a national surveillance network in 2016 and 2017. The randomisation was done to include three age groups of <5 years, 5-14 years, and >15 years. Only some metadata related to the subjects (age and gender) were known. The analysis was performed in two batches which included total nucleic acid extraction, followed by individual library preparation for DNA and RNA and sequencing on Illumina iSeq100. The generated genomic data were interpreted using CZID and confirmed with polymerase chain reaction.

Results

Human alphaherpesvirus 2 and Enterovirus B were seen in two of the ninety samples. These hits were confirmed by qPCR and seminested PCR respectively. Most of the other samples were marred by low sample quality, lack of clinical metadata, storage issues, and possible ambiguous collection procedures.

Conclusion

From this study, two documented, causative agents were revealed through metagenomic next generation sequencing. Insufficiency of clinical metadata, process controls, and absence of standard operating procedure to collect and store samples in nucleic acid protectants impeded the study and incorporated ambiguity while correlating the identified hits to infections. Therefore, there is dire need of implementing stringent collection procedures for sample collection, including proper process controls. Despite the challenging conditions, this study highlights the usefulness of mNGS to investigate diseases with unknown etiologies, such as AES, and guide the development of adequate clinical management algorithms and outbreak investigations in low-middle income settings.
Keywords: acute encephalitis syndrome, enterovirus, Human alphaherpes virus, low-middle-income country, metagenomic next generation sequencing
INTRODUCTION

Acute Encephalitis Syndrome (AES) is defined by acute onset of fever and a change in mental status (including symptoms such as confusion, disorientation, coma, or inability to talk) and/or new onset of seizures (excluding simple febrile seizures) in a person of any age at any time of year.¹ This term was coined by World Health Organization (WHO) in 2008.¹ Globally, based on various studies, the incidence of AES has ranged from 3.5 to 7.4 per 100,000 patients-years, with a higher incidence among children.²

The patients suffering from AES usually present acute onset of fever and altered sensorium. This is followed by rapidly worsening clinical conditions and death.³ The survivors can suffer from long term health issues, including neurological sequelae.³,⁴ The etiologies behind AES can be grouped under infective and non-infective categories, with the infective category comprising of a broad range of organisms (bacteria, virus, parasites).²,⁵ The causative agents of AES also vary with season and geographic location.⁶ Research has shown that the etiologies of AES remain unknown in 68-75% of the cases, while Japanese encephalitis (JE) constitutes about 15% of the cases.⁷,⁸,⁹,¹⁰ The landscape of AES, in terms of etiology, has changed in India as well, where outbreak investigations and surveillance studies have increasingly reported non-JEV etiologies.¹¹

In Nepal, JE is majorly associated with mortality and morbidity among children.¹² Therefore, since 2004, the Ministry of Health and Population of Nepal, supported by the Office of Infection Prevention Division, World Health Organisation (WHO), has integrated JE surveillance with Acute Flaccid Paralysis, Neonatal Tetanus, and Measles in its National Surveillance Network.¹³ Until 2011, over 23,000 AES cases were reported by the surveillance network.¹⁴ Due to a lack of knowledge in etiology, AES cases are only tested for JE and clinical management is performed based on this result. The incidence of undiagnosed AES etiology contributes to a high rate of death and morbidity.¹⁴ There could be several etiologies, including vaccine preventable etiologies, that cause acute encephalitis, which upon identification could direct public health efforts for prevention, including expanded use of vaccines or addressing gaps in vaccine coverage. Herpes Simplex Virus (HSV), Varicella-Zoster Virus (VSV), Enterovirus, Adenovirus, and Rubella, as well as emerging pathogens such as Nipah, Chandipura and Chikungunya have all been reported as causative viral agents of AES, while Neisseria meningitidis, Streptococcus pneumoniae, Listeria sp, and Brucella have been reported as causative bacterial agents.²,¹⁵,¹⁶,¹⁷
While molecular methods such as PCR require prior genetic information on causative agents, genomic methods such as metagenomic Next Generation Sequencing (mNGS) can simultaneously identify minute amounts of infections and co-infections of varying origin in a single investigation and assist in the investigation of transmission of such infections. With the recent dramatic decrease in sequencing costs, this technology provides access to genomic information in a scale that can be implemented to fill gaps in routine clinical practice and address epidemiological questions. In addition to identification (identifying genotypes, virulence or pathogenesis), NGS provide information epidemiological investigation (comparative genomics, phylogenetic analysis).

This study, employing mNGS to explore the infective etiologies behind AES, complements a growing number of studies that have used a similar approach to investigate encephalitis, including in a LMIC context. The identification of such etiologies is an important step in developing effective prevention and treatment measures which in turn will reduce disability and morbidity in LMICs and beyond.

**METHODOLOGY**

**Sample Collection and Selection**

The investigation included a random selection of 90 retrospective cerebrospinal fluid (CSF) samples that were collected by WHO-IPD (World Health Organisation-Immunization Preventable Diseases), throughout Nepal, as a part of the National AES Surveillance Network in collaboration with FWD (Family Welfare Department) in 2016 and 2017. These samples had been tested for JE at NPHL (National Public Health Laboratory) and stored at low but undocumented temperature. For this study, only those samples that tested negative for JE were selected. A randomisation was done to include three age groups of <5 years, 5-14 years, and >15 years. Only some metadata related to the subjects (age and gender) were known. Each sample was provided with unique study codes to maintain privacy.

**Nucleic Acid Extraction and mNGS**

Total Nucleic Acid was extracted from the CSF samples using Zymo Quick-DNA/RNA™ Pathogen MiniPrep (R1042).

The total nucleic acid samples were aliquoted into two sub samples for RNA and DNA library preparation, respectively. The library preparations were done using NEB Library Prep Kit Ultra II RNA for RNA (New England Biolabs, E7770S) library preparation, and NEB Library Prep Kit Ultra II FS
DNA for DNA library preparation (New England Biolabs, E7805S). The library preparation for the first 30 samples was done in a single batch (at Chan Zuckerberg Biohub, USA) while the remaining 60 samples were done in three batches of 20 samples each (at Dhulikhel Hospital Kathmandu University Hospital, Nepal). Negative extraction and library preparation controls were included in each batch. The library preparation included 10ng of input nucleic acid, followed by fragmentation, adapter ligation, cleanup (Solid Phase Reversible Immobilization beads), barcoding and amplification of library for 12-16 cycles. With subsequent library preparations, quality control was done using agarose gel electrophoresis and Tapestation 4200 platform from Agilent Technologies and later by qPCR using Kapa Illumina Library Amplification (KK2702) and Quantitation Complete Kit (KK4923). It was made sure that the length of DNA in the libraries was around 350-400bp and had concentration >1nM. In RNA library preparation, ERCC (External Control Controls Consortium, 4456740) RNA Spike-in controls were used.

The libraries that passed quality control filters were pooled and run on an Illumina iSeq100 sequencer. The sequencing was performed for 2x146bp length using custom unique dual indices of 12 bp length. 5% PhiX was added as an internal control for sequencing. The loading concentration of pooled libraries was maintained at 100-120pM.

**Data Analysis**

The analysis was performed on the CZ ID (formally known as IDSeq) platform developed by Chan Zuckerberg Initiative and CZ Biohub. CZID accepts raw sequencing data, perform host and quality filtration, followed by execution of assembly based alignment pipeline. The samples are analysed based on number of reads per million (Number of reads aligning to the taxon in the NCBI NR/NT database, per million reads sequenced), reads (Number of reads aligning to the taxon in the NCBI NT/ NR database), contig number (Number of assembled contigs aligning to the taxon in the NCBI NT/ NR database), id% and z-score. The samples were also visualized using a rpm heatmap where samples and controls are cross-matched against each other. Respective background models were created, from negative extraction and library preparation control, for RNA and DNA Libraries.

**PCR Confirmation**

Human alphaherpes virus confirmation was done by qPCR (KAPA HiFi HotStart Ready Mix) using two primer sets: established primers (FP: 5’TGCAGTTTACGTATAACCACATACAGC 3’ and RP: 5’ AGCTGCGGGCCTCGTT 3’) and self-designed primers (FP: 5’ GACTCAAACACGTGCACCAC
The primer sets were designed using NCBI primer blast and GeneScript, then checked with Beacon Designer Free and SnapGene Viewer. Similarly, for confirmation of Enterovirus, modified protocol with established primers from Enterovirus Surveillance Guidelines were used to perform semi-nested Polymerase Chain Reaction (snPCR). The protocol followed visualisation of the bands, for confirmation, in agarose (1.5%) gel electrophoresis.

**Ethics**
This study was ethically cleared from Nepal Health Research Council (NHRC) under id: 903 – 2019.

**RESULTS**

**Subject Metadata**
The samples selected for this study were banked, retrospective CSF samples collected in 2016 and 2017 with limited metadata such as age and gender. Out of the 90 subjects, 31 (34.4%) were female while the age distribution has been presented below:

<table>
<thead>
<tr>
<th>SN</th>
<th>Age Group</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;5 years</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>5-14 years</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>&gt;15 years</td>
<td>53</td>
</tr>
</tbody>
</table>

The median age of the subjects infected with AES was 20 years (IQR: 4-79 years)

**Sample collection**
The samples were collected in glass bottles without any preservative and transported to NPHL where they, first, had been tested for JE and subsequently stored at undocumented but low temperature. As these samples were collected in 2016 and 2017 and banked, negative controls were not available during collection and transportation.

**Nucleic Acid Extraction and mNGS**
The extracted nucleic acid had a concentration ranging from too low to detect to 222 ng/ul. As the analysis was done in two sets. Each set was processed for DNA and RNA Library Preparation and has
been presented accordingly. Out of 90 samples, only two samples showed confirmed hits from Enterovirus B and Human alphaherpesvirus 2, respectively.

**mNGS of RNA Libraries**

The results from RNA libraries showed some distinct organisms hit in CZID, and also provided a broad picture of the landscape of taxa across the samples. The following are heatmaps generated from through RNA library preparation.

**Figure 1:** Heatmap depicting top hits from sequencing of RNA library for S01 to S90. The organisms (at the genus level) that were seen in the samples are shown on the left. The names of the samples are on the top of the heat map. The samples marked by green are water controls during extraction (NEC) and library preparation (NLC). The heatmap was generated using the threshold of NT rPM (nucleotide reads per million) $\geq 10$ and NT L (alignment length in basepairs: length of the aligned sequence) $\geq 50$.

**Figure 2:** Heatmap depicting hits for Enterovirus (unnamed genus taxon 12059 and 138949) from sequencing of AES_S47_RNA. The names of the samples are on the top of the heat map. The samples marked by green are water controls during extraction (NEC) and library preparation (NLC). The heatmap was generated using the threshold of NT rPM (nucleotide reads per million) $\geq 10$ and NT L (alignment length in basepairs: length of the aligned sequence) $\geq 50$.

In figures 1 and 2, we can see top hits of organisms in the heat map that shows various organisms which are seen at similar levels in the water controls as well. Nevertheless, Pseudomonas genus is seen...
in all of the samples including few negative controls. There was similar trend, in both sets, with other organism such as Sphingomonas, Acinetobacter, Escherichia and others.

Interestingly, only AES_S47_RNA showed a hit to Enterovirus B. This hit was particular to sample 47 and not seen in any negative controls. The metrics such as rPM of 359,409.1 (provides information of the abundance of a specific microbe within the sample), NT L (depicts the length of aligned sequence in base pair), Z score of 99 (shows the significance of any hit compared to the background), coverage visualization (assess breadth and depths of reads) and id of 85.8% signify that the organism hit is highly similar to the reference organism. The figure below shows the abundance of Enterovirus B in Sample 47 (NT rPM >=10 and NT  L >=50). The coverage breadth of this hit was 98.7% with depth of 700.4x as seen in Figure 3.

Figure 3: Coverage visualisation of Enterovirus in AES_S47_RNA from CZID, depicting the coverage metrics of the contigs generated for the particular hit along with coverage depth and width. The hit was visualised with the threshold of NT rPM >=10 and NT  L >=50.

mNGS of DNA Libraries

In mNGS of DNA libraries, hits were observed for Human alphaherpesvirus 2 [AES_S28_DNA] from the first set. The same sample showed hit for Human alphaherpesvirus 1, but in a very low abundance. Additionally, background contaminations (laboratory and hospital) were seen in the water controls in this DNA sequencing result as well. Similar to RNA Libraries, most of the samples showed hits for Sphingomonas spp., Pseudomonas spp, or Acinetobacter spp. Nevertheless, the following figure depicts the result of the hit where there were 2,598.6 rPM for Human alphaherpesvirus 2.
Figure 4: Result from CZID depicting the details of top hits in AES_S28_DNA along with various metrics related to the hit.

The hits were visualised with the threshold of $NT_{rPM} >= 10$ and $NT_{L} >= 50$.

However, due to lower coverage, contig visualization was not available for this hit.

PCR Confirmation

Confirmation of Human alphaherpesvirus 2

Out of the two primer sets used, the established primers fared better providing Ct value of 23.11 for Human alphaherpesvirus 2.

Confirmation of Enterovirus B

After completion of snPCR for Enterovirus, the band was seen between 700-800bp after first amplification and between 300-400bp after final amplification. This confirmed the presence of Enterovirus as per the Enterovirus Surveillance Guidelines.

DISCUSSION

Demography of Acute Encephalitis Syndrome

Most of the subjects suffering from acute encephalitis syndrome (AES) were young male population of median age 20 years. This gender distribution was concurrent to previous studies done, in Nepal, on epidemiology of AES. It has been observed that AES affects individuals from both gender and all ages, however, most of the studies have been done in younger population, as they pose high risk due to lack of developed antibodies. Another study done in Nepal also observed the young median age (19 years) for AES, while others observed older population.

Metagenomic Next Generation Sequencing

In this study, out of the 90 samples tested, most (n=88) of them could not be specified as specific hits. This was due to high level of background contaminants resulting in low confidence in calling organism
hits within the experimental samples. Nevertheless, two samples showed confirmed hits for Enterovirus B and Human alphaherpesvirus 2, respectively. The absence of causative agent in remaining samples indicate that either the samples did not have intact nucleic to start with or were degraded.\textsuperscript{16,39} Extended storage in uncontrolled conditions could also be a likely contributing factor for the degradation.

As per the result of mNGS, the high Z score (99) for Enterovirus B shows that hit for the organism is present significantly in our sample, when compared to the background. The average length of alignment (as shown by L metrics) is long (L=7258.7), which confirms for a good local alignment to reference.\textsuperscript{40} The id\% is also higher (85.8\%) meaning that the organism is highly similar to the reference organism in the database. Additionally, when the genome coverage is seen in detail, we can see that our sequenced genome depicts good coverage breadth and depth (depth of 700x and breadth of 98.7\%), which is the range and uniformity of sequencing coverage for the particular hit.\textsuperscript{39} The presence of ENVB was also confirmed through snPCR followed by visualization of product size specific for all enteroviruses.\textsuperscript{29}

The hit for alphaherpesvirus 2 was considered significant because it was not present in the control samples at the thresholds used to analyse the sample (high Z score of 100\%, L value of 128.9, id\% of 99.9\%) considered reliable.\textsuperscript{24,27,41} The low contig value, for this hit, could be because of the organism being present at such a low abundance that the sequencer did not sequence enough reads to generate a contig. The contig value is dependent upon the total number of reads and the size of organism’s genome.\textsuperscript{42} Additionally, the decreased sensitivity of mNGS due to low abundance of pathogen has been studied for CSF.\textsuperscript{43} Several methods have been reported that can be used to increase the abundance of pathogen sequences or remove the unwanted host sequences.\textsuperscript{44,45} Nevertheless, as this genus is associated with encephalitis, the sample was taken further for analysis.\textsuperscript{46,47} During confirmation, the lower Ct value of 23.11 indicates presence of alphaherpesvirus 2, a known causative agent, in the sample.

**Enterovirus B and Human Alphaherpesvirus 2**

Enterovirus B is a known causative agent of encephalitis.\textsuperscript{16,48,49,50} Enteroviruses are named by their transmission-route through the intestine.\textsuperscript{51} Studies have shown that enterovirus can cause various diseases in the nervous system, including aseptic meningitis, acute paralysis, encephalitis, meningo-encephalomyelitis among others, in children.\textsuperscript{52,53,54} Interestingly, studies in India have linked Enterovirus, among other pathogens, to AES, by various studies.\textsuperscript{55,56,57} For instance, Enterovirus
outbreak was first reported from Uttar Pradesh, India in 2006 with seasonal outbreaks with high fatality occurring for several years. Southern Nepal borders with UP, India and due to open borders and similar climate, it is no surprise that Enterovirus was found in AES infected CSF sample from Nepal. Additionally, some studies in Nepal have reported Enterovirus as possible etiology of AES for Nepal.

Similarly, Human alphaherpesvirus 2 is known to cause encephalitis in neonates and immunocompromised patients. Herpes simplex encephalitis (HSE) has significant morbidity and mortality, even with early diagnosis and treatment. HSV is found to be one of the predominant causes of AES in the western world. Among herpes simplex encephalitis, the vast majority of the encephalitis is caused by HSV-1, with HSV-2 being the etiology in less than 10% of the cases. Studies in India and Nepal have reported the presence of HSV-2 as causative agent of encephalitis, with varying range of incidence.

Clinical Data and Process Control

However, due to lack of proper clinical metadata, the presence of Enterovirus B and Human alphaherpes 2 virus could not be clinically correlated. Clinical metadata such as onset of fever, date of infection, fatality, WBC counts, adjoining infection, etc are vital to correspond with the presence of infections. Additionally, usual environmental contaminants such as Sphingomonas spp., Pseudomonas spp, or Acinetobacter spp were seen. For instance, Sphingomonas are widely distributed in nature, having been isolated from many different land and water habitats, as well as from plant root systems, clinical specimens, and other sources. This is essentially due to their ability to survive in low concentrations of nutrients. Background contaminations of laboratory and hospital origin were also seen in the water controls. With appropriate use of background or negative controls, a background model can be created and subsequently subtracted from the results.

Collection Procedures

The lack of identification of causative agent in other 88 samples could be because the majority of samples were degraded prior to extraction. The samples that were analysed were as old as 2016 and 2017, and were collected without stringent or consistent guidelines. They could have gone through numerous freeze and thaw cycles. Moreover, the sample could have gone through degradation prior to storage depending upon how they were transported from the site of collection. The causative agents
could also have left the cerebrospinal fluid prior to collection. Collection of samples in nucleic acid
protectant such as Zymo RNA/DNA Shield would have protected the nucleic acid from degradation
after sampling.76,77

We postulate that the results of this study should not be corroborated as limitations as the CSF samples
analysed were not collected specifically for mNGS. The lack of stringent collection procedures and
sample quality are likely limiting to the analysis, resulting in ambiguous interpretation of some
samples. Additionally, the sequencing was done in Illumina iSeq100 which has a maximum of
approximately 4 million reads per run and can only accommodate a certain number of organisms with
adequate coverage breadth and depth.78 Additionally, host depletion and pathogen enrichment methods
can be applied for samples with low pathogen abundance.44,45

CONCLUSION

From this study, two documented, causative agents were revealed through metagenomic next
generation sequencing and subsequently confirmed by PCR. Insufficiency of clinical metadata, process
controls, and absence of standard operating procedure to collect in nucleic acid protectant incorporates
ambiguity when correlating identified pathogens to infections. Therefore, there is a dire need of
implementing stringent collection procedures for sample collection, including proper process controls.
Additionally, delicate samples such as CSF should be collected in a protectant and transported in a
controlled and sterile environment. Collection of clinical metadata such as WBC count, primary
diagnosis, discharge type and presence of other known organisms are vital for the analysis.
Identification and investigation of etiologies behind AES is essential for developing clinical
management algorithms, improving surveillance with region-specific treatment and prevention policy
as well as outbreak investigation. We do not expect the administration of mNGS as a regular diagnostic
tool but rather an investigational and exploration tool to identify causative etiologies and develop
molecular methods (such as qPCR) for diagnosis.

Declaration of competing interest

We declare that we have no potential conflicts of interest.

Acknowledgements

We thank Bill and Melinda Gates Foundation, Grand Challenges Explorations Grant for the support.
We appreciate the guidance from Chan Zuckerberg Biohub, San Francisco. We express our gratitude to
Family Welfare Department and National Public Health Laboratory (NPHL) for supporting the study by providing the samples. We also thank WHO-IPD for providing data for the study.

REFERENCES:


18. Patricia J Simner, Steven Miller, Karen C Carroll, Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases, Clinical Infectious Diseases, Volume 66, Issue 5, 1 March 2018, Pages 778–788


76. Thraenhart O, Jursch C. Virucidal activity if the nucleic acid preservation product “DNA/RNA Shield” against the murine parvovirus (MVM) at 20°C. 2018. Eurovir.