Establishing most suitable NGS approach for EV Surveillance

UK Health Security Agency

K. Fuentes-Cabrejo[1], S. Hang[1], Y. Mohammadi[1], D. Bibby[2], C. Celma[1]

[1] Enteric Virus Unit. Reference Services. Colindale. UKHSA.

[2]Antiviral Unit. Reference Services. Colindale. UKHSA.

INTRODUCTION

Enteroviruses (EV) can infect the respiratory and gastrointestinal tract and are commonly transmitted through the faecal-oral route.[1] Infections can cause from mild illness to severe clinical disease complications including meningoencephalitis, myelitis, paralysis, myocarditis, sepsis-like syndrome, respiratory disease, and acute hepatitis mainly in young children. EVs belong to the *Picornaviridae* family and their ssRNA genome consists of a single open-reading frame (ORF) between 7.2kb and 8.5kb in length, four structural proteins and seven non-structural proteins. EV are classified into genogroups and genotypes based on their VP1 capsid protein sequence.[2] EV-A71 and EV-D68 are considered genotypes of clinical importance as they have been associated with large outbreaks, severe respiratory disease infections, and fatal cases. EV as all RNA viruses are well known for its high mutation rate and genomic recombination. [3]

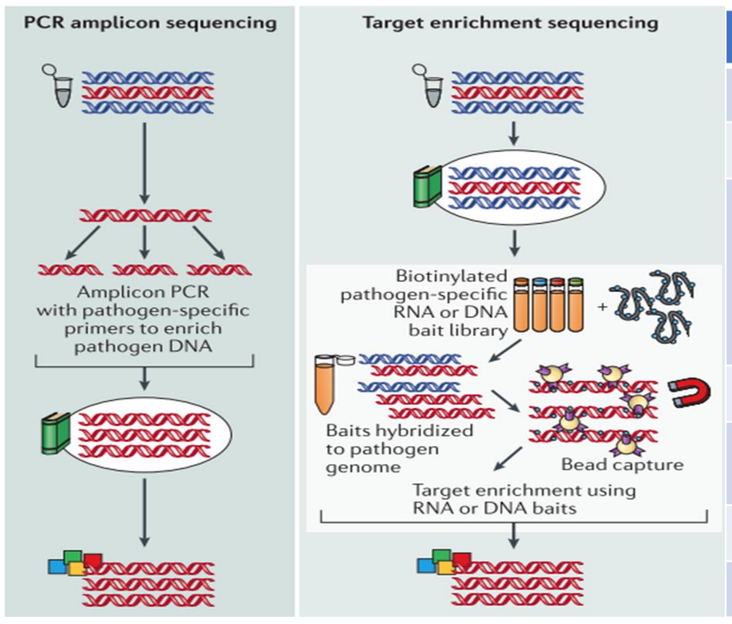
The Enteric Virus Unit as part of the UKHSA Reference Services, is the National Reference Laboratory for EV surveillance. EV referred samples are genotyped by PCR amplification of the VP1 gene (partial) and Sanger sequencing. However, EV co-infection cases, emergent strains, genomic mutations, and recombinant events remain undetected using this method. Through the application next generation sequencing (NGS) technologies, it has been possible to identify new EV genotypes[4], correlate strains with clinical symptoms[5], and investigate outbreaks[6].

AIM: Explore NGS technologies to strengthen current EV surveillance workflow. Through full or partial EV genome sequence analysis broaden our knowledge about genome diversity and investigate its circulation in the UK.

METHODS

EV Sample selection criteria

- *Archived original material or extracted RNA from 2016 to 2022 (excluding 2020).*
- *Isolates (cell culture).*
- *Different sample types including: biopsy, blood, faeces, CSF, respiratory, vesicle fluid and swabs.* ***Only*** *respiratory derived samples tested in Targeted sequence capture method.*
- *Confirmed positive samples by Real Time PCR (Ct<32) .*
- *Genotyped by Sanger sequencing.*
- *Untypable samples that required further investigation.*



PCR Amplicon-based[8]	Targeted sequence capture[9]
2 days	4 days
Extracted RNA (QIASymphony-QIAGEN)	RNA Extraction (Nuclisens-Easymag bioMerieux).
RT primer (specific EV primers, RNAseOUT) RNA denaturation and snap-cool step RT Reaction (MMX buffer, SSI HiFi, RNAseOUT) PCR 1 (EV specific primers)	DNA digestion Library preparation (cDNA synthesis, PCR1, ribosomal cDNA cleavage, PCR2) Library pooling, hybridisation and capture-based target enrichment. Pooling of captured libraries.
Clean-up: AMPure XP Beads and 80% ethanol.	Clean-up: AMPure XP Beads and 80% ethanol.
QC: Gel electrophoresis. Qubit dsDNA HS	QC: Real Time PCR, Glomax, Qubit dsDNA HS, Bioanalyzer
Illumina sequencing (including library preparation)	Illumina sequencing
Data analysis: in-house EVU pipeline	Data analysis: in-house EVU pipeline

RESULTS

PCR Amplicon based NGS

A total of 232/467 (49%) archived genomic RNA samples have been sequenced. Absence of PCR amplification on QC gel electrophoresis and no sequence data output were considered as NGS failures. Using PCR amplicon-based method described by Majumdar *et al*[8], commonly reported and unusual EV genotypes (highlighted in bold) were successfully sequenced in EVU. Cases of co-infection with more than one EV genotype or HRV (Human Rhinovirus) were also detected.

EV Genogroup	Total Sequenced	Strains
EV-A	21	*CV-A1, CV-A2, CV-A4, CV-A6, CV-A10, CV-A16, CV-A17,* ***EV-A76, EV-A90***
EV-B	30	*CV-A9, CV-B2, CV-B4, E-3, E-9, E-11, E-18,* ***EV-B80***
EV-C	11	***CV-A19,CV-A20, CV-A22, CV-A24, EV-C99,EV-C105***
EV-D	130	*EV-D68*
HRV	16	*HRV-A1, HRV-A21, HRV-A78, HRV-A80, HRV-C2, HRV-C6, HRV-C7, HRV-C9, HRV-C15, HRV-C24, HRV-C25, HRV-C31, HRV-C35, HRV-C39, HRV-C40, HRV-C56*
Co-infection (EV/EV)	17	*CV-A1/ E11, CV-A1/E-21/EV-A90, CV-A4/EV-D68, CV-A4/CV-A9/EV-D68, CV-A5/EV-D68, CV-A6/EV-D68, CV-A6/HRV-7, CV-6A/ CV-B2, CV-16/CV-A24, CV-A9/ EV-D68, CV-B2/EV-D86, CV-B3/ E-25, E-2/CV-A22*
Co-infection (EV/RV)	7	*HRV-A1/EV-D68, HRV-A21/EV-D68, HRV-A78/HRV-C2, HRV-A80/EV-D68, HRV-C2/EV-D68, HRV-C6/EV-D68, HRV-C39/EV-D68*
Grand Total	232	

Table 1. EV strain types sequenced using PCR Amplicon based method from 2021 to 2022.

EV-D68

During 2021 and 2022, an increased number of EV-D68 cases were reported in the UK with a total of 132 samples sequenced. EV-D68 was detected in different sample types including respiratory (n=87), faeces (n=6), skin/vesicle swabs (n=8), CSF (n=1) and unspecified location (n=30). Four phylogenetic clades have been described for EV-D68. Since 2016, clade B3 is the main circulating clade reported in the UK and was confirmed in 93% (130/140) of EV-D68 sequenced isolates.

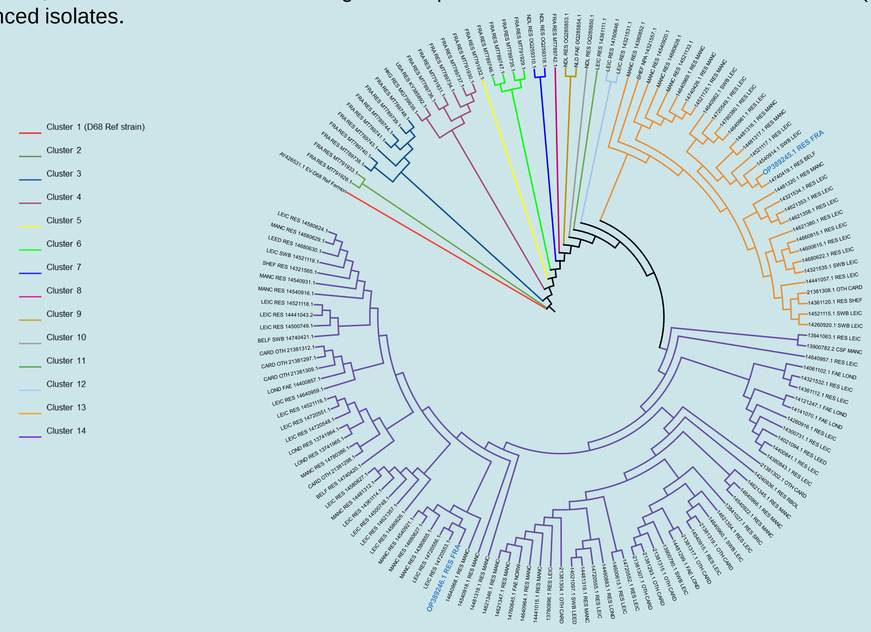


Figure 1. Neighbour-joining phylogeny tree showing EV-D68 Clade B3 UK sequenced samples, reference strain and international isolates (France: FRA, Hong Kong: HKG, United States: USA, Netherlands: NDL). Strain name indicates isolate ID/Accession number, sample type and sender.

Targeted sequence capture

Archived respiratory derived samples and cell culture isolates were tested. Complete NGS Sequence capture workflow was performed following *Manso et al*[9] protocol. Hybridization step using Agilent probe was tested using manufacturers specifications. Similar number of samples were tested on each experiment. Relative abundance of target virus in extracts and enriched libraries was determined by Real Time PCR. Cycle threshold (*Ct*) increase, and mass quantification results were similar in both experiments. Genomic mass quantification values of captured libraries using Agilent EV probe are relatively low. Despite having similar fragment size in the captured libraries, quality of traces does not display normal distribution. Fluorescent values are significantly higher in TWIST compared to Agilent.

Hybridisation method	Sample Type	No Samples	RT-qPCR Extract (Ct)	qPCR Library (Ct)	Δ Ct	Glomax (ng/uL)	Quantity in the Pool (500ng)	Qubit ng/ul	Average Fragment Size
Agilent	Clinical	35	28.9	15.7	13.1	24.7	100	6.2	559
TWIST	TC	32	16.7	7.2	9.4	24.2	350	17.8	570

Table 2. Sample specifications and QC results in EV NGS targeted capture experiments.

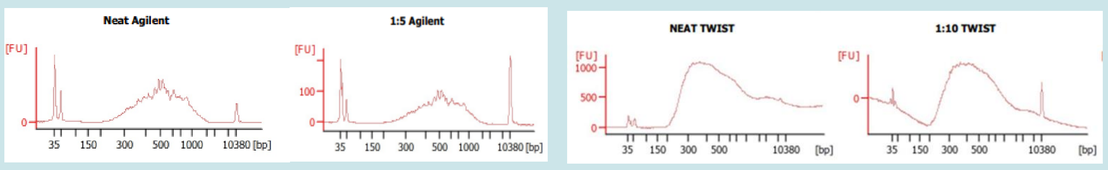


Figure 2. Captured libraries average fragment size distribution using Agilent and TWIST EV capture probes. Base pairs (bp X-axis) and Fluorescent Units (FU-Y axis).

Sequencing

Samples captured using Agilent and TWIST EV probes, were sequenced on MiSeq Illumina platform (Illumina V2 cartridge 300 cycles). Number of sequences retrieved vs samples submitted varied between the two methods Agilent (10/35) and TWIST (32/32). A higher number of reads per position was obtained using TWIST probe compared to Agilent.

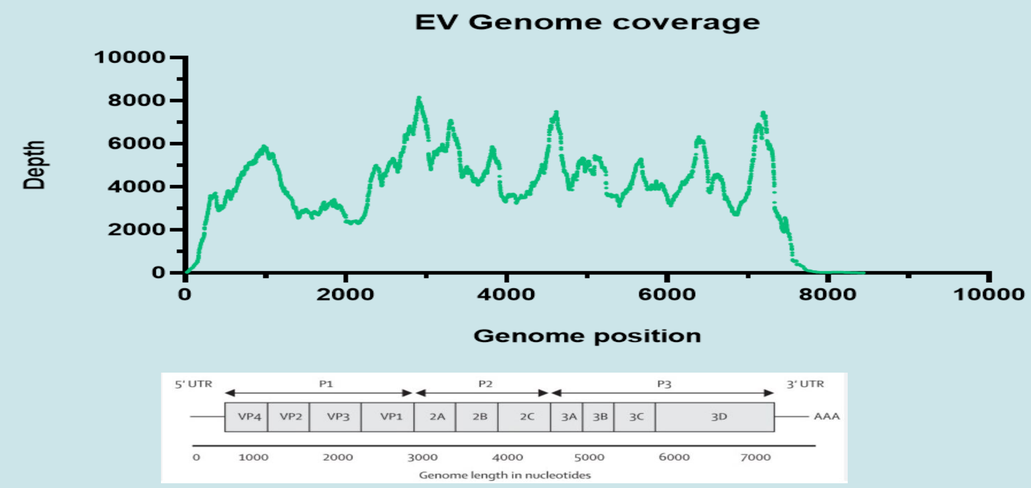


Figure 3. Average genome coverage (number reads/position) of samples captured using TWIST EV probe. EV genome schematic showing length in nucleotides aligned with genome coverage.

DISCUSSION and CONCLUSIONS

Sequencing results obtained using two different EV NGS approaches show potential applicability to our surveillance workflow. PCR Amplicon-based method has produced consistent results compared to existing Sanger sequencing. Despite NGS failures, the workflow was able to successfully sequence one unusual genotype (i.e.EV-C105) in a sample that could not be genotyped using our current testing workflow. EV/EV and EV/HRV co-infection cases were also sequenced using this approach. Phylogenetic analysis of EV-D68 (clade B3) UK sequenced samples, reference strains and international isolates shows 14 different clusters. Majority of UK samples are distributed in clusters 13 and 14. Two isolates from France showed close relatedness to UK respiratory samples. Through NGS, in depth investigation during outbreak and peak season could provide useful information about transmission and recombinant events within clades. Constant primer sequences update is crucial to be able to detect novel and unusual circulating strains.

A direct comparison between the EV custom designed probes from two manufacturers Agilent and TWIST could not be established. Different sample type, EV probe design and hybridization methods were used on separate experiments. Despite Agilent and TWIST capture methods did not meet the QC criteria stablished by *Manso et al* [9] protocol for Illumina sequencing, TWIST capture probe will be used for future experiments. This method is well established as a reference service and EVU would like to embed into this workflow to ease processing and downstream Bioinformatic analysis. Initial virus genome sequencing results using TWIST captured EV respiratory samples have shown satisfactory and reliable results compared to our Sanger sequencing. Additional optimisation and genomic data analysis is required to establish QC parameters that will help us achieve consistent genome coverage and reproducibility between runs. Different sample types, minimum genomic mass and sample concentration for Illumina sequencing will need to be adjusted specifically for EV testing.

Each method (PCR Amplicon based and Targeted Sequence capture) have shown potential applicability and satisfactory results however additional validation and optimisation work is required to establish the most suitable, cost-effective EV NGS method to support current surveillance strategy in the future. A robust workflow should be implemented to sequence relevant EV positive referred samples. Novel, unusual and pathogenic EV strains of clinical importance should be closely monitored by sequencing to respond to any public health threats.

ACKNOWLEDGEMENTS

EVU and AVU team members, David Bibby, Hodan Mohamed and Carmen Manso for the training, guidance and support during EV targeted sequence capture NGS experiments.

REFERENCES

1. *PALACIOS, G., et al. 2005. Enteroviruses as agents of emerging infectious diseases. Journal of NeuroVirology, 11*
2. *ISAACS, S. R., et al. 2018. Amplification and next generation sequencing of near full-length human enteroviruses for identification and characterisation from clinical samples. Sci Rep,* ***8,*** *11889.*
3. *MUSLIN, C.,et al. 2019. Recombination in Enteroviruses, a Multi-Step Modular Evolutionary Process. Viruses, 11.*
4. *VAN LEER-BUTER, C. et al. 2016. Newly Identified Enterovirus C Genotypes, Identified in the Netherlands through Routine Sequencing of All Enteroviruses Detected in Clinical Materials from 2008 to 2015. J Clin Microbiol, 54, 2306-14.*
5. *KAWADA, J. et al 2016. Identification of Viruses in Cases of Pediatric Acute Encephalitis and Encephalopathy Using Next-Generation Sequencing. Sci Rep,* ***6,*** *33452.*
6. *STELZER-BRAID, S. et al. 2020. Next generation sequencing of human enterovirus strains from an outbreak of enterovirus A71 shows applicability to outbreak investigations. J Clin Virol, 122, 104216.*
7. *HOULDCROFT, C. J., et al. 2017. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol,* ***15,*** *183-192.* ***Image 1.***
8. *MAJUMDAR M, et al. 2018. Detection by Direct Next Generation Sequencing Analysis of Emerging Enterovirus D68 and C109 Strains in an Environmental Sample from Scotland. Front Microbiol. Aug 21.*
9. *MANSO CF, et al. 2020. Technical Validation of a Hepatitis C Virus Whole Genome Sequencing Assay for Detection of Genotype and Antiviral Resistance in the Clinical Pathway. Front. Microbiol. 11:576572.*