Diagnosis and analysis of unexplained cases of childhood encephalitis in Australia using metagenomic next-generation sequencing

Ci-Xiu Li[1,2]†, Rebecca Burrell[2,3]†, Russell C Dale[4], Alison Kesson[2,3], Christopher C Blyth[5,6,7], Julia E Clark[8,9], Nigel Crawford[10], Cheryl A. Jones[2,3,11]†, Philip N. Britton[2,3,11]†, Edward C. Holmes[1,2]†* on behalf of the Australian Childhood Encephalitis study investigators.

[1]School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, NSW, Australia.

[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The University of Sydney, NSW, Australia.

[3]The Children's Hospital at Westmead, Westmead, NSW, Australia.

[4]Kids Neuroscience Centre, Children's Hospital at Westmead Clinical School, Faculty of Medicine, University of Sydney, NSW, Australia.

[5]Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute and School of Medicine, University of Western Australia, Nedlands, WA, Australia.

[6]Department of Infectious Diseases, Perth Children's Hospital, Nedlands, WA, Australia.

[7]Department of Microbiology, PathWest Laboratory Medicine WA, QEII Medical Centre, Nedlands, WA, Australia.

[8]Infection Management, Queensland Children's Hospital, Brisbane, QLD, Australia.

[9]School of Clinical Medicine, Childrens Health Queensland Clinical Unit, University of Queensland, QLD, Australia.

[10]Murdoch Children's Research Institute, Royal Children's Hospital Flemington Road, Parkville, VIC 3052 Australia.

[11]Kids Research, Sydney Children's Hospitals Network (Westmead), Westmead, NSW, Australia,

*Correspondence to:
Prof. Edward C. Holmes
Marie Bashir Institute for Infectious Diseases & Biosecurity,
School of Life & Environmental Sciences and School of Medical Sciences
The University of Sydney | Sydney | NSW | 2006 | Australia
T +61 2 9351 5591; E edward.holmes@sydney.edu.au

†Contributed equally to the paper.

Abstract

Encephalitis is most often caused by a variety of infectious agents, the identity of which is commonly determined through diagnostic tests utilising cerebrospinal fluid (CSF). Immune-mediated disorders are also a differential in encephalitis cases. We investigated the clinical characteristics and potential aetiological agents of unexplained encephalitis through metagenomic next-generation sequencing of residual clinical samples of multiple tissue types and independent clinical review. A total of 43 specimens, from both sterile and non-sterile sites, were collected from 18 encephalitis cases with no cause identified by the Australian Childhood Encephalitis study. Samples were subjected to total RNA sequencing to determine the presence and abundance of potential pathogens, to reveal mixed infections, pathogen genotypes, and epidemiological origins, and to describe the possible aetiologies of unexplained encephalitis. From this, we identified five RNA and two DNA viruses associated with human infection from both non-sterile (nasopharyngeal aspirates, nose/throat swabs, urine, stool rectal swab) and sterile (cerebrospinal fluid, blood) sites. These comprised two human rhinoviruses, two human seasonal coronaviruses, two polyomaviruses and one picobirnavirus. With the exception of picobirnavirus all have been previously associated with respiratory disease. Human rhinovirus and seasonal coronaviruses may be responsible for five of the encephalitis cases reported here. Immune-mediated encephalitis was considered clinically likely in six cases and RNA sequencing did not identify a possible pathogen in these cases. The aetiology remained unknown in nine cases. Our study emphasises the importance of respiratory viruses in the aetiology of unexplained child encephalitis and

suggests that the routine inclusion of non-CNS sampling in encephalitis clinical guidelines/protocols could improve the diagnostic yield.

Author Summary

Encephalitis is caused by both infectious agents and auto-immune disorders. However, the aetiological agents, including viruses, remain unknown in around half the cases of encephalitis in many cohorts. Importantly, diagnostic tests are usually based on the analysis of cerebrospinal fluid which may limit their utility. We used a combination of meta-transcriptomic sequencing and independent clinical review to identify the potential causative pathogens in cases of unexplained childhood encephalitis. Accordingly, we identified seven viruses associated with both sterile and non-sterile sampling sites. Human rhinovirus and seasonal coronaviruses were considered as most likely responsible for five of the 18 encephalitis cases studied, while immune-mediated encephalitis was considered the cause in six cases, and we were unable to determine the aetiology in nine cases. Overall, we demonstrate the role of respiratory viruses as a cause of unexplained encephalitis and that sampling sites other than cerebrospinal fluid is of diagnostic value.

Keywords. Unexplained encephalitis, meta-transcriptomics, non-sterile sites, virus, pathogen discovery

Introduction

Encephalitis, defined as inflammation of the brain tissue, is caused by a broad range of infectious agents, including bacteria, viruses, fungi and parasites, as well as a number of auto-immune disorders [1, 2]. However, in 40% to 60% of cases in many cohorts the aetiological agents remain unknown [3, 4]. Investigation is challenging because clinical presentation may be atypical and the differential diagnosis may include unusual and opportunistic pathogens – especially if the patient is immunosuppressed - or a non-infective cause. The conventional means of infectious encephalitis diagnosis are specific PCR assays on central nervous system samples including cerebrospinal fluid (CSF) and brain biopsy, as well as immunohistochemical and serological assays [4]. Failure to detect a causative agent by such routine laboratory diagnostics may reflect a lack of diagnostic tests for rare, novel or divergent pathogens, limited volume of CNS samples, and overlapping clinical presentation caused by infectious and non-infectious processes [5].

The incidence of all-cause childhood encephalitis (including infection-associated encephalopathy) is estimated to be between 3.8 and 5.0 per 100,000 population aged ≤14 years [6]. The leading causes are picornaviruses (enteroviruses and parechoviruses), herpes simplex viruses 1 and 2, influenza (infection associated encephalopathy), bacterial meningo-encephalitis and immune-mediated causes such as acute disseminated encephalo-myelitis (ADEM) and anti-N-methyl-D-aspartate receptor (NMDAr) encephalitis [7]. The burden of disease is considerable, with 49% admitted to ICU, a case-fatality rate of 5%, and 27% of cases showing neurodisability at discharge from hospital [7].

Metagenomic next-generation sequencing (mNGS) has successfully identified a broad range of infectious agents in a range of clinical syndromes [8-10] and is gradually being established as a powerful and reliable diagnostic platform [8]. Indeed, mNGS has identified an increasing number of novel or unexpected pathogens associated with encephalitis [11]. Total RNA sequencing – meta-transcriptomics – may be especially powerful as it provides a simple way to characterise all the actively transcribing microbes in a sample and estimate their abundance [12-14]. Not only does total RNA sequencing identify the RNA viruses present in a sample, but also those DNA viruses, bacteria, parasites and fungi that are actively transcribing RNA [15].

The Australian Childhood Encephalitis (ACE) study has comprehensively identified, collected data, and reviewed cases of this severe syndrome nationally through active sentinel hospital surveillance since 2013, and requested banking of salvaged laboratory biospecimens from cases [7]. Herein, we describe the use of total RNA sequencing on samples of differing tissue types obtained from 18 cases of childhood encephalitis categorised as having unknown cause. Notably, several common respiratory viruses, including human seasonal coronaviruses and rhinoviruses, were identified in samples outside of the CNS and potentially caused the disease process in these cases.

Results

Clinical evaluation

A total of 18 cases, representing patients from four Australian states (Figure 1A), were chosen for study. These cases had previously been reviewed and confirmed to have encephalitis using the Brighton criteria and International Encephalitis Consortium (IEC) definitions [7], but with unknown causality (Table 1). The sex ratio of patients was 1.57:1 male to female, with ages ranging from neonates to twelve years (average, 4.5 years) such that only pre-adolescent children were considered (Table 1).

Of the 18 unknown encephalitis patients, abnormal neuroimaging (MRI/CT)) was found in 14/18, and EEG findings consistent with encephalitis were observed in 10/18 (Table 1). Most patients had fever (13/18) and lethargy/drowsiness (11/18). Irritability was recorded in 6/18 patients, ataxia/unusual behaviour in 7/18, decreased level of consciousness (LOC) in 5/18, confusion in 4/18, headache in 6/18, vomiting in 5/18, poor feeding in 4/18 and status epilepticus in 2/18 (Table 1). In addition, diarrhoea, cough, rash, convergent squint, coryza, eye deviation, dysarthria, facial droop, urinary retention and hyperreflexia were observed in one or two cases (Table 1). One patient was reported with respiratory illness in the two weeks prior to admission. Three patients were transferred to an intensive care unit (ICU) within 24 hours of admission. One patient with severe combined immunodeficiency (SCID) died. Two patients showed clinical improvement with corticosteroid treatment.

Table 1. Demographic and clinical characteristics of the 18 patients studied here.

				CNS Inflammation					**Reduced Responsiveness**						
Case No.	Sex/Age	Respiratory symptoms	Gastro-intestinal symptoms	Fever	CSF pleocytosis	CSF WCC	MRI changes	EEG changes	Absent environment	Absent eye	Decreased arousability	Seizure LOC	*Focal neurological signs	Brighton criteria	International encephalitis consortium criteria
1	F/3.2mo		Y	Y	Y	11	Y	Y	Y	Y	Y	Y	Y	Level 2	Probable
2	M/7.1yo		Y	Y	Y	501		N/A	Y				Y	Level 2	Possible
3[#]	M/1.1yo		Y	Y		3	Y		Y	Y	Y	Y	Y	Level 2	Probable
4	M/7.4yo		Y	Y		2	N/A	Y	Y		Y			Level 2	Possible
5	M/9.5yo			Y		0		Y	Y	Y	Y		Y	Level 2	Possible
6	M/2.6yo			Y		1		N/A	Y	Y	Y		Y	Level 3	Possible
7	F/1.1yo		Y		Y	21	Y	Y	Y	Y	Y	Y	Y	Level 2	Probable
8[†]	M/8.5yo			Y	Y	29	Y	N/A	Y	Y	Y		Y	Level 2	Probable
9	F/3.8wks			Y	Y	65	Y	Y	Y		Y		Y	Level 2	Possible
10[†]	M/9.9yo		Y		Y	29	Y	N/A			Y			Level 2	Probable
11	M/3mo				Y	72	Y	N/A	N/A	Y			Y	Level 3	Probable
12[†]	M/3.6yo		Y			0	Y	N/A					Y	Level 3	Probable
13	F/3.4yo	Y	Y	Y	Y	21	Y		Y	Y	Y	Y	Y	Level 2	Probable
14	M/12.6yo	Y	Y		Y	66		N/A	Y	Y	Y	Y	Y	Level 2	Possible
15	M/8.1yo			Y		6	Y	Y	Y	Y	Y		Y	Level 2	Probable
16	F/8.6mo				Y	31	Y	Y	Y	Y	Y	Y	Y	Level 2	Probable
17[##]	F/3.3yo			Y	N/A	N/A	N/A				Y	Y	Y	Level 3	Probable
18	F/7.8yo			Y	Y	7	Y	Y	Y	Y	Y	Y	Y	Level 2	Probable

Y = Yes, N/A = Not Done. All cases had an initial expert panel diagnosis of Unknown Encephalitis. * Y is indicative of cases with one or more of the following: focal seizure, cranial nerve abnormality, visual impairment, motor weakness, altered deep tendon reflexes, cerebellar dysfunction, abnormal movement, or any other focal neurological signs. Respiratory symptoms included cough/coryza and increased work of breath requiring mechanical ventilation for both cases, gastrointestinal symptoms include diarrhoea and/or vomiting.

[#]This patient was with severe combined immunodeficiency (SCID) and the outcome was death.

[##]This patient was reported with respiratory illness in the previous 2 weeks before admission.

†These patients were transferred to intensive care unit (ICU) within 24 hours of admission.

Diagnostic testing for a range of viral agents by PCR on CSF, blood/plasma, sputum and stool samples were performed in the local hospitals. All were negative with the exception of one positive detection of rotavirus from a stool sample, and one detection each of rhinovirus and coronavirus (OC43) in respiratory samples (Table S1). None were considered significant to the clinical presentation. Similarly, serological tests were negative (not consistent with acute infection) for all cases (Table S1). Some cases were also tested for antibodies against ganglioside, muscle specific tyrosine kinase (MUSK), acetylcholine receptor (AChR), N-Methyl-D-Aspartate receptor (NMDAR), voltage-gated potassium channel (VGKC) and neuromyelitis optica (NMO). All were negative (Table S1).

A multidisciplinary expert team (PNB, RD, AK, CAJ) re-reviewed clinical presentation, available diagnostic testing using published criteria for assigning causation in encephalitis and criteria for clinically diagnosing autoimmune encephalitis [16, 17]. Following this review, nine were considered to likely have infectious causes, six immune-mediated causes, and three could not be further classified (Table 2).

Table 2. Possible pathogen identification by metagenomic next-generation sequencing (mNGS) and final diagnosis.

Case No.	Initial Diagnosis	Independent Clinical Review	Virus	Specimen	Abundance (RPM)	mNGS conclusion	Final Diagnosis
1	Unknown encephalitis	Possible infectious	Human Rhinovirus A9	NPA	8399	Possible	Possible infectious encephalitis - Human Rhinovirus A9
2	Unknown encephalitis	Unclear cause					Encephalitis – unknown cause
3	Unknown encephalitis	Possible infectious	Human coronavirus OC43 WU Polyomavirus	NPA NPA	523673 1548	Possible/ indeterminate	Possible infectious encephalitis - Human coronavirus OC43
4	Unknown encephalitis	Possible infectious					Encephalitis – unknown cause
5	Unknown encephalitis	Possible immune mediated					Possible immune mediated encephalitis
6	Unknown encephalitis	Unclear cause					Encephalitis – unknown cause
7	Unknown encephalitis	Possible infectious					Encephalitis – unknown cause
8	Unknown encephalitis	Possible immune mediated					Possible immune mediated encephalitis
9	Unknown encephalitis	Possible infectious	Human coronavirus OC43	Blood	82	Possible	Possible infectious encephalitis - Human coronavirus OC43
10	Unknown encephalitis	Possible immune mediated	JC polyomavirus	Urine	80	Unlikely	Possible immune mediated encephalitis
11	Unknown encephalitis	Possible infectious	Human Rhinovirus B52 Human Rhinovirus B52	NPA Stool	454711 87	Possible/ indeterminate	Possible infectious encephalitis - Human Rhinovirus B52

12	Unknown encephalitis	Possible immune mediated					Possible immune mediated encephalitis
13	Unknown encephalitis	Unclear cause					Encephalitis – unknown cause
14	Unknown encephalitis	Possible infectious					Encephalitis – unknown cause
15	Unknown encephalitis	Possible immune mediated	Human coronavirus OC43	NPA	170	Unlikely	Possible immune mediated encephalitis
16	Unknown encephalitis	Possible infectious					Encephalitis – unknown cause
17	Unknown encephalitis	Possible infectious	Human coronavirus HKU1 Human coronavirus HKU1 Human picobirnavirus	Rectal Swab Respiratory Stool	6060 2939 2124	Possible Possible Unlikely	Possible infectious encephalitis - Human coronavirus HKU1
18	Unknown encephalitis	Possible immune mediated					Possible immune mediated encephalitis

16

Meta-transcriptomic exploration of potential infectious microbes

A total of 43 specimens, comprising CSF, blood, serum, urine, stool, rectal swabs (rectal SW), nasal swabs (respiratory SW), throat swabs (throat SW), nasopharyngeal aspirates (NPA) and endotracheal aspirates (ETA), were utilised in metagenomic testing (Figure 1B). All 43 samples were individually examined using meta-transcriptomics, generating 2.77 billion raw paired-end reads in total (between 12.2 and 81.0 million reads per sample) (Table S2). For each library, 48.46 to 85.79% of the reads were retained after removal of low complexity and redundant reads (Figure S1, Table S2), and 1.10 to 80.19% of these reads were subsequently retained after removal of human reads (Figure S1, Table S2). The resultant sequence reads and assembled contigs were annotated against NCBI reference databases, revealing a number of microbes, including potential pathogens. These are described in more detail below.

Detection of viral sequences in clinical samples of encephalitis

Blastx comparisons against the nr database identified at least 7 virus species (5 RNA viruses and 2 DNA viruses) associated with human infection (Figure 2). The total virus positive rate of specimens was 23% (10/43), while the total virus positive rate of patients was 39% (7/18). All the viruses identified are known to be associated with overt disease, with the exception of a picobirnavirus present in one stool sample that may represent a bacteriophage [18]. Among these viruses, human coronavirus OC43 (HCoV-OC43) was present in three cases, while human coronavirus HKU1 (HCoV-HKU1), human rhinovirus A (HRVA), human rhinovirus

B (HRVB), JC polyomavirus (JCPyV) and WU polyomavirus (WUPyV) were detected in one case each. Finally, HCoV-HKU1 was detected in both respiratory swab and rectal swab of case 17, while HRVB was detected in both NPA and stool of case 11 (Figure 2). CSF was available for testing from 9/18 cases and we observed a relative absence of possible virus sequence detections amongst these CSF samples.

Virus expression levels were quantified by estimating their relative abundance (RPM, reads per million). Accordingly, the highest virus abundance was 523,673 RPM (HCoV-OC43, case 03 in NPA), followed by 454,710 RPM (HRV-B52, case 11 in NPA) (Figure 2, Table 2). Viruses with greater than 1,000 RPM (> 0.1% of total reads) included HCoV-HKU1 (case 17, in rectal swab and respiratory swab), HRV-A9 (case 01, in NPA), HPBV (case 17, in stool) and WU polyomavirus (case 03, in NPA) (Figure 2, Table 2). The abundance level of HCoV OC43-was greater than 100 RPM in case 15 (throat swab) and lower than 100 RPM in case 09 (blood) (Figure 2, Table 2).

To identify specific viral genotypes/lineages and their epidemiological origins, phylogenetic analyses were performed using the resulting complete or partial virus genomes (Table S3). This revealed that all HCoV-HKU1, HCoV-OC43, WUPyV and JCPyV sequences determined here were >99.5% identical to the most closely related sequences available on publicly available data bases (Table S3), belonging to genotypes B, G, 3b and 2B, respectively (Figure3A, 3B, 3E and 3F). Two sequences of human rhinovirus B52 - HRV-B52/11-4818/VIC/AU/2019 and HRV-B52/11-4817/VIC/AU/2019 - detected in an NPA and stool specimen from case 11 shared 100% sequence identity with each other yet

were relatively distant to known viruses (92.8% nt identity) (Figure 3C, Table S3), such that it likely represents a new genotype of this virus. The human rhinovirus A9 identified here, HRV-A9/01-14618/NSW/AU/2019, exhibited 96.5% sequence identity to the most closely publicly available sequence (Figure 3D, Table S3).

One picobirnavirus was identified from one stool sample (case 17), and the near complete sequences of two segments were obtained (Figure 4B). Segment 1 contains two open reading frames, with ORF2 encoding the viral capsid protein and ORF1 encoding a hypothetical protein with unknown function, while segment 2 contains an ORF encoding the viral RNA-dependent RNA polymerase (RdRp) gene (Figure 4B). The conserved motif of the ribosomal binding site (RBS) nucleotide sequence (AGGAGG) is present upstream of the ORF2 of segment 1 and RdRp of segment 2 (Figure 4B). Two copies of the conserved ExxRxxNxxxE aa motif are identifiable un ORF2 of segment 1(Figure 4B). Phylogenetic analysis based on RdRp showed that HPBV/17-26618/WA/AU/2019 is related to primate PBVs within genogroup 1 (Figure 4A), and the RdRp shared 72.4% amino acid sequence identity with the corresponding protein of macaque PBV (AVD54068) (Table S3).

Characterization of infecting bacteria

We characterised microbial taxonomic clades belonging to seven phyla and 74 species (Figure 5A). The dominant bacterial families identified were *Enterobacteriaceae* (63.64%, estimated using MetaPhlAn2), followed by *Veillonellacae* (10.18%), *Enterococcaceae* (6.07%), *Pseudomonadaceae* (5.89%), *Clostridiales family xi incertae* (5.30%),

Streptococcaceae (2.75%), and *Bacteroidaceae* (1.45%) (Figure 5). As *Escherichia coli*, unclassified *Escherichia* and unclassified *Veillonella* were observed in almost every library (Figure 6), they were considered unlikely to be infectious agents. *Enterococcus faecalis*, as part of intestinal microbiota and an opportunistic pathogen, was detected in two sterile sites: blood (case 06) and CSF (case 14), although with extremely low abundance (1 and 14 RPM; 53 and 549 reads, respectively) (Figure 6). In other cases, the bacteria identified were known human colonisers or of indeterminate pathogenic status and were observed in non-sterile sites making them unlikely pathogens (Figure 6). No bacterial sequences detected were considered potentially pathogenic.

Correlation between results from clinical evaluation and mNGS

Among those cases with likely infectious causes following clinical evaluation, seven harboured viruses at non-sterile sites (NPA, urine, stool, throat swab, rectal swab and respiratory swab) and one in a sterile site (blood) (Table 2). Case 03, a 1-year-old boy with severe combined immunodeficiency (SCID), had the highest abundance of human coronavirus OC43 (523,673 RPM) and medium abundance of Wu polyomavirus (1548RPM) in NPA, indicating a possible synergistic interaction between these two viruses. Although human coronavirus OC43 has not usually been considered a pathogen capable of causing encephalitis, it's very high abundance and clinically likely infectious cause makes it a possible pathogen in this case. Human rhinoviruses (HRV-A9 and HRV-B52), that are commonly associated with respiratory diseases, were detected in NPA of case 01 and 11, with

high to extremely high abundance (8399-454,711 RPM) (Figure 2, Table 2), respectively. These cases were clinically considered likely infectious, suggesting that the two rhinoviruses identified may be the causative pathogens. JC polyomavirus and human coronavirus OC43 were detected in cases 10 and 15 at relatively low abundance (80 and 170 RPM, respectively). Although both these viruses have previously been associated with encephalitis [19, 20], their low abundance and potential constitutive expression (JCV) make them less likely to be pathogens in these two cases. Importantly, although HCoV-OC43 was not highly abundant in case 09, it was detected in blood (sterile site), which indicates likely pathogenicity in this case (Figure 2, Table 2). In addition, a high abundance of human coronavirus HKU1 were detected in both respiratory swab and rectal swab of case 17, who was reported with a respiratory illness in the before admission, suggesting potential pathogenicity in this case.

Among those cases with likely immune-mediated cause following clinical evaluation, viruses were detected in only two: case 10 (JCV) and case 15 (HCoV-OC43). However, in both cases the viruses were at very low abundance and present at non-sterile sites and so were considered unlikely pathogens. The relative lack of pathogen sequences among those cases with likely immune-mediated causes was notable.

Discussion

Identification of a causative agent in infectious encephalitis is most validly achieved by obtaining a specimen directly at the site of inflammation, such as CSF or brain biopsy tissue,

122 although the latter is seen as a last resort as sampling comes with associated risks [4].
123 Therefore, in most cases, CSF is considered the best surrogate specimen for the assessment of
124 neurological disorders [21]. Unlike routine laboratory methods that are limited to the
125 detection of known or related agents, mNGS offers unprecedented opportunities to detect a
126 broad range of microbes including their relative abundance and has recently been employed
127 successfully in detecting a diverse array of potential pathogens not previously associated with
128 CNS disease [22], including astroviruses [3, 23, 24] and arenavirus [25] in cases of
129 encephalitis in immunocompromised and transplant-associated patients. Also of note was that
130 a pegivirus, although considered non-pathogenic, was detected using mNGS in a patient with
131 encephalitis [26], and this technique has recently detected parasitic worms, bacteria, and
132 fungi in CSF specimens from patients with CNS infection of unknown aetiology [5, 11, 27-
133 29].
134 Importantly, in instances where a CSF sample is unavailable or mNGS testing of CSF
135 fails to detect a pathogen, other sites and samples including respiratory tract (nasal and throat
136 swabs, NPA), gastrointestinal tract (faeces, rectal swab), urine or blood may provide an
137 additional indirect opportunity to test for the presence of potential pathogens that might have
138 transferred in the brain [4]. Further, for some well-established CNS pathogens, CSF testing
139 for viral nucleic acid shows low sensitivity. For example, flaviviruses and enterovirus A71
140 have been detected in throat swabs, stool or urine from patients with clinical encephalitis, yet
141 were absent in CSF or plasma/serum collected concurrently [30-32]. Among the cases with
142 CSF available for testing (9/18 cases) here, none showed the presence of a possible viral

143 pathogen. In contrast, four respiratory viruses – human rhinovirus A9, human rhinovirus B52,
144 human coronavirus OC43 and human coronavirus HKU1 – were detected in respiratory and
145 gastrointestinal samples (NPA, respiratory swab and rectal swab) from four cases, all of
146 which were in high abundance (>1,000 RPM), indicating that they are experiencing active
147 replication in these patients. Further, these viruses were detected in patients considered likely
148 to have an infectious aetiology in an independent clinical evaluation and so are considered
149 potential pathogens.

150 Human coronaviruses are respiratory viruses infrequently reported with neuroinvasive
151 properties in both mice and humans [20, 33, 34]. One study suggested that coronavirus
152 infection in the central nervous system is as common as in respiratory tract, although with
153 distinct features [35]. HCoV-OC43 has been identified in CSF and brain tissue in isolated
154 cases of children with acute encephalomyelitis or ADEM [20, 36]. In mice, HCoV-OC43 has
155 been shown to directly invade the CNS via the olfactory route, indicating that intranasal
156 infection may play a role in propagating viruses [20, 37]. Our detection of HCoV in NPA and
157 respiratory samples in high abundance suggests they are possible causative agents, and that
158 the olfactory pathway needs further consideration as a route, albeit infrequent, for
159 neuroinvasion in humans [33, 38]. The potential role of human coronaviruses in acute
160 neurological disease has been further highlighted by the growing evidence of neurological
161 disease of multiple phenotypes as infrequent clinical presentation of SARS-CoV-2 infection
162 [39-42].

Rhinoviruses (genus *Enterovirus*) are the cause of various respiratory illnesses [38]. Unlike some members within the same genus (polioviruses, echovirus, coxsackieviruses, enteroviruses) often implicated in encephalitis, rhinoviruses are discounted as potential causes of encephalitis in consensus criteria [43]. However, although rhinoviral CNS infection is rare, there are reports of human rhinovirus A and B in respiratory specimens associated with acute neurological disease including acute cerebellitis and meningitis, respectively, and one rhinovirus (unknown species) was detected in CSF in a case of sepsis-like illness [44-46]. Further, enterovirus D68 is acknowledged as an emerging neurotropic enterovirus with predominant replication in the respiratory tract [47, 48], suggesting that members of the *Enterovirus* genus may have neuropathic potential more broadly. Although the two species of rhinovirus detected in this study were identified in in nasopharyngeal aspirates, on the basis of clinical evaluation and mNGS analysis they are considered possible causative agents. Targeted assays of the CSF or antibody tests could be beneficial for pathogen determination in future cases.

Another possible explanation for the pathogenicity of these respiratory viruses at non-CSF sites is that they represent para-infectious encephalitis resulting from indirect CNS pathogenicity. A variety of severe encephalopathy/encephalitis syndromes have been described in association with viruses such as influenza, adenovirus and human herpes virus-6, and it is possible that other respiratory viruses could produce disease via these mechanisms [38, 49-53]. It is clear that larger studies are needed to better understand the role of non-CNS samples in the detection of candidate pathogens amongst cases of encephalitis with otherwise

184 unknown cause. These studies should involve the testing of non-CNS samples in combination
185 with molecular and serological CSF and serum investigations.

186 The frequency and range of bacterial reads identified in almost all specimens analysed
187 highlights the challenge of applying and interpreting mNGS particularly to non-sterile
188 samples. Most reads corresponded to known human colonisers and/or organisms with
189 questionable pathogenicity. An exception was *E. coli* that was detected in sterile site samples
190 from several cases, but discounted given that reads were present in all samples analysed.
191 Similarly, *E. faecalis* was present in two sterile site samples in two cases, but at very low
192 abundance in sterile site specimens in two cases, and we were unable to confirm these
193 infections by PCR. Notably, all the cases studied received empiric antibiotic treatment with a
194 third-generation cephalosporin to which enterococci are inherently resistant. Concurrent or
195 preceding antibiotic usage is an important consideration in mNGS studies where treatment
196 may bias detection to resistant bacterial species [9].

197 An additional insight from this study is the apparent utility of clinical definitions of
198 immune-mediated encephalitis. Among the unknown encephalitis cases under enhanced
199 evaluation here, six were considered likely to be immune-mediated. Despite a similar number
200 and spectrum of samples available for mNGS, we found no possible pathogens using mNGS
201 among these six cases compared with possible pathogens in five of nine cases clinically
202 considered likely to be infectious by an independent clinical review panel. The frequency of
203 likely immune-mediated encephalitis in this case series of encephalitis of unknown cause
204 emphasises the need for further research to identify diagnostic biomarkers.

205 It is not possible to definitively determine that the viruses identified using mNGS here
206 were the causes of the encephalitis in each individual case studied. However, the combination
207 of enhanced clinical evaluation with mNGS appears a fruitful way forward in evaluation of
208 cases of unknown encephalitis [9].
209

210 # Methods

211 ## Ethics statement

212 This study was performed under the ethical approval of the Sydney Children's Hospitals
213 Network Human Research Ethics Committee (HREC/13/SCHN/191).
214

215 ## Enrolment criteria and case definition

216 The cases selected in this study all met the case definition of encephalitis [54]. This included
217 patients aged <= 14 years admitted to hospital with encephalopathy, altered consciousness
218 that persisted for longer than 24 hours, including lethargy, irritability, or a change in
219 personality and behaviour, and at least two of the following criteria: (i) fever or history of
220 fever (≥38°C) during the presenting illness, (ii) seizures and/or focal neurological, (iii)
221 abnormal findings on electroencephalographic (EEG) consistent with encephalitis, and (iv)
222 abnormal results of neuroimaging (CT or MRI) suggestive of encephalitis (Table 1). In
223 addition, each of the study subjects tested negative with routine diagnostic protocols.
224 All samples were collected from diagnostic laboratories between March 2014 and
225 December 2016. All specimens were originally collected by trained clinicians using sterile

swabs or aseptic technique and were immediately sent to local laboratories where diagnostic testing was undertaken. Following diagnostic testing, residual samples were stored at -20°C. Following collection for research purposes, informed consent was obtained from parents/guardians for research testing and specimens were then placed into -80°C storage locally and transferred in batches on dry ice to a central specimen repository where they were stored in a -80°C freezer.

RNA extraction, library construction and sequencing

Samples were subjected to RNA extraction using the RNeasy plus universal kit (QIAGEN, Chadstone Centre, Victoria, Australia). The concentration and quality of final extractions were examined using a NanoDrop spectrophotometer (ThermoFisher Scientific, USA). The Trio RNA-Seq kit (NuGEN Technologies, USA) was used for library preparation in all cases as it targets low concentration RNA samples (as low as 500 pg) [55]. Paired-end (150 bp) sequencing of these RNA library was performed using the Illumina NovaSeq platform. All library preparation and sequencing steps were performed by Australian Genome Research Facility (AGRF), Melbourne.

Microbial characterization

For each library/case, quality control was performed and adaptor sequences and low-quality/low-complex reads were removed using BBmap (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/) and CD-HIT-DUP [56]. Human reads were

247 removed by mapping to the human genome. Non-human reads were then compared against
248 reference virus database downloaded from GenBank and the NCBI non-redundant protein
249 database using Blastn and Diamond [57] blastx, respectively, with e-value thresholds of $1 \times$
250 10^{-10} and 1×10^{-4}. Blast hits were then annotated by taxonomy. Reads were *de novo* assembled
251 using Megahit [58, 59] from the virus-positive library in which they were identified based on
252 the blast procedure described above. In cases with low genome coverage, reads were directly
253 mapped to the sequence of a close relative, and a consensus genome was obtained from the
254 mapped reads. To exclude contamination due to index hopping, for each virus only those
255 present at >0.1% of the highest viral abundance were considered true positives. Sequence
256 specific primers were designed, and RT-PCR and Sanger sequencing were performed to
257 verify and confirm the presence of viruses with highly similar sequences that appeared in
258 multiple libraries.

259 Virus abundance were estimated as Reads Per Million (RPM) [55], calculated by using
260 the relation "mapped reads / total reads * one million". For picobirnavirus, open reading
261 frames were predicted using orffinder (https://www.ncbi.nlm.nih.gov/orffinder/) and gene
262 annotated against the conserved domain database (CDD) [60].

263 The resulting complete or partial virus genomes (Table S3), as well as related
264 background viruses from GenBank, were aligned using MAFFT version 7 [61] and subjected
265 to phylogenetic analysis using the maximum likelihood method available in PhyML 3.0 [62]
266 employing the General Time Reversible (GTR) model of nucleotide substitution with a
267 gamma distribution of among-site rate variation and 1000 bootstrap replicates.

Bacterial taxonomic profiling was initially performed using MetaPhlAn2 [63] with default parameters, mapping the non-human sequence reads to a set of ~1 million unique clade-specific marker genes from ~13,500 bacterial and archaeal species.

Data availability

All non-human sequence reads generated here have been deposited on the NCBI Sequence Read Achieve (SRA; BioProject accession PRJNA633210).

Clinical evaluation

All 18 cases had previously undergone clinical review as part of the Australian Childhood Encephalitis study and been determined to have an unknown cause [7]. A clinical evaluation panel was convened comprising a clinical paediatric neurologist (RD), paediatric infectious disease physician (CJ), clinical epidemiologist (PB) and clinical microbiologist (AK) who re-evaluated each case in terms of clinical presentation, available diagnostic testing and response to empiric therapies using published criteria for assigning causation in encephalitis [43] and clinically diagnosing autoimmune encephalitis [16]. This panel was blinded to the mNGS results.

Logic model for determining pathogenic potential of mNGS detections

We applied a logic model (Figure S2) to determine the likely significance of mNGS results based upon the nature of the sample from which mNGS detection occurred, the organism

289 abundance and pathogenicity, and accounting for the enhanced clinical evaluation findings to

290 arrive at a final clinical diagnosis.

Acknowledgments

We acknowledge The University of Sydney HPC service (Artemis) for providing high-performance computing resources that have contributed to the research reported in this paper.

Author Contributions

Conceptualization: Cheryl A. Jones, Philip N. Britton, Edward C. Holmes.

Data curation: Ci-Xiu Li, Rebecca Burrell.

Formal analysis: Ci-Xiu Li, Rebecca Burrell, Philip N. Britton, Edward C. Holmes.

Funding acquisition: Cheryl A. Jones, Philip N. Britton, Edward C. Holmes.

Investigation: Russell C. Dale, Alison Kesson, Christopher C Blyth, Julia E Clark, Nigel Crawford, Cheryl A. Jones

Methodology: Ci-Xiu Li, Rebecca Burrell

Resources: Rebecca Burrell, Alison Kesson, Christopher Blyth, Julia Clark, Nigel Crawford.

Writing - original draft: Ci-Xiu Li.

Writing - review & editing: Ci-Xiu Li, Rebecca Burrell, Russell C. Dale, Christopher C Blyth Julia E Clark, Cheryl A. Jones, Philip N. Britton, Edward C. Holmes.

References

1. Granerod J, Tam CC, Crowcroft NS, Davies NWS, Borchert M, Thomas SL. Challenge of the unknown: A systematic review of acute encephalitis in non-outbreak situations. Neurology. 2010; 75: 924-932. doi: 10.1212/WNL.0b013e3181f11d65. PMID: 20820004.

2. Singh TD, Fugate JE, Rabinstein AA. The spectrum of acute encephalitis Causes, management, and predictors of outcome. Neurology. 2015; 84: 359-366. doi: 10.1212/Wnl.0000000000001190. PMID: 25540320.

3. Brown JR, Morfopoulou S, Hubb J, Emmett WA, Ip W, Shah D, et al. Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. Clin Infect Dis. 2015; 60: 881-888. doi: 10.1093/cid/ciu940. PMID: 25572899.

4. Kennedy PGE, Quan PL, Lipkin WI. Viral encephalitis of unknown cause: current perspective and recent advances. Viruses. 2017; 9: 138. doi: 10.3390/v9060138. PMID: 28587310.

5. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu GX, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med. 2014; 370: 2408-2417. doi: 10.1056/NEJMoa1401268. PMID: 24896819.

6. Britton PN, Khoury L, Booy R, Wood N, Jones CA. Encephalitis in Australian children: contemporary trends in hospitalisation. Arch Dis Child. 2016; 101: 51-56. doi: 10.1136/archdischild-2015-308468. PMID: 26475868.

7. Britton PN, Dale RC, Blyth CC, Clark JE, Crawford N, Marshall H, et al. Causes and clinical features of childhood encephalitis: a multicenter, prospective cohort study. Clin Infect Dis. 2020; 70: 2517-2526. doi: 10.1093/cid/ciz685. PMID: 31549170.

8. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. Annu Rev Pathol. 2019; 14: 319-338. doi: 10.1146/annurev-pathmechdis-012418-012751. PMID: 30355154.

9. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med. 2019; 380: 2327-2340. doi: 10.1056/NEJMoa1803396. PMID: 31189036.

10. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019; 20: 341-355. doi: 10.1038/s41576-019-0113-7. PMID: 30918369.

11. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, et al. Chronic meningitis investigated via metagenomic next-generation sequencing. JAMA Neurol. 2018; 75: 947-955. doi: 10.1001/jamaneurol.2018.0463. PMID: 29710329.

12. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev Microbiol. 2018; 16: 410-422. doi: 10.1038/s41579-018-0029-9. PMID: 29795328.

13. Shi M, White VL, Schlub T, Eden JS, Hoffmann AA, Holmes EC. No detectable effect of *Wolbachia* wMel on the prevalence and abundance of the RNA virome of *Drosophila*

346 *melanogaster*. Proc Biol Sci. 2018; 285: 20181165. doi: 10.1098/rspb.2018.1165. PMID:
347 30051873.

348 14. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate
349 RNA virosphere. Nature. 2016; 540: 539-543. doi: 10.1038/nature20167. PMID:
350 27880757.

351 15. Shi M, Zhang YZ, Holmes EC. Meta-transcriptomics and the evolutionary biology of
352 RNA viruses. Virus Res. 2018; 243: 83-90. doi: 10.1016/j.virusres.2017.10.016. PMID:
353 29111455.

354 16. Graus F, Titulaer MJ, Balu R, Benseler S, Bien CG, Cellucci T, et al. A clinical approach
355 to diagnosis of autoimmune encephalitis. Lancet Neurol. 2016; 15: 391-404. doi:
356 10.1016/S1474-4422(15)00401-9. PMID: 26906964.

357 17. Cellucci T, Van Mater H, Graus F, Muscal E, Gallentine W, Klein-Gitelman MS, et al.
358 Clinical approach to the diagnosis of autoimmune encephalitis in the pediatric patient.
359 Neurol Neuroimmunol Neuroinflamm. 2020; 7: e663. doi:
360 10.1212/NXI.0000000000000663. PMID: 31953309.

361 18. Krishnamurthy SR, Wang D. Extensive conservation of prokaryotic ribosomal binding
362 sites in known and novel picobirnaviruses. Virology. 2018; 516: 108-114. doi:
363 10.1016/j.virol.2018.01.006. PMID: 29346073.

19. Tan CS, Koralnik IJ. Progressive multifocal leukoencephalopathy and other disorders caused by JC virus: clinical features and pathogenesis. Lancet Neurol. 2010; 9: 425-437. doi: 10.1016/S1474-4422(10)70040-5. PMID: 20298966.

20. Morfopoulou S, Brown JR, Davies EG, Anderson G, Virasami A, Qasim W, et al. Human coronavirus OC43 associated with fatal encephalitis. N Engl J Med. 2016; 375: 497-498. doi: 10.1056/NEJMc1509458. PMID: 27518687.

21. Ellul M, Solomon T. Acute encephalitis - diagnosis and management. Clin Med. 2018; 18: 155-159. doi: 10.7861/clinmedicine.18-2-155. PMID: 29626021.

22. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. J Infect. 2018; 76: 225-240. doi: 10.1016/j.jinf.2017.12.014. PMID: 29305150.

23. Sato M, Kuroda M, Kasai M, Matsui H, Fukuyama T, Katano H, et al. Acute encephalopathy in an immunocompromised boy with astrovirus-MLB1 infection detected by next generation sequencing. J Clin Virol. 2016; 78: 66-70. doi: 10.1016/j.jcv.2016.03.010. PMID: 26991054.

24. Quan PL, Wagner TA, Briese T, Torgerson TR, Hornig M, Tashmukhamedova A, et al. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. Emerg Infect Dis. 2010; 16: 918-925. doi: 10.3201/eid1606.091536. PMID: 20507741.

25. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med. 2008; 358: 991-998. doi: 10.1056/NEJMoa073785. PMID: 18256387.

26. Tuddenham R, Eden JS, Gilbey T, Dwyer DE, Jennings Z, Holmes EC, et al. Human pegivirus in brain tissue of a patient with encephalitis. Diagn Microbiol Infect Dis. 2020; 96: 114898. doi: 10.1016/j.diagmicrobio.2019.114898. PMID: 31753519.

27. Guo LY, Li YJ, Liu LL, Wu HL, Zhou JL, Zhang Y, et al. Detection of pediatric bacterial meningitis pathogens from cerebrospinal fluid by next-generation sequencing technology. J Infect. 2019; 78: 323-337. doi: 10.1016/j.jinf.2018.12.001. PMID: 30550819.

28. Fan S, Qiao X, Liu L, Wu H, Zhou J, Sun R, et al. Next-generation sequencing of cerebrospinal fluid for the diagnosis of neurocysticercosis. Front Neurol. 2018; 9: 471. doi: 10.1016/j.clineuro.2020.105752. PMID: 29971042.

29. Fan S, Ren H, Wei Y, Mao C, Ma Z, Zhang L, et al. Next-generation sequencing of the cerebrospinal fluid in the diagnosis of neurobrucellosis. Int J Infect Dis. 2018; 67: 20-24. doi: 10.1016/j.ijid.2017.11.028. PMID: 29196276.

30. Tsai JD, Tsai HJ, Lin TH, Chang YY, Yang SH, Kuo HT. Comparison of the detection rates of RT-PCR and virus culture using a combination of specimens from multiple sites for enterovirus-associated encephalomyelitis during enterovirus 71 epidemic. Jpn J Infect Dis. 2014; 67: 333-338. doi: DOI 10.7883/yoken.67.333. PMID: 25241681.

31. Ooi MH, Solomon T, Podin Y, Mohan A, Akin W, Yusuf MA, et al. Evaluation of different clinical sample types in diagnosis of human enterovirus 71-associated hand-foot-and-mouth disease. J Clin Microbiol. 2007; 45: 1858-1866. doi: 10.1128/JCM.01394-06. PMID: 17446325.

32. Perez-Velez CM, Anderson MS, Robinson CC, McFarland EJ, Nix WA, Pallansch MA, et al. Outbreak of neurologic enterovirus type 71 disease: a diagnostic challenge. Clin Infect Dis. 2007; 45: 950-957. Epub 2007/09/21. doi: 10.1086/521895. PMID: 17879907.

33. Dube M, Le Coupanec A, Wong AHM, Rini JM, Desforges M, Talbot PJ. Axonal transport enables neuron-to-neuron propagation of human coronavirus OC43. J Virol. 2018; 92: e00404-18. doi: 10.1128/JVI.00404-18. PMID: 29925652.

34. Arbour N, Day R, Newcombe J, Talbot PJ. Neuroinvasion by human respiratory coronaviruses. J Virol. 2000; 74: 8913-8921. Epub 2000/09/12. doi: 10.1128/jvi.74.19.8913-8921.2000. PMID: 10982334.

35. Li YY, Li HP, Fan RY, Wen B, Zhang J, Cao XY, et al. Coronavirus infections in the central nervous system and respiratory tract show distinct features in hospitalized children. Intervirology. 2016; 59: 163-169. doi: 10.1159/000453066. PMID: 28103598.

36. Yeh EA, Collins A, Cohen ME, Duffner PK, Faden H. Detection of coronavirus in the central nervous system of a child with acute disseminated encephalomyelitis. Pediatrics. 2004; 113: e73-76. doi: 10.1542/peds.113.1.e73. PMID: 14702500.

422 37. Jacomy H, St-Jean JR, Brison E, Marceau G, Desforges M, Talbot PJ. Mutations in the
423 spike glycoprotein of human coronavirus OC43 modulate disease in BALB/c mice from
424 encephalitis to flaccid paralysis and demyelination. J Neurovirol. 2010; 16: 279-293. doi:
425 10.3109/13550284.2010.497806. PMID: 20642316.

426 38. Desforges M, Le Coupanec A, Dubeau P, Bourgouin A, Lajoie L, Dube M, et al. Human
427 coronaviruses and other respiratory viruses: underestimated opportunistic pathogens of
428 the central nervous system? Viruses. 2019; 12: 14. doi: 10.3390/v12010014. PMID:
429 31861926;.

430 39. Paterson RW, Brown RL, Benjamin L, Nortley R, Wiethoff S, Bharucha T, et al. The
431 emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory
432 findings. Brain. 2020; 143: 3104-3120. doi: 10.1093/brain/awaa240. PMID: 32637987.

433 40. Lindan CE, Mankad K, Ram D, Kociolek LK, Silvera VM, Boddaert N, et al.
434 Neuroimaging manifestations in children with SARS-CoV-2 infection: a multinational,
435 multicentre collaborative study. Lancet Child Adolesc Health. 2020; 5: 167-177. doi:
436 10.1016/S2352-4642(20)30362-X. PMID: 33338439.

437 41. Kremer S, Lersy F, Anheim M, Merdji H, Schenck M, Oesterle H, et al. Neurologic and
438 neuroimaging findings in patients with COVID-19: A retrospective multicenter study.
439 Neurology. 2020; 95: e1868-e82. doi: 10.1212/WNL.0000000000010112. PMID:
440 32680942.

441 42. Almqvist J, Granberg T, Tzortzakakis A, Klironomos S, Kollia E, Ohberg C, et al.
442 Neurological manifestations of coronavirus infections - a systematic review. Ann Clin
443 Transl Neurol. 2020; 7: 2057-2071. doi: 10.1002/acn3.51166. PMID: 32853453.

444 43. Granerod J, Cunningham R, Zuckerman M, Mutton K, Davies NW, Walsh AL, et al.
445 Causality in acute encephalitis: defining aetiologies. Epidemiol Infect. 2010; 138: 783-
446 800. doi: 10.1017/S0950268810000725. PMID: 20388231.

447 44. Harvala H, McIntyre CL, McLeish NJ, Kondracka J, Palmer J, Molyneaux P, et al. High
448 detection frequency and viral loads of human rhinovirus species A to C in fecal samples;
449 diagnostic and clinical implications. J Med Virol. 2012; 84: 536-542. doi:
450 10.1002/jmv.23203. PMID: 22246843.

451 45. Pelkonen T, Roine I, Anjos E, Kaijalainen S, Roivainen M, Peltola H, et al.
452 Picornaviruses in cerebrospinal fluid of children with meningitis in Luanda, Angola. J
453 Med Virol. 2012; 84: 1080-1083. doi: 10.1002/jmv.23304. PMID: 22585725.

454 46. Hazama K, Shiihara T, Tsukagoshi H, Matsushige T, Dowa Y, Watanabe M. Rhinovirus-
455 associated acute encephalitis/encephalopathy and cerebellitis. Brain Dev. 2019; 41: 551-
456 554. doi: 10.1016/j.braindev.2019.02.014. PMID: 30850156.

457 47. Blomqvist S, Savolainen C, Raman L, Roivainen M, Hovi T. Human rhinovirus 87 and
458 enterovirus 68 represent a unique serotype with rhinovirus and enterovirus features. J
459 Clin Microbiol. 2002; 40: 4218-4223. doi: 10.1128/jcm.40.11.4218-4223.2002. PMID:
460 12409401.

461 48. Lauinger IL, Bible JM, Halligan EP, Aarons EJ, MacMahon E, Tong CY. Lineages, sub-
462 lineages and variants of enterovirus 68 in recent outbreaks. PLoS One. 2012; 7: e36005.
463 doi: 10.1371/journal.pone.0036005. PMID: 22536453.

464 49. Mizuguchi M, Yamanouchi H, Ichiyama T, Shiomi M. Acute encephalopathy associated
465 with influenza and other viral infections. Acta Neurol Scand. 2007; 115: 45-56. doi:
466 10.1111/j.1600-0404.2007.00809.x. PMID: 17362276.

467 50. Mizuguchi M, Abe J, Mikkaichi K, Noma S, Yoshida K, Yamanaka T, et al. Acute
468 necrotising encephalopathy of childhood: a new syndrome presenting with multifocal,
469 symmetric brain lesions. J Neurol Neurosurg Psychiatry. 1995; 58: 555-561. doi:
470 10.1136/jnnp.58.5.555. PMID: 7745402.

471 51. Britton PN, Dale RC, Blyth CC, Macartney K, Crawford NW, Marshall H, et al.
472 Influenza-associated encephalitis/encephalopathy identified by the Australian Childhood
473 Encephalitis study 2013-2015. Pediatr Infect Dis J. 2017; 36: 1021-1026. doi:
474 10.1097/INF.0000000000001650. PMID: 28654561.

475 52. Takanashi J. Two newly proposed infectious encephalitis/encephalopathy syndromes.
476 Brain Dev. 2009; 31: 521-528. doi: 10.1016/j.braindev.2009.02.012. PMID: 19339128.

477 53. Bohmwald K, Galvez NMS, Rios M, Kalergis AM. Neurologic alterations due to
478 respiratory virus infections. Front Cell Neurosci. 2018; 12: 386. doi:
479 10.3389/fncel.2018.00386. PMID: 30416428.

480 54. Granerod J, Ambrose HE, Davies NWS, Clewley JP, Walsh AL, Morgan D, et al. Causes
481 of encephalitis and differences in their clinical presentations in England: a multicentre,
482 population-based prospective study. Lancet Infect Dis. 2010; 10: 835-844. doi:
483 10.1016/S1473-3099(10)70222-X. PMID: 20952256.

484 55. Li CX, Li W, Zhou J, Zhang B, Feng Y, Xu CP, et al. High resolution metagenomic
485 characterization of complex infectomes in paediatric acute respiratory infection. Sci Rep.
486 2020; 10: 3963. doi: 10.1038/s41598-020-60992-6. PMID: 32127629.

487 56. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. CD-HIT: accelerated for clustering the next-
488 generation sequencing data. Bioinformatics. 2012; 28: 3150-3152. doi:
489 10.1093/bioinformatics/bts565. PMID:23060610.

490 57. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND.
491 Nat Methods. 2015; 12: 59-60. doi: DOI 10.1038/nmeth.3176. PubMed PMID:
492 25402007.

493 58. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A
494 fast and scalable metagenome assembler driven by advanced methodologies and
495 community practices. Methods. 2016; 102: 3-11. doi: 10.1016/j.ymeth.2016.02.020.
496 PMID: 27012178.

497 59. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node
498 solution for large and complex metagenomics assembly via succinct de Bruijn graph.

499 Bioinformatics. 2015; 31: 1674-1676. doi: 10.1093/bioinformatics/btv033. PMID:
500 25609793.

501 60. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, Geer LY, et al.
502 CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43: D222-D226.
503 doi: 10.1093/nar/gku1221. PMID: 25414356.

504 61. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale
505 multiple sequence alignments. Bioinformatics. 2018; 34: 2490-2492. doi:
506 10.1093/bioinformatics/bty121. PMID: 29506019.

507 62. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New
508 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
509 performance of PhyML 3.0. Syst Biol. 2010; 59: 307-321. doi: 10.1093/sysbio/syq010.
510 PMID:20525638.

511 63. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al.
512 MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015; 12:
513 902-903. doi: 10.1038/nmeth.3589. PMID: 26418763.

514

515 **Figure legends**
516

517 **Fig. 1. Geographic context of the study.** (A) Maps showing the residential postal codes
518 within Australia of patients with clinical diagnosis as encephalitis reported at the Children's
519 Hospital at Westmead (CHW) between March 2014 and September 2016. Red star indicates
520 one case and dark red star indicates two cases. (B) The bar chart showing the diversity of
521 sample types collected in four states (NSW, QLD, VIC and WA) in Australia, coloured by
522 sample types.
523

524 **Fig. 2. Diversity and abundance of viruses identified in this study.** The bar chart shows
525 number of virus species in each sample: RNA viruses (blue) and DNA viruses (yellow). The
526 heatmap shows the abundance level of different virus species within each library. The
527 abundance level of reads was normalized to unique reads mapped per million input reads
528 (RPM). HCoV-OC43, human coronavirus OC43; HCoV-HKU1, human coronavirus HKU1;
529 HRV-A9, human rhinovirus A9; HRV-B52, human rhinovirus B52; Echo6, echovirus E6;
530 HCoV-HKU1, human coronavirus HKU1; HPBV, human picobirnavirus; JCPyV, JC
531 polyomavirus; WUPyV, WU polyomavirus.
532

533 **Fig. 3. Phylogenetic relationships and intra-specific diversity and of viruses identified by**
534 **mNGS in this study.** Viruses identified as part of this project are marked in red and viruses
535 from the same type/subgroup are shaded light red, whereas those representing background

phylogenetic diversity are shown in black. All horizontal branch lengths are scaled to the number of nucleotide substitutions per site, and trees are mid-point rooted for clarity.

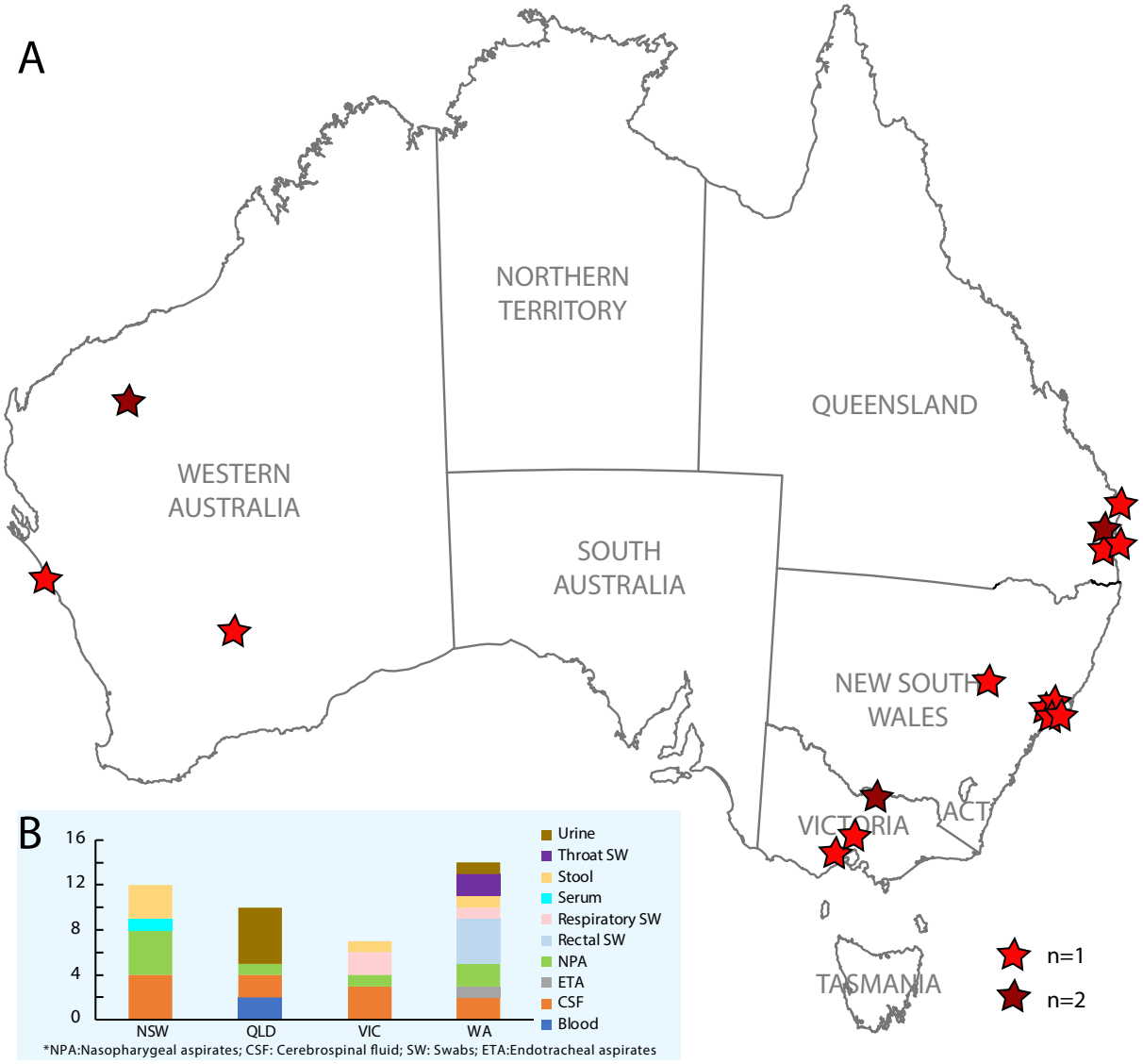
Fig. 4. Analysis of the novel picobirnavirus. (A) Phylogenetic relationships of the RNA-dependent RNA polymerase (RdRp) proteins of the novel human picobirnavirus identified from this study and representative picobirnaviruses. (B) Genome organization of the novel human picobirnavirus from one stool sample. Segment 1 encodes two hypothetical proteins (ORF1 and ORF2), with arrows standing for directions of uncomplete ORFs. Segment 2 encodes the viral RNA-dependent RNA polymerase (RdRp). The ribosome binding sites (RBS) and the ExxRxNxxxE amino acid (aa) motifs are marked.

Fig. 5. Detailed analysis of the bacteria identified in this study. (A) Bacterial taxa identified using mNGS. Each dot represents a taxonomic entity. From the inner to outer circles, the taxonomic levels range from kingdom to species. Different colored dots indicate different taxonomic levels according to the color key shown. Numbers in parentheses indicate the total number of unique taxonomies detected at each level. (B) Profiling the relative abundance of the bacteria identified here. Cladogram representing the predominant family and top 16 bacterial species present in these cases. Circle size depicts microbial abundance. Abundance levels (reads per million total reads) were estimated using MetaPhlAn2. The taxonomic tree was visualized using GraPhlAn.

557 **Fig. 6. Heatmap of the abundance level for the bacteria identified in each case studied**

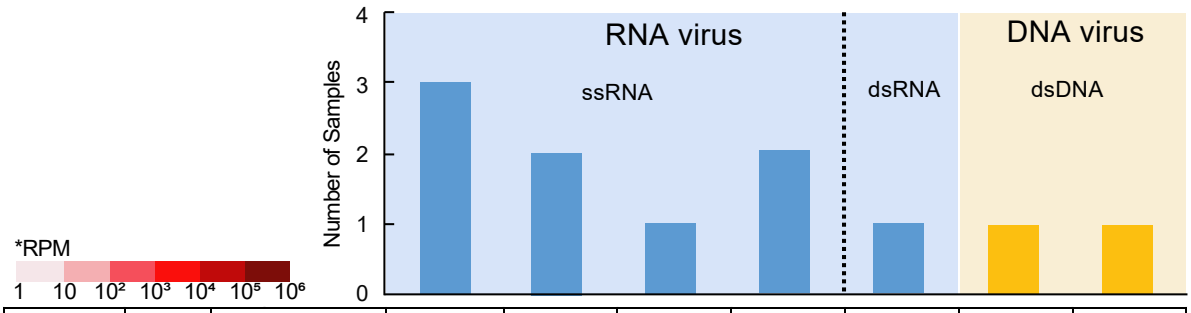
558 **here.** The abundance level of microbial reads was normalized to RPM, and the estimation

559 was performed at the species level.



560

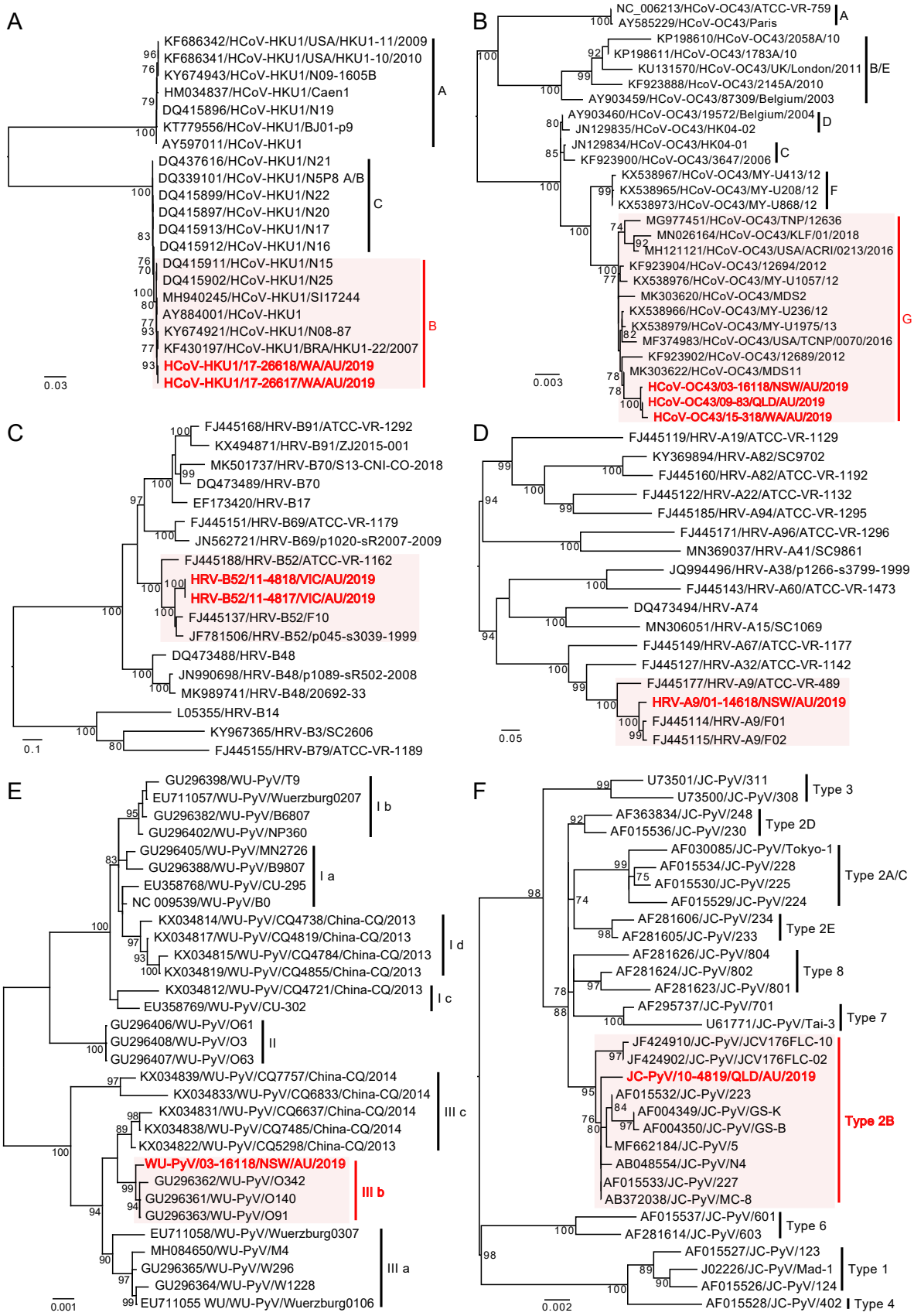
561 **Figure 1**



State	Cases	Sample type	HCoV-OC43	HCoV-HKU1	HRVA9	HRVB52	HPBV	JCPyV	WUPyV
NSW	01	NPA							
		Serum							
		Stool							
		Stool							
		Stool							
	02	NPA							
	03	NPA							
	04	CSF							
		NPA							
	05	CSF							
		CSF							
		CSF							
QLD	06	Blood							
		Urine							
		Urine							
	07	NPA							
	08	Urine							
		Urine							
		CSF							
	09	Blood							
	10	Urine							
		CSF							
VIC	11	NPA							
		Stool							
	12	CSF							
	13	CSF							
		Respiratory SW							
	14	CSF							
		Respiratory SW							
WA	15	CSF							
		Rectal SW							
		Throat SW							
		Urine							
	16	CSF							
		NPA							
		Rectal SW							
	17	NPA							
		Rectal SW							
		Respiratory SW							
		Stool							
	18	ETA							
		Rectal SW							
		Throat SW							

*NPA:Nasopharygeal aspirates; CSF: Cerebrospinal fluid; SW: Swabs; ETA:Endotracheal aspirates

Figure 2



565

566 **Figure 3**

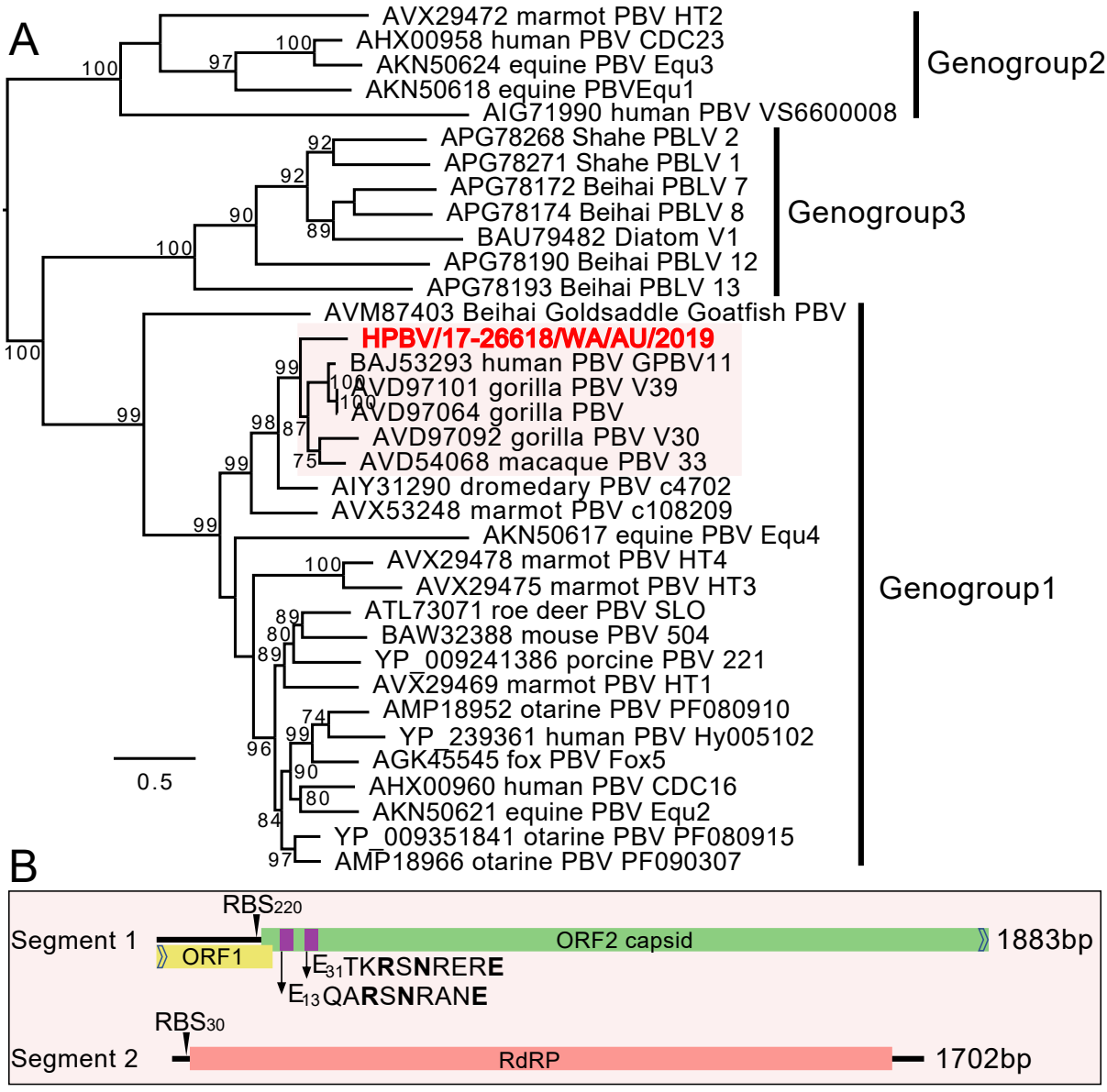


Figure 4

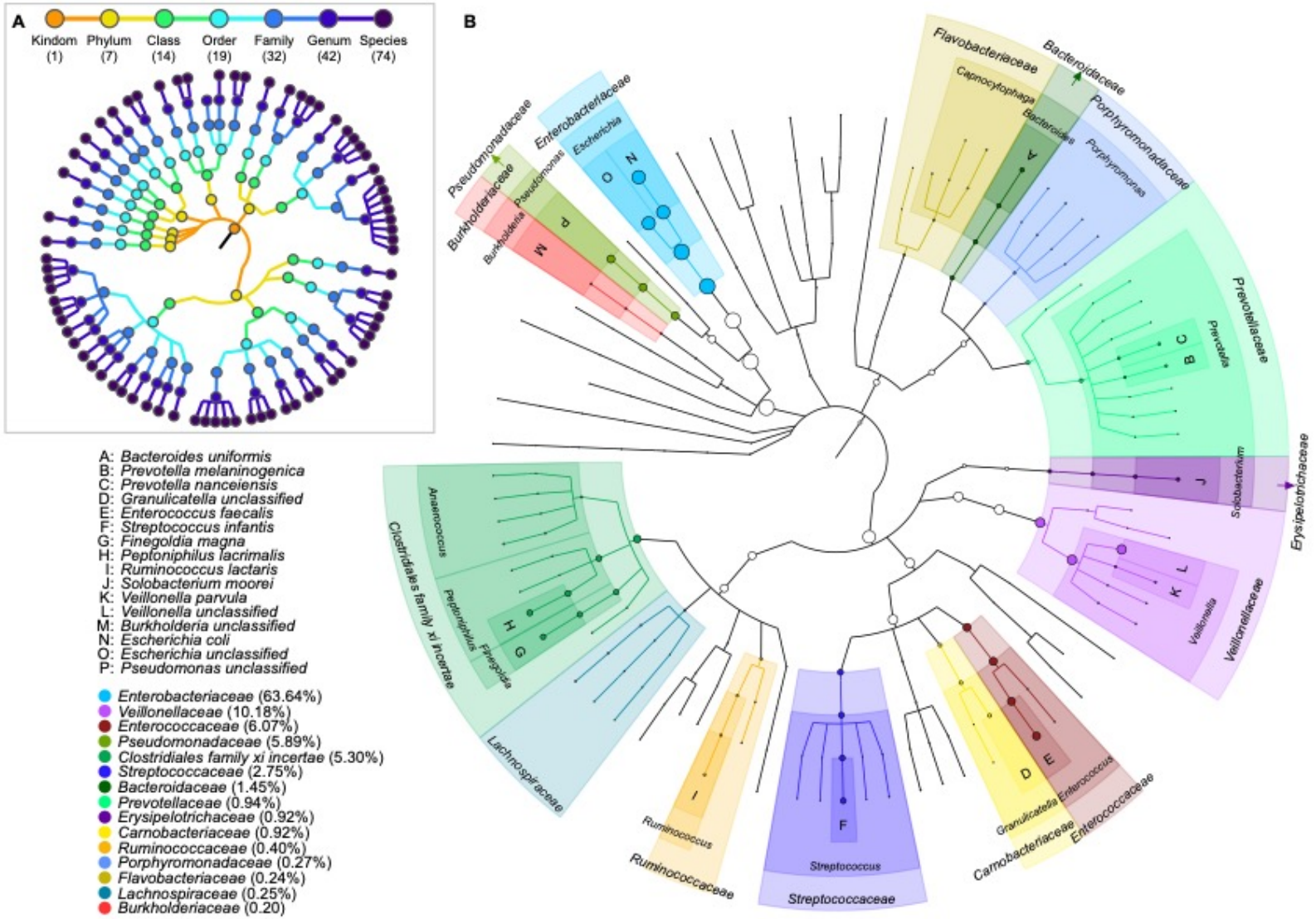
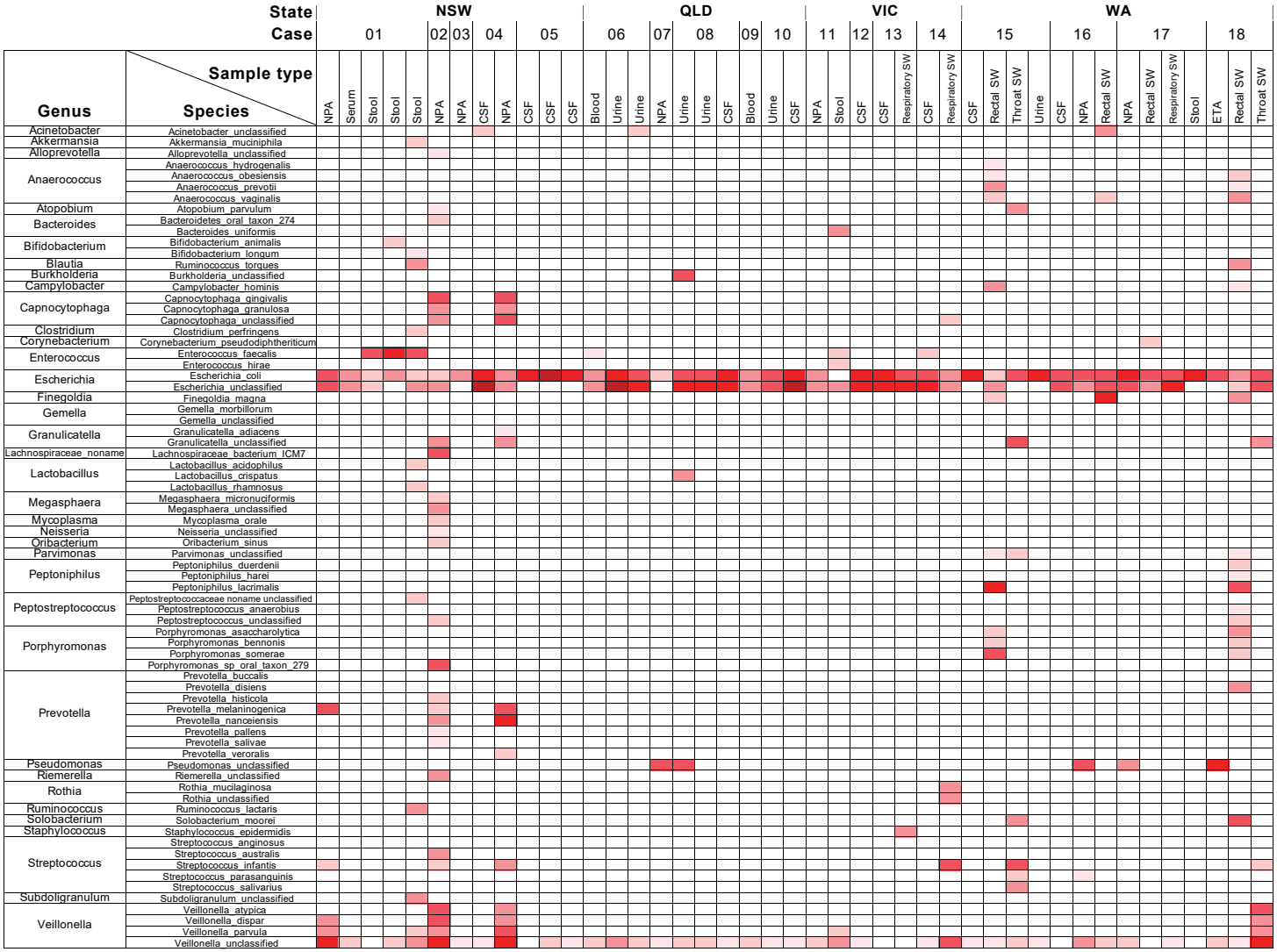

A
Kindom (1)
Phylum (7)
Class (14)
Order (19)
Family (32)
Genum (42)
Species (74)
B
A: Bacteroides uniformis
B: Prevotella melaninogenica
C: Prevotella nanceiensis
D: Granulicatella unclassified
E: Enterococcus faecalis
F: Streptococcus infantis
G: Finegoldia magna
H: Peptoniphilus lacrimalis
I: Ruminococcus lactaris
J: Solobacterium moorei
K: Veillonella parvula
L: Veillonella unclassified
M: Burkholderia unclassified
N: Escherichia coli
O: Escherichia unclassified
P: Pseudomonas unclassified
Enterobacteriaceae (63.64%)
Veillonellaceae (10.18%)
Enterococcaceae (6.07%)
Pseudomonadaceae (5.89%)
Clostridiales family xi incertae (5.30%)
Streptococcaceae (2.75%)
Bacteroidaceae (1.45%)
Prevotellaceae (0.94%)
Erysipelotrichaceae (0.92%)
Carnobacteriaceae (0.92%)
Ruminococcaceae (0.40%)
Porphyromonadaceae (0.27%)
Flavobacteriaceae (0.24%)
Lachnospiraceae (0.25%)
Burkholderiaceae (0.20)
Flavobacteriaceae
Capnocytophaga
Bacteroidaceae
Bacteroides
Porphyromonadaceae
Porphyromonas
Prevotellaceae
Prevotella
Solobacterium
Erysipelotrichaceae
Veillonellaceae
Veillonella
Enterococcaceae
Enterococcus
Granulicatella
Carnobacteriaceae
Streptococcaceae
Streptococcus
Ruminococcaceae
Ruminococcus
Lachnospiraceae
Clostridiales family xi incertae
Anaerococcus
Peptoniphilus
Finegoldia
Enterobacteriaceae
Escherichia
Pseudomonadaceae
Pseudomonas
Burkholderiaceae
Burkholderia

Figure 5



*NPA:Nasopharygeal aspirates; CSF: Cerebrospinal fluid; SW: Swabs; ETA:Endotracheal aspirates

Figure 6

Supporting Information Captions

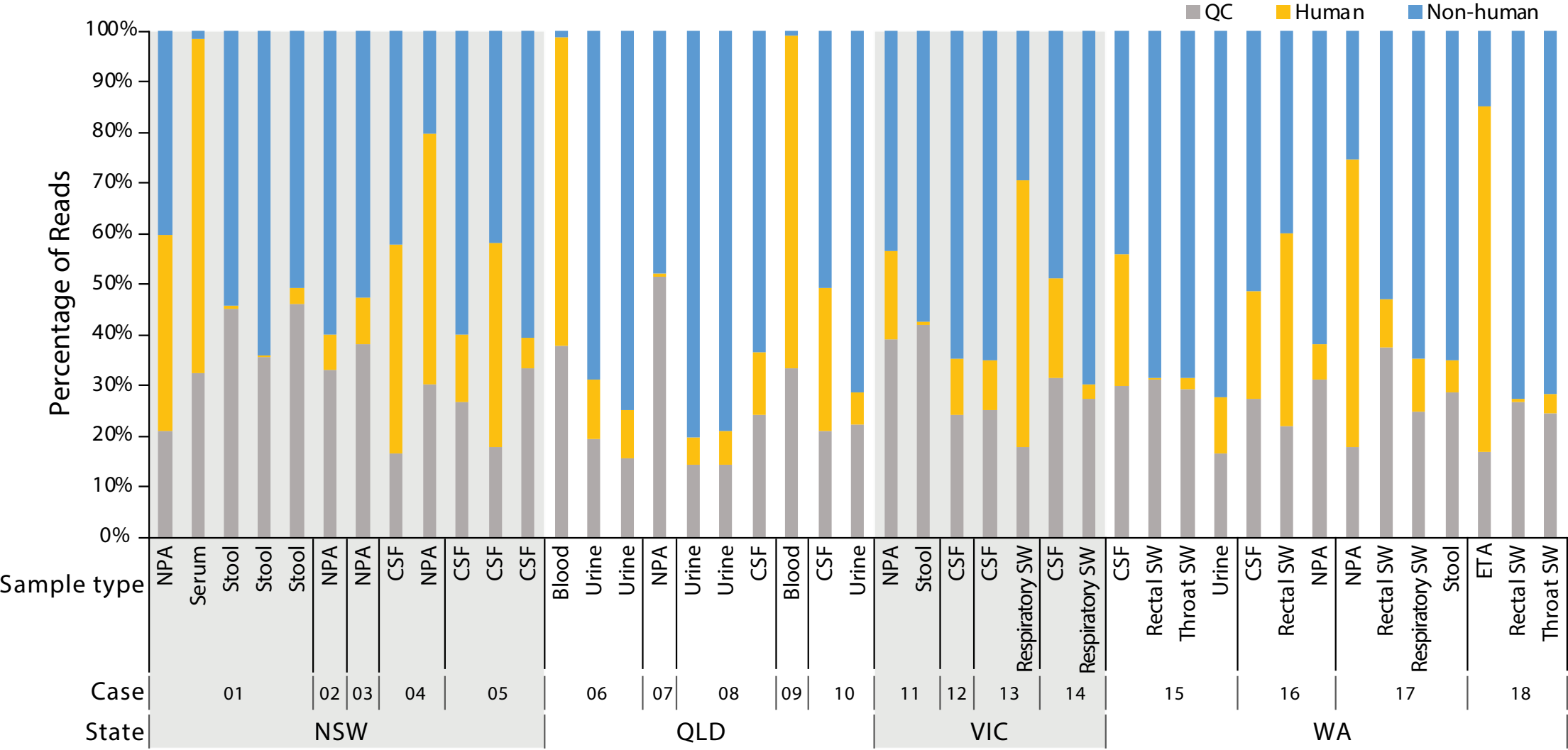
Fig S1. Summary of taxonomy and abundance from meta-transcriptomic sequencing of the libraries from each patient. The bar chart shows the proportion of assigned taxa according to the keys provided. The panel shows the proportion of reads that were assigned as quality controlled (QC), human and non-human in each library.

Fig S2. Logic model to determine potential the pathogenicity of the viruses, bacteria and fungi identified here. #mNGS = metagenomic next-generation sequencing. *Sterile sites included blood, serum, and CSF; non-sterile sites included nasopharyngeal aspirates, endotracheal samples, throat and nasal swabs, as well as stool and rectal swabs. Infectious and immune-mediated syndromes were defined by the advanced clinical expert panel. ^Abundance was determined to be high if sequence reads were above 10000 RPM, or low if below 1000 RPM. In the event of detection in multiple samples sterile samples taken priority.

Table S1. Summary of laboratory tests performed on the encephalitis cases.

Table S2. Information on the RNA sequencing libraries generated in this study.

Table S3. Identity of viruses identified in this study with the most closely related sequence available on public sequence databases.



*NPA:Nasopharygeal aspirates; CSF: Cerebrospinal fluid; SW: Swabs; ETA:Endotracheal aspirates

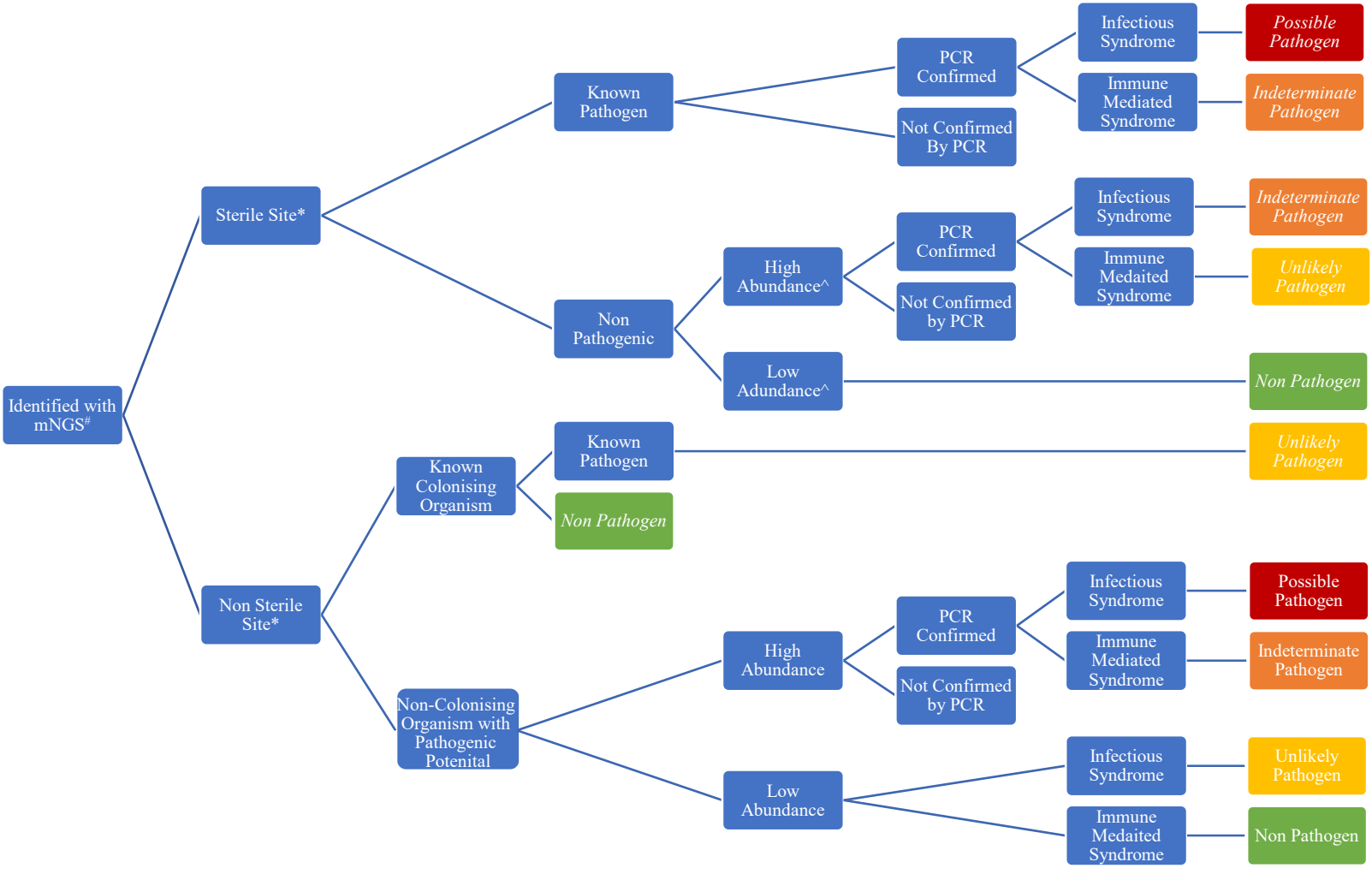

Identified with mNGS#
Sterile Site*
Known Pathogen
PCR Confirmed
Infectious Syndrome
Possible Pathogen
Immune Mediated Syndrome
Indeterminate Pathogen
Not Confirmed By PCR
Non Pathogenic
High Abundance^
PCR Confirmed
Infectious Syndrome
Indeterminate Pathogen
Immune Medaited Syndrome
Unlikely Pathogen
Not Confirmed by PCR
Low Adundance^
Non Pathogen
Non Sterile Site*
Known Colonising Organism
Known Pathogen
Unlikely Pathogen
Non Pathogen
Non-Colonising Organism with Pathogenic Potenital
High Abundance
PCR Confirmed
Infectious Syndrome
Possible Pathogen
Immune Mediated Syndrome
Indeterminate Pathogen
Not Confirmed by PCR
Low Abundance
Infectious Syndrome
Unlikely Pathogen
Immune Medaited Syndrome
Non Pathogen

Table S1. Summary of laboratory tests performed on the encephalitis cases.

Case No.	Infectious Disease																					Immune Mediated		
	CSF Culture			CSF PCR										Serology		Sputum PCR	Gastro-Intestinal							
	Bacterial	*Fungal*	*Mycoplasma*	HSV	EV	HPeV	VZV	EBV	CMV	HHV	N. Meningitidis	*M. pneumoniae*		Acute EBV	M. pnuemoniae IgM		Rotavirus	Adenovirus	Enterovirus	Bacterial Culture		CSF NDMA IgG	Serum VGKC IgG	Oligoclonal Banding
1	N			N	N											P*				N				N
2				N	N						N					N								
3	N			N	N					N				N		P#				N				N
4	N			N	N											N			N					
5	N			N	N									N	N		N			N				N
6	N	N	N	N	N									N	N	N						N		N
7	N	N	N	N	N	N		N			N	N		N	N									E
8	N	N	N	N	N		N	N	N					N	N	N						N		P
9	N	N	N	N	N											N								
10	N		N	N	N							N				N						N		N
11	N			N	N		N		N	N		N				N	N	N	N	N				
12	N			N	N							N		N										N
13	N			N	N	N	N	N		N		N		N		N	P			N		N	N	N
14	N			N	N		N					N				N								N
15	N			N	N	N	N	N	N	N	N	N		N					N					
16	N			N	N	N	N		N	N	N			N					N					
17															N	N	N	N						
18				N	N		N				N	N		N	N							N		

N = Negative, P = Positive, E = Equivocal, Blank = Untested. HSV = Herpes Simplex Virus, EV = Enterovirus, HPeV = Human Parechovirus, VZV = Varicella Zoster Virus, EBV = Epstein-Barr Virus, CMV = Cytomegalovirus, HHV = Human Herpesvirus. Negative for acute EBV = No EBV IgM, and/or pos EBV NA IgG. * Rhinovirus, # Coronavirus, ** Stool tested >day 7 of admission, no diarrhoea. Oligoclonal banding is positive if present only in CSF or is higher in CSF than in serum.

Table S2. Information on the RNA sequencing libraries generated in this study.

Case No.	Libraries	Sample types	Paired reads (150bp)	Data	QC	Total human	Total non-human
01	14618NPA	NPA	68,331,435	20.64 Gb	20.95%	38.68%	40.37%
	1469	Serum	62,439,715	18.86 Gb	32.38%	65.91%	1.71%
	14617	Stool	72,597,211	21.92 Gb	45.14%	0.42%	54.43%
	14619	Stool	76,093,344	22.98 Gb	35.56%	0.37%	64.07%
	14618S	Stool	81,029,503	24.47 Gb	45.94%	3.32%	50.74%
02	14918	NPA	69,918,956	21.12 Gb	32.94%	7.13%	59.93%
03	16118	NPA	64,014,451	19.33 Gb	37.97%	9.36%	52.67%
04	1972	CSF	55,650,564	16.81 Gb	16.53%	41.33%	42.14%
	19718	NPA	61,763,723	18.65 Gb	30.09%	49.55%	20.36%
05	2051	CSF	12,207,214	3.69 Gb	26.55%	13.48%	59.98%
	2052	CSF	65,679,279	19.84 Gb	17.69%	40.40%	41.91%
	2053	CSF	45,372,654	13.70 Gb	33.33%	5.92%	60.75%
06	4	Blood	80,788,125	24.40 Gb	37.62%	61.03%	1.35%
	419	Urine	65,728,328	19.85 Gb	19.32%	11.88%	68.80%
	420	Urine	59,130,993	17.86 Gb	15.69%	9.48%	74.84%
07	1419	NPA	70,769,034	21.37 Gb	51.54%	0.41%	48.05%
08	6019	Urine	63,645,388	19.22 Gb	14.21%	5.60%	80.19%
	6020	Urine	66,076,819	19.96 Gb	14.36%	6.46%	79.18%
	631	CSF	64,218,139	19.39 Gb	24.16%	12.35%	63.49%
09	83	Blood	76,472,229	23.09 Gb	33.29%	65.60%	1.10%
10	4819	Urine	69,409,747	20.96 Gb	22.08%	6.58%	71.34%
	481	CSF	63,814,144	19.27 Gb	20.83%	28.25%	50.91%
11	4818	NPA	70,749,343	21.37 Gb	39.21%	17.34%	43.45%
	4817	Stool	69,714,937	21.05 Gb	41.82%	0.62%	57.56%
12	561	CSF	61,044,714	18.44 Gb	24.18%	11.16%	64.66%
13	951	CSF	30,481,591	9.21 Gb	25.17%	9.84%	64.99%
	9518	Respiratory Swab	62,643,416	18.92 Gb	17.66%	52.71%	29.64%
14	1031	CSF	60,509,053	18.27 Gb	31.36%	19.83%	48.81%
	10318	Respiratory Swab	71,156,865	21.49 Gb	27.28%	2.80%	69.92%
15	31	CSF	59,215,458	17.88 Gb	29.78%	26.16%	44.05%
	317	Rectal Swab	74,947,448	22.63 Gb	30.98%	0.31%	68.70%
	318	Throat Swab	72,626,978	21.93 Gb	29.09%	2.49%	68.41%
	003X	Urine	56,431,288	17.04 Gb	16.60%	11.15%	72.25%
16	881	CSF	52,399,272	15.82 Gb	27.15%	21.51%	51.34%
	8818	NPA	62,379,902	18.84 Gb	31.18%	6.98%	61.84%
	8817	Rectal Swab	71,705,712	21.66 Gb	21.80%	38.09%	40.11%
17	26619	NPA	66,107,489	19.96 Gb	17.72%	56.84%	25.45%
	26617	Rectal Swab	79,261,749	23.94 Gb	37.49%	9.57%	52.94%
	26618R	Respiratory Swab	69,315,594	20.93 Gb	24.91%	10.32%	64.77%
	26618S	Stool	61,180,294	18.48 Gb	28.56%	6.52%	64.92%
18	65419	ETA	66,718,955	20.15 Gb	16.71%	68.52%	14.77%
	65417	Rectal Swab	69,257,149	20.92 Gb	26.66%	0.68%	72.66%
	65418	Throat Swab	67,246,928	20.31 Gb	24.50%	3.82%	71.69%
Total			2,770,245,130	836.61 Gb			

Table S3. Identity of viruses identified in this study with the most closely related sequence available on public sequence databases.

Virus	Closest relative	Identity	Genes or genome alignments used in the phylogenetic analysis	
			Domain	Alignment length (bp)
HCoV-OC43/03-16118/NSW/AU/2019	MK303622/HCoV-OC43/MDS11	99.73%	Spike protein	4433
HCoV-OC43/09-83/QLD/AU/2019	MK303622/HCoV-OC43/MDS11	99.73%	Spike protein	4433
HCoV-OC43/15-318/WA/AU/2019	MK303622/HCoV-OC43/MDS11	99.65%	Spike protein	4433
HCoV-HKU1/17-26618/WA/AU/2019	MH940245/HCoV-HKU1/SI17244	99.75%	Partial Spike protein	2446
HCoV-HKU1/17-26617/WA/AU/2019	MH940245/HCoV-HKU1/SI17244	99.75%	Partial Spike protein	2446
HRV-A9/01-14618/NSW/AU/2019	FJ445114/HRV-A9/F01	96.48%	Near complete genome	6808
HRV-B52/11-4818/VIC/AU/2019	FJ445137/HRV-B52/F10	92.82%	Near complete genome	6513
HRV-B52/11-4817/VIC/AU/2019	FJ445137/HRV-B52/F10	92.82%	Near complete genome	6513
HPBV/17-26618/WA/AU/2019	AVD54068/Macaque PBV 33	72.44%(aa)	RdRp	524 amino acids
JC-PyV/10-4819/QLD/AU/2019	AF015533/JC-PyV/227	99.79%	Partial genome	2378
WU-PyV/03-16118/NSW/AU/2019	GU296361/WU-PyV/O140	99.82%	Near complete genome	5230