SUPPLEMENTARY APPENDIX

A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the United States from 2012-2014: a retrospective cohort study

Alexander L. Greninger, M.D., Ph.D.[1,2¶], Samia N. Naccache, Ph.D.[1,2¶], Kevin Messacar, M.D.[3], Anna Clayton[4], B.S., M.P.H., Guixia Yu, B.S.[1,2], Sneha Somasekar, B.S.[1,2], Scot Federman, B.A.[1,2], Doug Stryke, B.S.[1,2], Christopher Anderson, B.S[4], Shigeo Yagi, Ph.D.[4], Sharon Messenger, Ph.D. [4], Debra Wadford, Ph.D. [4], Dongxiang Xia, M.D., Ph.D.[4], James P. Watt, M.D., M.P.H. [4], Keith Van Haren, M.D.[5], Samuel R. Dominguez, M.D., Ph.D.[3], Carol Glaser, D.V.M., M.D.[4], Grace Aldrovandi, M.D.[6], and Charles Y. Chiu, M.D., Ph.D.[1,2]*

Affiliations:
[1]University of California, San Francisco, CA, USA
[2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA
[3]Children's Hospital Colorado and University of Colorado School of Medicine, Aurora, CO, USA
[4]California Department of Public Health, Richmond, CA, USA
[5]Lucile Packard Children's Hospital at Stanford University, Palo Alto, CA, USA
[6]Children's Hospital Los Angeles and University of Southern California, Los Angeles, CA, USA

[¶]These authors contributed equally to the manuscript.

*corresponding author
Charles Chiu
Department of Laboratory Medicine
University of California, San Francisco, San Francisco, CA
charles.chiu@ucsf.edu

TABLE OF CONTENTS

SUPPLEMENTARY METHODS

Clinical cohorts and samples

De-identified clinical data and samples from patients with acute flaccid myelitis (AFM), encephalitis, or enterovirus D68 (EV-D68)-associated upper respiratory infection (URI) were collected by the California Department of Public Health (CDPH) between 2009 and 2014. Stool, whole blood, serum, nasopharyngeal swab / wash or oropharyngeal swab (NP/OP), and cerebrospinal fluid (CSF) samples were analyzed (Table S1).

Coded rectal swab, NP/OP, whole blood, serum, and CSF samples from a cluster of 3 patients with AFM were collected at Children's Hospital Los Angeles from August to October 2014 and analyzed (Table S1). In addition, CSF from 16 consecutive patients presenting with clinical aseptic meningitis or encephalitis in August 2014 and testing positive for EV by a Focus Simplexa assay were screened for EV-D68.

Coded blood, NP/OP, and CSF samples from 8 patients from a previously reported cluster of AFM cases at Children's Hospital Colorado (CHCO) from September to October of 2014[1, 2], and one additional patient from CHCO outside of this cluster were analyzed.

Nucleic acid extraction

Total nucleic acid was extracted from 100-400 µL of clinical sample on a robotic Qiagen EZ1 instrument using the Qiagen EZ1 Virus Mini Kit v2.0 (Qiagen, Valencia, CA).

RT-PCR screening

All nucleic acid extracts were screened for the presence of EV by heminested reverse transcription polymerase chain reaction (RT-PCR) (Table 2). The first round uses primers targeting the 5'-untranslated region (5'-UTR) of EV and rhinoviruses (RV)[3], followed by a heminested PCR to boost sensitivity. If EV-D68 is documented by Sanger sequencing of the resulting product, another screening heminested RT-PCR reaction targeting the EV-D68 VP1 gene was run as an independent confirmatory test for EV-D68. A third heminested RT-PCR reaction was then run to recover the sequence of the EV-D68 VP1 gene. For all 3 heminested RT-PCR reactions, the first and second rounds of amplification were conducted using the Qiagen One-Step RT-PCR Kit in 25 µL total reaction volume (4 µL Q solution, 4 µL 5X buffer, 1 µL dNTP, 1 µL enzyme), with 12 pmol of each primer (Table S2). 2 µL of nucleic acid were used as a template for the 1st round reaction, while 1.25 µL of 1st round material were then used as template for the 2nd round. Conditions for the heminested RT-PCR were as follows:

- 5'-UTR heminested RT-PCR, 1st round: 94°C, 30s / 51°C, 30 s / 72°C, 30 s
- 5'-UTR heminested RT-PCR, 2nd round: 94°C, 30s / 53°C, 30s / 72°C, 30s
- VP1 heminested RT-PCR (screening or sequencing), 1st round: 94°C, 30s / 55°C, 30s / 72°C, 2 min
- VP1 heminested RT-PCR (screening or sequencing), 2nd round: 94°C, 30s / 52°C, 30s / 72°C, 2 min

Viral quantification

As the 5'UTR RT-PCR was found to be more sensitive for EV-D68 detection than the VP1 RT-PCR, a single round qRT-PCR using primers targeting the 5'-UTR of EV/RV[4] (Table S1 and Figure S1) to quantify viral genome copies in EV-D68-positive samples. EV-D68 copy number was calculated by generating a standard curve of diluted EV-D68 amplicon from 10^8 to 0.1 copies, which exhibited linearity across the 10^8 to 10^1 range. For EV-D68 quantification, the Qiagen Quantitect RT-PCR Kit was used according to the manufacturer's instructions at a final optimal primer concentration of 1 µM.

Construction of metagenomic NGS libraries from CSF samples

From the 25 AFM patients in the current study, CSF samples from 14 patients were available for analysis by metagenomic NGS. Amplified cDNA was generated directly from extracted RNA as previously described[5-8]. Metagenomic NGS libraries were then constructed either from randomly amplified cDNA using a modified TruSeq protocol[5] or random hexamer cDNA synthesis using a Nextera XT protocol (Illumina) with 2 ng of input cDNA[6-8]. A CSF sample from a patient with EV-A71-associated encephalitis was used as a positive control, and analyzed in parallel with CSF samples from AFM patients.

Construction of metagenomic NGS libraries from respiratory samples

To decrease background from host and microbial flora, extracted RNA from NP/OP samples was treated with DNase, followed by Direct-zol RNA purification (Zymo Research) according to the manufacturer's instructions. After generation of amplified cDNA, metagenomic NGS libraries were constructed using a Nextera XT protocol as previously described[6-8].

EV-D68 genome sequencing by Sanger sequencing

For EV-D68 genome sequence recovery and to bridge gaps in the assemblies from shotgun and probe-enriched metagenomic next-generation sequencing (NGS), RT-PCR was performed using the Qiagen One-Step RT-PCR Kit in 20 µL total volume with 2 µL of extracted RNA and 12 pmol of each primer (Table S2). PCR conditions were 94°C, 30 s / 53°C, 30 s / 72°, 1 min with 40 cycles of

amplification. PCR products were visualized by 2% agarose gel electrophoresis and extracted for Sanger sequencing using the forward and reverse primers.

EV-D68 genome sequencing by targeted probe enrichment and NGS

XGen lockdown 120 nucleotide (nt) probes (IDT Technologies) were designed using ArrayOligoSelector[9] to tile across all fully sequenced EV-D68 genomes in the National Center for Biotechnology Information (NCBI) GenBank database, and were curated for redundancy using CD-HIT[10] at a 90% cutoff to a final set of 227 probes. Seven NGS libraries were constructed from cDNA as previously described[7], and enrichment was performed using the XGen lockdown protocol (IDT Technologies) and SeqCap EZ Hybridization and Wash Kit (Roche) . Probe-enriched NGS libraries were sequenced as a 250 nt paired-end run on an Illumina MiSeq. EV-D68 sequences were retrieved by BLASTn alignment to the Fermon strain at an E-value cutoff of 10^{-8}, followed by read mapping and consensus sequence generation in Geneious v6.1.8[11].

Validation of NGS libraries

NGS library concentration and average fragment size were determined using the KAPA Library Quantification Kit (Kapa Biosystems, Foster City, CA) and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA). NGS was performed on an Illumina MiSeq (150 nt single-end or paired-end reads) or HiSeq 2500 sequencer (100 or 150 nt paired-end reads).

Pathogen detection by metagenomic NGS

Sequencing data were analyzed for pathogen presence using the SURPI ("sequence-based ultrarapid pathogen identification") computational pipeline[6], which identifies viruses, bacteria, fungi, and parasites by computational subtraction of human host sequences, followed by nucleotide and translated nucleotide (protein) alignment to all microbial sequences in the National Center for Biotechnology Information (NCBI) GenBank database (June 2014). Raw SURPI output was filtered for misalignments by BLASTn analysis of sequences with unique assigned GenBank identifier (gi) numbers using an E-value threshold cutoff of 10^{-8}. Microbial reads were taxonomically classified to the appropriate rank (family, genus, or species) by use of a in-house developed classification algorithm incorporating the SNAP nucleotide aligner (v1.0beta.14)[12]. Summary read count tables were generated for pathogenic viruses, bacteria, and eukaryotic organisms outside of the phylum *Chordata* or kingdom *Viridiplantae*.

Phylogenetic and molecular clock analyses

Multiple sequence alignments were performed in Geneious v8.0 using MUSCLE[13]. Phylogenetic trees were constructed in Geneious using the MrBayes algorithm[14] at default parameters with the 1962 Fermon strain of EV-D68 as an outgroup. Molecular clock analyses were performed using BEAST v1.8.1 as previously described[16, 17].

Accession numbers

EV-D68 sequences generated in this study have been deposited in GenBank (EV-D68 VP1: accession number KM892497, KM892498, KM892502, KP126908, KP126909, KP126910, KP126912, EV-D68 genomes, full or partial: KM892499−KM892501, KP100792−KP100796 , KP126911, KP322752).

NGS data with human sequences removed have been deposited in the NIH Sequence Read Archive (accession numbers pending).

SUPPLEMENTARY RESULTS

Interpretation of viral sequences in NGS data

The ability to detect viral infections from CSF samples by metagenomic NGS at the depth of sequencing performed (13 to 117 million reads) was demonstrated in the current study by detection of EV-A71 in CSF from a patient with an unknown encephalitis (Table 2, Figure S3, and Table S1).

NGS libraries constructed from NP/OP samples were treated with DNase following nucleic acid extraction to reduce background from the human host and bacterial flora. As this protocol reduces sensitivity of detection and speciation for non-viral microbes (i.e. bacteria, fungi, and parasites), only viral sequences are shown for the NP/OP samples. The ability to detect DNA viruses is also impacted by the use of DNAse and we cannot exclude the possibility that our data is biased by reduced sensitivity for detection of DNA viruses.

Viral sequences in the metagenomic NGS data were interpreted as follows. Read count tables (Table S4) were generated reporting the number of reads found to align to a unique viral species for each indexed clinical sample. If reads were found to align to only one unique viral species within a family, the read counts corresponding to that species were directly reported in Tables 2 and S1. If reads to multiple viral species within the same genus or family were detected, read counts reported in Table 2 and S1 were obtained directly from the SURPI-generated coverage maps (Figure S3), which selects the best viral species hit on the basis of percent mapped read coverage[6]. Only reads from known animal or human viruses were reported in Tables 2 and S1. In addition, viral reads attributed to known, previously characterized sources of contamination were excluded from Tables 2 and S1, as follows:

Laboratory contamination from PCR amplicons**.** Sparse reads (n=7) corresponding to simian adenovirus E were found sporadically and were traced back to simian adenoviruses previously sequenced at University of California, San Francisco[18, 19]. Sparse reads to influenza A/B were also rarely found and mapped only to the NA/HA segments used for clinical laboratory typing. As both clinical laboratory testing as well as confirmatory follow-up PCR testing to recover influenza sequences from other regions were negative, and as these reads had been seen before in other NGS datasets, the influenza A/B reads were attributed to laboratory contamination.

Reagent contamination. Reads corresponding to bacterial phages in the *Inoviridae*, *Myoviridae*, *Podoviridae*, and *Siphoviridae* families were found across all samples and are common contaminants of NGS preparations.

Cross-contamination in the flow cell during NGS. Despite the use of dual indexes, we have observed rare cross-barcode contamination, which occurs when a sample contains a large proportion of reads corresponding to a single pathogen. In October 2014, a previously unknown respiratory sample positive for human rhinovirus A24 (US/CO/14-95) was sequenced in parallel with other CSF and respiratory samples, generating >18 million rhinovirus reads (50% of the total reads), and also cross-contaminating the other samples with inadvertent rhinovirus reads. When the sequencing run was repeated without including US/CO/14-95, no rhinovirus A24 read were observed in any of the remaining libraries, proving that the rhinovirus A24 reads in other samples were due to cross-contamination (Table S4).

Interpretation of sequences from divergent viruses in NGS data

In a CSF sample from an AFM patient testing EV-D68-negative (US/CA/14-6010), 52 reads corresponding to a divergent parvovirus closest in identity to bovine parvovirus were found, generating 3 assembled contiguous sequences (contigs). These contigs were identical to those found in fetal bovine serum from the laboratory (data not shown). Bovine parvovirus reads have previously been described as contaminants of fetal bovine serum[20].

Interpretation of bacterial sequences in NGS data

No sequences corresponding to credible neuropathogenic bacteria or mycobacteria were found in CSF from 14 AFM patients. Two CSF samples (US/CA/14-6067 and US/CA/14-6070) were highly contaminated from sequences to multiple bacterial and fungal organisms, so interpretation of results from those reads was not possible. No signatures corresponding to potentially neuropathogenic *Bartonella, Brucella, Rickettsia,* or *Coxiella* bacteria were detected at a threshold cutoff of ≥10 reads. Sparse reads were observed to clinically relevant bacteria that are also common environmental contaminants, including *Elizabethkingia* LA1-18, *Acinetobacter haemolyticus, Enterococcus faecium*, and *Enterobacter*, but these bacterial species, if truly present in CSF, should have been culture-positive and have never been associated with AFM, thus deemed most likely to be due to environmental laboratory contamination.

Interpretation of eukaryotic (fungal/parasitic) sequences in NGS data

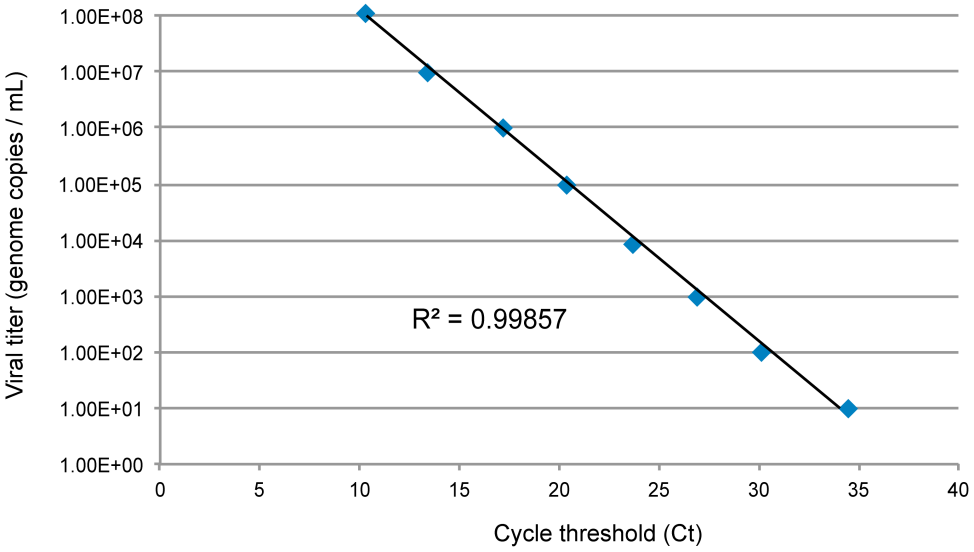
We only included reads mapping to potentially pathogenic families of fungi or eukaryotic parasites, as follows: *Ancylistaceae, Ancylostoma, Ascaris, Aspergillus, Babesia, Balantidium , Basidiobolaceae, Baylisascaris, Blastocystis, Blastomyces , Brachiola, Brugia, Candida, Clonorchis, Coccidioides, Cryptococcus, Cunninghamellaceae, Cyclospora, Dientamoeba, Diphyllobothrium, Dracunculus, Echinococcus, Encephalitozoon, Entamoeba, Enterobius, Enterocytozoon, Epidermophyton, Fasciola, Fasciolopsis, Fonsecaea, Geotrichum, Giardia, Gnathostoma, Histoplasma, Hortaea, Hymenolepis, Isospora, Leishmania, Malassezia, Metagonimus, Microsporidium, Mucoraceae, Naegleria, Necator, Nosema, Onchocercidae, Paracoccidioides, Paragonimus, Piedraia, Plasmodium, Pleistophora, Pneumocystis, Rhinosporidium, Saksenaeaceae, Sarcoptes, Schistosoma, Sporothrix, Strongyloides , Syncephalastraceae, Taenia, Thamnidiaceae, Toxocara, Toxoplasma, Trachipleistophora, Trichinella, Trichomonas, Trichophyton, Trichosporon, Trichuris, Trypanosoma, Vittaforma,* and *Wuchereria.* All putative eukaryotic sequences within these families were manually checked using BLASTn at a 10^{-8} E-value cutoff against the NCBI GenBank database (August 2014).

The only eukaryotic sequences corresponding to a potential human pathogen were from *Malassezia* spp. in 5 CSF samples from AFM patients from Colorado (Table S8)[21]. *Malassezia* spp. are considered part of the normal flora of the skin[21], so these reads were attributed to inadvertent contamination from skin flora during lumbar puncture and not deemed significant. In addition, all routine fungal cultures of CSF from AFM patients, which would support growth of *Malassezia* spp., were negative (Table 2).

SUPPLEMENTARY FIGURES

Figure S1. Quantification of EV-D68 viral genome copy number. (A) A standard curve was generated to correlate copy number with cycle threshold (Ct) in a SYBR-Green RT-PCR assay targeting the EV 5'UTR. The titer in viral copies per mL (y-axis) is plotted against the Ct (x-axis). **(B)** Table of viral copy numbers corresponding to EV-D68-positive samples.

A



B

Patient	Disease	Clinical sample type	Ct	Copy # / µL	Adjusted copy # / mL
US/CA/12-5641	AFM	NP swab	29·7	179	53,575
US/CA/12-5837	AFM	NP swab	32·1	37	11,141
US/CA/14-6067	AFM	NP wash (8/18/14)	28·4	435	32,605
		NP/OP swab (8/18/14)	34·3	8	2,396
US/CA/14-6070	AFM	OP swab (8/24/14)	29·8	172	12,863
		NP/OP swab (8/25/14)	30·2	131	39,293
		OP swab (8/26/14)	32·8	23	6,998
		blood (8/25/14)	33·1	19	5,565
		stool (8/26/14)	34·7	6	458
US/CA/14-6089	AFM	OP swab	28·4	441	132,190
US/CA/14-6092	AFM	NP swab	27·2	1,016	304,906
US/CA/14-6100	AFM	NP wash (10/6//14)	24·9	4,789	1,436,842
		NP wash (10/6/14)	23·1	15,474	1,160,572
		OP swab (10/6/14)	28·8	328	24,567
US/CO/13-60	AFM	NP swab	26·4	1,731	519,290
US/CO/14-86	AFM	NP swab	22·0	31,828	9,548,532
US/CO/14-93	AFM	NP swab	24·9	4,569	1,370,626
US/CO/14-94	AFM	NP swab	31·3	61	18,223
US/CO/14-98	AFM	NP swab	no Ct	n/a	n/a
US/CA/11-1767	Encephalitis	NP swab	16·2	1,663,602	499,080,509
US/CA/14-R2	URI	NP swab	30·8	86	6,425
US/CA/14-R1	URI	NP swab	22·0	32,260	2,419,529
US/CA/14-6100SIB	URI	NP swab (10/8/14)	31·8	45	3,410
		OP swab (10/8/14)	26·9	1,219	91,440
US/CA/10-786	URI	culture supernatant	15·9	2,064,039	619,211,828
US/CA/09-871	URI	NP swab	23·4	12,813	3,843,900

Figure S2. Molecular clock analysis of EV-D68 by VP1 gene sequence. All 180 complete EV-D68 VP1 sequences available in GenBank as of Dec 2014, including the 17 new EV-D68 VP1 gene sequences in this study (boldface), were aligned using MUSCLE[13], and phylogenetic trees were constructed using the MrBayes algorithm[14]. EV-D68 strains from AFM patients are grouped together in a novel clade (clade B1) and include sequences from patients with severe respiratory illness from the 2014 outbreak. AFM cases are marked in red text, encephalitis in orange text, and respiratory illness only in blue text. Labeled nodes show the divergence-time estimates in years from January 1st, 2015. Abbreviations: nt, nucleotide.

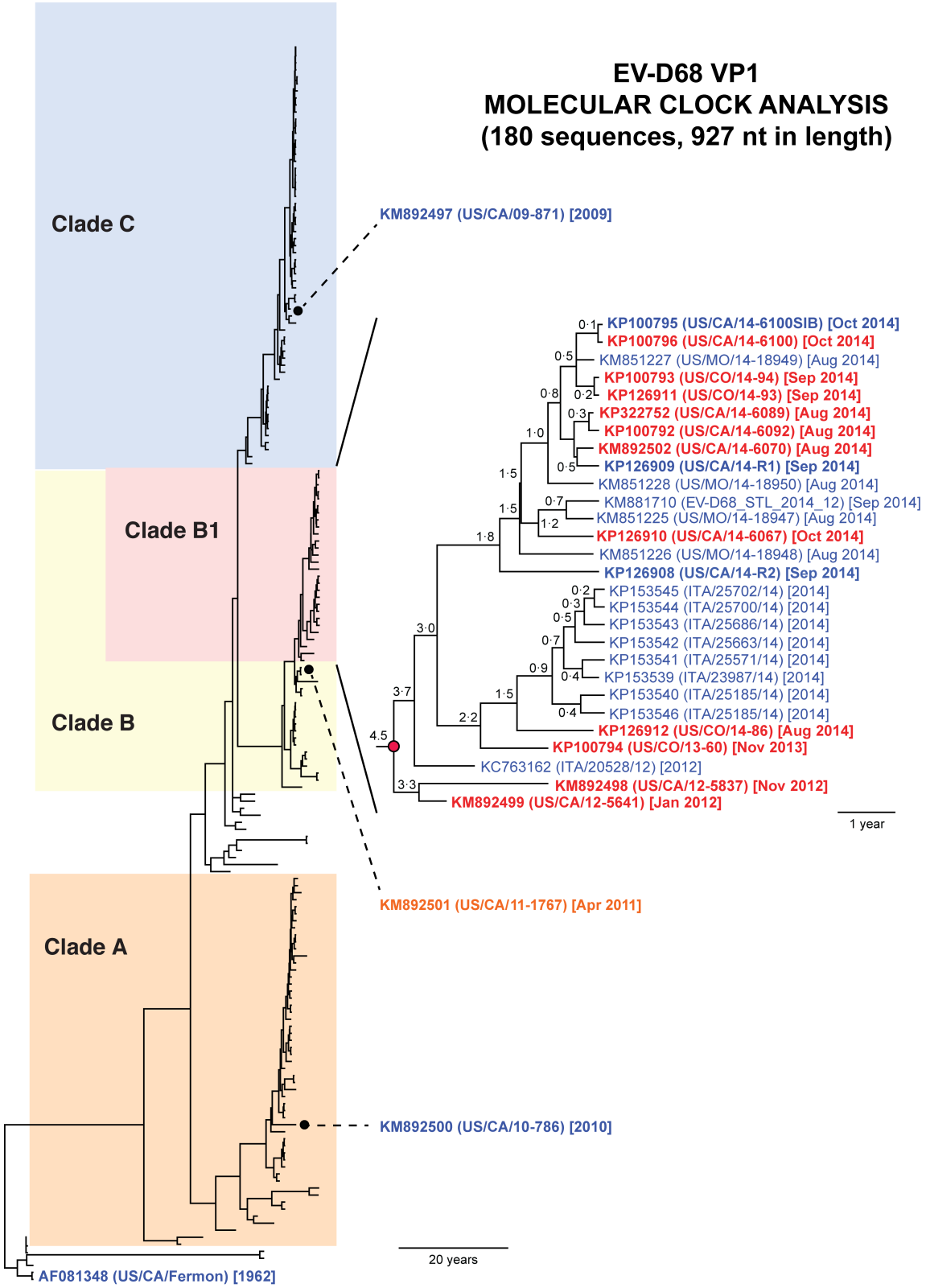
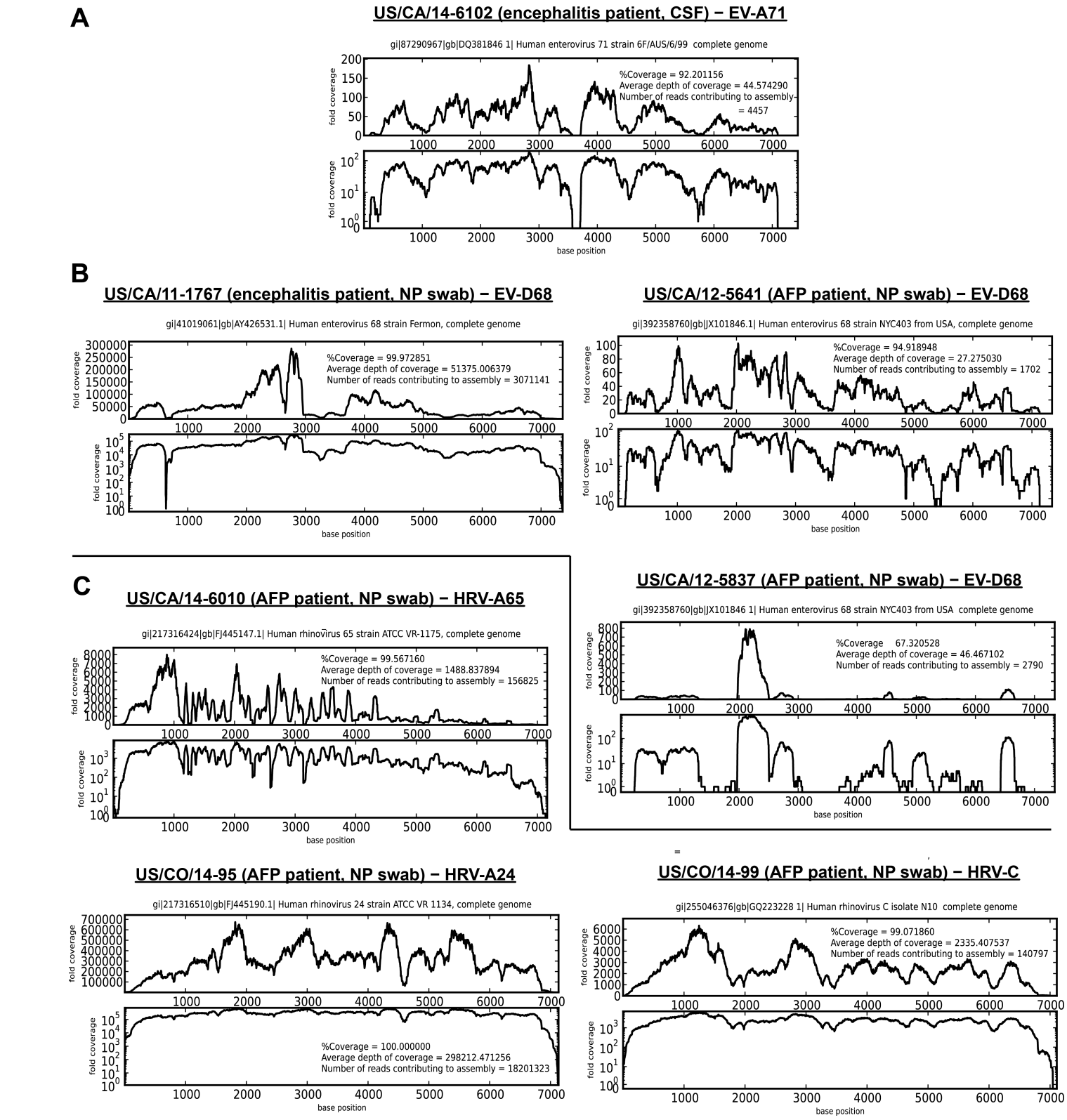


Figure S3. Molecular clock analysis of EV-D68 by VP1 gene sequence (large PDF). All 180 complete EV-D68 VP1 sequences available in GenBank as of Dec 2014, including the 17 new EV-D68 VP1 gene sequences in this study, were aligned using MUSCLE[13], and phylogenetic trees were constructed using the MrBayes algorithm[14]. Each branch is labeled with the GenBank accession number, and nodes display the divergence-time estimates in years from January 1st, 2015.

Figure S3.pdf

Figure S4. Genome coverage plots of enteroviruses and rhinoviruses. Coverage plots were automatically generated from metagenomic NGS data using the SURPI computational pipeline[6], and the y-axes are displayed using both a linear and logarithmic scale. The closest species match in the GenBank database, percent coverage, average depth of coverage, and number of reads contributing to the assembly are shown. **(A)** EV-A71 in CSF from a patient with encephalitis. **(B)** EV-D68 in NP/OP swabs from patients with AFM. **(C)** Rhinoviruses detected in NP/OP swabs with individual patients with AFM.



SUPPLEMENTARY TABLES

Table S1. Microbiological testing of AFM and EV-D68-positive patients (expanded)

TableS1.xlsx

Table S2. Primers used for EV-D68 detection and sequencing.

Purpose	PCR Test	Name	Sequence
screening	**Enterovirus/Rhinovirus 5'UTR [heminested RT-PCR]** Rd 1: DK001-Hemi-F / DK004-Outer-R Rd2: DK479-Inner-R / DK004-Outer-R	DK001-Hemi-F	CAAGCACTTCTGTTTCCC
		DK004-Outer-R	CACGGACACCCAAAGTAGT
		DK479-Inner-R	ATTCAGGGGCCGGAGGA
screening	**EV-D68 VP1 [heminested RT-PCR]** Rd 1: EVD68-628F-Hemi / EVD68-868R-Outer Rd 2: EVD68-628F-Hemi / EVD68-785R-Inner	EVD68 628F-Hemi	GATGGCTTTGCYGGATTTGAGAA
		EVD68 868R-Outer	TGAGYGCATTTGGTGCTCTTTCT
		EVD68 785R-Inner	GCCCATGCTTTTATGTGTTTAGG
quantification	**Enterovirus/Rhinovirus 5'UTR [SYBR-Green qRT-PCR]**	EV-KaresF	CGGCCCCTGAATGCGGCTAA
		EV-KaresR	GAAACACGGACACCCAAAGTA
sequencing	**EV-D68 VP1-2A [heminested RT-PCR]** Rd 1: VP1-Inner-F / VP1-Outer-R Rd 2: VP1-Inner-F / VP1-Inner-R	VP1-Outer-R	AGAGATTGTTTCTGTAAACCCTGCTC
		VP1-Inner-F	ACGTTGGCTATGTCACTTGTT
		VP1-Inner-R	TATAAGCAGTAATGCCTTGTTCC
viral genome sequencing	**EV-D68 [RT-PCR]**	EVD68 1F	TTAAAACAGCCTTGGGGTTGTTC
		EVD68 950R	GTAATACTCTATCACTGTAGCC
		EVD68 774F	GGATCTCATATCACATACAATCAG
		EVD68 1790R	CACTTCTAGCATGTTRCGGAC
		EVD68 1620F	GCTCCAATGTGTTGTGAGTT
		EVD68 2830R	CATTGCTTGAAGTGTCAAGTCAG
		EVD68 2530F	GAAGCCATACAAACTCGCAC
		EVD68 3560R	GTACCTTGCYGGGTAATATTC
		EVD68 3530F	GAAGGCCCAGGGATYCAATG
		EVD68 4410R	CATGTATTATTAAACAAACCGGTTC
		EVD68 4380F	CAGTTCAAGTCCAAATCTCGC
		EVD68 5650R	GCACAGCRTCATTRTAATCATCC
		EVD68 5490F	CATCTGTTGGAGAAACCATTTAC
		EVD68 6590R	CAGTGTARTCAAAAGCAAAGATTTC
		EVD68 6360F	CCTTTTGTAACYTTTGTAAAAGATG
		EVD68 7180R	CAAAATTCATTRTARGCCTCCTC
		EVD68 7070F	CAGTRATGCCAATGAAGGAGATAC
		EVD68oligodT-tag	GCTCGCGAGCGCGTTTAAACGCGCA-CGCGTTTTTTTTTTTTTTTTTTVN
viral genome sequencing (filling gaps from metagenomic NGS	**EV-D68 [RT-PCR]**	EVD68-7F	CAGCCTTGGGGTTGTTCCC
		EVD68-170R	AGTGCTTGCTCATGAGAGCGC
		EVD68-5110F	CAGTGGATTCTCAAGAAGTTAGGG
		EVD68-5644R	CTCATGTCTGGGCAGAAAATGTC
		EVD68-6746F	AATGGTGGAATGCCCTCTGGTT
		EVD68-7058R	GATGTATGAGAAAGGGGTATTGATC

Table S3. Summary table of NGS read counts.

TableS3.xlsx

Table S4. Summary table of viral read counts

TableS4.xlsx

Table S5. Summary table of bacterial read counts

TableS5.xlsx

Table S6. Summary table of fungal and parasitic read counts.

TableS6.xlsx

REFERENCES

1. Messacar K, Schreiner TL, Maloney JA, et al. A cluster of acute flaccid paralysis and cranial nerve dysfunction temporally associated with an outbreak of enterovirus D68 in children in Colorado, USA. *Lancet* 2015.
2. Pastula DM, Aliabadi N, Haynes AK, et al. Acute neurologic illness of unknown etiology in children - colorado, august-september 2014. *MMWR Morbidity and mortality weekly report* 2014; 63(40): 901-2.
3. Imamura T, Suzuki A, Lupisan S, et al. Detection of enterovirus 68 in serum from pediatric patients with pneumonia and their clinical outcomes. *Influenza and other respiratory viruses* 2014; 8(1): 21-4.
4. Kares S, Lonnrot M, Vuorinen P, Oikarinen S, Taurianen S, Hyoty H. Real-time PCR for rapid diagnosis of entero- and rhinovirus infections using LightCycler. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 2004; 29(2): 99-104.
5. Lee D, Das Gupta J, Gaughan C, et al. In-depth investigation of archival and prospectively collected samples reveals no evidence for XMRV infection in prostate cancer. *PloS one* 2012; 7(9): e44954.
6. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome research* 2014; 24(7): 1180-92.
7. Naccache SN, Peggs KS, Mattes FM, et al. Diagnosis of Neuroinvasive Astrovirus Infection in an Immunocompromised Adult With Encephalitis by Unbiased Next-Generation Sequencing. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2015.
8. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *The New England journal of medicine* 2014; 370(25): 2408-17.
9. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, DeRisi JL. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome biology* 2003; 4(2): R9.
10. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; 22(13): 1658-9.
11. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012; 28(12): 1647-9.
12. Zaharia M, Bolosky WJ, Curtis K, et al. Faster and more accurate sequence alignment with SNAP. *arXiv* 2011; 1111.5572.
13. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 2004; 32(5): 1792-7.
14. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001; 17(8): 754-5.
15. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 1987; 4(4): 406-25.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 2012; 29(8): 1969-73.
17. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 2014; 345(6202): 1369-72.
18. Chen EC, Yagi S, Kelly KR, et al. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS pathogens* 2011; 7(7): e1002155.
19. Chiu CY, Yagi S, Lu X, et al. A novel adenovirus species associated with an acute respiratory outbreak in a baboon colony and evidence of coincident human infection. *mBio* 2013; 4(2): e00084.
20. Gagnieur L, Cheval J, Gratigny M, et al. Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture. *Biologicals : journal of the International Association of Biological Standardization* 2014; 42(3): 145-52.
21. Jagielski T, Rup E, Ziolkowska A, Roeske K, Macura AB, Bielecki J. Distribution of Malassezia species on the skin of patients with atopic dermatitis, psoriasis, and healthy volunteers assessed by conventional and molecular identification methods. *BMC dermatology* 2014; 14: 3.