bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Intra- and interpatient evolution of enterovirus D68 analyzed by whole-genome deep sequencing

Robert Dyrdak,[1,2] Monika Mastafa,[1,2] Emma B. Hodcroft,[3,4] Richard A. Neher,[3,4] and Jan Albert[1,2]
[1]*Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden*
[2]*Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden*
[3]*Biozentrum, University of Basel, Basel, Switzerland*
[4]*Swiss Institute of Bioinformatics, Basel, Switzerland*

(Dated: March 4, 2019)

Worldwide outbreaks of enterovirus D68 (EV-D68) in 2014 and 2016 have caused serious respiratory and neurological disease. To investigate diversity, spread, and evolution of EV-D68 we performed near full-length deep sequencing in 54 samples obtained in Sweden during the 2014 and 2016 outbreaks. In most samples, intrapatient variability was low and dominated by rare synonymous variants, but three patients showed evidence of dual infections with distinct EV-D68 variants from the same subclade. Interpatient evolution showed a very strong temporal signal, with an evolutionary rate of 0.0039 ± 0.0001 substitutions per site and year. Phylogenetic trees reconstructed from the sequences suggest that EV-D68 was introduced into Stockholm several times during the 2016 outbreak. Putative neutralization targets in the BC and DE loops of the VP1 protein were slightly more diverse within-host and tended to undergo more frequent substitution than other genomic regions. However, evolution in these loops did not appear to have been driven the emergence of the 2016 B3-subclade directly from the 2014 B1-subclade. Instead, the most recent ancestor of both clades was dated to 2009. The study provides a comprehensive description of the intra- and interpatient evolution of EV-D68, including the first report of intrapatient diversity and dual infections. The new data along with publicly available EV-D68 sequences are included in an interactive phylodynamic analysis on `nextstrain.org/enterovirus/d68` to facilitate timely EV-D68 tracking in the future.

INTRODUCTION

Enterovirus D68 (EV-D68) has been recognized as an emerging pathogen after recent worldwide outbreaks of serious respiratory and neurological disease. The virus was discovered in 1962 in California, USA in four children with pneumonia (Schieble *et al.*, 1967). Prior to 2014, EV-D68 was reported only sporadically with a total of 699 confirmed cases in Europe, Africa, and southeast Asia between 1970 to 2013 (as reviewed by Holm-Hansen *et al.* (2016)). But in 2014, 2,287 of cases with EV-D68 infection were reported, mainly in North America and Europe, but also in southeast Asia and Chile (Holm-Hansen *et al.*, 2016). Following the 2014 outbreak, another wave of EV-D68 infections was observed in 2016 with reports of outbreaks in several parts of the world, including Europe (Barnadas *et al.*, 2017; Dyrdak *et al.*, 2016; Knoester *et al.*, 2017; Piralla *et al.*, 2018), USA (Messacar *et al.*, 2017; Wang *et al.*, 2017), Argentina (Ruggieri *et al.*, 2017), and Taiwan (Wei *et al.*, 2018).

Even though EV-D68 was rarely reported before the recent outbreaks, a high prevalence of neutralizing antibodies have reported in samples collected before 2014 (Vogt and Crowe Jr, 2018). This indicates that infection with EV-D68 has been common also before the outbreaks in 2014 and 2016.

The main clinical presentation of EV-D68 infection is mild to severe respiratory symptoms (Holm-Hansen *et al.*, 2016), but in rare cases EV-D68 can also cause a poliomyelitis-like disease termed acute flaccid myelitis (AFM) (Dyda *et al.*, 2018; Messacar *et al.*, 2018). An increase of this otherwise unusual clinical presentation coincided with the outbreak of EV-D68 in 2014. In animal studies, strains from 2014 were more virulent than the prototype Fermon strain isolated in 1969 (Hixon *et al.*, 2017b; Zhang *et al.*, 2018a).

EV-D68, as other enteroviruses, has a genome size of approximately 7.5 kb and codes for a polyprotein, which is processed into four structural (VP1-4), and seven non-structural proteins (2A-C, 3A-D). Phylogenetically EV-D68 is divided into three clades (A, B and C) (Tokarz *et al.*, 2012), and the B clade is further subdivided into three sub-clades or lineages (B1, B2 and B3) (Gong *et al.*, 2016). The 2014 and 2016 outbreaks were caused by viruses belonging to the B1 and B3 subclades, respectively (Barnadas *et al.*, 2017; Dyrdak *et al.*, 2016;

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Knoester *et al.*, 2017; Piralla *et al.*, 2018; Wang *et al.*, 2017; Wei *et al.*, 2018; Zhang *et al.*, 2016). Similar to other RNA viruses, EV-D68 has been reported to have a high substitution rate with $6.2 \cdot 10^{-3}$ (Tokarz *et al.*, 2012) or $5.12 \cdot 10^{-3}$ (Ny *et al.*, 2017) substitutions per site per year in the highly variable VP1 gene.

Mechanisms of immune protection and immune escape for EV-D68 are not well understood. Neutralizing antibodies have been shown to be protective in animal models (Dai *et al.*, 2018; Hixon *et al.*, 2017a,b; Zhang *et al.*, 2018a,b). Based on homology with the neutralizing sites NIm-IA and NIm-IB on human rhinovirus 14 (HRV-14), regions on the BC and DE loops in EV-D68 have been proposed as putative epitopes for neutralizing antibodies (Liu *et al.*, 2015). However, also other sites in VP1 and other proteins may be of importance. Thus, amino acid variability in regions flanking the VP1 loops were reported to correlate with differences in neutralisation sensitivity between the prototype Fermon strain and strains from 2014 (Zhang *et al.*, 2015). For coxsackie virus B4, it was shown that a deletion in the BC loop reduced the neutralizing effect of CVB4-specific antisera (McPhee *et al.*, 1994). VP2 and VP3 may also contain neutralizing epitopes as demonstrated in early poliovirus studies (Minor *et al.*, 1986) and as recently reviewed for the enterovirus A genus (Fang and Liu, 2018). Little is known about the contribution of cellular immunity to the clearance and evolution of EV-D68 (and other enteroviruses).

Here, we report deep sequencing of near full-length genomes on the Illumina next-generation sequencing (NGS) platform and analyze intra- and interpatient evolution of EV-D68 during the recent outbreaks. We show that within-host diversity is typically low, but that dual infection is not uncommon, suggesting high infection incidence during the outbreaks. To facilitate exploration of the data and future analyses, we combined the whole genome sequences reported here with publicly available genomes and implemented an interactive visualization in `nextstrain` available at nextstrain.org/enterovirus/d68.

MATERIALS & METHODS

Study population and samples

For this study we used 54 nasopharyngeal and 5 other respiratory samples from two earlier studies on EV-D68 outbreaks in Stockholm in 2014 and 2016 (Dyrdak *et al.*, 2016, 2015). The 59 samples were drawn from 54 patients. Six patients had been sampled in 2014, and 48 patients (53 samples) in 2016. Thirty of the patients were female. The median age was 3.6 years (range 2 months – 63 years). Five patients were sampled twice during their acute illness (0, 1, 1, 2 and 7 days apart).

All, but one, samples were submitted from hospitals in Stockholm. However, we cannot exclude the possibility that a few samples might be from patients who had been transferred for care to Stockholm from other parts of Sweden.

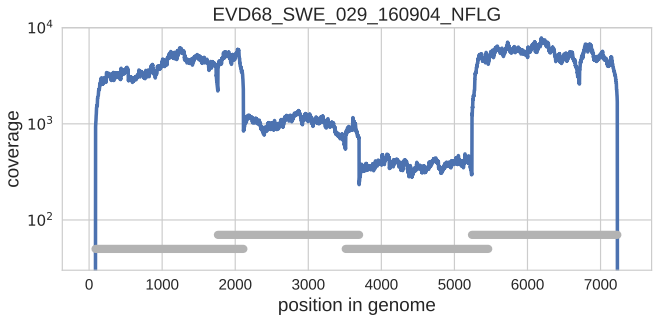

FIG. 1 **Example of coverage across the EV-D68 genome in a representative clinical sample.** Note that the coverage varies somewhat in different parts of the genome due slightly different quantity of amplified DNA from the four overlapping amplicons in the material that was used for sequencing library preparation. The four overlapping are indicated by the horizontal grey bars.

Details about the patient demographics are given in supplementary table S2.

The collection dates of the samples from the outbreak in 2016 ranged between August 19 to September 13. The samples had tested positive for EV-D68 at the Department of Clinical Microbiology at Karolinska University Hospital, by EV-D68 specific realtime PCR and/or partial VP4/VP2 sequencing. For details see (Dyrdak *et al.*, 2016, 2015). A few EV-D68-positive samples from the earlier studies were not included in this study because no sample material remained or because the virus levels were very low as evaluated by the cycle threshold (Ct) values in the EV-D68 realtime PCR. The Ct values for the included samples ranged from 13.9 to 33.9 (median 22.3). Details about the study samples are given in supplementary table S3.

The study was reviewed and approved by the Regional Ethical Review Board in Stockholm, Sweden (registration no. 2017/1317–32).

Near full-length EV-D68 amplification

Primers were designed to cover almost the entire EV-D68 genome in four overlapping fragments (F1–F4) that were approximately 2,000 base pairs long (Table S1). An alignment of published EV-D68 genomes was used to design forward and reverse primers targeting highly conserved regions of the EV-D68 genome, with similar melting temperatures, and with minimal tendency for hairpin and primer-dimer formation. Inner nested primers for fragment F1 were designed for template quantification.

RNA was manually extracted using the RNeasy Lipid Tissue Mini Kit (Qiagen Cat. No. 74804). For each na-

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sopharyngeal sample, 200 μl was taken and mixed with 1 ml Qiazol, 200 μl chloroform and 5 μl carrier-RNA. After centrifugation and several washing steps, each aliquot was eluted with 100 μl RNase-free water and were either stored at -70°C or used directly for PCR amplification.

Each sample was amplified in duplicate by one-step RT-PCR for each of the four overlapping fragments. The RT-PCR mixture contained: 11 μl of RNA template, 0.5 μM of forward and reverse PCR primer, 1 ng/μl random hexamers, 1 μl Superscript III RT/Platinum Taq HiFi Enzyme mix (Invitrogen, Stockholm, Sweden), 25 μl 2x reaction mix, and RNase-free water to a total volume of 50 μl. The PCR cycling profile was: $\times$1: 30 min at 50°C, 2 min at 94°C (cDNA synthesis); $\times$30: 15 sec at 94°C, 30 sec at 50°C, 90 sec at 68°C (PCR-amplification); $\times$1: 5 min at 68°C, ∞ at 4°C (final extension).

To approximately quantify the number of input EV-D68 RNA templates for sequencing, a dilution series of the RNA (1:10, 1:100, 1:1000) was amplified in duplicate with nested primers for fragment F1. EV-D68 RNA templates were enumerated by the Poisson distribution formula. The outer PCR is described above. The nested PCR mixture contained: 2.5 μl of product from the outer PCR, 0.2 μM of forward and reverse PCR primer, 1 unit of Platinum Taq HiFi (Invitrogen, Stockholm, Sweden), 5 μl 10x reaction mix, and RNase-free water to a total volume of 50 μl. The cycling profile was: $\times$1: 2 min at 94°C, (denaturation); $\times$30: 15 sec at 94°C, 30 sec at 50°C, 90 sec at 68°C (PCR-amplification); $\times$1: 6 min at 68°C, at 6 min (final extension).

Duplicates of each fragment were pooled and purified with AGENCOURT AMPure XP PCR purification kit and quantified with Qubit assays (Q32851, Life Technologies). Purified DNA from each fragment were diluted to the same concentration, pooled and sent to the Clinical Genomics Unit at Science for Life Laboratory (SciLifeLab, Stockholm, Sweden) for sequencing.

DNA library preparation and sequencing

The DNA preparations were quantified using Quant-iTTM dsDNA High-Sensitivity Assay Kit and Tecan Spark 10M plate reader. 1 μl of DNA (approximately 0.5–2.0 ng/μl) was used in the tagmentation reaction using Nextera chemistry (Illumina) to yield fragments larger than 150 bp. The tagmented library underwent 11 cycles of PCR with single-end indexed primers (IDT Technologies) followed by purification using Seramag beads according to the protocol. The library was quantified using Quant-iTTM dsDNA High-Sensitivity Assay Kit and Tecan Spark 10M plate reader and pair-end (2$\times$101 bp) sequenced to a depth of 100,000 – 1,000,000 reads per sample on a HiSeq 2500 Illumina sequencer. Base calling and demultiplexing was done using bcl2fastq v1.87, without allowing any mismatch in the index sequence.

Quality controls experiments

Quality controls were prepared from two samples with low Ct values (SWE_027_160829 and SWE_037_160829, having Ct values 14.62 and 19.27, respectively). The samples were diluted 1:100 in a pool of virus-negative nasopharyngeal samples and stored at -70°C in 200 μl aliquots. The enterovirus RNA concentration of the samples was quantified by limiting dilution in 10 replicas of ten-fold dilutions, from 1:10 to 1:100,000 with an in-house enterovirus realtime-PCR (Tiveljung-Lindell *et al.*, 2009). Based on Poisson calculations, the samples were estimated to have 4,400 and 240 EV-D68 RNA copies per μl, respectively, prior to dilution. The two samples were RNA extracted, PCR amplified and sequenced twice in two separate runs. The number of EV-D68 RNA templates in each PCR reaction was approximately 48,400 and 2,640, respectively.

Read filtering, mapping and analysis

Sequencing reads were trimmed using TrimGalore (Krueger, 2015; Martin, 2011) and mapped against the sequence `KX675261.1` using `bwa mem` (Li and Durbin, 2009). We used the `pysam` wrapper to samtools to generate a pile-up of all reads and quantify insertions relative to the reference. Custom python scripts were used to generate consensus sequences and quantify intrasample diversity from the pile-ups. All these steps were chained using the workflow engine `snakemake` (Köster and Rahmann, 2012). The entire pipeline is available `github.com/neherlab/EV-D68_analysis_Dyrdak_2019` and uses a code from a separate project `github.com/neherlab/SVVC`. Commit `81068ee` was used to analyze the data along with commit `a2cbb35` of `SVVC`. To analyze linkage disequilibrium, we calculated genotype frequencies p_{12} at pairs of positions that were covered by a read pair more than 100-fold and calculated their correlation $D_{12} = p_{12} - p_1 p_2$ (aka linkage disequilibrium). This correlation was normalized such that maximal linkage corresponds to $D_{12} = 1$ and complete repulsion to $D_{12} = -1$.

The consensus sequences for successfully sequenced samples have been deposited in GenBank (accession numbers MH674111–MH674166 and MH844544).

The scripts that analyze diversity in pile-ups and generate most of the the figures of this study are available at `github.com/neherlab/EV-D68_analysis_Dyrdak_2019`.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

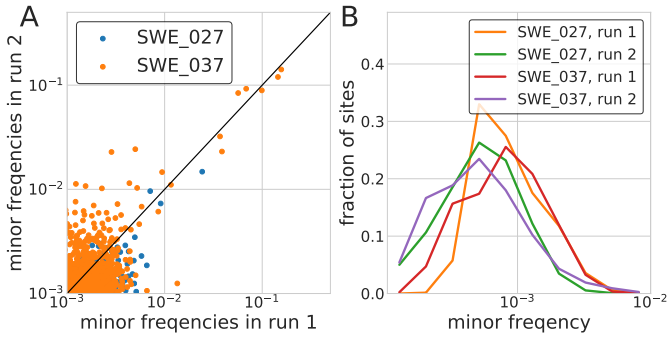

FIG. 2 **Accuracy of minor variant frequencies.** Panel A shows the consistency of minor variant frequencies across two replicate extraction, RT-PCR, and sequencing of two samples. Minor variation was relatively consistently recovered down to a frequency of about 1%. Panel B shows the distribution of the frequencies of non-consensus calls across all sites covered in excess of 2,000-fold. At the majority of sites, no variation above 0.001 is observed. Note that this variation includes within sample variation, RT-PCR errors, and sequencing errors.

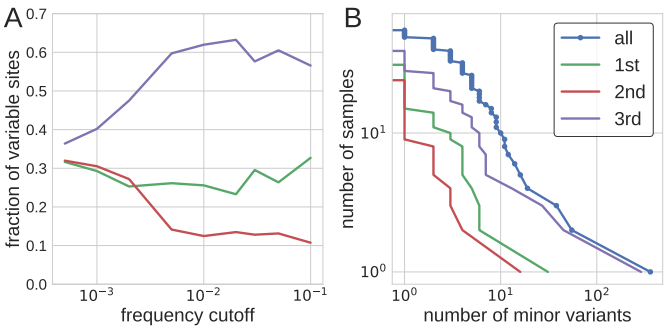

FIG. 3 **Within sample diversity.** Panel A shows the distribution of the number of variable sites in coding regions among different coding positions. At very low frequency cutoffs, variable sites are approximately equally distributed in codons. The distribution rapidly changes once the frequency cutoff exceeds 0.3%, beyond which variable sites are found predominantly at 3rd codon positions and more rarely at 1st and 2nd codon positions. Panel B shows inverse cumulative distribution of the number of samples that have more than a certain number of sites that are variable above a level of 3%. Separate distributions are shown for all sites and for 1st, 2nd, and 3rd codon position in the coding region of the genome.

Phylodynamics analysis

Sequences generated in this study were combined with sequences and metadata of all EV-D68 genomes with a length of at least 6,000 bp available on Virus Pathogen Resource (ViPR) (Pickett *et al.*, 2011) as of 2018-09-02. The combined data set was analyzed using `nextstrain` (Hadfield *et al.*, 2018) and snakemake (Köster and Rahmann, 2012). The `augur` pipeline was run using the aligner MAFFT (Katoh *et al.*, 2002), the tree-builder (Nguyen *et al.*, 2015), and the phylodynamic package TreeTime (Sagulenko *et al.*, 2018). The repository detailing the analysis pipeline is available at `https://github.com/neherlab/enterovirus_nextstrain`. Sub-genogrouping of sequences was done using the Enterovirus Genotyping Tool Version 1.0 (Kroneman *et al.*, 2011). The resulting analysis is visualized with `auspice` and is available at `nextstrain.org/enterovirus/d68`.

Five sequences were excluded from the phylodynamic analysis because they were deemed problematic. The following three sequences are within a few mutations from other sequences sampled several years earlier or later: EV-D68/Homo_sapiens/USA/U797/2007 is identical to strain EV-D68/Homo_sapiens/USA/C7791/2014; EV_D68/Homo_sapiens/USA/O622a/2012 is very similar to strains sampled in 2003; EV_D68/Homo_sapiens/USA/MO78/2009 is very similar to strains sampled in 2014 in the USA. Given the high evolutionary rate of enteroviruses, such stasis over many years is very unlikely (we expect about 30 changes per year) and these sequences are likely dated incorrectly. The case CA/RESP/10_786 is less clear, but the sequence is several standard deviations less diverged from the root than expected given its collection date. USA/TX/2014_19267 is likely a recombinant sequence.

Mixed effects model of evolution (MEME) analysis (Murrell *et al.*, 2012) using the DataMonkey package (Weaver *et al.*, 2018) was used to analyze for episodic diversifying selection.

RESULTS

Fifty-four of 59 samples yielded a coverage of at least 100x in all four fragments and were included in further analyses. Of the excluded samples, two did not yield sequence in any fragment (Ct values 30.99 and 33.89) and three samples lacked complete sequence in one of four fragments (Ct values 20.58, 22.95, and 32.31). All five excluded samples were from 2016. Fig. 1 shows the coverage plot for a representative sample. The sequencing coverage for each sample is presented in supplementary table S3.

Reproducibility of identification of intrasample single nucleotide variants (iSNVs)

In addition to consensus sequences, deep population sequencing delivers information about intrasample diversity. Low genome template input, biased amplification, RT-PCR errors and sequencing errors can all potentially skew or inflate intrasample diversity estimate (Zanini *et al.*, 2017). To control for such artifacts, we processed two high-titer patient samples in duplicate (see Methods). The accuracy of the detection and quantification

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

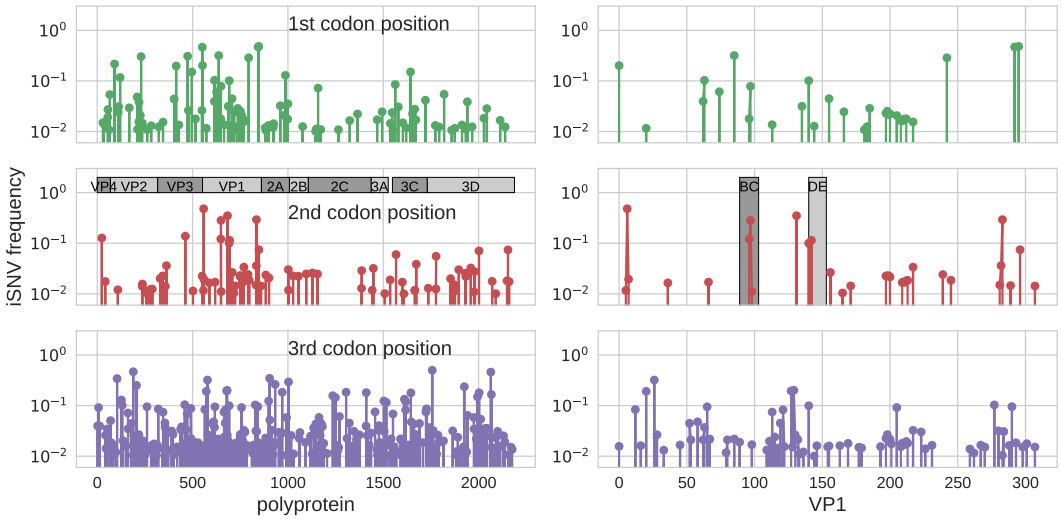

FIG. 4 **EV-D68 intrapatient variability across the genome and codon positions.** The three rows show how intrapatient variability is distributed across the EV-D68 genome for 1st, 2nd, and 3rd codon position, respectively. The panel on the left show iSNVs along the entire polypeptide, the panels on the right zoom into VP1 and highlight the BC and DE loops. This figure includes data from one sample per patient (the first sample, for patients sampled twice), but excludes the three dual infected samples (see main text).

of iSNVs was evaluated by comparing results from the two runs (Fig. 2A). Minor variants down to below 1% were consistently detected in both runs. Also, the estimated frequency of individual iSNVs showed good consistency between the two runs at frequencies above approximately 1% (Panel A), but a few iSNVs had markedly higher frequencies in one of the replicates (run 2 of sample SWE_037).

At the majority of sites, we observed background of minor variation at frequencies well below 0.1% (see Fig. 2B).

Reproducibility of detection of iSNVs was also investigated in the five patients were sampled twice during their acute illness (0, 1, 1, 2 and 7 days apart). Supplementary figure S3 shows that frequencies of iSNVs above the 1% level were largely concordant across the two consecutive samples, especially in patient SWE_037 (for whom the first sample was also sequenced twice) and less so in patient SWE_012. In patient SWE_021, with seven days between the samples, there was a tendency for an increased variation in the second sample Unfortunately, the short time between the samples prevented us from estimating the rate of intrapatient evolution or other longitudinal aspects of intrapatient evolution of EV-D68.

Analysis of intrasample variability in all 54 patient samples also provided information about our ability to detect true biological variation. Fig. 3A shows the distribution of sites with variation above a certain frequency cutoff among the three codon positions. The distribution was approximately even between codon sites at very low cutoffs below 10^{-3} and then rapidly changed around approximately 0.3% before plateauing at 1%.

Collectively, our results show that a majority of iSNVs at frequencies >1% represent true biological variation, whereas variation <0.1% mostly is due to amplification and sequencing errors. Between 0.1% and 1% an increasing proportion of variation represents true iSNVs. The level of accuracy achieved is comparable to what we reported earlier for a similar sequencing strategy for HIV-1 (Zanini *et al.*, 2017, 2016) and other viruses (Grubaugh *et al.*, 2018).

Intrapatient variability

As described above intrasample variability exceeding 1% likely represents true biological variation.

Approximately 60% of the variable sites above this level occurred at the 3rd codon position and around 30% and 10% in the 1st and 2nd codon positions, respectively (see Fig. 3A). Consequently, intrapatient variability is mostly synonymous (10%, 0%, and 95% of 1st, 2nd, and 3rd positions, respectively, admit synonymous mutations in our mapping reference).

Figure 3B shows that most (44 of 54) samples had fewer than 10 sites that displayed variability above a level of 3%. However, three samples had more than 20 variable sites, which suggested dual infections (see below). In the remaining 51 samples, both synonymous and non-synonymous substitutions were mostly scattered across the genome (Fig. 4). However, there was some clustering of minor variation in 1st and 2nd codon positions in the structural proteins (VP4 – VP1) including putative targets for neutralizing antibodies in BC and DE loops of VP1 (Liu *et al.*, 2015). The details about complete

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

amino acid substitutions and minor variation in these regions are given in supplementary tables S4 and S5.

Evidence of dual infections

As mentioned above, three samples (SWE_007, SWE_045, and SWE_046) drawn from three different patients showed more than 20 iSNVs. In all three samples, most minor variants were detected at a frequency around 10% (Fig. 5, left). This suggested the presence of two distinct variants; a major variant constituting around 90% of the virus population in the sample and a minor variant constituting around 10%. To investigate the origin of these minor variants we constructed a "minor" consensus sequence for each sample by exchanging the strict majority rule consensus nucleotide by the minor variant at positions where they exceeded a frequency of 1%. A phylogenetic tree analysis showed that the major and minor sequences were monophyletic in all samples except the three samples with indications of dual infections. For these samples the major and minor consensus sequences occupied different positions in the tree, see supplementary figure S1. To further investigate these putative dual infections we analyzed the linkage of iSNVs (Fig. 5, right). Almost all iSNVs occurred in complete linkage (LH=1) with neighboring iSNVs, which provides strong evidence for the presence of two distinct variants, rather than rapid evolution of intrapatient variation starting from a single transmitted variant.

Phylodynamics of EV-D68

The interpatient evolution and phylogenetic relationship of the EV-D68 variants from Sweden and 509 published EV-D68 sequences was investigated using the `nextstrain` platform available at `https://nextstrain.org/enterovirus/d68` (Hadfield *et al.*, 2018), which allows users to interactively explore the dataset. A screen shot is displayed in Fig. 6.

The `nextstrain` phylodynamic analysis showed that all Swedish EV-D68 variants from the 2016 outbreak belonged to the B3 lineage of EV-D68, which agrees with our earlier findings based on VP2/VP4 gene sequences (Dyrdak *et al.*, 2016). Within the B3 lineage our new Swedish sequences were interspersed with international sequences, primarily from the US, but monophyletic clusters of Swedish sequences were also observed (Fig. 6, right). A bootstrap analysis showed that most branches were well supported (Fig. S4).

These findings indicate that during the worldwide 2016 outbreak EV-D68 was introduced into Stockholm several times, and that these separately introduced variants thereafter spread locally.

Interestingly, the tree indicated that the 2016 B3 subclade did not evolve directly from the 2014 B1 subclade. Instead, the most recent common ancestor (MRCA) of the B1 and B3 subclades was estimated to have existed in first half of 2009 (90% CI Dec 2008 – Jul 2009). Many of the proximal sequences in the B1 and B3 subclades were sampled in East Asian countries, but this should be interpreted with caution since sampling probably has been incomplete and geographically biased. A tangle-tree analysis suggested recombination within subclades, but not between clades or subclades (supplementary figure S2), but also this finding should be interpreted with caution since there was limited intrasubclade variation.

Interpatient evolution

A plot of root-to-tip distances of EV-D68 sequences vs. time showed an extremely strong temporal signal (Fig. 7). We estimated an evolutionary rate of $3.8 \cdot 10^{-3} \pm 10^{-4}$ substitutions per site and year for the entire genome and $4.2 \cdot 10^{-3} \pm 4 \cdot 10^{-4}$ for the VP1 region, which is slightly lower than previous estimates (Ny *et al.*, 2017; Tokarz *et al.*, 2012)). 87% of all substitutions were synonymous.

We used `nextstrain` to investigate diversity across the EV-D68 genome in the sample containing the genomes reported in this study and database genomes. Ancestral sequences were reconstructed using the phylogeny in Fig. 6, and the number of transition events for each codon were counted along the tree. Fig. 8 shows that amino acid substitution events were observed in all EV-D68 proteins and varied between 0 and 27 events per codon. The five positions with the highest number of substitution events were: codon 234 in protein VP3 (27 events), 273 in 2C (20 events), 22 in 2A (15 events), 60 in VP3 (15 events), and 277 in 2C (14 events). None of these positions have been reported to be neutralizing epitopes. Codon 277 in 2C and codon 1 in VP1 showed evidence of episodic diversifying selection as per MEME analysis ($p < 0.05$).

The BC and DE loops in the VP1 protein showed relatively high diversity, especially in positions 95, 97, 98, 142, 143 and 148, suggesting that immune selection may have impacted on EV-D68 evolution (see Fig. 8, bottom panel, and for details see `nextstrain`).

We also used `nextstrain` to track on which branches in the EV-D68 phylogeny amino acid substitutions had evolved and been fixated. The branches leading to the MRCA of the B1 and B3 clades had many amino acid substitutions, including substitutions at codons 90, 95, 97, 98, 145 and 148 in the BC and DE loops. In contrast, no amino acid substitutions in the loops were observed on branches separating the B1 and B3 subclades. Similarly, only very few substitutions in the loops were observed within the B1 and B3 subclades.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

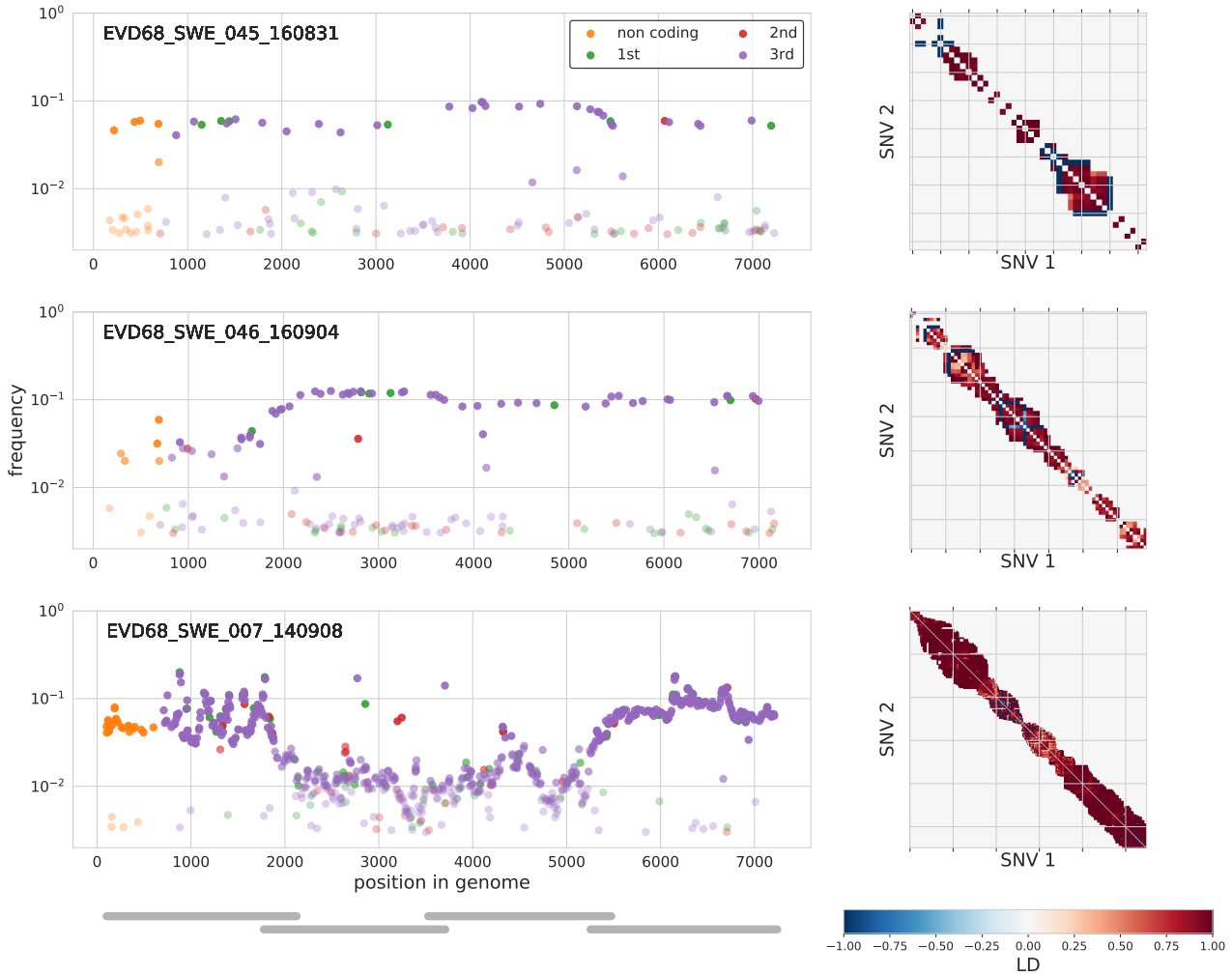

FIG. 5 **Minor iSNVs across the EV-D68 genome for three samples with putative dual infections.** These three samples had more than 20 highly covered sites with minor variants in excess of 3% frequency. On the left, iSNVs are colored by codon position or as non-coding, and given opacity depending on their frequency. The majority of these variants were at 3rd positions and had similar frequencies across the amplicons, indicated by grey bars in the lower left panel. iSNVs in three amplicons (amplicon 1 of sample SWE_046 and amplicons 2 and 3 of sample SWE_007), however, were found at substantially lower frequency, possibly due to primer mismatches. The three panels on the right are indexed by iSNV order on the genome and show linkage disequilibrium between iSNVs (>1%) close enough to each other that they were covered at least 100-fold by the same sequencing read. Almost all of these variants at in complete linkage (dark red). A number of iSNVs just above 1% in sample SWE_045 and SWE_046 are likely variants in the background of the dominant variant. Those are in complete 'anti-linkage' with neighboring iSNVs are $\sim$ 10% (dark blue).

DISCUSSION

We have performed a comprehensive investigation of the intra- and interpatient evolution of EV-D68 using new deep near full-length sequences from 54 Swedish patients sampled during the 2014 and 2016 outbreaks and available database sequences. The new data and previously published sequences are available on `nextstrain` to facilitate real-time tracking and further study of EV-D68 evolution and spread. Intrapatient variability was mostly low and dominated by rare synonymous substitutions, consistent with deep sequencing studies on influenza virus infections (Debbink *et al.*, 2017). However, three patients showed evidence of dual infections. The phylodynamic analysis indicated EV-D68 was introduced into Stockholm several times during the 2016 outbreak and then spread locally.

Three patients showed evidence of dual infection with two different EV-D68 variants. Even though co-infections with different enteroviruses has been reported for poliovirus (PV) and non-polio enteroviruses (NPEV) (Isaacs *et al.*, 2018; Melnick *et al.*, 1951; Parks *et al.*, 1967), this is to our knowledge the first report of dual infections with EV-D68. Dual infections during the 2016

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

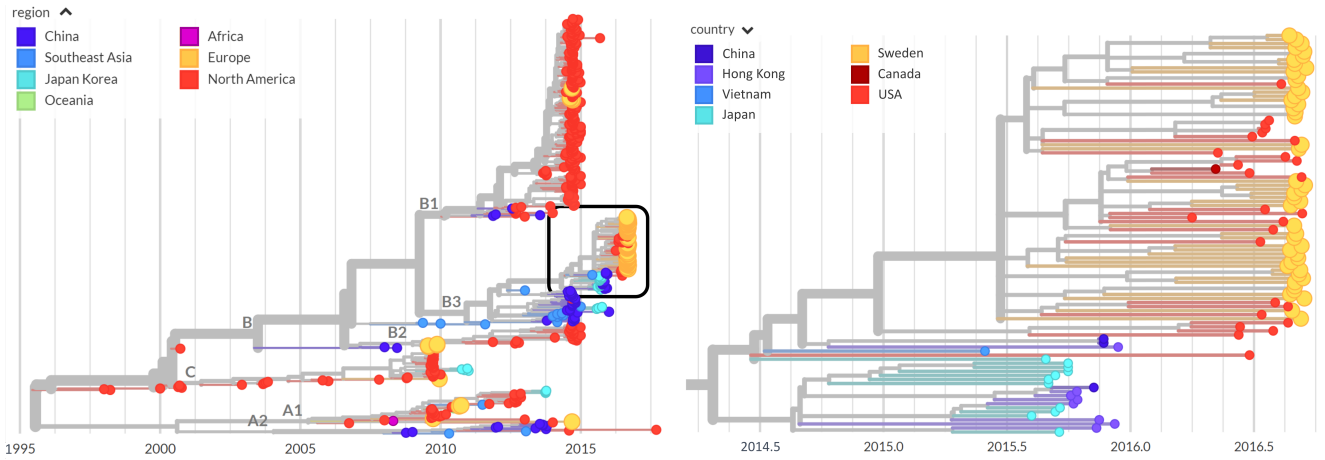

FIG. 6 **Phylogenetic analysis.** The left panel shows a phylogenetic tree as rendered by `nextstrain` with the sequences labelled by region. Those from Europe, including those from Sweden, are colored in orange. The right panel is a zoomed-in view of the boxed area on the left-hand panel colored by country and highlighting the Swedish sequences from the 2016 outbreak (orange). Swedish sequences often cluster together, but are interspersed with sequences from Canada (dark red) and the USA (red), implying multiple introductions that then spread locally.

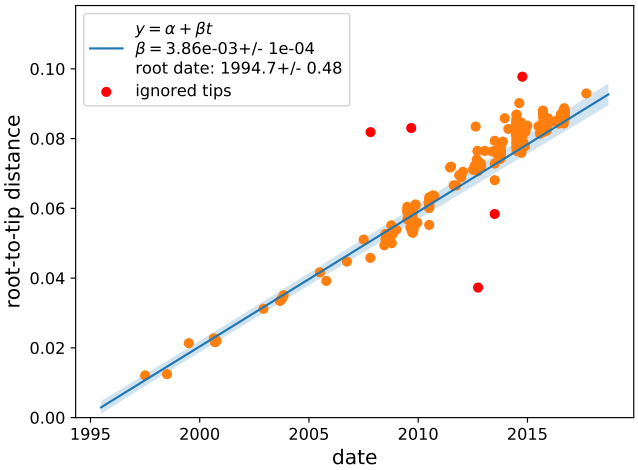

FIG. 7 **Temporal signal.** A scatter plot of root-to-tip distance vs time indicates an extremely strong temporal signal with an evolutionary rate of $\mu = 0.0039 \pm 0.0001$ substitutions per site and year.

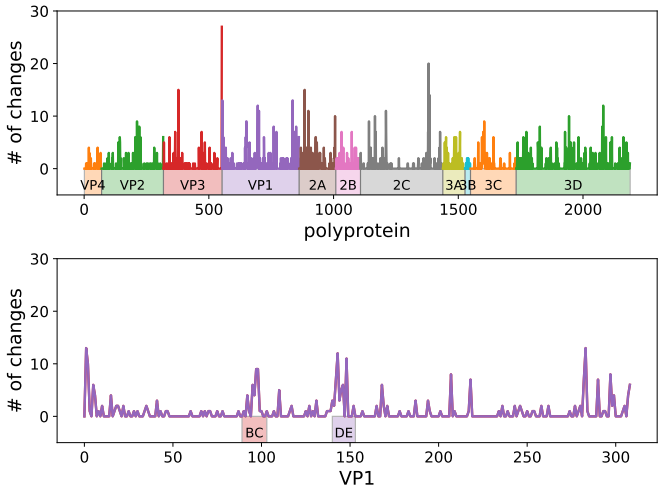

FIG. 8 **Interpatient diversity of EV-D68.** Amino acid substitution events on the phylogeny in Fig. 6 left were enumerated using `nextstrain`. Top panel shows the number of such events for each codon along the entire coding sequence. Bottom panel shows events in VP1, where peaks were observed in the BC and DE loops, as well as both ends of the protein.

outbreak are not completely surprising in view of the high EV-D68 incidence during the outbreak (Dyrdak *et al.*, 2016) and the fact that the present study showed multiple introductions of the B3 subclade into Stockholm. We also found indications of intrasubclade recombination, but not between clades or subclades. Similar findings have been reported in three recent studies (Ny *et al.*, 2017; Tan *et al.*, 2015; Yip *et al.*, 2017). It is expected that recombination primarily would occur within subclades since recombination requires co-infection with two or more virus variants and EV-D68 variants that circulate during an outbreak or season tend to be closely related.

The EV-D68 genome was amplified in four overlapping amplicons that were sequenced on the Illumina HiSeq platform. Our results indicate that variation down to a frequency of 1% represented "true" iSNVs, which were reproducibly detected and quantified. In contrast, most variation below 0.1% represented noise from cDNA synthesis, PCR and sequencing. It would be interesting to compare the performance of our sequencing protocol with the pan-enterovirus protocol recently published by (Isaacs *et al.*, 2018). In the 54 samples that were sequenced to sufficient depth we observed that most nu-

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cleotides were conserved. Among iSNVs detected above the 1% level, most occurred at the 3rd codon position, i.e. were usually synonymous. The finding that intrapatient variation was limited and mostly synonymous was expected since infection usually appeared to have been established by a single virion and EV-D68 is an acute infection where is limited time for the virus to diversify within the infected host.

Albeit there is a bias introduced by sampling and typing being done only at certain centras, the phylodynamic analysis showed that the splits between EV-D68 clades and subclades occurred several years ago. For instance, the MRCA of the B1 and B3 subclades was estimated to have existed in 2009. Enterovirus infections, like influenza, has a seasonal pattern. However, in contrast to influenza, which infects people of all age groups and has a ladder-like phylogenetic pattern that indicates a strong immune selection, EV-D68 infects predominantly children, and its phylogenetic pattern does not suggest continuous immune escape (Grenfell *et al.*, 2004). In line with this, the branches leading to the B1 and B3 subclades did not have amino acid substitutions in the BC and DC loops, which are putative targets for neutralizing antibodies. However, it should be acknowledged that there is limited knowledge about targets for humoral and cellular immunity in EV-D68.

Our phylodynamic analysis suggested that the recent global outbreaks of EV-D68 might have been preceded by low-level circulation of EV-D68 in East Asia. However, this pattern should be interpreted with much caution since the number of available full-length EV-D68 sequences is low and the sequences have not been systematically collected in time or space. We hope that the `nextstrain` platform will allow for continuous collection and sharing of EV-D68 sequence data, which will be needed to obtain a better understanding of how new successful EV-D68 variants evolve and spread. Better knowledge about humoral and cellular immunity against EV-D68 is also needed.

ACKNOWLEDGEMENTS

We gratefully acknowledge expert sequencing service and advice by Karin Sollander, Cecilia Svensson and Valtteri Wirta at the Clinical Genomics Unit at Science for Life Laboratory (SciLifeLab), and Lina Thebo and Eva Eriksson at the Karolinska Institute. We would also like to thank the entire nextstrain team.

REFERENCES

Barnadas, C., S. E. Midgley, M. N. Skov, L. Jensen, M. W. Poulsen, and T. K. Fischer (2017), Journal of Clinical Virology **93**, 40.

Dai, W., C. Zhang, X. Zhang, P. Xiong, Q. Liu, S. Gong, L. Geng, D. Zhou, and Z. Huang (2018), Vaccine **36** (5), 653.

Debbink, K., J. T. McCrone, J. G. Petrie, R. Truscon, E. Johnson, E. K. Mantlo, A. S. Monto, and A. S. Lauring (2017), PLOS Pathogens **13** (1), e1006194.

Dyda, A., S. Stelzer-Braid, D. Adam, A. A. Chughtai, and C. R. MacIntyre (2018), Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin **23** (3), 10.2807/1560-7917.ES.2018.23.3.17-00310.

Dyrdak, R., M. Grabbe, B. Hammas, J. Ekwall, K. E. Hansson, J. Luthander, P. Naucler, H. Reinius, M. Rotzén-Östlund, and J. Albert (2016), Eurosurveillance **21** (46).

Dyrdak, R., M. Rotzén-Östlund, A. Samuelson, M. Eriksson, and J. Albert (2015), Infectious diseases **47** (10), 734.

Fang, C.-Y., and C.-C. Liu (2018), Expert review of vaccines (just-accepted).

Gong, Y.-N., S.-L. Yang, S.-R. Shih, Y.-C. Huang, P.-Y. Chang, C.-G. Huang, K.-C. Kao, H.-C. Hu, Y.-C. Liu, and K.-C. Tsao (2016), Medicine **95** (31).

Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes (2004), science **303** (5656), 327.

Grubaugh, N. D., K. Gangavarapu, J. Quick, N. L. Matteson, J. G. D. Jesus, B. J. Main, A. L. Tan, L. M. Paul, D. E. Brackney, S. Grewal, N. Gurfield, K. K. V. Rompay, S. Isern, S. F. Michael, L. L. Coffey, N. J. Loman, and K. G. Andersen (2018), bioRxiv , 383513.

Hadfield, J., C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R. A. Neher (2018), Bioinformatics 10.1093/bioinformatics/bty407.

Hixon, A. M., P. Clarke, and K. L. Tyler (2017a), The Journal of infectious diseases **216** (10), 1245.

Hixon, A. M., G. Yu, J. S. Leser, S. Yagi, P. Clarke, C. Y. Chiu, and K. L. Tyler (2017b), PLoS pathogens **13** (2), e1006199.

Holm-Hansen, C. C., S. E. Midgley, and T. K. Fischer (2016), The Lancet Infectious Diseases **16** (5), e64.

Isaacs, S. R., K. W. Kim, J. X. Cheng, R. A. Bull, S. Stelzer-Braid, F. Luciani, W. D. Rawlinson, and M. E. Craig (2018), Scientific reports **8** (1), 11889.

Katoh, K., K. Misawa, K.-i. Kuma, and T. Miyata (2002), Nucleic Acids Research **30** (14), 3059.

Knoester, M., E. H. Schölvinck, R. Poelman, S. Smit, C. L. Vermont, H. G. Niesters, and C. C. Van Leer-Buter (2017), Emerging infectious diseases **23** (1), 140.

Köster, J., and S. Rahmann (2012), Bioinformatics **28** (19), 2520.

Kroneman, A., H. Vennema, K. Deforche, H. Avoort, S. Penaranda, M. Oberste, J. Vinjé, and M. Koopmans (2011), Journal of Clinical Virology **51** (2), 121.

Krueger, F. (2015), "TrimGalore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files," .

Li, H., and R. Durbin (2009), Bioinformatics **25** (14), 1754.

Liu, Y., J. Sheng, A. Fokine, G. Meng, W.-H. Shin, F. Long, R. J. Kuhn, D. Kihara, and M. G. Rossmann (2015), Science **347** (6217), 71.

Martin, M. (2011), EMBnet.journal **17** (1), 10.

McPhee, F., R. Zell, B.-Y. Reimann, P. H. Hofschneider, and R. Kandolf (1994), Virus research **34** (2), 139.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Melnick, J. L., A. S. Kaplan, E. Zabin, G. Contreras, and N. W. Larkum (1951), Journal of Experimental Medicine **94** (6), 471.

Messacar, K., E. J. Asturias, A. M. Hixon, C. Van Leer-Buter, H. G. M. Niesters, K. L. Tyler, M. J. Abzug, and S. R. Dominguez (2018), The Lancet. Infectious Diseases 10.1016/S1473-3099(18)30094-X.

Messacar, K., C. C. Robinson, K. Pretty, J. Yuan, and S. R. Dominguez (2017), Journal of Clinical Virology **92**, 39.

Minor, P. D., M. Ferguson, D. M. Evans, J. W. Almond, and J. P. Icenogle (1986), Journal of General Virology **67** (7), 1283.

Murrell, B., J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. K. Pond (2012), PLoS genetics **8** (7), e1002764.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh (2015), Molecular Biology and Evolution **32** (1), 268.

Ny, N. T. H., N. T. Anh, V. T. T. Hang, L. A. Nguyet, T. T. Thanh, D. Ha, N. Minh, D. Ha, A. McBride, H. Tuan, *et al.* (2017), Wellcome open research **2**, 41.

Parks, W. P., L. T. Queiroga, and J. L. Melnick (1967), American journal of epidemiology **85** (3), 469.

Pickett, B. E., E. L. Sadat, Y. Zhang, J. M. Noronha, R. B. Squires, V. Hunt, M. Liu, S. Kumar, S. Zaremba, Z. Gu, *et al.* (2011), Nucleic acids research **40** (D1), D593.

Piralla, A., N. Principi, L. Ruggiero, A. Girello, F. Giardina, E. De Sando, S. Caimmi, S. Bianchini, G. L. Marseglia, G. Lunghi, *et al.* (2018), Journal of Clinical Virology **99**, 91.

Ruggieri, V., M. I. Paz, M. G. Peretti, C. Rugilo, R. Bologna, C. Freire, S. Vergel, and A. Savransky (2017), european journal of paediatric neurology **21** (6), 884.

Sagulenko, P., V. Puller, and R. A. Neher (2018), Virus Evolution **4** (1), 10.1093/ve/vex042.

Schieble, J. H., V. L. Fox, E. H. Lennette, *et al.* (1967), American journal of epidemiology **85** (2), 297.

Tan, Y., F. Hassan, J. E. Schuster, A. Simenauer, R. Selvarangan, R. A. Halpin, X. Lin, N. Fedorova, T. B. Stockwell, T. T.-Y. Lam, *et al.* (2015), Journal of virology , JVI.

Tiveljung-Lindell, A., M. Rotzén-Östlund, S. Gupta, R. Ullstrand, L. Grillner, B. Zweygberg-Wirgart, and T. Allander (2009), Journal of medical virology **81** (1), 167.

Tokarz, R., C. Firth, S. A. Madhi, S. R. Howie, W. Wu, S. Haq, T. Briese, W. I. Lipkin, *et al.* (2012), Journal of general virology **93** (9), 1952.

Vogt, M. R., and J. E. Crowe Jr (2018), Journal of the Pediatric Infectious Diseases Society **7** (suppl_2), S49.

Wang, G., J. Zhuge, W. Huang, S. M. Nolan, V. L. Gilrane, C. Yin, N. Dimitrova, and J. T. Fallon (2017), Scientific Reports **7** (1), 1242.

Weaver, S., S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond (2018), Molecular biology and evolution **35** (3), 773.

Wei, H.-Y., T.-K. Yeh, J.-Y. Hsieh, I.-P. Lin, and J.-Y. Yang (2018), Journal of Microbiology, Immunology and Infection.

Yip, C. C., J. Y. Lo, S. Sridhar, D. C. Lung, S. Luk, K.-H. Chan, J. F. Chan, V. C. Cheng, P. C. Woo, K.-Y. Yuen, *et al.* (2017), International journal of molecular sciences **18** (5), 1065.

Zanini, F., J. Brodin, J. Albert, and R. A. Neher (2017), Virus Research Deep sequencing in virology, **239**, 106.

Zanini, F., J. Brodin, L. Thebo, C. Lanz, G. Bratt, J. Albert, and R. A. Neher (2016), eLife Sciences **4**, e11282.

Zhang, C., X. Zhang, W. Dai, Q. Liu, P. Xiong, S. Wang, L. Geng, S. Gong, and Z. Huang (2018a), Viruses **10** (2), 58.

Zhang, C., X. Zhang, W. Zhang, W. Dai, J. Xie, L. Ye, H. Wang, H. Chen, Q. Liu, S. Gong, *et al.* (2018b), Emerging microbes & infections **7** (1), 3.

Zhang, Y., J. Cao, S. Zhang, A. J. Lee, G. Sun, C. N. Larsen, H. Zhao, Z. Gu, S. He, E. B. Klem, *et al.* (2016), Virus evolution **2** (1).

Zhang, Y., D. D. Moore, W. A. Nix, M. S. Oberste, and W. C. Weldon (2015), Journal of Clinical Virology **69**, 172.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPLEMENTARY INFORMATION

TABLE S1 Primers for near full-length sequencing of EV-D68 in four overlapping fragments (F1–F4). The nested primers for F1 were used for template quantification.

Primers for fragments 1-4 (F1–4) and nesting primer (F1)	
Primers (length)	Sequence
F1 (2108 bp)	
E68_F1_fw_63–87	TCG GTA CCT TTG TAC GCC TGT TTT A
E68_F1_rev_2146–2170	TAC TAG ATT GYA RTC CAA ART CCC A
F2 (1962 bp)	
E68_F2_fw_1772–1796	CAA CTC CAG ARA TGC ACA TIC CAG G
E68_F2_rev_3757–3733	GTT CCA TAR CAT CAG TAT CTA ACC A
F3 (2001 bp)	
E68_F3_fw_3524–3548	ACC CRA TTT GYT TTG AAG GYC CAG G
E68_F3_rev_5524–5500	TGT AAA TGR TTT CTC CAA CAG ATG C
F4 (2015 bp)	
E68_F4_fw_5252–5276	CCA CCT TTG TAT CAA TAG CIG GTG T
E68_F4_rev_7291–7267	CTA ACC ATT TCC GTC TAA GAC TAG A
Nesting	
F1 (1547 bp)	
E68_F1_fw2_303–327	CGT GTG TAG CTT GGG TCG ATG AGT C
E68_F1_rev2_1849–1825	TAA TCT CCA TCA TTG ATT CCA CTT G

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE S2 Breakdown of patient age, sex, and sample material.

Age (years)	$n =$
< 1	8
1–5	27
6–18	10
> 18	9
Total	**54**
Sex	
F	30
M	24
Total	**54**
Sample material	
Nasopharyngeal (aspirate)	31
Nasopharyngeal (secretion)	23
Tracheal (secretion)	3
Other respiratory	2
Total	**59**

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE S3: Details about study samples.

Case no.	Sampling date	GenBank ID	Coverage	CT EV-PCR	CT RV-PCR
001_14	2014–09–05	MH674113	$\geq 10^3$	33.80	0
002_14	2014–09–08	MH674115	$\geq 10^3$	31.47	0
004_14	2014–09–16	MH674116	$\geq 10^3$	25.59	0
005_14	2014–08–14	MH674111	$\geq 10^3$	24.00	24.86
006_14	2014–08–14	MH674112	$\geq 10^3$	16.77	19.67
007_14	2014–09–08	MH674114	$\geq 10^3$	17.62	17.40
007_16	2016–09–09	MH674135	$\geq 10^3$	20.97	26.99
008_16	2016–09–09	MH674136	$\geq 10^3$	30.94	37.45
009_16	2016–09–11	MH674137	$\geq 10^3$	31.8	32.54
010_16	2016–09–09	MH674160	$\geq 10^3$	22.94	26.09
011_16	2016–09–09	MH674161	$\geq 10^3$	21.20	22.86
012_16	2016–09–11	MH674163	$\geq 10^3$	21.97	23.75
012_16	2016–09–13	MH674164	$\geq 10^3$	23.33	24.46
013_16	2016–09–14	MH674165	$\geq 10^3$	23.93	24.67
017_16	2016–08–23	MH674118	Missing	20.58	23.92
019_16	2016–08–24	MH674119	$\geq 10^3$	19.26	19.55
020_16	2016–08–25	MH674120	$\geq 10^3$	21.15	22.20
021_16	2016–08–28	MH674121	$\geq 10^3$	15.73	18.52
021_16	2016–09–04	MH674129	$\geq 10^2$	31.68	0
022_16	2016–08–28	MH674122	$\geq 10^3$	31.60	0
023_16	2016–08–30	MH674123	$\geq 10^3$	23.76	24.95
024_16	2016–08–30	MH674124	$\geq 10^3$	15.75	17.81
024_16	2016–08–31	MH674154	$\geq 10^3$	19.91	21.23
025_16	2016–08–30	MH674125	$\geq 10^3$	25.63	27.51
026_16	2016–08–31	MH674126	Missing	22.95	23.76
027_16	2016–08–29	MH844544	$\geq 10^3$	18.62	20.38
028_16	2016–09–04	MH674128	$\geq 10^3$	22.34	23.90
029_16	2016–09–04	MH674130	$\geq 10^2$	20.46	23.00
030_16	2016–08–19	MH674140	$\geq 10^3$	21.90	24.49
031_16	2016–08–22	MH674141	$\geq 10^3$	13.91	14.99
032_16	2016–08–25	MH674142	$\geq 10^3$	16.39	16.37
033_16	2016–08–26	MH674143	$\geq 10^3$	25.63	25.58
034_16	2016–08–29	MH674146	$\geq 10^3$	20.98	23.20
035_16	2016–08–29	MH674147	$\geq 10^2$	27.82	29.35
036_16	2016–08–29	MH674148	$\geq 10^3$	18.71	19.95
037_16	2016–08–29	MH674149	$\geq 10^3$	19.27	21.56
037_16	2016–08–30	MH674150	$\geq 10^3$	23.37	25.55
038_16	2016–08–30	MH674151	$\geq 10^3$	20.20	21.91
039_16	2016–08–31	MH674153	$\geq 10^3$	17.57	19.78
039_16	2016–08–31	MH674127	$\geq 10^3$	20.79	20.91
040_16	2016–09–01	MH674155	$\geq 10^3$	15.34	15.52
041_16	2016–09–01	MH674156	$\geq 10^3$	26.74	28.64
042_16	2016–09–02	MH674157	$\geq 10^3$	22.27	22.97
044_16	2016–09–10	MH674162	$\geq 10^2$	28.77	30.29
045_16	2016–08–31	MH674152	$\geq 10^3$	24.43	25.70
046_16	2016–09–04	MH674131	$\geq 10^3$	22.15	24.76
047_16	2016–09–13	MH674139	$\geq 10^2$	28.70	21.22
048_16	2016–08–22	MH674138	$\geq 10^3$	28.69	28.62
049_16	2016–08–23	MH674117	$\geq 10^3$	19.77	21.53
050_16	2016–08–29	MH674145	$\geq 10^3$	20.68	22.88
051_16	2016–09–05	MH674132	Missing	32.31	33.59
052_16	2016–09–08	MH674159	$\geq 10^3$	25.57	0
053_16	2016–09–16	MH674166	$\geq 10^3$	27.38	28.48
054_16	2016–09–07	MH674133	$\geq 10^3$	25.59	27.44
055_16	2016–09–07	MH674134	$\geq 10^3$	21.94	25.74
056_16	2016–09–05	MH674158	$\geq 10^3$	17.81	19.30
057_16	2016–08–29	MH674144	$\geq 10^3$	18.62	20.38
058_16	2016–09–16	N/A	Fail	33.89	0
059_16	2016–09–11	N/A	Fail	30.99	25.98

Note that some patients were sampled at two different time points. Coverage: $\geq 10^3$ = at least x1000 over all amplicons; $\geq 10^2$ = at least x100 over all amplicons; Missing = at least one amplicon with $< 10^2$ coverage; Fail = No amplicon. EV-PCR = realtime-PCR for enterovirus. RV-PCR = realtime-PCR for rhinovirus.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE S4 Amino acid sequence and substitutions in the putative targets for neutralizing antibodies in the BC-loop in the VP1 protein (codon 89–105, consensus sequence KDHTSSAAQADKNFFKW).

Sample			Codon position and consensus amino acid			
			96	98	99	100
Case no.	Collection date	Sequence variant	A	Q	A	D
007_16	2016–09–09	Major	-	-	T	-
007_16	2016–09–09	Minor	-	-	T	-
012_16	2016–09–11	Major	-	-	T	-
012_16	2016–09–11	Minor	-	-	T	-
012_16	2016–09–13	Major	-	-	T	-
012_16	2016–09–13	Minor	-	-	T	-
021_16	2016–09–04	Minor	-	R	-	-
022_16	2016–08–28	Major	-	-	T	-
022_16	2016–08–28	Minor	-	-	T	-
023_16	2016–08–30	Major	-	-	T	-
025_16	2016–08–30	Minor	-	-	-	H
027_16	2016–09–01	Major	-	-	T	-
027_16	2016–09–01	Minor	-	-	T	-
028_16	2016–09–04	Minor	-	R	-	-
031_16	2016–08–22	Major	-	-	T	-
031_16	2016–08–22	Minor	-	-	T	-
033_16	2016–08–26	Minor	-	-	-	H
036_16	2016–08–29	Major	-	-	T	-
036_16	2016–08–29	Minor	-	-	T	R
038_16	2016–08–30	Minor	-	-	V	-
039_16	2016–08–31	Minor	-	R	-	-
053_16	2016–09–16	Major	-	-	T	-
053_16	2016–09–16	Minor	-	-	T	-
054_16	2016–09–07	Major	T	-	-	-
054_16	2016–09–07	Minor	T	-	-	-

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE S5 Amino acid sequence and substitutions in the putative targets for neutralizing antibodies in the DE-loops in the VP1 protein (codons 140–152 consensus sequence: NGSSNNTYVGLPD).

Sample			Codon position and consensus amino acid	
			144	147
Case no.	Collection date	Sequence variant	S	T
007_16	2016–09–09	Minor	-	A
010_16	2016–09–09	Major	-	A
010_16	2016–09–09	Minor	-	A
012_16	2016–09–11	Major	N	-
012_16	2016–09–11	Minor	N	-
012_16	2016–09–13	Major	N	-
012_16	2016–09–13	Minor	N	-
013_16	2016–09–14	Major	-	A
013_16	2016–09–14	Minor	-	A
023_16	2016–08–30	Major	N	-
023_16	2016–08–30	Minor	N	-
035_16	2016–08–29	Major	-	A
035_16	2016–08–29	Minor	-	A
036_16	2016–08–29	Major	N	-
036_16	2016–08–29	Minor	N	-
050_16	2016–08–29	Major	-	A
050_16	2016–08–29	Minor	-	A
055_16	2016–09–07	Major	G	-
055_16	2016–09–07	Minor	G	-

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. S1 Tree with minor sequences in dually infected patients

FIG. S1 **Phylogenetic placing of minor variants in dually infected patients.** The figures hows a mid-point rooted ML tree of all consensus sequences obtained in this study (in blue) along with minority sequences from three samples from dually infected individuals (in orange, red, and green), and a sample of publicly available sequences to illustrate some EVD68 clades (labelled on internal nodes).

The minor variants do not cluster tightly with any other sample but rather look like representative samples from the populations. Minor variant sequence of SWE_007_140908 is slightly problematic since variants in the amplicons F2 and F3 cannot be cleanly separated from sequencing error. To control for this, the region from 2,000-5,500bp in this sequence was replaced with 'N's.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. S2 Tangle-tree analysis for possible recombinants.

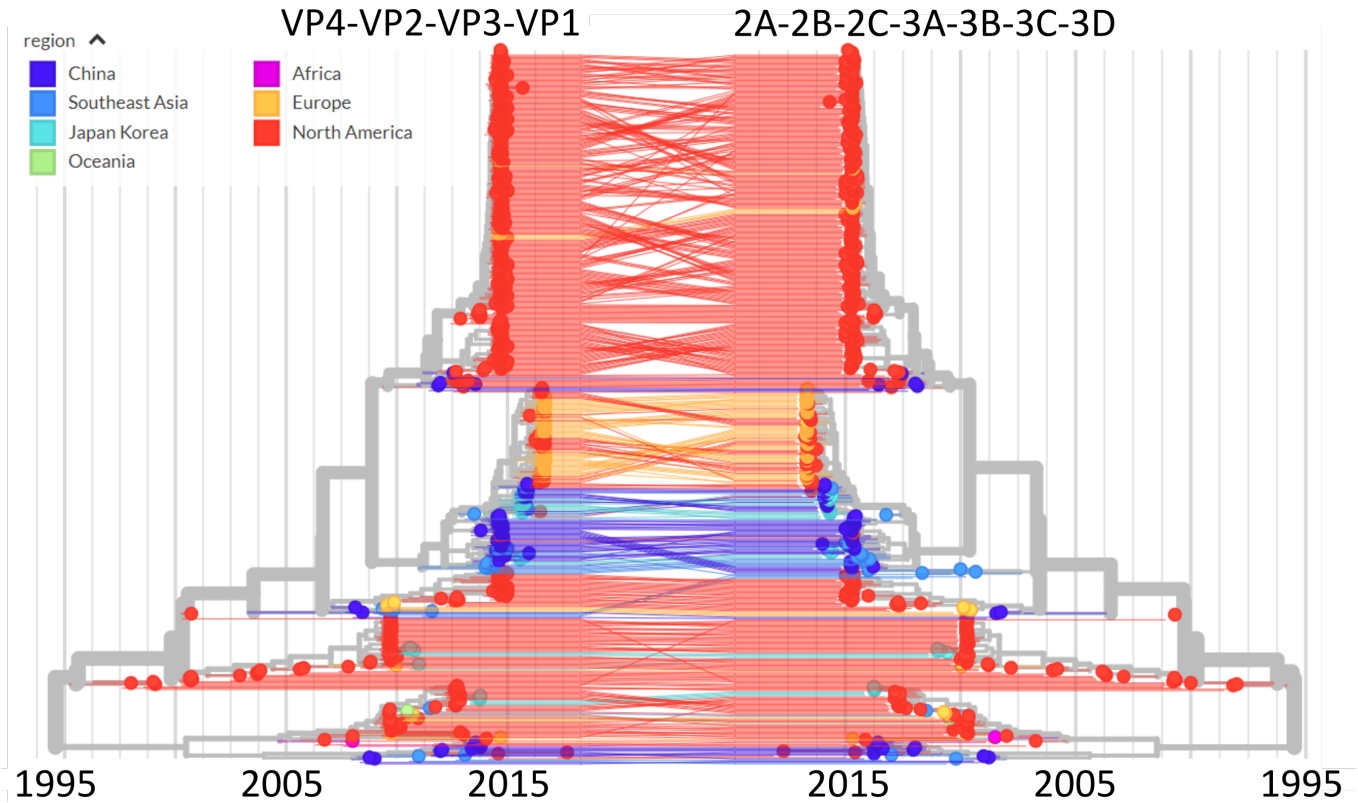

FIG. S2 **Tangle-tree analysis for possible EV-D68 recombinants.** Phylogenetic trees of sequences coding for VP4-VP1 (5' region) and 2A-3D (3' region) are congruent at the level of major clades, suggesting little inter-clade recombination. Trees were calculated using the nextstrain analysis pipeline and visualized using nextstrain's auspice.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. S3 Minor variant consistency of in repeatedly sampled individuals.

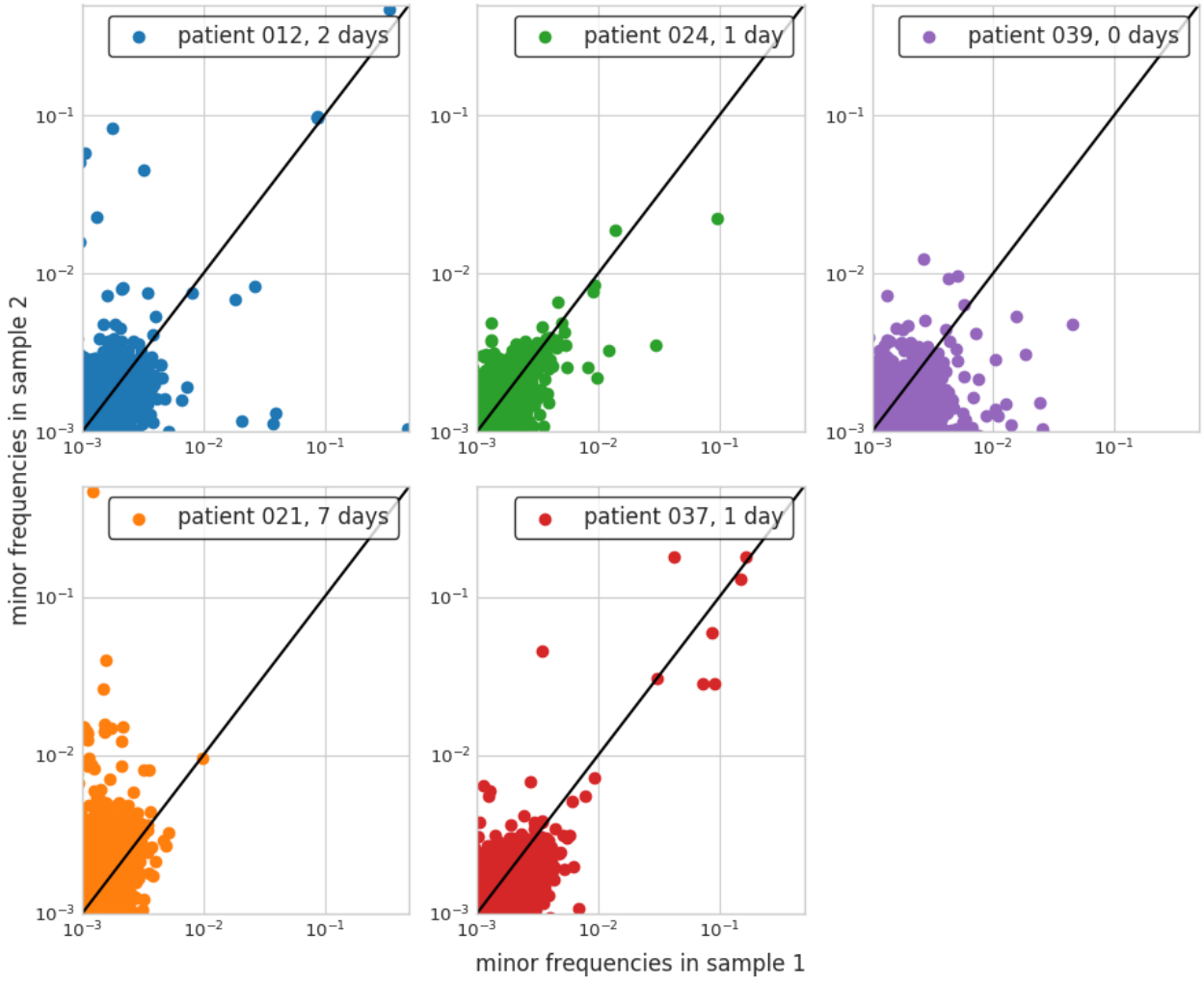

FIG. S3 **Minor variant consistency in repeatedly sampled individuals.** Each panel shows a scatter plot of minor variant frequencies in two samples from the same individual, sampled between 0 and 7 days apart. As expected, most iSNVs are at frequencies below one percent but some samples harbor iSNVs in the range of 10%. Their frequencies are largely concordant for samples 1 day apart, but deviate in the samples that are 2 or 7 days apart.

bioRxiv preprint doi: https://doi.org/10.1101/420836; this version posted March 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. S4 Analysis of multiple introductions into the Sweden during the 2016 outbreak

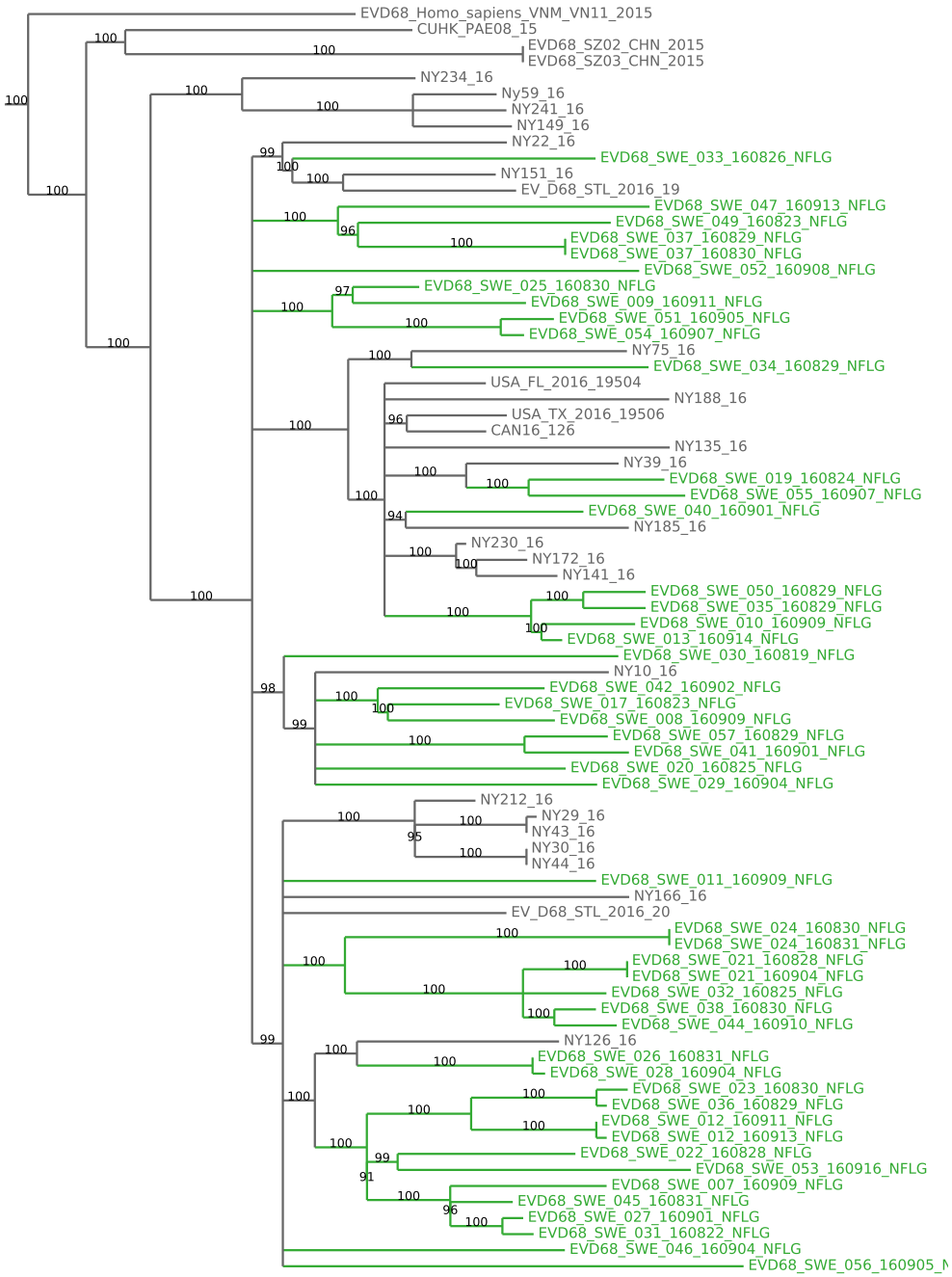

FIG. S4 **Swedish samples are not monophyletic.** This figures show a tree in which all branches with less than 90% bootstrap support are collapsed (bootstrap values are shown on remaining branches) and monophyletic clusters of exclusively Swedish samples are colored in green. Even after reducing the tree to well supported branches, Swedish samples are interspersed with North American samples. There are at least eight "Sweden only"-clades that are separated from other Swedish clades by international isolates.