

# Evolution and spatio-temporal dynamics of Enterovirus A71 subgenogroups in Vietnam

Nguyen Thi Thanh Thao<sup>1\*</sup>, Celeste Donato<sup>2\*,\$</sup>, Vu Thi Huyen Trang<sup>1</sup>, Nguyen Trung Kien<sup>1</sup>,  
Phạm Mai Thuy Trang<sup>1</sup>, Tran Quoc Khanh<sup>1</sup>, Dang Thi Nguyet<sup>1</sup>, October Sessions<sup>2</sup>, Hoang  
Quoc Cuong<sup>1</sup>, Phan Trong Lan<sup>1</sup>, Vu Thi Que Huong<sup>1</sup>, H. Rogier van Doorn<sup>3</sup>, Dhanasekaran  
Vijaykrishna<sup>2,4,#</sup>

<sup>1</sup>Immunology and Microbiology Department, Pasteur Institute, Ho Chi Minh City, Vietnam

<sup>2</sup>Duke-NUS Graduate Medical School, 169857, Singapore

<sup>3</sup>Oxford University Clinical Research Unit, Ha Noi, Vietnam

<sup>4</sup>Biomedicine Discovery Institute and Department of Microbiology, Monash University,  
Melbourne, Victoria 3800, Australia

\*These authors contributed equally

# Corresponding author: Vijaykrishna Dhanasekaran vijay.dhanasekaran@monash.edu;

Phone: +613 9905 5415

\$Current address: Celeste Donato, Biomedicine Discovery Institute and Department of  
Microbiology, Monash University, Melbourne, Victoria 3800, Australia.

**Running Title:** Enterovirus A71 evolution in Asia

© The Author 2017. Published by Oxford University Press for the Infectious Diseases Society of  
America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com.

## **Abstract**

**Background:** Enterovirus A71 (EV-A71) is the major cause of severe hand, foot and mouth disease and viral encephalitis in children across the Asia-Pacific region, including in Vietnam which has experienced a high burden of disease in recent years. Multiple subgenogroups (C1, C4, C5 and B5) concurrently circulate in the region with a large variation in epidemic severity. The relative differences in their evolution and epidemiology were examined within Vietnam and globally.

**Methods:** A total of 752 VP1 gene sequences were analysed (413 generated in this study combined with 339 obtained from GenBank), collected from patients in 36 provinces in Vietnam during 2003–2013 along with epidemiological metadata. Globally representative VP1 gene datasets of subgenogroups were used to co-estimate time-resolved phylogenies and relative genetic diversity to infer virus origins and regional transmission network.

**Results:** Despite frequent virus migration between countries, the highest genetic diversity of individual subgenogroups was maintained independently for several years in specific Asian countries representing genogroup-specific sources of EV-A71 diversity.

**Conclusion:** This study highlights a persistent transmission network of EV-A71, with specific Asian countries seeding other countries in the region and beyond, emphasising the need for improved EV-A71 surveillance and detailed genetic and antigenic characterisation.

**Key words:** Enterovirus A71; Vietnam; Hand Foot and Mouth Disease; Phylogenetics

**summary:**

Southeast Asia and Vietnam in particular experience a significant burden of Enterovirus A71 (EV-A71) disease, which frequent outbreaks reported. This study reveals the endemicity, importation and exportation of strains within the region informing public health measures and vaccination strategies.

Accepted Manuscript

## INTRODUCTION

Enterovirus A71 (EV-A71) (species *Enterovirus A*, genus *Enterovirus*; family *Picornaviridae*) along with several other Enterovirus A genotypes typically cause a self-limiting hand, foot and mouth disease (HFMD) [1]. EV-A71 infection can also present as neurological disease including aseptic meningitis, acute flaccid paralysis, and brain stem encephalitis with neurogenic pulmonary oedema, which can be fatal [1, 2]. Although recognised in humans since the 1960s [3, 4], EV-A71 has emerged in the past two decades as a major cause of epidemic HFMD and viral encephalitis. The Western Pacific Regional Office of the World Health Organization established a HFMD reporting system in 2011 to report biweekly case numbers from China, Hong Kong, Macao, Japan, the Republic of Korea, Singapore and Vietnam ([http://www.wpro.who.int/emerging\\_diseases/HFMD](http://www.wpro.who.int/emerging_diseases/HFMD)).

Two major EV-A71 genogroups (B and C) circulate in Asian countries with a wide variation in epidemic severity [5], whereas genogroup A, first detected in 1969, is now largely extinct and genogroup D is thought to be endemic in India [6] and E and F in parts of Africa [7]. Until 1997, the evolution of B and C genogroups were characterised by the emergence and serial replacement of subgenogroups: B2 replaced B1 that circulated for several years and C2 replaced C1 in epidemic countries [3]. However, the subsequent emergence of the severe form of EV-A71 in Asia coincided with the detection of multiple new subgenogroups (B0-B5, C0-C5), many of which continue to circulate. Countries experienced outbreaks with a 2–3-year cyclical pattern [5] thought to be due to the accumulation of susceptible infants lacking prior exposure to the virus, sufficient to maintain an epidemic [8]. Epidemic peaks are often observed in late spring to early summer in several Asian countries suggesting temperature and humidity may influence transmission [9]. Furthermore, large epidemics are often associated with the introduction of antigenically distinct subgenogroups into a population, replacing older subgroups that were endemic for many years [3].

EV-A71 has been reported in Vietnam since 2003 [10], however severe and widespread epidemics [11-13] were only observed following the introduction of genogroups C4 in 2011 and B5 in 2012 resulting in >300,000 cases and 225 fatalities during 2011–2013 [14]. Epidemiological parameters [12] and country-level migration patterns [13] of these epidemics are well understood. Outbreaks in Vietnam prior to 2011 were caused by C5, with the largest reported epidemic in 2005 associated with 173 cases; 51 with neurological complications and 3 fatalities [10]. Since 2011, HFMD has been classified by the Vietnamese Ministry of Health as a severe infectious disease with outbreak potential, which requires all hospitals to report cases weekly through the national communicable disease surveillance system [11].

Our aim was to understand the differences in the long-term evolution and epidemiology of four EV-A71 subgenogroups C1, C4, C5, and B5 that were detected in Vietnam during 2003–2013. Using a globally representative dataset we inferred subgenogroup specific variation in migration patterns, temporal changes in genetic diversity and natural selection.

## **METHODS**

### **Clinical information and specimen collection**

Samples including stool, throat swab, vesicle swab or CSF (for severe cases) collected from patients with symptoms of HFMD during 2003–2013, in 29 provinces in central and southern Vietnam, were sent to the Institute Pasteur in Ho Chi Minh City for virus characterisation.

Anonymized sample metadata included patient age, date and province of sample collection, severity score as defined by the Vietnam Ministry of Health [15] and disease outcome (Supp Table 1). This study was approved by the Ethics Committees of the Institute Pasteur in Ho Chi Minh City.

### **Virus detection, isolation and sequencing**

Viral RNA was extracted from 6799 stool, throat swab, vesicle swab or CSF using the QIAamp viral RNA extraction kit (QIAGEN), and tested by RT-PCR using previously described VP1 specific primers [10]. The RT-PCR positive samples from 3200 patients were passaged in human rhabdomyosarcoma (RD) (ATCC CCL136) and African green monkey kidney (Vero) (ATCC CCL81) cell lines. To provide longitudinal data 413 samples were selected for sequencing (Figure 1a). RNA was extracted from cell culture supernatant (as described above) and the VP1 gene amplified by RT-PCR using the Access RT-PCR System (Promega Co., USA) with primers 2349F (5'-GCYTAYATAATAGCAYTGGCGGCAGC-3') and 3393R (5'-GGCGGTTRACCACYCTDAAGTTGCCCAC-3'). Each reaction was incubated in a thermal cycler for 45 min at 48 °C, 2 min at 94 °C, 35 cycles of 94 °C for 10 sec, 50 °C for 10 sec, 65 °C for 1 min and then at 65 °C for 5 min. Amplicons were confirmed by 1% agarose gel electrophoresis and purified with a Wizard SV gel and PCR clean up System (Promega). The complete VP1 gene was sequenced utilizing the primers 2349F, 2757F (5'-GCHAAYTGGGAYATAGACATAAC-3'), 2780R (5'-CCCTRTATCTGTATTGDCC-3') and 3393R and analysed by ABI 3130XL (Applied Biosystems Inc.), and were deposited in GenBank (KU887762 - KU888174).

## Phylogenetic analysis

Viruses sequenced in this study were analyzed with EV-A71 sequences (complete VP1) obtained from GenBank (300 C1, 2803 C4, 43 C5 and 384 B5). Metadata for GenBank entries from Vietnam (n=339), were obtained from submission authors where possible. Multiple sequence alignments were automated using MAFFT v.7 [16]. Alignments were screened for recombinant viruses using SBP (Single Break Point) and GARD (Genetic Algorithm for Recombination Detection) [17] accessed through the DataMonkey webserver of HyPhy [18], and GENECONV, MaxChi, Chimaera, SiScan and 3Seq, using the Recombination Detection Program v4.46 [19].

Maximum likelihood (ML) trees were constructed using RAxML v8.0 [20] applying the General Time Reversible nucleotide substitution model with a gamma distribution of among-site rate variation (GTR+G+I). Support for individual nodes were estimated with 1,000 rapid bootstrap replicates. Trees were visualised and annotated using FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). TempEst v1.4 [21] was used to plot root-to-tip genetic distances, and sequences not conforming to a linear evolutionary pattern were discarded. Time-measured evolutionary histories were reconstructed using Bayesian phylogenetic inference in BEAST v1.8.2 [22] using the GTR+G substitution model under a relaxed uncorrelated lognormal molecular clock to account for varied evolutionary rates among lineages [23] and a GMRF (Gaussian Markov Random Fields) Bayesian Skyride coalescent tree prior [24]. Three independent Markov chain Monte Carlo chains were run for 100 million steps and sampled every 10,000<sup>th</sup> generation, with the first 10% discarded as burn-in. Convergence and mixing of the chains were inspected using Tracer v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>); all continuous parameters yielded effective sample sizes greater than 200. A maximum clade credibility (MCC) tree was summarized using

TreeAnnotator v1.7.4 [22]. Tracer v1.5 was used to plot the relative genetic diversity of individual datasets.

### **Site-specific and internal vs. external branch selection.**

Selection pressures were investigated using the SLAC (Single-Likelihood Ancestor Counting), FEL (Fixed Effects Likelihood), iFEL (internal branches FEL), MEME (Mixed Effects Model of Evolution) and FUBAR (Fast, Unconstrained Bayesian AppRoximation) (p-value, <0.01 or posterior threshold 0.9 to minimize false-positives) [25], available through the DataMonkey webserver of HyPhy [18]. The internal and external branch-wise  $d_N/d_S$  ratios were estimated using the two-model ratio in CodeML [26].

## **RESULTS**

**Subgenogroup prevalence in Vietnam.** We sequenced the complete VP1 gene (891 nucleotides) of 413 EV-A71 cell culture isolates from patients in 29 provinces collected during 2003–2013 (Figure 1a). An additional 339 complete VP1 genes collected during 2005, 2008, and 2011–2013 from 24 provinces in Vietnam were acquired from GenBank. Overall, 752 VP1 gene sequences from a total of 36 provinces were included, with the majority collected from the South East and Mekong River Delta regions (Figure 1a, Supp Table 1, Supp Figure 1). ML trees were used to designate all Vietnamese EV-A71 sequences to four subgenogroups: C1 (n=4), C4 (n=459), C5 (n=187) and B5 (n=102) (Figure 1b). Two viruses were identified as recombinants (Supp Figure 2) and excluded from all subsequent analysis.



**Patient demography and severity.** Patient age was known for 713/752 samples. C4 infections were associated with moderately older children (median age = 2.0) compared to B5 (1.5) and C5 (1.5) with no significant change over time (Table 1). Clinical severity scores were known for 696/752 patients and C4 infections were associated with the highest proportion of severity score 4 diagnosis (Table 1). Survival outcome was only known for the 413 patients included in this study, with 35 recorded deaths associated with C4 (n=33) and C5 (n=2). Samples from this study arose from hospital-based syndromic surveillance that focussed on collecting samples from cases with severity of 2b and higher, whilst previous studies included less severe cases in sample collection [11-13]. The differing criteria for sample collection may have biased these results.

**Spatio-temporal diversity of EV-A71.** We co-estimated time-resolved phylogenies and relative genetic diversity to infer virus origins and regional transmission network of the four subgenogroups detected in Vietnam (Figure 2, 3).

*Subgenogroup C1.* C1 has circulated extensively around the world since the first detection in Australia in 1986 and caused serial epidemics in Australia, Malaysia and Singapore [1, 27] (Figure 2). Our estimates of the time of most recent common ancestor (TMRCA) of the global C1 subgenogroup in mid 1984 (mean TMRCA 1984.5, 95% highest posterior density (HPD), 1984.0–1985.1) (Table 2) was comparable to prior estimates of a TMRCA during 1983 (95% HPD, 1982.2–1984.2) [28]. Based on available sequence data, prior to 1996, C1 viruses were predominantly reported in Australia, Europe and the USA [29] (Supp Figure 3a). We observed a distinct change in phylogenetic structure during 1996 (Figure 2) coinciding with a bottleneck in the relative genetic diversity of C1 viruses (Figure 3). This signifies a period when, with the exception of a lineage that was introduced into Malaysia in 1996, all other C1 lineages became globally extinct. All C1 viruses that attained sustained circulation in Europe and were sporadically detected elsewhere in Southeast Asia were

derived from this sustaining Malaysian lineage (Figure 2, Supp Figure 3a). Four C1 viruses were detected in Ho Chi Minh City and neighbouring Tay Ninh province during 2005 suggesting limited transmission or that C1 circulated largely sub-clinically. These viruses clustered with European sequences suggesting a single direct introduction into Vietnam followed by limited circulation in the population.

*Subgenogroup C4.* The C4 subgenotype, further divided into C4a and C4b, diverged as early as the mean TMRCA during mid-to-late 1996 (1996.7, HPD, 1996.2–1997.0) (Table 2, Figure 2). Following the first report from Japan in 1997 [30], C4b was predominantly detected in China, with periodic detection in Taiwan, Thailand and Europe (Supp Figure 3b). Viruses belonging to the second lineage (C4a) were predominantly detected since 2007 from China, Japan, South Korea, Taiwan, Thailand, Cambodia, Vietnam and parts of Europe (Supp Figure 3b). The C4 tree exhibits a slender dominant tree trunk reflecting the serial extinction of lineages after each epidemic period (Figure 2, Supp Figure 3b), however since 2007 coinciding with the emergence of severe and widespread EV-A71 epidemics in China, C4a viruses have diverged into multiple co-circulating lineages. The relative genetic diversity of global C4 increased rapidly as strains first emerged followed by a decade of minor fluctuations, representing regional outbreaks, and an overall gradual increase in relative genetic diversity until 2011 (Figure 3). The largest increase in relative genetic diversity was between 2010 and 2012 as viruses from the 2010 outbreak spread into Vietnam and Cambodia, and reached Taiwan and Europe. The similarity of the relative genetic diversity plots of the C4 global and China C4a datasets (Figure 3) and the prominence of China in the trunk of the C4 phylogeny (Figure 2, Supp Figure 3b) strongly suggests an epidemic source for the C4 viruses within China.

Three C4 lineages were detected in Vietnam during 2004–2013. The C4b viruses detected during 2004–2006 formed two separate lineages (Supp Figure 3c), closely associated with strains from China (Shenzhen, Guangdong) detected during 2001–2004, and Taiwan in 2004, respectively (Supp Figure 3b). Multiple introductions of C4a were detected in Vietnam since 2011 (mean TMRCA 2008.1, HPD, 2007.6–2008.5), however the majority formed a monophyletic clade, derived from viruses circulating in China during 2010 (Figure 2, Supp Figure 3b). There was sustained circulation of this lineage within Vietnam, with the South East and Mekong River Delta provinces acting as the sources of viruses that spread north. There was a rapid increase in genetic diversity of C4a circulating in Vietnam peaking in mid-2011, followed by a moderate bottleneck and small period of constant genetic diversity from late 2012–2013 (Figure 3).

*Subgenogroup C5.* Since the detection of C5 in association with an outbreak in southern Vietnam in 2003 [10], this subgenogroup has been continuously detected in Vietnam with decreasing prevalence since 2012 (Figure 1b). Eleven C5 sequences available from other countries, including Taiwan, Thailand, Singapore and Finland (Supp Figure 3d) appear to be derived from Vietnam. Sampling dates suggest recent introduction of these viruses (Figure 2). These results suggest that C5 has been endemic to Vietnam over the past decade with periodic introduction to other Asian countries. However, there appears to be a lack of success of C5 outside of Vietnam as suggested by limited and sporadic detection (Figure 2, Supp Figure 3d). Within Vietnam, the South East and Mekong River Delta regions appears to act as the endemic source of C5 that spread to the central and northern provinces (Supp Figure 3e). The TMRCA of the global C5 subgenogroup was estimated at early-to-mid 2001 (2001.3, HPD, 2000.3–2002.1), which is similar to previously reported estimates of 2002.8 (HPD, 2001.8–2003.6) [28] (Table 2). Due to the endemic nature of C5 in Vietnam, the relative genetic diversity estimated for the global dataset is largely derived from Vietnam with

minor biannual fluctuations. The largest increase in diversity was associated with the 2005 outbreak in Vietnam (Figure 3).

*Subgenogroup B5.* B5 is exclusive to Asia with numerous outbreaks reported in Japan, Malaysia, Brunei, Singapore and Taiwan [31-34]. The TMRCA was estimated during early 2002 (2002.2, HPD, 2001.9–2002.5), similar to previous reports [28] (Table 2). The strong geographic clustering of B5 (Figure 2, Supp Figure 3f), suggest endemicity of diverged lineages in multiple countries in Asia. In comparison to other subgenogroups, the B5 viruses exhibit greater divergence between the spatially separated lineages (Supp Figure 3f). In particular, since 2007 we observed two major spatially separated variants of B5 in Malaysia and Taiwan, respectively. Viruses from both lineages have been detected in Vietnam. The global B5 diversity had a marked increase until 2006 followed by cyclical peaks and bottlenecks associated with outbreaks in Malaysia in 2003 [32], Brunei in 2006 [31], and in Thailand [33] and Taiwan [34] in 2012 (Figure 3).

Vietnamese B5 viruses shared a common ancestor circulating in early 2011 (2011.0, HPD, 2010.6–2011.3), just prior to their detection in the same year. The majority of viruses circulating during the 2012-2013 outbreak evolved from a lineage circulating in Malaysia suggesting a direct transmission between countries, and a second, smaller, localized cluster was introduced from Taiwan (Supp Figure 3f). B5 viruses circulated Vietnam-wide with sustained transmission between the southern and central provinces and limited introduction and circulation in the northern provinces, mainly limited to Hanoi (Supp Figure 3g).

### Differential selection pressures between B and C subgenogroups.

The mean nucleotide substitution rate of the C4 subgenogroups ( $6.6 \times 10^{-3}$ ) was higher than other global subgenogroups, including C1, ( $4.3 \times 10^{-3}$ ), C5 ( $4.1 \times 10^{-3}$ ) and B5 ( $4.6 \times 10^{-3}$ ) (Table 2), and those estimated previously for the B and C genogroups [28], however these rates fall within ranges described for RNA viruses [35]. Residue 98 was under positive selection in all C subgenogroups, while residue 145 was under positive selection in C1 and B5 (Table 2). Among the Vietnam lineages, residue 43 of the B5 dataset was identified to be under positive selection. A comparison of branch-wise  $d_N/d_S$  ratios of internal branches (ancestral nodes) and external branches (tips of the tree) show that there was a greater fixation of amino acids in the B5 dataset (internal vs. external  $d_N/d_S$  ratio = 1.5) than the C subgenogroups (0.62–0.75). The passage of the viruses from these studies may have biased these results.

## DISCUSSION

Our analysis of the genetic and epidemiological data provides key insights into the complex transmission network of EV-A71 throughout Southeast Asia, the effect of which extends beyond the region. We observed that subgenogroups attained long-term endemicity in specific countries that acted as genogroup-specific epidemic sources, including C4 in China, C5 in Vietnam and B5 in Taiwan and Malaysia, whereas the epidemic source of C1 could not be ascertained. These Asian lineages have been introduced to Europe, revealing the importance the region plays in the global dominance of EV-A71. We also observe key differences in the evolution and epidemiology of the dominant B5 and C4 lineages, with a greater fixation of amino acids in B5 likely owing to the diversifying phylogeny. Taken together, the endemicity of subgenogroups in different countries of Southeast Asia with periodic mixing and co-circulation facilitate the continuous transmission of diverse EV-A71 viruses with the potential for recombination and emergence of novel subgenogroups.

The epidemiology of EV-A71 in Vietnam drastically changed from 2011 with switching of subgenogroup predominance observed for the first time, associated with increased severity and number of infected children [11-13, 15, 36]. The marked decrease in C5 detection after nine years of sustained circulation coincided with the emergence of the C4 and B5 subgenotype introduced into Vietnam. Genogroup switching has been observed in other countries in Asia, in particular Malaysia, Japan and Taiwan, and European countries including the Netherlands [3]. However, subgenogroup switching is not always evident; in China C4 circulated exclusively for a number of years and continually seed new strains into global circulation [5].

Population immunity, within countries and globally, shapes the evolution of EV-A71. Our analysis revealed that subgenogroups exhibit differing evolutionary patterns with B5 that predominantly circulated in Malaysia, Taiwan and Vietnam under a greater selection pressure than C4 (Table 1). However, we found that the C4 viruses introduced to Vietnam were under selection pressure (data not shown), similar to B5. This evolutionary pattern indicates that EV-A71 likely evolves as antigenic variants and this has significant implications for vaccine design, as over time vaccine formulations may need to be updated as new antigenic variants emerge.

Residue 43 of VP1 was identified under positive selection in the Vietnam B5 dataset and is located within a proposed immunodominant IgM and IgG epitope identified by EV-A71 antisera [37]. The change from E to K in a subset of isolates in 2013 suggests the emergence of an antigenic variant. Residue 145 of VP1 was identified under positive selection in the C1 and C4 global datasets, as well as being previously identified in B2, B3, and B4 datasets [28]. The selection of this residue across subgenogroups suggests

particular importance in immune recognition or pathogenesis. Residue 145 is positioned at the canyon surface of the capsid and alterations in this residue influence receptor binding and thus may influence pathogenicity. The amino acid change G/Q has been proposed as a more virulent phenotype [38]. Based on binding profile, EV-A71 is classified into two distinct phenotypes', P-selectin glycoprotein ligand-1 ((PSGL-1 binding) (PB)) and PSGL-1-non-binding (non-PB) viruses [39]. Residue 145 plays a critical role in PSGL-1 binding as this site acts as a molecular switch to change the binding profile [40]. Residue 145G and 145Q exhibit the PB phenotype and those with 145E have a non-PB phenotype, regardless of the subgenogroup. A non-human primate model suggested there is a strong *in vivo* selection of 145E variants, associated with the development of viremia and neuropathogenesis. 98.8% of the C4, 94.6% of C5 and 61.8% of B5 viruses had 145E residues, partly reflecting the differing severity observed in this study, with more severe disease associated with C4 and C5 infections compared to B5. Residue 98, reportedly in conjunction with residue 145, has also been associated with pathogenesis in this model [41]. However, these results have to be taken in consideration with possible mutations associated with adaption to cell culture and primates [41].

Like other enteroviruses, EV-A71 is permissive to recombination events. The two recombinant strains identified in this study appeared to be sporadic events. Sequencing of the complete genome of Vietnamese isolates may identify more recombination events.

## CONCLUSION

This study highlights the endemic sources of EV-A71 subgenogroups which seed strains into the rest of Southeast Asia and beyond. There is a persistent transmission network occurring in Southeast Asia with the sustained transmission of B5 highlighting the extensive pathways evident between countries in close proximities. Sustained pathways between more distant

countries are evident in the transmission of C5 from Vietnam to Taiwan and the reciprocated transmission of B5. The movement of adults is possibly contributing to the rapid transmission of EV-A71 in Asia and globally as adults with subclinical or asymptomatic infections likely act as a reservoir and contribute to the dissemination of EV-A71 in the population [42]. A greater understanding of the transmission dynamics of EV-A71 in Asia and the role adults play in the dissemination of strains is vital to inform public health strategies and vaccine design. Furthermore, subgenogroup switching periodically observed in many countries along with the often-observed ladder-like phylogeny from these data suggests temporal patterns of lineage replacement (akin to antigenic drift for influenza) that has significant implications for control and vaccine development [43].



### **Conflict of interest**

The authors do not have a commercial or other association that might pose a conflict of interest.

### **Funding statement.**

CD is supported by an Australian National Health and Medical Research Council Peter Doherty Biomedical Fellowship. CD and DV are supported by the Duke-NUS Signature Research Program funded by the Agency of Science, Technology and Research, Singapore and the Ministry of Health, Singapore. HRvD is funded by the Wellcome Trust of Great Britain. NTTT, VTHT, TQK, PTL, VTQH, HQC, NTK, PMTT and PTN are funded by Ministry of Health, Vietnam.

### **Acknowledgements.**

We thank all laboratories, physicians, epidemiologists as well as the Preventive Medicine Centers and Provincial Hospital in the south of Vietnam for their cooperation and advice in collecting these data.

### **Conference presentation:**

- **Virus Genomics and Evolution**, Cambridge, United Kingdom, June 2016  
*“Evolutionary dynamics of Enterovirus 71 in Vietnam and Malaysia”*
- **Hand, Foot & Mouth Disease International Conference**, Singapore, July 2016  
*“Evolutionary dynamics of Enterovirus 71 in Vietnam and Malaysia”*

**Corresponding author:** Vijaykrishna Dhanasekaran

[vijay.dhanasekaran@monash.edu](mailto:vijay.dhanasekaran@monash.edu)

Department of Microbiology

19 Innovation Drive

Monash University

Clayton 3800

Phone: [+613 9905 5415](tel:+61399055415)

[Fax: +613 9902 9222](tel:+61399029222)

**Alternative corresponding author:** [Celeste Donato](#)

Department of Microbiology

19 Innovation Drive

Monash University

Clayton 3800

Phone: [+613 9905 0263](tel:+61399050263)

[Fax: +613 9902 9222](tel:+61399029222)

### Figure Legends:

**Figure 1.** Distribution of Vietnam EV-A71 strains analysed in this study. **a)** Sample distribution from provinces in Vietnam over time. **b)** Subgenogroup distribution over time.

**Figure 2.** Maximum clade credible trees of EV-A71 subgenogroups. Nodes correspond to mean TMRCA. Branches are coloured by country of isolation. MCC trees with visible tip labels and 95% HPD intervals are presented as supplementary Figures 4 a-g.

**Figure 3.** The relative genetic diversity of EV-A71 subgenogroups. Bayesian skyline plots were derived for global strains and a subset isolated in Vietnam. A measure of genetic diversity is given on the y-axis with the 95% HPD shown in solid colour and the median as a dashed line.

**Table 1. Patient demography and severity of EV71 in Vietnam**

		B5 n=102	C1 n=4	C4 n=459	C5 n=187	Total n=752
Age (years)	Mean	1.8	2.5	2.1	1.9	
	Median	1.5	2.5	2.0	1.5	
	Range	0.08–6	2–3	0.08–12	0.08–9	
	Unknown	23	0	12	4	
Severity score <sup>1</sup>	1	1	2	7 (1.5%)	21 (11.2%)	31
	2	0	0	0	7 (3.7%)	7
	2a	5 (4.9%)	0	40 (8.7%)	9 (4.8%)	54
	2b	46	2	179 (40%) (6 fatal)	85 (45.5%)	312
	3	24	0	161 (35.1%) (5 fatal)	51 (27.3%)	236
	4	2 (2%)	0	47 (10.2%) (21 fatal)	3 (1.6%) (2	52
	Meningitis	0	0	1 (0.2%)	3 (1.6%)	4
	Mortality	0	0	32 (94.1%)	2 (5.8%)	34
	Unknown	24	0	24 (5.2%) (1 fatal)	8 (4.3%)	56

<sup>1</sup>The clinical severity score was defined following the Vietnamese Ministry of Health clinical grading system [15].

**Table 2. Nucleotide substitution rates and selection pressures of Global EV-A71 subgenogroups**

Dataset		TMRCAs [95% HPD]	Rate $\times 10^{-3}$ [95% HPD]	VP1 residues under positive selection	mean $d_N/d_S$ <sup>1</sup>	Internal/external $d_N/d_S$ ratio <sup>2,3</sup>
<b>C1</b>	<b>Global</b> (n=269)  1986 – 2012	27.464 (1984.5 = Jul 1984)  [28.0280, 26.9423]	4.3551  [3.7845, 4.9348]	98, 145	0.0611	0.6814
<b>C4</b>	<b>Global</b> (n=453)  1997 – 2014.0	17.343 (1996.7 = Sep 1996)  [17.8511, 17.0336]	6.546  [5.6562, 7.516]	98	0.0564	0.7474
<b>C5</b>	<b>Global</b> (n=177)  2003.4 – 2012.2	10.910 (2001.3 = May 2001)  [11.8583, 10.0593]	4.1136  [3.7845, 4.9465]	98	0.0616	0.6270
<b>B5</b>	<b>Global</b> (n=351)  2003.0 – 2014.8	12.532 (2002.2 = Mar 2002)  [12.8537, 12.2306]	4.553  [3.7945, 5.3734]	145	0.0877	<b>1.461</b>

<sup>1</sup>  $d_N/d_S$  estimated using SLAC. <sup>2</sup>Estimated using CodeML. <sup>3</sup>The dominant monophyletic lineage in each tree was analysed and each dataset was randomly subsampled to 80 to avoid any bias due to inconsistent sampling size across datasets.

Accepted Manuscript

## References

1. Wong SSY, Yip CCY, Lau SKP, Yuen KY. Human enterovirus 71 and hand, foot and mouth disease. *Epidemiol Infect* **2010**; 138:1071-89.
2. McMinn PC. Enterovirus vaccines for an emerging cause of brain-stem encephalitis. *N Engl J Med* **2014**; 370:792-4.
3. van der Sanden S, Koopmans M, Uslu G, van der Avoort H, Dutch Working Group for Clinical Virology. Epidemiology of enterovirus 71 in the Netherlands, 1963 to 2008. *J Clin Microbiol* **2009**; 47:2826-33.
4. Schmidt NJ, Lennette EH, Ho HH. An apparently new enterovirus isolated from patients with disease of the central nervous system. *J Infect Dis* **1974**; 129:304-9.
5. Solomon T, Lewthwaite P, Perera D, Cardoso MJ, McMinn P, Ooi MH. Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis* **2010**; 10:778-90.
6. Saxena VK, Sane S, Nadkarni SS, Sharma DK, Deshpande JM. Genetic diversity of enterovirus A71, India. *Emerg Infect Dis* **2015**; 21:123-6.
7. Bessaud M, Razafindratsimandresy R, Nougairede A, et al. Molecular comparison and evolutionary analyses of VP1 nucleotide sequences of new African human enterovirus 71 isolates reveal a wide genetic diversity. *PLoS One* **2014**; 9:e90624.
8. NikNadia N, Sam IC, Rampal S, et al. Cyclical patterns of hand, foot and mouth disease caused by enterovirus A71 in Malaysia. *PLoS Negl Trop Dis* **2016**; 10:e0004562.
9. Koh WM, Bogich T, Siegel K, et al. The epidemiology of hand, foot and mouth disease in Asia: a systematic review and analysis. *Pediatr Infect Dis J* **2016**; 35:e285-300.
10. Tu PV, Thao NT, Perera D, et al. Epidemiologic and virologic investigation of hand, foot, and mouth disease, southern Vietnam, 2005. *Emerg Infect Dis* **2007**; 13:1733-41.

11. Nguyen NT, Pham HV, Hoang CQ, et al. Epidemiological and clinical characteristics of children who died from hand, foot and mouth disease in Vietnam, 2011. *BMC Infect Dis* **2014**; 14:341.
12. Geoghegan JL, Tan le V, Kuhnert D, et al. Phylodynamics of enterovirus A71-associated hand, foot, and mouth disease in Viet Nam. *J Virol* **2015**; 89:8871-9.
13. Donato C, Hoi le T, Hoa NT, et al. Genetic characterization of enterovirus 71 strains circulating in Vietnam in 2012. *Virology* **2016**; 495:1-9.
14. World Health Organization. Emerging disease surveillance and response: hand, foot, and mouth disease (HFMD). Available at: [http://www.wpro.who.int/emerging\\_diseases/HFMD/en](http://www.wpro.who.int/emerging_diseases/HFMD/en). Accessed 25 October 2016.
15. Khanh TH, Sabanathan S, Thanh TT, et al. Enterovirus 71-associated hand, foot, and mouth disease, Southern Vietnam, 2011. *Emerg Infect Dis* **2012**; 18:2002-5.
16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**; 30:772-80.
17. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* **2006**; 23:1891-901.
18. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **2010**; 26:2455-7.
19. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **2015**; 1: vev003.
20. Stamatakis A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**; 30:1312-3.
21. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2016**; 2:vev007.



22. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **2012**; 29:1969-73.
23. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **2006**; 4:e88.
24. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* **2008**; 25:1459-71.
25. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **2005**; 22:1208-22.
26. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **2007**; 24:1586-91.
27. McMinn P, Lindsay K, Perera D, Chan HM, Chan KP, Cardoso MJ. Phylogenetic analysis of enterovirus 71 strains isolated during linked epidemics in Malaysia, Singapore, and Western Australia. *J Virol* **2001**; 75:7732-8.
28. Tee KK, Lam TT, Chan YF, et al. Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *J Virol* **2010**; 84:3339-50.
29. Mirand A, Schuffenecker I, Henquell C, et al. Phylogenetic evidence for a recent spread of two populations of human enterovirus 71 in European countries. *J Gen Virol* **2010**; 91:2263-77.
30. Hosoya M, Kawasaki Y, Sato M, et al. Genetic diversity of enterovirus 71 associated with hand, foot and mouth disease epidemics in Japan from 1983 to 2003. *Pediatr Infect Dis J* **2006**; 25:691-4.
31. AbuBakar S, Sam IC, Yusof J, et al. Enterovirus 71 outbreak, Brunei. *Emerg Infect Dis* **2009**; 15:79-82.
32. Chua KB, Kasri AR. Hand foot and mouth disease due to enterovirus 71 in Malaysia. *Virol Sin* **2011**; 26:221-8.
33. Mauleekoonphairoj J, Vongpunsawad S, Puenpa J, Korkong S, Poovorawan Y. Complete genome sequence analysis of enterovirus 71 isolated from

- children with hand, foot, and mouth disease in Thailand, 2012-2014. *Virus Genes* **2015**; 51:290-3.
34. Luo ST, Chiang PS, Chung WY, et al. Reemergence of enterovirus 71 epidemic in northern Taiwan, 2012. *PloS One* **2015**; 10:e0116322.
35. Jenkins GM, Rambaut A, Pybus OG, Holmes EC. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* **2002**; 54:156-65.
36. Thoa le PK, Chiang PS, Khanh TH, et al. Genetic and antigenic characterization of enterovirus 71 in Ho Chi Minh City, Vietnam, 2011. *PloS One* **2013**; 8:e69895.
37. Gao F, Wang YP, Mao QY, et al. Enterovirus 71 viral capsid protein linear epitopes: identification and characterization. *Virol J* **2012**; 9:26.
38. Li R, Zou Q, Chen L, Zhang H, Wang Y. Molecular analysis of virulent determinants of enterovirus 71. *PloS One* **2011**; 6:e26237.
39. Nishimura Y, Shimojima M, Tano Y, Miyamura T, Wakita T, Shimizu H. Human P-selectin glycoprotein ligand-1 is a functional receptor for enterovirus 71. *Nat Med* **2009**; 15:794-7.
40. Nishimura Y, Lee H, Hafenstein S, et al. Enterovirus 71 binding to PSGL-1 on leukocytes: VP1-145 acts as a molecular switch to control receptor interaction. *PLoS Pathog* **2013**; 9:e1003511.
41. Kataoka C, Suzuki T, Kotani O, et al. The role of VP1 amino acid residue 145 of enterovirus 71 in viral fitness and pathogenesis in a cynomolgus monkey model. *PLoS Pathog* **2015**; 11:e1005033.
42. Zeng M, El Khatib NF, Tu S, et al. Seroepidemiology of enterovirus 71 infection prior to the 2011 season in children in Shanghai. *J Clin Virol* **2012**; 53:285-9.
43. Huang SW, Kiang D, Smith DJ, Wang JR. Evolution of re-emergent virus and its impact on enterovirus 71 epidemics. *Exp Biol Med* **2011**; 236:899-908

**Figure 1**

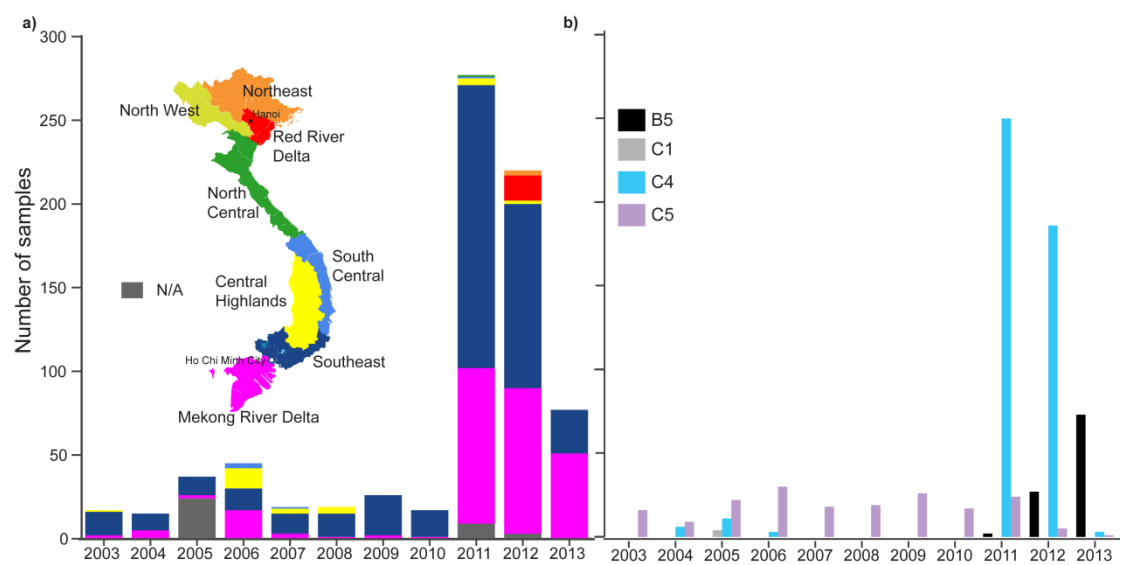


Figure 2

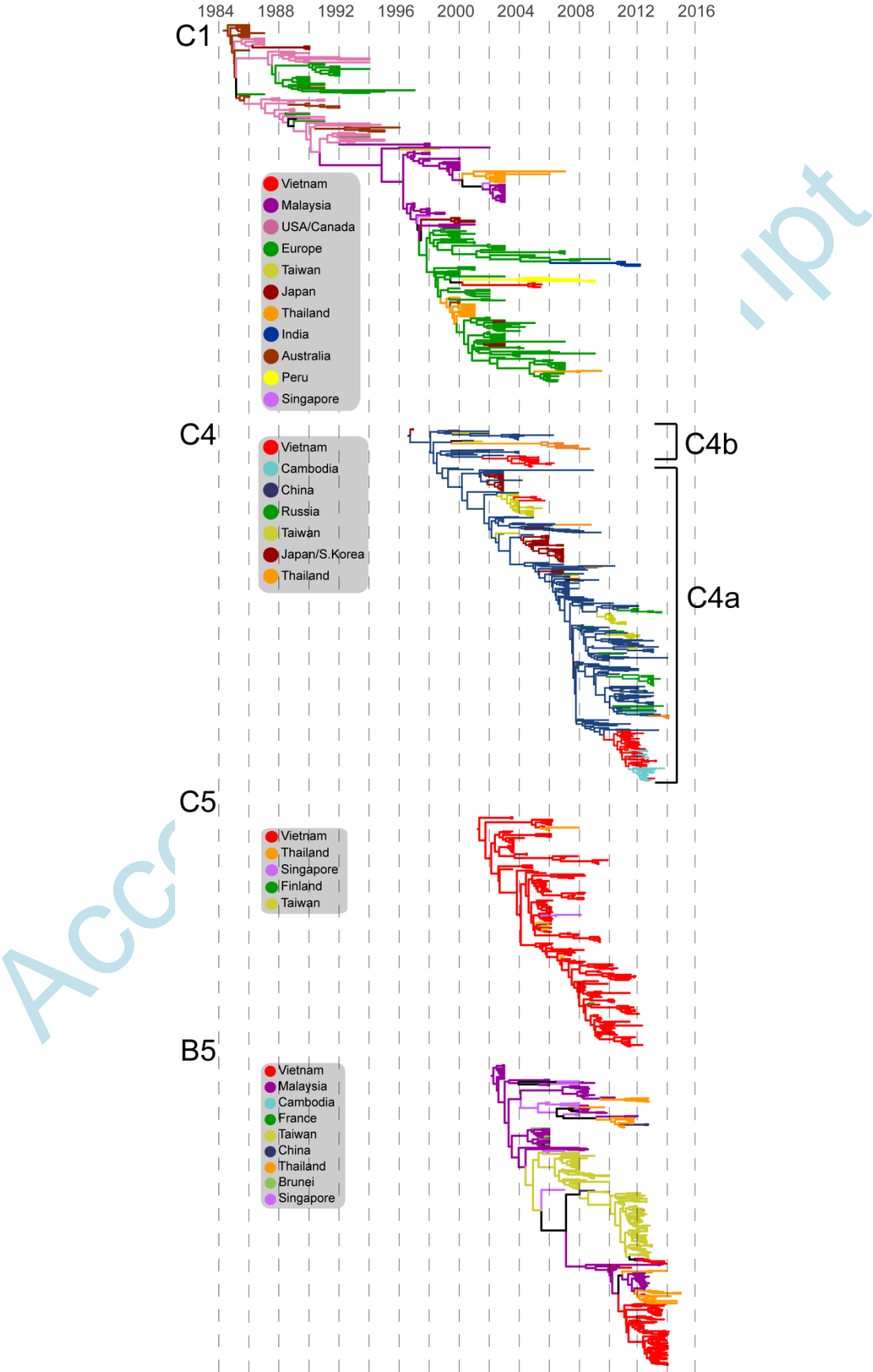


Figure 3

