Molecular epidemiology and the evolution of human coxsackievirus A6

Jiratchaya Phuenpa,[1] Sompong Vongpunsawad,[1] Riikka Osterback,[2] Matti Waris,[2] Eva Eriksson,[3] Jan Albert,[3] Sofie Midgley,[4] Thea K Fischer,[4] Anna M Eis-Hübinger,[5] María Cabrerizo,[6] Eleanor Gaunt,[7] Peter Simmonds,[7,8] and Yong Poovorawan[1]

[1]Center of Excellence in Clinical Virology, Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[2]Department of Virology, University of Turku, and Department of Clinical Virology, Turku University Hospital, 20520 Turku, Finland

[3]Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, 17177 and Department of Clinical Microbiology, Karolinska University Hospital, 17177, Stockholm, Sweden

[4]Department of Microbiological Diagnostics & Virology, Artillerivej 5, DK-2300 Copenhagen, Denmark

[5]Institute for Virology, University of Bonn Medical Center, Sigmund-Freud-Str. 25, D-53105 Bonn, Germany

[6]Instituto de Salud Carlos III, 28220 Majadahonda, Madrid, Spain

[7]Infection and Immunity Division, Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK

[8]Nuffield Department of Medicine, University of Oxford, Oxford, OX1 3SY, UK

*Corresponding author:

Prof. Yong Poovorawan, MD.,

Center of Excellence in Clinical Virology, Department of Pediatrics,

Faculty of Medicine, Chulalongkorn University, Bangkok, 10330 Thailand

E-mail: Yong.P@chula.ac.th. Telephone: +662-256-4909, Fax: +662-256-4929

1   Abstract: 201 words

2   Text: 3456 words

3   Figure: 6

4   Table: 2

5   Supplemental Figure: 1

6   Supplemental Table: 2

7   References: 45

8   Pages   22

9

10

11

12

13

14

15

16

17

18

19

20

21

Abstract

Published reports have described a dramatically increased incidence of coxsackievirus A6 (CV-A6) infections associated with hand, foot and mouth disease (HFMD) since their occurrence in 2008 in Finland. The aim of our study was to investigate the relationship between these disease outbreaks with the evolutionary dynamics of CV-A6 and the appearance of novel recombinant forms (RFs) of the virus. One hundred and fifty-one virus variants collected between 2013 and 2014 from different countries (Germany, Spain, Sweden, Denmark, and Thailand) were genotyped by using nested RT-PCR and sequencing of the VP1 capsid and 3D polymerase gene. Based on the analysis of the VP1 gene, the substitution rates of CV-A6 was estimated at 8.1×10^{-3} substitutions/site/year. There was an increasing likelihood between CV-A6 genome recombination and VP1 sequence divergence, with an estimated half-life of the RFs of 3.1 years. Bayesian phylogenetic analysis of the data showed that recently occurring recombination groups (RF-E, -F, -H, -J and -K) shared a common ancestor (RF-A). Recombination breakpoints were frequently observed between the 2A-2C gene and the 5´ untranslated region. This study revealed the potential for new CV-A6 variants to emerge and potentially modify disease outcomes of this major etiologic agent for HFMD affecting children worldwide.

Introduction

Coxsackievirus A6 (CV-A6) is a group of genetically diverse RNA viruses belonging to the *Picornaviridae* family and the enterovirus (EV) species. The enterovirus species contains important pathogenic viruses such as poliovirus, hepatitis A virus, and EV 71 viruses (Knowles, 2012). CV-A6 has been classified in the genus *Enterovirus* species A which currently comprises a total of 20 types. EVs possess a positive-stranded RNA genome of approximately ~7.400 nucleotides (or 7.4 kb), encased by a highly structured icosahedral

1   capsid. The viral genome is translated into a large polyprotein that is subsequently cleaved

2   into structural (VP1 to VP4) and non-structural proteins (2A to 2C and 3A to 3D). The VP1 is

3   the immunodominant structural protein and contains the most important serotype-specific

4   neutralization epitopes. The degree of similarity of nucleotides and amino acid sequences of

5   the VP1 region provides the primary tool for identification and assignment of new types

6   within a species; novel EV variants showing less than 75% nucleotide sequence identity are

7   classified as new types (Oberste *et al.*, 1999).

8   Infection with CV-A6 is frequently asymptomatic, but may also manifest as influenza

9   like syndrome or as a self-limiting hand, foot and mouth disease (HFMD). Previously,

10   HFMD typically occurred in young children and was predominantly caused by enterovirus 71

11   (EV-A71) and CV-A16. HFMD has only sporadically associated with other members of EV

12   species A (Puenpa *et al.*, 2011; Schuffenecker *et al.*, 2011; Wu *et al.*, 2010). More recently,

13   CV-A6 was first identified as the cause of epidemic HFMD with unusual clinical

14   manifestation in Finland in 2008 (Osterback *et al.*, 2009). Since then, HFMD cases

15   associated with CV-A6 have been reported in several countries in Europe, Asia and USA

16   (Flett *et al.*, 2012; Fujimoto *et al.*, 2012; Hayman *et al.*, 2014; Montes *et al.*, 2013; Puenpa *et

17   al.*, 2013; Tan *et al.*, 2015), reflecting the rapid emergence of a more pathogenic form of this

18   species A serotype. Lately, CV-A6 has been the most common EV type identified in several

19   outbreaks of HFMD that have typically been characterized by a more severe clinical

20   spectrum (eczema coxsackium and erosive lesions) as well as subsequent benign

21   onychomadesis (Ben-Chetrit *et al.*, 2014; Buttery *et al.*, 2015; Chatproedprai *et al.*, 2015;

22   Feder *et al.*, 2014; Lott *et al.*, 2013; Sinclair *et al.*, 2014; Wei *et al.*, 2011; Yasui *et al.*, 2013).

23   Since 2008, the predominant causative pathogen of HFMD in Thailand was either

24   EV-A71 or CV-A16, which, however changed from year to year. In 2012, a large-scale

25   outbreak of HFMD took place in Thailand with the highest incidence rate in a decade.

1   Meanwhile, CV-A6 had replaced EV-A71 and CV-A16 as the major EV types, and was

2   identified in 33.5% of surveillance samples in Thailand in 2012 (Puenpa *et al.*, 2013).

3   Moreover, there was a high number of CV-A6-infected patients presented with influenza like

4   symptoms (Puenpa *et al.*, 2014).

5   Recently, (Gaunt *et al.*, 2015) described a potential correlation between the

6   emergence of new CV-A6 variants with altered clinical phenotypes and the appearance of

7   novel recombinant forms (RFs) of the virus. The latter possessed phylogenetically distinct

8   3Dpol region sequences likely acquired from other species A serotypes through

9   recombination. They found that several different RFs have circulated over the past decade

10   worldwide. Notably, the emergence of the new recombinant form (RF-H) has been

11   associated with clinically reported cases of HFMD presenting as eczema herpeticum in

12   Edinburgh in 2014.

13   How widely RF-H and its predecessor, RF-A and other recently emerged RFs

14   circulate in Europe is unknown, nor is it clear whether the greater community circulation and

15   spiking incidence of HFMD in Thailand reflects the more widely spread of such

16   recombinants. To address these questions, we utilized molecular epidemiology by analyzing

17   VP1 sequence divergence and 3Dpol sequence grouping of a total of 151 CV-A6 isolates and

18   samples collected in several European countries (Germany, Spain, Sweden and Denmark)

19   and Thailand between 2013 and 2014.

20

21   **Materials and Methods**

22   **Samples.**

23   A total of 151 anonymized CV-A6-positive clinical samples from Denmark,

24   Germany, Spain, Sweden and Thailand collected between years 2013 and 2014 were selected

1   for genetic characterization. Sequences were named using the following convention: two-

2   letter country code/sample code/3Dpol clade/year of collection (e.g. DK/M22416/G/2014 for

3   sample code M22416 referred from Denmark, collected in 2014, and belonging to the 3Dpol

4   clade G). 3Dpol clade assignments were identified from the results of phylogeny (see

5   below). Sequences obtained in the current study were supplemented with 79 previously

6   published nucleotide sequences of complete genomes from 6 countries, including the CV-A6

7   prototype strain Gdula. The respective samples were collected between 1949 and 2014.

8   Details of the sequences used in the study including the countries of origin, sampling dates

9   and accession numbers are provided in Supplementary Table S1.

10   **Amplification of VP1 and 3Dpol regions.**

11   Nested reverse transcription-PCRs (RT-PCRs) were carried out using newly designed

12   primers listed in Table S2 to amplify a 1,331-bp region of the VP1 gene and a 1,200-bp

13   region of the 3Dpol gene. The reverse transcription and first round PCR were performed

14   using Superscript III One-Step RT-PCR system with Platinum Taq High Fidelity according to

15   the manufacturer's instructions (Invitrogen, Carlsbad, CA). The following conditions were

16   used: heating at 43 ˚C for 1 h, then 20 cycles of 53 ˚C for 1 min and 55 ˚C for 1 min,

17   followed by 70 ˚C for 15 min and 94 ˚C for 2 min. PCR cycles included 40 cycles of 94 ˚C

18   (30 s), 50 ˚C (30 s), and 68 ˚C (1 min 45 s) and a final extension at 68 ˚C for 5 min. One

19   microliter of the first-round reaction was used as template in the second-round PCR

20   containing second-round primers and GoTaq DNA polymerase (Promega, Madison, WI) for

21   30 cycles of 94 ˚C for 30 s, 50 ˚C for 30 s and 72 ˚C for 90 s, followed by a final extension

22   step for 5 min at 72 ˚C. Sequencing of the amplification products was carried out directly in

23   both directions using ABI 7200 BigDye capillary sequencing (Applied Biosystems,

24   Framingham, MA).

1   **Complete genome sequencing.**

2   The full-length genome of 39 CV-A6 variants representing four of the recombination

3   groups were sequenced (RF-A, n = 21; RF-F, n = 10; RF-G, n = 2; and RF-H, n = 6) using

4   primer sets (Table S2) designed according to CV-A6-specific nucleotide sequences

5   downloaded from GenBank. cDNA was synthesized with Superscript III reverse

6   transcriptase (Invitrogen, Carlsbad, CA) and random hexamer primers and used as the

7   template for nested PCR. The cycling conditions were identical to the second-round PCR

8   mentioned above.

9   **Sequence analysis.**

10   Sequences were formatted, assembled and aligned using the SSE 1.2 sequence editor

11   package (www.virus-evolution.org) (Simmonds, 2012). Phylogenetic trees were constructed

12   using the Maximum-likelihood method and determined best-fit model of nucleotide

13   substitution for each sequence data set using a correction value in the model selection

14   procedure implemented in the MEGA program (v6) (Tamura *et al.*, 2013). In the analyses,

15   Kimura two-parameter (K2P) with invariant sites (I) was selected as the substitution model

16   for VP1 and 5UTR; K2P with I and gamma distribution (I-) for 3Dpol; K2P with I- for VP4/2

17   and 1000 bootstraps.

18   Rates of evolution and molecular clock phylogeny (Drummond & Rambaut, 2007)

19   were estimated from VP1 gene sequences using the Bayesian Markov Chain Monte Carlo

20   (MCMC) method implemented in BEAST, version 1.8.0 (Drummond *et al.*, 2012). Bayesian

21   MCMC analyses were performed using a relaxed log-normal molecular clock model. The

22   dynamic among study populations was estimated by performing a constant size and

23   exponential growth models. Each Bayesian MCMC analysis was for 100,000,000 iterations

24   and sampled every 10,000 states and discarding 10% of the chain as burn-in. Convergence of

1   the chains and effective sample sizes of the estimates were checked using Tracer

2   (http://beast.bio.ed.ac.uk/Tracer). The resulting tree of each run was summarized using Tree

3   Annotator, and the maximum clade credibility tree was visualized with

4   TreeAnnotator/FigTree.

5   **Nucleotide sequence accession numbers.**

6   All newly generated sequences obtained in this study were submitted to GenBank and

7   were assigned the accession numbers KX212338 - KX212678.

8

9   **Results**

10   **Phylogeny of CV-A6 VP1 and 3Dpol genome regions**

11   CV-A6 variants associated with HFMD presenting with eczema herpeticum in

12   Edinburgh in 2014 were clustered in a novel appearance clade (RF-H) distinct from other

13   previously characterized recombination groups assigned in the 3D region (Gaunt *et al.*,

14   2015). These variants were also phylogenetically distinct in the VP1 region (Gaunt *et al.*,

15   2015). In the current study we have extended the analysis of VP1 and 3D regions to CV-A6

16   variants identified in wider geographical area including Denmark ($n = 22$), Germany ($n = 4$),

17   Spain ($n = 14$), Sweden ($n = 6$) and Thailand ($n = 105$). CV-A6 sequences for the VP1 region

18   between positions 2485 and 3816 (numbering based on the Gdula prototype strain; GenBank

19   accession number AY421764) were obtained.

20   VP1 sequences from Denmark and Spain (2014) clustered within the same lineage

21   (lineage I) with those analysed previously from the patients with eczema herpeticum

22   presenting in 2014 in Edinburgh but distinct from previous outbreaks in Taiwan (lineage II),

23   Thailand (lineage III), China (lineage VI) and Finland (lineage V) (Fig. S1).

1   In order to investigate the occurrence of recombination, sequences from the 3Dpol

2   region were obtained and analyzed for each of the 151 isolates included in the VP1 analysis.

3   The data set was supplemented by 39 (near-) complete genome sequences of CAV6 variants

4   from Denmark (n = 8), Germany (n = 3), Spain (n = 3) and Thailand (n = 25). CAV6 3Dpol

5   clades comprising groups A, B, C, D, E, F, G, H, I, J and K were designated through

6   identification of bootstrap-supported clades as described previously for other enteroviruses

7   (McWilliam Leitch *et al.*, 2009; McWilliam Leitch *et al.*, 2012; McWilliam Leitch *et al.*,

8   2010). Designations follow the nomenclature established in our previous analysis of

9   Edinburgh CV-A6 isolates (Gaunt et al., 2014). The majority of variants in the study

10   clustered within two of the previously assigned recombinant forms, RF-A (99/151) and RF-F

11   (37/151) (Table 1). Four Danish variants clustered within clade RF-H, as did two of the

12   variants from Spain. Three variants from Denmark and Spain were assigned as RF-G (Fig.

13   1b). No sequences within this study clustered within the previously described RF-B, RF-C,

14   RF-D and RF-E. None of the CV-A6 3Dpol groups contained sequences from other species

15   A serotypes.

16   Phylogenetic analysis of the VP1, VP4/2 and the 5´UTR produced broadly similar

17   groupings of CV-A6 variants in each region (Fig. 1a, 2a and 2b). However, there were

18   several examples of incongruent groupings that indicate the existence of further

19   recombination within CV-A6. For example, RF-E variants grouped with RF-A and –H in

20   VP1 region but changed position in the 5´UTR and VP4/2. In contrast to VP1, RF-A were

21   interspersed in the 5´UTR and VP4/2 regions with the member of RF-J. The phylogenetic

22   group formed of RF-K was monophyletic in the VP1, VP4/2 and the 5´UTR regions but the

23   only one exception grouped with RF-D. These findings indicate the occurrence of multiple

24   recombinant events between VP1, VP4/2 and the 5'UTR.

25

1 **Sequence diversity of CV-A6**

2   Sequence divergence in VP1 provides a proxy measure for the time of divergence of

3 CV-A6 variants from which estimates of the life spans of individual recombinant forms can

4 be estimated. By pairwise comparison of each CV-A6 variant with each other, there was a

5 precise correlation between VP1 sequence distances and assignment to the same or a different

6 RF group in 3Dpol; a divergence threshold of 0.1 divided the two sets of comparisons (Fig. 3

7 and 4a).

8   CV-A6 VP1 variability was restricted primarily to synonymous sites, which indicates

9 that most changes in sequences occurred through neutral drift. Through molecular clock

10 analysis, the nucleotide substitution rate and times to the most recent common ancestor

11 (tMRCAs) of different regions and the assignment of 3Dpol clades of the greatest RFs (RF-A

12 and RF-F) were estimated. The substitution rate of the whole data set of all VP1 sequences

13 was estimated to be 8.1×10^{-3} substitutions/site/year (high-probability distribution [HPD]

14 range, 6.0×10^{-3} to 10.5×10^{-3}); the date that the MRCA of all CV-A6 clusters existed was

15 estimated to be 1947 (HPD, 1940 to 1949). The estimated date of the MRCA of RF-A was

16 1999 (HPD, 1995 to 2003); since then, RF-A lineage has diverged into five recombination

17 groups (RF-E, -F, -H, -J and -K). The VP1 lineage containing the RF-G recombinant form

18 was first documented in 2011 and thereafter underwent the recombination event that

19 exchanged the 3Dpol region sequences between 2004 and 2011. VP1 lineage 1 contains CV-

20 A6 variants belonging to RF-F with the recombination event dated between 2009 and 2012.

21 A further RF group (H) that appeared in 2013 probably recombined between 2011 and 2013.

22   Using the substitution rates estimated from VP1 sequence analysis, an estimation of

23 approximate half-lifes of CV-A6 lineages were calculated by combining the mean sequence

24 divergence in VP1 at the 50% recombinant frequency threshold (0.05) with the substitution

1   rate in VP1 (8.1×10^{-3} substitutions/site/year). The RF half-lifes of CV-A6 was

2   approximately 3.1 years, thus within the range estimated from previous analyses of other

3   enterovirus serotypes (Fig. 4b).

4   To determine more precisely the timescale of recombination events underlying the

5   appearance of each RF, datasets of VP1 gene sequences were analysed using the Bayesian

6   Markov Chain Monte Carlo (MCMC) method to generate time-correlated phylogeny (Fig. 5).

7   While earlier recombination events could not be reconstructed in any detail due to inadequate

8   sampling of CV-A6 before 2008, variants collected after this date were monophyletic falling

9   into three further lineages with estimated dates of splitting betweeen 2004 and 2005. The

10   oldest lineage comprised purely RF-G samples while the other two lineages contained

11   samples belonging to RF-A. One comprised solely RF-A and persisted for at least 11 years

12   (2005 to 2015). The other lineage contains RF-A sequences and samples isolated

13   subsequently belonging to other RF groups (E, F, H, J and K). Variants within lineage 2 (RF-

14   A) were those detected in the first HFMD outbreak in Finland in 2008 along with variants

15   detected subsequently in Europe and Asia over the following 1 to 5 years. In lineage 1, the

16   oldest variants were those originally described in Asia (Thailand and Japan) before 2012 and

17   then spread into Europe and Asia between 2013 and 2014. Within this lineage, several

18   separate recombination events generated RFs including an RF-F group described in Thailand

19   in 2012 (Puenpa *et al.*, 2013). The RFs (F, G, H and K) were monophyletic and likely

20   originated from single recombination events, unlike RF-E and RF-J which were detected in

21   more than one VP1 lineage. With the exception of RF-F which was isolated from both

22   Europe and Asia, other RFs groups were detected only from Europe or Asia; G and H

23   (Europe), E (Taiwan), J and K (China).

1   Having identified the likely time course and direction of the recombination events,

2   divergence scan analyses were performed between the ancestral RF-A sequences with

3   complete genome sequences generated in the current study (RF-F, -G, and -H) and the

4   previous study (RF-J and RF-K) recombinant forms to identify recombination breakpoints

5   (Fig. 6). The sharp increase in sequence divergence at various points in the P2 region

6   provided evidence for the occurrence of separate, individual recombination events for each

7   RF. The first breakpoint was founded at the 2A protein-encoding region around nucleotide

8   position 3500 (RF-G). The next breakpoints were located in 2B and at the border between 2B

9   and 2C regions (RF-F, -J and -H). The last breakpoint of RF-K can be recognized at

10   nucleotide position 5000 and included the 3' part of the 2C region.

11   **Discussion**

12   CV-A6 infection represents a significant global public health concern and has become

13   one of the most frequently detected enteroviral cause of HFMD. This study sought to

14   understand how widespread the new recombinant forms of CV-A6 have become in Europe

15   and other geographical regions through analysis of the molecular epidemiology and dynamics

16   of recombination of CV-A6 variants collected within the last 8 years. Differences in the

17   phylogenetic trees of VP1 and 3Dpol regions provided a strong evidence of recombination in

18   several strains. Specifically, three CV-A6 variants did not cluster within the same

19   recombinant group on the VP1 phylogenetic tree. The TW/00141/2007 strain was assigned

20   as RF-E, but grouped with RF-B. The CHN/P143/2013 strain was assigned as RF-J, but

21   grouped with RF-A. Finally, the CHN/CC13/2013 strain was assigned into RF-K, but

22   grouped with RF-D. The patterns of phylogenetic discordance observed from this study

23   suggests the occurrence of recombination in CV-A6 consistent with previous findings for

24   other enteroviruses (Cabrerizo *et al.*, 2014; Calvert *et al.*, 2010; McIntyre *et al.*, 2010;

1 McWilliam Leitch *et al.*, 2009; McWilliam Leitch *et al.*, 2012; McWilliam Leitch *et al.*,

2 2010).

3   As previously described (McWilliam Leitch *et al.*, 2009), the RF lifespans of CV-A6

4 were calculated by measuring the relationship between VP1 divergence and the likelihood of

5 recombination. The RF half-life of CV-A6 was closer to that estimated previously for E-30

6 (3.1 years), much shorter than that of EV-A71 and E-11 whereas the value was higher those

7 of E-6 and E-9 (Cabrerizo *et al.*, 2014; McWilliam Leitch *et al.*, 2009; McWilliam Leitch *et*

8 *al.*, 2012; McWilliam Leitch *et al.*, 2010). Based on analysis of Bayesian MCMC methods,

9 the rate of evolutionary change for CV-A6 in the VP1 region of 8.1×10^{-3} substitutions per

10 site per year (Table 2) fell within the middle of the range described for other enteroviruses, as

11 examples, CV-A24 (1.2×10^{-3}), CV-A21 (3.1×10^{-3}), EV-C96 (3.3×10^{-3}), EV-C99 (3.7×10^{-3}

12 $^{-3}$), CV-B5 (4.2×10^{-3}), E-11 (4.8×10^{-3}), EV-D68 (4.93×10^{-3}), EV-D70 (5×10^{-3}), CV-A4

13 (5.5×10^{-3}), E-9 (5.8×10^{-3}), EV-A71 ($3.1\text{-}7.2 \times 10^{-3}$), CV-A2 ($8.3 \times 10^{-3}$), E-30 ($8.8 \times 10^{-3}$),

14 CV-A16 (9.1×10^{-3}), poliovirus (10.3×10^{-3}), E-6 (11.2×10^{-3}), CV-A10 (14.1×10^{-3})

15 (Cabrerizo *et al.*, 2014; Gullberg *et al.*, 2010; Jenkins *et al.*, 2002; Jorba *et al.*, 2008;

16 Linsuwanon *et al.*, 2012; Lukashev *et al.*, 2014; McWilliam Leitch *et al.*, 2009; McWilliam

17 Leitch *et al.*, 2012; McWilliam Leitch *et al.*, 2010; Mirand *et al.*, 2010; Smura *et al.*, 2014;

18 Takeda *et al.*, 1994; Tee *et al.*, 2010; Yip *et al.*, 2013; Zhang *et al.*, 2010).

19   Phylogenetic reconstruction was used to analyze trait evolution such as temporal and

20 geographical correlates of individual recombination events. The most frequently detected

21 recombinant form, RF-A, showed decades-long circulation and was the ancestor of five

22 separate recombinant groups (RF-E, F, H, K and J) that have emerged in the past 5-10 years.

23 The first outbreak of CV-A6-HFMD in Finland in 2008 was associated with RF-A with the

24 subsequent appearance of this RF across of Europe and Asia between 2013 and 2014. The

1   more recent emergence of RFs in Europe and Asia between 2013 and 2014 originated from

2   descendants of multiple VP1 lineages that have diverged from RF-A variants circulating in

3   Asia (Thailand and Japan) between 2008 and 2010. Recombination events have previously

4   been recorded widely in other enteroviruses and were found to play a significant role in the

5   evolution of these genomic viruses and the recombination breakpoints detected in this study

6   (2A - 2C regions) are well-known recombination hotspots in enteroviruses (Lukashev *et al.*,

7   2005). Regarding CV-A6, breakpoints within VP3 and between 5´UTR and VP1 have been

8   detected in the genomes of RF-E variants in Taiwan (Gaunt *et al.*, 2015).

9   This study reports a detailed, multi-centre investigation of the emergence of an

10   enterovirus serotype associated with epidemics of HFMD. It catalogues the complexity of the

11   evolutionary processes associated with its geographical expansion and the occurrence of a

12   number of recombination events each involving replacement of close to complete non-

13   structural gene blocks at varying times since the founder recombinant form, RF-A was first

14   described in 2008 (Osterback *et al.*, 2009). Non-structural (NS) region sequences of most RFs

15   have not been described in association with other species A serotypes including any of the

16   RFs described for EV-A71 (McWilliam Leitch *et al.*, 2012). However, there was some

17   evidence for a limited degree of re-circulation within the recombination pool of NS region

18   sequences; RFs -E and -K appear at two different positions in the VP1 phylogenetic tree (Fig.

19   1a). Furthermore, a limited degree of sharing of NS regions sequences between different

20   species B serotypes has been documented (Bailly *et al.*, 2009).

21   Nevertheless, the typical pattern for an RF was its rapid emergence, variable

22   penetrance into the sampled virus population and relatively rapid extinction, within years

23   rather than decades, based on average recombination half-lives documented for CV-A6 and

24   other EV types (Fig. 4b). This patterns is well attested in the turnover of RFs of CV-A6.

25   While we remain relatively ignorant of the reasons for RF turnover, whether driven by

1   immunological, host adaptive factors or transmissibility, or alternatively whether it occurs as

2   a consequence of population bottlenecks and replacements without a fitness component,

3   molecular epidemiological studies such as these will be of value in gaining a longer term,

4   better understanding of the nature of enterovirus evolution and their clinical outcomes.

Acknowledgements

References

11 **Bailly, J. L., Mirand, A., Henquell, C., Archimbaud, C., Chambon, M., Charbonne, F., Traore, O. &
12   Peigue-Lafeuille, H. (2009).** Phylogeography of circulating populations of human echovirus
13   30 over 50 years: nucleotide polymorphism and signature of purifying selection in the VP1
14   capsid protein gene. *Infect Genet Evol* **9**, 699-708.
15 **Ben-Chetrit, E., Wiener-Well, Y., Shulman, L. M., Cohen, M. J., Elinav, H., Sofer, D., Feldman, I.,
16   Marva, E. & Wolf, D. G. (2014).** Coxsackievirus A6-related hand foot and mouth disease: skin
17   manifestations in a cluster of adult patients. *J Clin Virol* **59**, 201-203.
18 **Buttery, V. W., Kenyon, C., Grunewald, S., Oberste, M. S. & Nix, W. A. (2015).** Atypical
19   Presentations of Hand, Foot, and Mouth Disease Caused by Coxsackievirus A6--Minnesota,
20   2014. *MMWR Morb Mortal Wkly Rep* **64**, 805.
21 **Cabrerizo, M., Trallero, G. & Simmonds, P. (2014).** Recombination and evolutionary dynamics of
22   human echovirus 6. *J Med Virol* **86**, 857-864.
23 **Calvert, J., Chieochansin, T., Benschop, K. S., McWilliam Leitch, E. C., Drexler, J. F., Grywna, K., da
24   Costa Ribeiro, H., Jr., Drosten, C., Harvala, H., Poovorawan, Y., Wolthers, K. C. &
25   Simmonds, P. (2010).** Recombination dynamics of human parechoviruses: investigation of
26   type-specific differences in frequency and epidemiological correlates. *J Gen Virol* **91**, 1229-
27   1238.
28 **Chatproedprai, S., Tempark, T., Wanlapakorn, N., Puenpa, J., Wananukul, S. & Poovorawan, Y.
29   (2015).** Unusual skin manifestation of hand, foot and mouth disease associated with
30   coxsackievirus A6: cases report. *Springerplus* **4**, 362.
31 **Drummond, A. J. & Rambaut, A. (2007).** BEAST: Bayesian evolutionary analysis by sampling trees.
32   *BMC Evol Biol* **7**, 214.
33 **Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012).** Bayesian phylogenetics with BEAUti
34   and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973.
35 **Feder, H. M., Jr., Bennett, N. & Modlin, J. F. (2014).** Atypical hand, foot, and mouth disease: a
36   vesiculobullous eruption caused by Coxsackie virus A6. *Lancet Infect Dis* **14**, 83-86.

Flett, K., Youngster, I., Huang, J., McAdam, A., Sandora, T. J., Rennick, M., Smole, S., Rogers, S. L., Nix, W. A., Oberste, M. S., Gellis, S. & Ahmed, A. A. (2012). Hand, foot, and mouth disease caused by coxsackievirus a6. *Emerg Infect Dis* **18**, 1702-1704.

Fujimoto, T., Iizuka, S., Enomoto, M., Abe, K., Yamashita, K., Hanaoka, N., Okabe, N., Yoshida, H., Yasui, Y., Kobayashi, M., Fujii, Y., Tanaka, H., Yamamoto, M. & Shimizu, H. (2012). Hand, foot, and mouth disease caused by coxsackievirus A6, Japan, 2011. *Emerg Infect Dis* **18**, 337-339.

Gaunt, E., Harvala, H., Osterback, R., Sreenu, V. B., Thomson, E., Waris, M. & Simmonds, P. (2015). Genetic characterization of human coxsackievirus A6 variants associated with atypical hand, foot and mouth disease: a potential role of recombination in emergence and pathogenicity. *J Gen Virol* **96**, 1067-1079.

Gullberg, M., Tolf, C., Jonsson, N., Mulders, M. N., Savolainen-Kopra, C., Hovi, T., Van Ranst, M., Lemey, P., Hafenstein, S. & Lindberg, A. M. (2010). Characterization of a putative ancestor of coxsackievirus B5. *J Virol* **84**, 9695-9708.

Hayman, R., Shepherd, M., Tarring, C. & Best, E. (2014). Outbreak of variant hand-foot-and-mouth disease caused by coxsackievirus A6 in Auckland, New Zealand. *J Paediatr Child Health* **50**, 751-755.

Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* **54**, 156-165.

Jorba, J., Campagnoli, R., De, L. & Kew, O. (2008). Calibration of multiple poliovirus molecular clocks covering an extended evolutionary range. *J Virol* **82**, 4429-4440.

Knowles, N. J., Hovi, T., Hyypia, T., King, A. M. Q., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Simmonds, P., Skern, T. & other authors (2012). Family Picornaviridae. In Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, pp. 855–880. *Edited by A M Q King, M J Adams, E B Carstens & E J Lefkowitz London: Academic Press*.

Linsuwanon, P., Puenpa, J., Suwannakarn, K., Auksornkitti, V., Vichiwattana, P., Korkong, S., Theamboonlers, A. & Poovorawan, Y. (2012). Molecular epidemiology and evolution of human enterovirus serotype 68 in Thailand, 2006-2011. *PLoS One* **7**, e35190.

Lott, J. P., Liu, K., Landry, M. L., Nix, W. A., Oberste, M. S., Bolognia, J. & King, B. (2013). Atypical hand-foot-and-mouth disease associated with coxsackievirus A6 infection. *J Am Acad Dermatol* **69**, 736-741.

Lukashev, A. N., Lashkevich, V. A., Ivanova, O. E., Koroleva, G. A., Hinkkanen, A. E. & Ilonen, J. (2005). Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. *J Gen Virol* **86**, 3281-3290.

Lukashev, A. N., Shumilina, E. Y., Belalov, I. S., Ivanova, O. E., Eremeeva, T. P., Reznik, V. I., Trotsenko, O. E., Drexler, J. F. & Drosten, C. (2014). Recombination strategies and evolutionary dynamics of the Human enterovirus A global gene pool. *J Gen Virol* **95**, 868-873.

McIntyre, C. L., McWilliam Leitch, E. C., Savolainen-Kopra, C., Hovi, T. & Simmonds, P. (2010). Analysis of genetic diversity and sites of recombination in human rhinovirus species C. *J Virol* **84**, 10297-10310.

McWilliam Leitch, E. C., Bendig, J., Cabrerizo, M., Cardosa, J., Hyypia, T., Ivanova, O. E., Kelly, A., Kroes, A. C., Lukashev, A., MacAdam, A., McMinn, P., Roivainen, M., Trallero, G., Evans, D. J. & Simmonds, P. (2009). Transmission networks and population turnover of echovirus 30. *J Virol* **83**, 2109-2118.

McWilliam Leitch, E. C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O. E., Koike, S., Kroes, A. C., Lukashev, A., Perera, D., Roivainen, M., Susi, P., Trallero, G., Evans, D. J. & Simmonds, P. (2012). The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71. *J Virol* **86**, 2676-2685.

McWilliam Leitch, E. C., Cabrerizo, M., Cardosa, J., Harvala, H., Ivanova, O. E., Kroes, A. C., Lukashev, A., Muir, P., Odoom, J., Roivainen, M., Susi, P., Trallero, G., Evans, D. J. & Simmonds, P. (2010). Evolutionary dynamics and temporal/geographical correlates of

recombination in the human enterovirus echovirus types 9, 11, and 30. *J Virol* **84**, 9292-9300.

Mirand, A., Schuffenecker, I., Henquell, C., Billaud, G., Jugie, G., Falcon, D., Mahul, A., Archimbaud, C., Terletskaia-Ladwig, E., Diedrich, S., Huemer, H. P., Enders, M., Lina, B., Peigue-Lafeuille, H. & Bailly, J. L. (2010). Phylogenetic evidence for a recent spread of two populations of human enterovirus 71 in European countries. *J Gen Virol* **91**, 2263-2277.

Montes, M., Artieda, J., Pineiro, L. D., Gastesi, M., Diez-Nieves, I. & Cilla, G. (2013). Hand, foot, and mouth disease outbreak and coxsackievirus A6, northern Spain, 2011. *Emerg Infect Dis* **19**.

Oberste, M. S., Maher, K., Kilpatrick, D. R. & Pallansch, M. A. (1999). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J Virol* **73**, 1941-1948.

Osterback, R., Vuorinen, T., Linna, M., Susi, P., Hyypia, T. & Waris, M. (2009). Coxsackievirus A6 and hand, foot, and mouth disease, Finland. *Emerg Infect Dis* **15**, 1485-1488.

Puenpa, J., Chieochansin, T., Linsuwanon, P., Korkong, S., Thongkomplew, S., Vichaiwattana, P., Theamboonlers, A. & Poovorawan, Y. (2013). Hand, foot, and mouth disease caused by coxsackievirus A6, Thailand, 2012. *Emerg Infect Dis* **19**, 641-643.

Puenpa, J., Mauleekoonphairoj, J., Linsuwanon, P., Suwannakarn, K., Chieochansin, T., Korkong, S., Theamboonlers, A. & Poovorawan, Y. (2014). Prevalence and characterization of enterovirus infections among pediatric patients with hand foot mouth disease, herpangina and influenza like illness in Thailand, 2012. *PLoS One* **9**, e98888.

Puenpa, J., Theamboonlers, A., Korkong, S., Linsuwanon, P., Thongmee, C., Chatproedprai, S. & Poovorawan, Y. (2011). Molecular characterization and complete genome analysis of human enterovirus 71 and coxsackievirus A16 from children with hand, foot and mouth disease in Thailand during 2008-2011. *Arch Virol* **156**, 2007-2013.

Schuffenecker, I., Mirand, A., Antona, D., Henquell, C., Chomel, J. J., Archimbaud, C., Billaud, G., Peigue-Lafeuille, H., Lina, B. & Bailly, J. L. (2011). Epidemiology of human enterovirus 71 infections in France, 2000-2009. *J Clin Virol* **50**, 50-56.

Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* **5**, 50.

Sinclair, C., Gaunt, E., Simmonds, P., Broomfield, D., Nwafor, N., Wellington, L., Templeton, K., Willocks, L., Schofield, O. & Harvala, H. (2014). Atypical hand, foot, and mouth disease associated with coxsackievirus A6 infection, Edinburgh, United Kingdom, January to February 2014. *Euro Surveill* **19**, 20745.

Smura, T., Blomqvist, S., Vuorinen, T., Ivanova, O., Samoilovich, E., Al-Hello, H., Savolainen-Kopra, C., Hovi, T. & Roivainen, M. (2014). The evolution of Vp1 gene in enterovirus C species sub-group that contains types CVA-21, CVA-24, EV-C95, EV-C96 and EV-C99. *PLoS One* **9**, e93737.

Takeda, N., Tanimura, M. & Miyamura, K. (1994). Molecular evolution of the major capsid protein VP1 of enterovirus 70. *J Virol* **68**, 854-862.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.

Tan, X., Li, L., Zhang, B., Jorba, J., Su, X., Ji, T., Yang, D., Lv, L., Li, J. & Xu, W. (2015). Molecular epidemiology of coxsackievirus A6 associated with outbreaks of hand, foot, and mouth disease in Tianjin, China, in 2013. *Arch Virol* **160**, 1097-1104.

Tee, K. K., Lam, T. T., Chan, Y. F., Bible, J. M., Kamarulzaman, A., Tong, C. Y., Takebe, Y. & Pybus, O. G. (2010). Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *J Virol* **84**, 3339-3350.

Wei, S. H., Huang, Y. P., Liu, M. C., Tsou, T. P., Lin, H. C., Lin, T. L., Tsai, C. Y., Chao, Y. N., Chang, L. Y. & Hsu, C. M. (2011). An outbreak of coxsackievirus A6 hand, foot, and mouth disease associated with onychomadesis in Taiwan, 2010. *BMC Infect Dis* **11**, 346.

Wu, Y., Yeo, A., Phoon, M. C., Tan, E. L., Poh, C. L., Quak, S. H. & Chow, V. T. (2010). The largest outbreak of hand; foot and mouth disease in Singapore in 2008: the role of enterovirus 71 and coxsackievirus A strains. *Int J Infect Dis* **14**, e1076-1081.

Yasui, Y., Makino, T., Hanaoka, N., Owa, K., Horikoshi, A., Tanaka, A., Suehiro, Y., Shimizu, H., Kanou, K., Kobayashi, M., Konagaya, M. & Fujimoto, T. (2013). A case of atypical hand-foot-and-mouth disease caused by coxsackievirus A6: differential diagnosis from varicella in a pediatric intensive care unit. *Jpn J Infect Dis* **66**, 564-566.

Yip, C. C., Lau, S. K., Lo, J. Y., Chan, K. H., Woo, P. C. & Yuen, K. Y. (2013). Genetic characterization of EV-A71 isolates from 2004 to 2010 reveals predominance and persistent circulation of the newly proposed genotype D and recent emergence of a distinct lineage of subgenotype C2 in Hong Kong. *Virol J* **10**, 222.

Zhang, Y., Wang, D., Yan, D., Zhu, S., Liu, J., Wang, H., Zhao, S., Yu, D., Nan, L., An, J., Chen, L., An, H., Xu, A. & Xu, W. (2010). Molecular evidence of persistent epidemic and evolution of subgenotype B1 coxsackievirus A16-associated hand, foot, and mouth disease in China. *J Clin Microbiol* **48**, 619-622.

Legend

Figure 1. Phylogenetic analysis of VP1 (a) and 3Dpol (b) sequences of CV-A6 for study subjects and those previously determined (RF-B, -C, -D, and -E). Each sequence is labelled by country of origin, sample code, 3Dpol clade assignment and year of collection. Dot colors indicate the recombination group assignments based on 3Dpol phylogeny. Evolutionary distances are scaled according to the bars in the upper left-hand corner of each tree (number of nucleotide substitutions per site).

Figure 2. Phylogenetic analysis of 5´UTR (a) and VP4/2 (b) sequences of CV-A6 for study subjects and those previously determined (RF-B, -C, -D, and -E). Each sequence is labelled by country of origin, sample code, 3Dpol clade assignment and year of collection. Dot colors indicate the recombination group assignments based on 3Dpol phylogeny. Evolutionary distances are scaled according to the bars in the upper left-hand corner of each tree (number of nucleotide substitutions per site).

Figure 3. Histogram of genetic distances between CV-A6 sequences. The histogram shows the distribution of pairwise distances between sequences from this study and published CV-A6 sequences in 3Dpol region. The dashed line represents the threshold between a member of the same clades and between different clades.

1 **Figure 4.** (A) Association between VP1 sequence divergence (shown on the x-axis) and the

2 proportion of recombinant comparisons. (B) Comparison of mean half-lifes of CV-A6 with

3 previously estimated by the same method.

4 **Figure 5.** A dated phylogeny of VP1 sequences of CV-A6 variants in this study and

5 published sequences. Branch colours label recombination groups in each clade.

6 **Figure 6.** Divergence scan of nucleotide sequence between RF-A with other recombination

7 groups (RF-F, H, G, K and J).

8 **Table 1.** Recombination groups (RF-A to -K) based on phylogenetic analysis of the 3Dpol

9 region**.**

10 **Table 2.** Rates of sequence change and TMRCAs by MCMC analysis.

11

1   **Table 1.** Recombination groups (RF-A to -K) based on phylogenetic analysis of the 3Dpol region**.**

Country	RF-A	RF-B	RF-C	RF-D	RF-E	RF-F	RF-G	RF-H	RF-J	RF-K	Total
All	105					37	3	6			151
Denmark	13					3	2	4			22
Germany	3					1					4
Spain	10					1	1	2			14
Sweden	6										6
Thailand	73					32					105

2

3

4

5

6

7

8

9

10

11

12

13

1   **Table 2.** Rates of sequence change and TMRCAs by MCMC analysis.

RF group and geographic set	RF	n[†]	Divergence[*]				dN/dS		MCMC (BEAST)			
			Nucleotide		Amino acid				Substitution rate (10^{-3})[#]		tMRCA[‡]	
			VP1	3Dpol	VP1	3Dpol	VP1	3Dpol	VP1	3Dpol	VP1	3Dpol
Whole data set												
All	All	244	0.052	ND	0.014	ND	0.035	ND	8.1 (6.0-10.5)	ND	68.2 (66.2-75.4)	ND
Individual RF group												
All	RF-A	160	0.046	0.047	0.012	0.020	0.036	0.057	10.4 (7.6-13.8)	11.0 (8.2-14.4)	15.9 (12.3-20.4)	12.9 (12.2-14.4)
All	RF-F	36	0.016	0.013	0.010	0.006	0.097	0.055	1.4 (0.03-3.0)	6.5 (3.6-10.2)	17.1 (3.0-42.1)	3.5 (2.7-4.4)

2

3   [*] Mean pairwise P distances. ND, not determined.

4   [†] Number of sequences in each set analyzed.

5   [#] Mean value with the HPD interval in parentheses.

6   [‡] Time before the present of the most recent common ancestor

7

8

9

10

11

12

1 **Supporting information legends**

2 **Supplementary Figure 1. Phylogenetic analysis of VP1 sequences of CV-A6 for study subjects and those previously determined (RF-B, -**

3 **C, -D, and -E).** The tree display the color code for different countries.

4 **Supplementary Table S1.** Sequences information

5 **Supplementary Table S2.** Primer set used for whole genome amplification by nested RT-PCR

6