Molecular epidemiology and the evolution of human coxsackievirus A6

Jiratchaya Puenpa, Sompong Vongpunsawad, Riikka Österback, Matti Waris, Eva Eriksson, Jan Albert, Sofie Midgley, Thea K. Fischer, Anna M. Eis-Hübinger, María Cabrerizo, Eleanor Gaunt, Peter Simmonds and Yong Poovorawan

1Center of Excellence in Clinical Virology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
2Department of Virology, University of Turku, 20520 Turku, Finland
3Department of Clinical Virology, Turku University Hospital, 20520 Turku, Finland
4Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, 17177 Stockholm, Sweden
5Department of Clinical Microbiology, Karolinska University Hospital, 17177 Stockholm, Sweden
6Department of Microbiological Diagnostics & Virology, Artillerivej 5, DK-2300 Copenhagen, Denmark
7Institute for Virology, University of Bonn Medical Center, Sigmund-Freud-Str. 25, D-53105 Bonn, Germany
8Instituto de Salud Carlos III, 28220 Majadahonda, Madrid, Spain
9Infection and Immunity Division, Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK
10Nuffield Department of Medicine, University of Oxford, Oxford OX1 3SY, UK

Coxsackievirus A6 (CV-A6) is a major aetiologic agent for hand, foot and mouth disease (HFMD) in recent years. HFMD outbreaks associated with CV-A6 resulted from the evolutionary dynamics of CV-A6 and the appearance of novel recombinant forms (RFs). To examine this, 151 variants collected in 2013 and 2014 from Germany, Spain, Sweden, Denmark and Thailand were genotyped for the VP1 capsid and 3Dpol genes. Analysis of the VP1 gene showed an increasing correspondence between CV-A6 genome recombination and sequence divergence (estimated substitution rate of $8.1 \times 10^{-3}$ substitutions site$^{-1}$ year$^{-1}$ and RF half-life of 3.1 years). Bayesian phylogenetic analysis showed that recent recombination groups (RF-E, -F, -H, -J and -K) shared a common ancestor (RF-A). Thirty-nine full-length genomes of different RFs revealed recombination breakpoints between the 2A–2C and the 5' UTRs. The emergence of new CV-A6 recombination groups has become widespread in Europe and Asia within the last 8 years.

Human enteroviruses are genetically diverse RNA viruses within the genus *Enterovirus* genus in the family *Picornaviridae* and are responsible for a wide spectrum of clinical manifestations (Knowles *et al.*, 2012). Human enteroviruses are divided into four species (A–D). Species A, of which coxsackievirus A6 (CV-A6) is a member, currently comprises 20 types (Pons-Salort *et al.*, 2015). CV-A6 possesses a positive-stranded RNA genome of approximately 7400 nucleotides encapsidated by a highly structured icosahedral capsid. The viral genome is translated into a large polyprotein that is subsequently cleaved into structural (VP1–VP4) and non-structural (NS) (2A–2C and 3A–3D) proteins (Whitton *et al.*, 2005). The degree of similarity of nucleotides and amino acid sequences of the VP1 region provides the primary tool for the identification and assignment of new types.
within a species, in which novel variants showing less than 75% nucleotide sequence identity are classified as new types (Oberste et al., 1999).

CV-A6 infections are typically mild and asymptomatic, whereas enterovirus A71 and coxsackievirus A6 are most often implicated in causing hand, foot and mouth disease (HFMD), a disease characterized by vesicular exanthema on the hands, feet, and oral mucosa (Puenpa et al., 2011; Schuffenecker et al., 2011; Wu et al., 2010). In 2014, atypical HFMD was linked to CV-A6 infection in children with erythematous papular rash resembling eczema herpeticum (Sinclair et al., 2014) as a result of newly emerging variants of CV-A6. Such novel recombinant forms (RFs) of the virus have been assigned into RF-A–H based on the 3D polymerase (3Dpol) phylogeny (Gaunt et al., 2015; McWilliam Leitch et al., 2012). It was subsequently determined that RF-H, which possessed phylogenetically distinct 3Dpol region sequences likely acquired from other human enterovirus species A serotypes through recombination, was largely responsible for the clinically unusual HFMD in Edinburgh, UK (Gaunt et al., 2015). Here, we further defined how recently emerged RFs including RF-H and its predecessor RF-A are circulating more widely and contributing to the increased incidence of HFMD elsewhere around the world.

We initially examined the VP1 sequence divergence and 3Dpol sequence grouping by analysing 151 CV-A6 strains from Denmark (n=22), Germany (n=4), Spain (n=14), Sweden (n=6) and Thailand (n=105) collected in 2013 and 2014 (Table S1, available in the online Supplementary Material). Nested reverse transcription PCR (RT-PCR) was performed using newly designed primers to amplify the VP1 and 3Dpol genes (Table S2), followed by sequencing. The reverse transcription and first-round PCR utilized the Superscript III One-Step RT-PCR system with Platinum Taq High Fidelity (Invitrogen), while the second-round PCR utilized GoTaq DNA Polymerase (Promega). Sequences were analysed using the SSE 1.2 sequence editor package (www.virus-evolution.org) (Simmonds, 2012) and phylogenetic tree reconstruction was performed with the MEGA program (v6) using the best-fit models and the maximum-likelihood method (Tamura et al., 2013). All newly generated sequences were deposited in the GenBank database under the accession numbers KX212338–KX212678.

The VP1 sequences from the CV-A6 identified in Denmark and Spain clustered within lineage I, the very same group of CV-A6 responsible for eczema herpeticum in Edinburgh in 2014 (Fig. S1). They were distinct from CV-A6 from Taiwan (lineage II), Thailand (lineage III), China (lineage IV) and Finland (lineage V). Examination of the sequences from the 3Dpol region enabled the designation of bootstrap-supported clades comprising groups A, B, C, D, E, F, G, H, I, J and K (Fig. 1, Table S3) in agreement with previous analysis (McWilliam Leitch et al., 2009, 2010, 2012). The majority of the CV-A6 strains clustered within two of the previously assigned recombinant forms RF-A (105/151) and RF-F (37/151). Meanwhile, four variants from Denmark grouped with RF-H, as did two variants from Spain. The additional three variants from Denmark and Spain comprised RF-G.

Strong evidence of recombination was demonstrated by the observed differences in the phylogenetic trees based on the 3Dpol region, for which RF groups were assigned, and VP1 region for several strains. For example, the TW/00141/E/2007 strain was assigned to RF-E by the 3Dpol phylogeny, but grouped with the two RF-B strains (FI/TS3FinTu81042/B/2008 and JP/Kyoto1/B/1999) in the VP1 tree. The 3Dpol-assigned RF-J strain CN/P143/J/2013 grouped with some of the RF-A strains. Finally, the CN/CC13/K/2013 is an RF-K strain, but instead appeared next to an RF-D strain CN/HN421/D/2011 in the VP1 phylogeny.

To enable examination of the sequence relationships in other parts of the genome, we obtained 39 nearly complete genome sequences of CV-A6 variants (Germany=3, Spain=3, Denmark=8 and Thailand=25) representing RF-A (n=21), RF-F (n=10), RF-G (n=2) and RF-H (n=6). Comparison of the phylogenetic trees of the VP1, 5’ UTR and VP4/2 regions yielded broadly similar groupings of CV-A6 variants (Figs 1 and S2). However, several instances of discordance in the trees provided additional evidence for potential CV-A6 recombination. For example, RF-E variants grouped with RF-A and -H in the VP1, but not in the 5’ UTR and VP4/2 regions. RF-A strains were interspersed in the 5’ UTR and VP4/2 trees with RF-J, which contrasts with their consistent grouping in the VP1 region. All of RF-K was monophyletic in the VP1, VP4/2 and the 5’ UTR regions except one which grouped with RF-D. Therefore, the patterns of phylogenetic discordance were consistent with recombination in CV-A6 and in agreement with previous findings for other enteroviruses (Cabrerizo et al., 2014; Calvert et al., 2010; McIntyre et al., 2010; McWilliam Leitch et al., 2009, 2010, 2012).

We estimated the rates of evolution and molecular clock phylogeny (Drummond & Rambaut, 2007) from VP1 gene sequences using the Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST (v1.8.0) (Drummond et al., 2012). Two independent runs involved constant and exponential growth as priors with a chain length of 100 million and a relaxed log-normal molecular clock model were analysed using the SRRD06 model of substitution. All other parameters were optimized during the burn-in period. Convergence of the chains and effective sample sizes of the estimates were checked using Tracer (http://beast.bio.ed.ac.uk/Tracer).

Since sequence divergence in VP1 provides a proxy measure for the time of divergence of CV-A6 variants from which estimates of the life spans of individual RFs can be derived as described previously for echovirus type 30 isolates (McWilliam Leitch et al., 2009), we evaluated the VP1 evolutionary divergence and the proportion of recombinant comparisons for variants with different 3Dpol groups. Pairwise comparison among CV-A6 variants showed precise correlation between VP1 sequence distances and
Fig. 1. Phylogenetic analysis of CV-A6 variants in different genome regions. Maximum-likelihood trees of (a) VP1 (positions 2485 and 3816 numbered based on the Gdula prototype strain, GenBank accession number AY421764), (b) 3Dpol (positions 5061 and 6364 numbered based on the Gdula prototype strain, GenBank accession number AY421764) and previously determined reference groups (RF-B–E). The optimal substitution models were Kimura two-parameter K2P with invariant sites (I) for VP1 and Kimura two-parameter with I and gamma distribution (I–G) for 3Dpol. Each sequence is identified by the country of origin, sample code, 3Dpol clade assignment and year of collection. Dot colours indicate the recombination group assignments based on 3Dpol phylogeny. Bars denote the evolutionary distance according to the number of nucleotide substitutions per site. Bootstrap consensus was inferred from 1000 replicates.
assignment to different RF group by the 3Dpol region (Fig. S3). Furthermore, an estimate of the approximate RF half-lives of CV-A6 lineages was calculated by combining the mean sequence divergence in VP1 at the 50% recombination frequency threshold (estimated at 0.05) with the substitution rate in VP1 \((8.1 \times 10^{-3}\) substitutions site\(^{-1}\) year\(^{-1}\)) (Table 1). This corresponds to a period of 6.17 years \((0.05/0.0081)\) of divergent evolution, or approximately 3.1 years from a common ancestor. This RF half-life of CV-A6 was very similar to that estimated previously for echovirus type 30, although higher than E6 and E9 but lower than EV71 and E11 (Cabrerizo et al., 2014; McWilliam Leitch et al., 2009, 2010, 2012).

To determine when CV-A6 recombination groups first appeared, we reconstructed the temporal phylogeny from VP1 sequences. Since variability of the VP1 sequence was restricted primarily to synonymous sites, most sequence changes likely occurred through neutral drift. Using molecular clock analysis, we estimated the nucleotide substitution rate and times to the most recent common ancestor (tMRCA) of different regions and the assignment of 3Dpol clades of the greatest RFs (RF-A and -F). The substitution rate of the whole data set of all VP1 sequences was estimated to be \(8.1 \times 10^{-3}\) substitutions site\(^{-1}\) year\(^{-1}\) [high-probability distribution (HPD) range, \(6.0 \times 10^{-3}–10.5 \times 10^{-3}\)]. The MRCA of all CV-A6 clusters likely first appeared in 1947 (HPD, 1940–1949), while that of RF-A probably emerged in 1999 (HPD, 1995–2003). The diversity was low for RF-F variants as demonstrated by their tight clustering at the top of the tree (Fig. 1, Table 1), whereas RF-A variants dispersed into other branches and therefore implied that they possessed higher substitution rate. The number of RF-F variants was only one-third that of RF-A and therefore might have contributed to the bias due to sampling size.

To determine more precisely the timescale of recombination events underlying the appearance of each RF, datasets of VP1 gene sequences were further analysed using the Bayesian MCMC method to generate time-correlated phylogeny (Fig. 2). While earlier recombination events could not be reconstructed in any detail due to inadequate sampling of CV-A6 before 2008, variants collected after this date were monophyletic and fell into three further lineages with estimated dates of splitting between 2004 and 2005. The oldest lineage was composed purely of RF-G samples, which first appeared in 2011, and thereafter underwent the recombination event that exchanged the 3Dpol region sequences between 2004 and 2011. The other two lineages contained samples belonging to RF-A. One was composed solely of RF-A and persisted for at least 11 years (2005–2015). The other lineage contained RF-A sequences and samples isolated subsequently, which belonged to other RF groups (E, F, H, J and K). In lineage 1, the oldest variants were those originally described in Asia (Thailand and Japan) before 2012 and the lineage variants then spread into Europe and Asia between 2013 and 2014. Within this lineage, CV-A6 variants belonged to RF-F with the

### Table 1. Rates of sequence change and tMRCA\(_S\) by MCMC analysis

<table>
<thead>
<tr>
<th>RF group and geographic set</th>
<th>n(^*)</th>
<th>Divergence(^†)</th>
<th>Nucleotide</th>
<th>Amino acid</th>
<th>Substitution rate (10(^{-3}) substitutions site(^{-1}) year(^{-1}))</th>
<th>tMRCA(^§)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole data set</td>
<td>244</td>
<td>0.052</td>
<td>ND</td>
<td>ND</td>
<td>8.1 (6.6–10.5)</td>
<td>ND</td>
</tr>
<tr>
<td>Individual RF group</td>
<td>160</td>
<td>0.047</td>
<td>0.0012</td>
<td>0.020</td>
<td>10.4 (7.6–13.8)</td>
<td>11.0 (8.2–14.4)</td>
</tr>
<tr>
<td>RF-A</td>
<td>140</td>
<td>0.046</td>
<td>0.0012</td>
<td>0.020</td>
<td>10.4 (7.6–13.8)</td>
<td>11.0 (8.2–14.4)</td>
</tr>
<tr>
<td>RF-F</td>
<td>36</td>
<td>0.016</td>
<td>0.0015</td>
<td>0.006</td>
<td>1.1 (0.003–3.0)</td>
<td>1.4 (0.003–3.0)</td>
</tr>
</tbody>
</table>

\(^*\) Number of sequences in each set analysed.

\(^†\) Mean value with the HPD interval in parentheses.

\(^§\) Time to the most recent common ancestor.

\(ND\), Not determined.

To determine more precisely the timescale of recombination events underlying the appearance of each RF, datasets of VP1 gene sequences were further analysed using the Bayesian MCMC method to generate time-correlated phylogeny (Fig. 2). While earlier recombination events could not be reconstructed in any detail due to inadequate sampling of CV-A6 before 2008, variants collected after this date were monophyletic and fell into three further lineages with estimated dates of splitting between 2004 and 2005. The oldest lineage was composed purely of RF-G samples, which first appeared in 2011, and thereafter underwent the recombination event that exchanged the 3Dpol region sequences between 2004 and 2011. The other two lineages contained samples belonging to RF-A. One was composed solely of RF-A and persisted for at least 11 years (2005–2015). The other lineage contained RF-A sequences and samples isolated subsequently, which belonged to other RF groups (E, F, H, J and K). In lineage 1, the oldest variants were those originally described in Asia (Thailand and Japan) before 2012 and the lineage variants then spread into Europe and Asia between 2013 and 2014. Within this lineage, CV-A6 variants belonged to RF-F with the
Fig. 2. Temporal phylogeny of VP1 sequences of CV-A6 variants in this study and published sequences. Branch colours denote recombination groups in each clade. Two RF-A lineages (1 and 2) are noted.
recombination event and dated between 2009 and 2012. The most recent group, RF-H, which appeared in 2013, probably recombined between 2011 and 2013. The RFs (F, G, H and K) were monophyletic and likely originated from single recombination events, unlike RF-E and -J which were detected in more than one VP1 lineage. With the exception of RF-F, which was isolated from both Europe and Asia, other RF groups were detected only from Europe or Asia: G and H (Europe), E (Taiwan) and J and K (China). Variants within lineage 2 (RF-A) were those detected in the first HFMD outbreak in Finland in 2008 along with variants detected subsequently in Europe and Asia over the following 1–5 years.

Having identified the likely time course and direction of the recombination events, divergence scan analyses were performed between the ancestral RF-A sequences with complete genome sequences generated in the current study (RF-F–H) and the RFs from previous study (RF-J and -K) to identify recombination breakpoints (Fig. S4). The sharp increase in sequence divergence at various points in the P2 region provided evidence for the occurrence of separate, individual recombination events for each RF. The first breakpoint was found at the 2A protein-encoding region around nucleotide position 3500 (RF-G). The next breakpoints were located in 2B and at the border between 2B and 2C regions (RF-F, -J and -H). The last breakpoint of RF-K can be recognized at nucleotide position 5000 and included the 3′ part of the 2C region.

This study reports a detailed, multi-centre investigation of the emergence of an enterovirus serotype associated with epidemics of HFMD. It catalogues the complexity of the evolutionary processes associated with its geographical expansion and the occurrence of a number of recombination events each involving replacement of close to complete NS gene blocks at varying times since the founder recombinant form RF-A was first described in 2008 (Osterback et al., 2009). NS region sequences of most RFs have not been described in association with other species A serotypes including any of the RFs described for enterovirus A71 (McWilliam Leitch et al., 2012). However, there was some evidence for a limited degree of re-circulation within the recombination pool of NS region sequences; RF-E and -K appeared at two different positions in the VP1 phylogenetic tree. Furthermore, a limited degree of sharing of NS region sequences between different species B serotypes has been documented (Bailly et al., 2009).

Phylogenetic reconstruction was used to analyse trait evolution such as temporal and geographical correlates of individual recombination events, which was previously incomplete. The most frequently detected recombinant form, RF-A, showed decades-long circulation and was the ancestor of five separate recombination groups (RF-E, -F, -H, -K and -J) that have emerged in the past 5–10 years. The HFMD outbreak in Finland in 2008 was associated with RF-A with the subsequent appearance of this RF across Europe and Asia between 2013 and 2014. The more recent emergence of RFs originated from descendants of multiple VP1 lineages that have diverged from RF-A variants circulating in Asia (Thailand and Japan) between 2008 and 2010. Enterovirus recombination events can play a significant role in the evolution, and breakpoints detected in this study (2A–2C regions) are well-known recombination hotspots (Lukashev et al., 2005). CV-A6 recombination breakpoints within VP3 and between 5′ UTR and VP1 have been detected in the genomes of RF-E variants in Taiwan (Gaunt et al., 2015).

In summary, the typical pattern for an RF was its rapid emergence, variable penetrance into the sampled virus population and relatively rapid extinction, within years rather than decades, based on the average recombination half-lives documented for CV-A6 and other EV types. These patterns are well-attested in the turnover of RFs of CV-A6. While we remain relatively ignorant of the reasons for RF turnover, whether driven by immunological, host adaptive factors or transmissibility, or alternatively whether it occurs as a consequence of population bottlenecks and replacements without a fitness component, molecular epidemiological studies will be of value in understanding the nature of enterovirus evolution and the clinical outcomes.

**Acknowledgements**

This study was funded by the National Research Council of Thailand, the Research Chair Grant from the National Science and Technology Development Agency, the Center of Excellence in Clinical Virology of Chulalongkorn University (GCE 58-014-30-004), King Chulalongkorn Memorial Hospital and the RGJ PhD program (PHD/0087/2554). This study was also supported by the Wellcome Trust (WT103767MA) and the Biotechnology and Biological Sciences Research Council (BB/J004324/1).

**References**


