A Proposal for the Classification and Nomenclature of Hand, Foot and Mouth Disease-Related Enteroviruses

Yajuan Dong
Henan Normal University
Zhenzhou Wan
Medical Laboratory of Taizhou Fourth People's Hospital
Shenwei Li
Shanghai International Travel Healthcare Center
Jian-Hua Wang
Guangzhou Institutes of Biomedicine and Health
Xia Jin
Shanghai Public Health Clinical Center
Guoying Yu
Henan Normal University
Chiyu Zhang (zhangcy1999@ips.ac.cn)
Institut Pasteur of Shanghai Chinese Academy of Sciences

Research Article

Keywords: Enterovirus, hand, foot and mouth disease, classification, genetic distance, recombination, EVA71, CVA16, CVA6, CVA10

Posted Date: June 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34737/v1

Abstract

Background: Enterovirus has diverged into many types, some of which cause hand, foot and mouth disease (HFMD) in children. The predominant enterovirus types associated with HFMD are EVA71, CVA16, CVA6 and CVA10. Subtyping of these enteroviruses is crucial to HFMD surveillance. Because of lacking proper and uniform criteria and being based on partial VP1 sequences, however, current classification resulted in some confusing and conflicting results.

Method: We reclassified EVA71, CVA16, CVA6 and CVA10 using a combined criteria of phylogenic relationship and genetic distance.

Results: Using the combined criteria, we classified EVA71 into seven genotypes of A–G, CVA16 and CVA6 into three subtypes of A-C, and CVA10 into nine subtypes/sub-subtypes of A-G, H1 and H2, and identified eight unclassified subtypes that lack genomic sequences. The mean genetic divergence was 15.5-33.8% between subtypes, 12-15% between sub-subtypes, and less than 12% within subtypes/sub-subtypes. In addition, we identified two new EVA71 inter-subtype recombinants RF01_CG and RF02_CG and demonstrated that EVA71 subtypes D and F and CVA10 subtype B experienced inter-subtype recombination events during early evolution.

Conclusions: The new nomenclature proposal provides a reasonable framework for proper classification of enteroviruses, which will be useful for epidemiological surveillance of HFMD, disease management, and vaccine development.

Background

Hand, foot and mouth disease (HFMD) is a common contagious disease of childhood. It is caused by infection with various non-polio and non-rhinovirus enteroviruses and characterized by fever and skin eruptions on the hands and feet, and vesicles in the mouth (1, 2). HFMD has been a major public health burden across the Asia-Pacific region with an estimated cases of over 1.38 million per year worldwide (2).

Enteroviruses belong to the *Picornaviridae*, a family of small, non-enveloped viruses with a positive-stranded RNA genome of approximately 7.4 kilobase in size (3), and it is highly divergent and hierarchically classified into 15 species including enterovirus A-L and rhinovirus A-C based on sequence identity and genome organization (3–5). Some enterovirus species are further classified into a large number of genotypes based on a genetic distance of over 25% at nucleotide level (4, 6). Distinct enterovirus genotypes often exhibit various biological properties, such as infectivity, transmissibility and pathogenesis, and they are involved in different diseases (1, 5, 7, 8). HFMD is mainly attributed to some genotypes of enteroviruses A and B (1, 2, 5).

Enterovirus A71 (EVA71) and Coxsackievirus A16 (CVA16) are the two most commonly detected enteroviruses among HFMD cases. However, the molecular epidemiology of HFMD-related enteroviruses are changing during the past decade, with a progressive increase of CVA6, CVA10 and other enterovirus types (8). Because of extraordinary degree of genetic diversity of circulating strains worldwide, dominant enterovirus types were often further divided into subtypes simply based on their phylogenetic relationships in partial VP1 fragment (9–18). Because of lack of distance-based criterion, current subtypes of enteroviruses appeared to be inconsistent and contradictory among researchers. In addition, the lack of the phylogenetic criterion of full-length genomic sequence might also result in misclassification of some reported subtypes just based on partial VP1 fragment. These problems complicate and hamper the investigations of epidemiology and pathogenicity of HFMD-related enteroviruses. The increasing complexities of HFMD-related enteroviruses raise an urgent need to re-evaluate the current nomenclature system (4, 19, 20). In this study, we performed systematical re-classifications of four dominant HFMD-related enteroviruses, EVA71, CVA16, CVA6 and CVA10, and proposed consensus subtyping criterion for enteroviruses.

Materials And Methods

Sequence collection

All available full-length genomic sequences of EVA71, CVA16, CVA10 and CVA6 were downloaded from the GenBank on November, 2019. According to the prototype strains of EVA71 (BrCr: U22521), CVA16 (G-10: U05876), CVA6 (Gdula: AY421764), and CVA10 (Kowalik: AY421767) in ICTV (The international Committee on Taxonomy of Viruses), the sequences with a length of shorter than 6800 nt were deleted. The full-length genomic sequences of more than 6800 nt and their complete VP1 sequences were subjected to the phylogenetic analyses. All available near-complete VP1 sequences of the four enteroviruses in GenBank were also downloaded on November, 2019.

Phylogenetic analyses

All sequence alignments were performed using MUSCLE implemented in MEGA-X. To classify four predominant HFMD-related enteroviruses EVA71, CVA16, CVA6, and CVA10, maximum likelihood (ML) trees were constructed based on the full-length genomic sequences using MEGA-X with 1000 bootstrap replications. The ML trees were also constructed using the complete VP1 sequences from genomic sequences, together with addition near-complete VP1 sequences from GenBank, which can form independent clades and represent potential new subtypes. The parameters used for the ML tree construction were General Time Reversible model (GTR) with Gamma Distributed With Invarient Sites (G + I), and partial deletion of sequence gaps. To minimize the calculation time in ML tree construction, only one representative sequence was included if there were two or more sequences sharing sequence identity of more than97% for EVA71, 98% for CVA6, and 99% for both CVA10 and CVA16. The number of sequences with high identity was shown after the name of each representative strain.

Recombination analysis

To detect potential recombination occurring in enteroviruses, bootscanning and similarity plot analyses were performed using SimPlot v.3.5.1 (21). The Kimura two-parameter substitution model with a transition/transversion ratio of 2 was selected in the analysis.

Results

In principle, a combination of phylogeny- and distance-based criteria in the analysis of genomic sequences is required for genotyping a virus (19, 20, 22). To define a subtype of enterovirus, we proposed a distance criterion of over 15% between subtypes at the genomic level and/or at least at the near-complete VP1 gene level (when genomic sequences are not available) together with a well-supported clade (over 75% bootstrap value support). If two or more well-supported clades are formed within the same subtypes, they can be further divided into sub-subtypes based on a cut-off of over 12% genetic distance. To reclassify the subtypes of four dominant HFMD-related enteroviruses EVA71, CAV16, CAV6 and CVA10, we preformed phylogenetic analyses of both full-length genomic sequences and near-complete VP1 gene sequences.

Phylogenetic classification of EVA71

All 922 near full-length genomic sequences of EVA71 available in GenBank were downloaded (on November, 2019). Of them, 913 with a length of more than 7000 nt were subject to sequence alignment. After removing highly homologous sequences with more than 97% sequence identity, 131 representative genomic sequences were used to construct maximum likelihood (ML) tree. The ML tree of the genomic sequences showed that the vast majority of EVA71 strains were clustered within three well-supported large clades (with 100% bootstrap value), and the others formed small clusters or independent phylogenetic branches (Fig. 1A). Together with the genetic distance data (Table 1), EVA71 was classified into seven subtypes, and designated as genotypes A to G (Fig. 1A), in keeping with previous nomenclature order of EVA71 (Fig. 1A). The three large clades correspond to subtypes B, C and G. Among the seven subtypes, subtypes E had only one available full-length genomic sequence. Interestingly, one strain (DL71/CHN/2012) that was clustered between subtypes C and B, was not defined as an independent subtype because it had about 13.4% genetic distances with subtypes C, and was highly suspected to be an inter-subtype recombinant.

Table 1
Mean genetic distance and standard error among genotypes of EVA71.

Subtype	A	B	C	D	E	F	G	Unclassified 1	Unclassified 2	Unclassified 3	Unclas 4
A	NA/0.01 ± 0.00	0.215 ± 0.014	0.207 ± 0.013	NA	0.234 ± 0.014	NA	0.208 ± 0.012	0.326 ± 0.018	0.218 ± 0.013	0.231 ± 0.014	0.222 0.012
B	0.235 ± 0.006	0.09 ± 0.00/0.09 ± 0.00	0.185 ± 0.010	NA	0.221 ± 0.011	NA	0.206 ± 0.010	0.317 ± 0.016	0.194 ± 0.011	0.207 ± 0.012	0.215 0.011
C	0.239 ± 0.006	0.215 ± 0.005	0.07 ± 0.00/0.06 ± 0.00	NA	0.197 ± 0.011	NA	0.135 ± 0.007	0.311 ± 0.017	0.200 ± 0.011	0.197 ± 0.012	0.204 0.011
D	0.226 ± 0.006	0.220 ± 0.004	0.203 ± 0.004	0.06 ± 0.00/NA	NA	NA	NA	NA	NA	NA	NA
E	0.225 ± 0.006	0.222 ± 0.005	0.238 ± 0.005	0.191 ± 0.005	NA/0.09 ± 0.01	NA	0.200 ± 0.010	0.320 ± 0.017	0.208 ± 0.011	0.195 ± 0.011	0.188 0.010
F	0.224 ± 0.007	0.226 ± 0.005	0.194 ± 0.005	0.192 ± 0.005	0.213 ± 0.005	NA/NA	NA	NA	NA	NA	NA
G	0.242 ± 0.005	0.235 ± 0.005	0.205 ± 0.004	0.180 ± 0.004	0.235 ± 0.005	0.197 ± 0.004	0.11 ± 0.00/0.10 ± 0.01	0.303 ± 0.016	0.202 ± 0.011	0.192 ± 0.010	0.208 0.010
Unclassified 1	NA	NA	NA	NA	NA	NA	NA	NA/NA	0.332 ± 0.019	0.324 ± 0.017	0.324 0.017
Unclassified 2	NA	NA	NA	NA	NA	NA	NA	NA	NA/0.09 ± 0.01	0.200 ± 0.011	0.205 0.011
Unclassified 3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA/0.11 ± 0.01	0.172 0.009
Unclassified 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA/0.1 0.01

The data obtained from full-length and near-complete VP1 sequences are shown in lower left and top right quarters, respectively. NA: not available. It means t there are only one, no sequence, or more than two completely identical sequences.

Whether complete VP1 sequences generated consistent phylogeny of EVA71 with the full-length genomic sequences and met the distance criterion of subtyping is critical to determine the accuracy of VP1 sequence-based classification, albeit it was previously widely used. To assess the performance of VP1 sequence-based classification, complete VP1 sequences from 131 full-length genomic sequences were subjected to further phylogenetic analysis, together with 31 additional near-complete VP1 sequences retrieved from GenBank. According to phylogeny-based and genetic distance-based criteria, five EVA71 subtypes A-D and G could also be classified based on complete VP1 sequences (Fig. 1B and Table 1). Apart from five identified subtypes, three well-supported

small clusters, as well as a single branch at the root of the tree were observed in the VP1 tree (Fig. 1B), and met the genetic distance criterion of being new subtypes (Table 1). Because of the lack of available full-length genomic sequences, they were defined as unclassified subtypes 1–4. Their nomenclatures need to be determined in the future.

It is worth noting that the strains of subtypes D and F did not form independent subtype clades in the VP1 tree, but were clustered within the clade of subtype G (Fig. 1B). The inconsistency of phylogeny between full-length genomic and near-complete VP1 sequences suggests that recombination event might have occurred during the evolution of EVA71 subtypes D and F. Bootscan analyses confirmed that subtype D was involved in recombination between subtypes G and C, and subtype F was involved in recombination among subtypes C, D and G (supplementary Fig. S1A and B). In particular, the proportion of permuted trees appeared to be relative low (less than 50%) in the majority of the genomic regions of both subtypes D and F, implying that the recombination events occurred in the distant past. The suspected recombinant DL71/CHN/2012 was found to belong to subtype G in the VP1 region. Bootscan analyses illuminated that DL71/CHN/2012 originated recombination between subtypes G and C, and had a mosaic genome structure of G-C-G-C-G-C-G (Fig. 2A). The recombination pattern was further confirmed by separate phylogenetic analyses (Supplementary Fig. S2). Furthermore, a subtype G strain VR1432/China/2009 was found to be clustered between subtypes C and G in the VP1 tree, suggesting the presence of recombination between subtypes C and G at least in VP1 region (Fig. 1). Bootscan and separate phylogenetic analyses confirmed that VR1432/China/2009 was a recombinant between subtypes C and G with a mosaic genome structure of C-G-C-G-C-G (Fig. 2B and Supplementary Fig. S3). The recombinant strains DL71/CHN/2012 and VR1432/China/2009 were defined as EVA71 recombinant form RF01_CG and RF02_CG, respectively.

On the basis of full-length genomic and/or near-complete VP1 sequences, EVA71 was classified into seven subtypes from A to G, as well as two recombinant forms RF01_CG and RF02_CG. Furthermore, four unclassified subtypes 1–4 were identified based on near-complete VP1 sequences. The mean inter-subtype genetic distances ranged from 18–24.2% at the genome level and 13.5–33.2% at VP1 gene level, and the mean within-subtype distance ranged from 6–11% at the genome level and 1–11% at VP1 gene level (Table 1).

Phylogenetic Classification Of Cva16

All 142 available near full-length genomic sequences of CVA16 were retrieved from GenBank (on November, 2019). Of them, 138 had a genomic sequence of more than 7000 nt, and were used in sequence alignment. After the removal of 36 identical sequences, 102 representative genomic sequences were used in the phylogenetic analysis. Based on the phylogenetic relationship and genetic distance criterion (Fig. 3A and Table 2), CAV16 was classified into A, B, and C subtypes. Although subtype C was defined, it had only one available genomic sequence each, and was called as single genomic sequence subtype. To find more support for the classification of single genomic sequence subtype, and to identify more potential subtypes, a phylogenetic analysis using 102 VP1 sequences of the representative CVA16 strains and 19 additional VP1 sequences was performed. The additional VP1 sequences were selected if they closely clustered with any single genomic sequence subtype to form a well-supported clade, or they were unable to cluster with other cluster within any defined subtypes (A-C) in a preliminary phylogenetic analysis with all available near-complete VP1 sequences (data not shown). In the VP1 tree, six additional VP1 sequences were found to form a well-supported clade together with the representative strain of subtype C with a 100% bootstrap value (Fig. 3B). The subtype C clade completely met the genetic distance criterion of subtyping (Table 2), supporting the classification of subtype C. Furthermore, another five additional sequences formed an independent clade with 87% bootstrap value between the clades of subtypes B and C (Fig. 3B). The genetic distance analysis showed that the new clade had a mean distance of 12.2% to the subtype B strains (Table 2). Because of the lack of genomic sequences, the new clade was marked as unclassified (Fig. 3). As a result, three subtypes A, B, and C were identified for CVA16. In addition, a potential subtype was identified based on the near-complete VP1 analysis. The subtype B had mean genetic distances of 29.4% and 16.3% to subtypes A and C, and the unclassified subtype had mean genetic distances of 29.8%, 0.122 and 16.4% to subtypes A, B and C, respectively (Table 2). According to the classification of CVA16, the vast majority of the circulating strains belonged to subtype B.

Table 2
Mean genetic distance and standard error among genotypes of CVA16.

Subtype	A	B	C	Unclassified
A	NA/NA	0.294 ± 0.015	0.313 ± 0.018	0.298 ± 0.016
B	0.257 ± 0.006	0.08 ± 0.00/0.07 ± 0.00	0.163 ± 0.010	0.122 ± 0.008
C	0.266 ± 0.005	0.208 ± 0.004	NA/0.05 ± 0.00	0.164 ± 0.011
Unclassified	NA	NA	NA	NA/0.07 ± 0.01

The data obtained from full-length and near-complete VP1 sequences are shown in lower left and top right quarters, respectively. NA: not available. It means that there are only one, no sequence, or more than two completely identical sequences.

Phylogenetic Classification Of Cva6

A total of 230 near full-length genomic sequences of CVA6 were available in GenBank (on November, 2019), and 229 of them had a genomic sequence of more than 7000 nt. After the removal of 160 identical sequences, 69 full-length genomic sequences were used in the phylogenetic analysis. Simultaneously, the VP1 sequences of the 69 representative strains were also subjected to the phylogenetic analysis together with 17 additional VP1 sequences that might support the subtype with single genomic sequence or form potential new subtypes. Based on the phylogenetic relationship and genetic distance criterion (Fig. 4A and Supplementary Table 3), CAV6 was classified into three subtypes (A, B and C) at genomic sequence level. Of note, two independent small clusters formed by 5 strains with more than 99% bootstrap value support, and an independent branch by a strain (NIV43883/India/2004) were observed in the VP1 tree

(Fig. 4B). These new clusters and branch were dispatched to unclassified subtypes 1, 2 and 3 (Fig. 4B and Table 3). Therefore, CVA6 was classified into three subtypes A-C. The mean inter-subtype genetic distances were all over 15% regardless of whether being analyzed at genomic or VP1 sequence levels (Table 3).

Table 3
Mean genetic distance and standard error among genotypes of CVA6.

Subtype	A	B	C	Unclassified 1	Unclassified 2	Unclassified 3
A	NA/0.00 ± 0.00	0.183 ± 0.015	0.206 ± 0.015	0.319 ± 0.020	0.185 ± 0.014	0.197 ± 0.015
B	0.228 ± 0.007	NA/NA	0.186 ± 0.014	0.307 ± 0.020	0.193 ± 0.014	0.167 ± 0.013
C	0.232 ± 0.006	0.185 ± 0.005	0.09 ± 0.00/0.07 ± 0.00	0.324 ± 0.019	0.191 ± 0.013	0.177 ± 0.012
Unclassified 1	NA	NA	NA	NA/NA	0.322 ± 0.019	0.338 ± 0.020
Unclassified 2	NA	NA	NA	NA	NA/0.07 ± 0.01	0.195 ± 0.014
Unclassified 3	NA	NA	NA	NA	NA	NA/0.10 ± 0.01

The data obtained from full-length and near-complete VP1 sequences are shown in lower left and top right quarters, respectively. NA: not available. It means that there are only one, no sequence, or more than two completely identical sequences.

Phylogenetic Classification Of Cva10

A total of 127 near full-length genomic sequences of CAV10 were downloaded from GenBank (on November, 2019), and 124 that had a genomic sequence of more than 6800 nt were used. After the removal of 22 identical sequences, 102 representative genomic sequences were used in the phylogenetic analysis. Nine subtypes/sub-subtypes (A-G and H1 and H2) were identified for CVA10 based on the phylogenetic and genetic distance analyses of genomic sequences (Fig. 5A and Table 4). There was only one representative genomic sequence available for subtypes A-B, and D-G. To confirm the classification of CVA10, phylogenetic analyses of VP1 sequences were performed using those from the representative strains and 19 additional VP1 sequences retrieved from GenBank. The classification of CVA10 was well supported by the VP1 analyses regardless of using either phylogeny or genetic distances (Fig. 5B and Table 4). In particular, additional representative VP1 sequences were found to support the definitions of single genomic sequence subtypes A, B, G and F. According to the classification of CVA10, the vast majority of the circulating strains belonged to subtypes C and H (H1 and H2).

Table 4
Mean genetic distance and standard error among genotypes of CVA10.

Subtype	A	B	C	D	E	F	G	H1	H2
A	NA/0.00 ± 0.00	0.292 ± 0.021	0.283 ± 0.020	0.269 ± 0.020	0.307 ± 0.023	0.289 ± 0.020	0.270 ± 0.019	0.299 ± 0.021	0.272 ± 0.019
B	0.240 ± 0.007	NA/0.05 ± 0.00	0.235 ± 0.018	0.236 ± 0.019	0.242 ± 0.019	0.230 ± 0.017	0.236 ± 0.017	0.212 ± 0.015	0.216 ± 0.015
C	0.250 ± 0.006	0.232 ± 0.007	0.04 ± 0.00/0.04 ± 0.00	0.142 ± 0.013	0.176 ± 0.014	0.206 ± 0.015	0.196 ± 0.014	0.206 ± 0.016	0.206 ± 0.014
D	0.243 ± 0.005	0.227 ± 0.007	0.170 ± 0.005	NA/NA	0.164 ± 0.015	0.186 ± 0.015	0.179 ± 0.015	0.203 ± 0.016	0.212 ± 0.016
E	0.249 ± 0.007	0.228 ± 0.006	0.177 ± 0.006	0.158 ± 0.006	NA/NA	0.193 ± 0.015	0.196 ± 0.016	0.210 ± 0.017	0.217 ± 0.016
F	0.270 ± 0.006	0.244 ± 0.006	0.226 ± 0.007	0.227 ± 0.006	0.232 ± 0.007	NA/0.07 ± 0.010	0.188 ± 0.014	0.207 ± 0.016	0.214 ± 0.016
G	0.259 ± 0.006	0.246 ± 0.008	0.224 ± 0.005	0.226 ± 0.006	0.219 ± 0.007	0.197 ± 0.006	NA/0.11 ± 0.010	0.188 ± 0.014	0.191 ± 0.014
H1	0.270 ± 0.006	0.234 ± 0.006	0.234 ± 0.005	0.235 ± 0.005	0.231 ± 0.006	0.198 ± 0.006	0.195 ± 0.006	0.08 ± 0.00/0.05 ± 0.00	0.145 ± 0.011
H2	0.265 ± 0.006	0.235 ± 0.005	0.231 ± 0.005	0.233 ± 0.005	0.230 ± 0.005	0.198 ± 0.005	0.191 ± 0.005	0.154 ± 0.004	0.12 ± 0.00/0.09 ± 0.010

The data obtained from full-length and near-complete VP1 sequences are shown in lower left and top right quarters, respectively. NA: not available. It means that there are only one, no sequence, or more than two completely identical sequences.

Interestingly, we found an inconsistent location of the clade of subtype B between the ML trees of genomic sequence and VP1 sequence (Fig. 5). The subtype B clade was located at the root of the ML tree of genomic sequences, but it was located at middle, and closely clustered with the clade of subtype H (including H1 and H2) in the VP1 sequence tree. This result implies that recombination events between subtype B and other subtypes have occurred during the evolution of subtype B strains. To confirm this hypothesis, we performed bootscan and phylogenetic analyses. The bootscan analysis showed that several genomic segments from subtypes H and F split the backbone of subtype A into seven segments (supplementary Fig. S1C), which was further confirmed by separate phylogenetic analyses (data not shown). Because subtype B was genetically closely related to subtype A, the subtype A strain used in the bootscan analysis can reflect the early strain of subtype B. Therefore, the bootscan analysis indicated that several subtypes H, F and C-related genomic segmented were inserted into the genomic backbone of subtype B by recombination during the evolution of subtype B.

Comparison Of The New Nomenclature With The Old Ones

According to the new classification, EVA71, CVA16, CVA6 and CVA10 were divided into 7, 3, 3 and 9 subtypes/sub-subtypes in alphabetical order, as well as 4, 1, 3 and 0 unclassified potential subtypes, respectively (Table 5). The major difference between the new and old nomenclatures of EVA71 was that previous C1-C3 and C5 were redefined as subtypes F and G, and C2 to subtype D. The dominant sub-subtype C4 was redefined as subtype C. Previous subtypes D, F and G were defined as unclassified subtypes because of the lack of genomic sequence. For CVA16, the changes were that previous C and B1 strains were merged into new sub-subtype B, and previous subtype D was redefined as subtype C. Previous nomenclature of CVA10 and CVA6 based on partial VP1 sequences seemed to be relatively confusing, and contained some misclassifications. The corresponding relationship of new and old nomenclatures of both enteroviruses are detailed in Table 5. For the ease of use of the new nomenclature system, GenBank accession numbers of reference strains and sequence alignment for each enterovirus are provided in Table 5 and supplementary file, respectively.

Table 5
Comparison of new and old nomenclatures of enteroviruses.

Enteroviruses	New nomenclature	Previous nomenclature	Reference strains
EVA71	A	A	U22521, GU434678, AB204853
	B	B1-B5	DQ341354, HQ189392, JF738001
	C	C4	HQ423142, AF302996, MG207963
	D	C2	MG013988, MG672479
	E	E	MG672478
	F	C1	MK800119
	G	C1-C3,C5	HQ647173, JN835312, DQ341359
	RF01_CG	C2	KF982854
	RF02_CG	C4	KC954664
	Unclassified 1	NA	KY115200*
	Unclassified 2	F	HG421068*, HG421069*
	Unclassified 3	G	KF906417*, KF906416*
	Unclassified 4	D	KF906421*, KF906419*, KF906425*
CVA16	A	A	U05876, JQ746659, EU812514
	B	C/B1a,b/B1a,b,c	JQ746677, JF738003, JQ746678
	C	D	MG957117, LT577761*,LT617115*
	Unclassified	B2	AB465370*, AM292455*, AM292461*
CVA6	A	A/B	AY421764, AF081297*
	B	NA	LR027552
	C	E/F/G	MH716144, JQ946054, LC126146
	Unclassified 1	A	KF412903*
	Unclassified 2	A/B/C	JQ364886*, KP143075*, LC421656*
	Unclassified 3	C/D/E	JN203517*, JQ364887*
CVA10	A	A/F	AY421767, AF081300*
	B	B/C/D	MH118041, KC879535*, HE572987*
	C	C/D/G	HQ728262, KP289406, KU578131
	D	NA	MF678312
	E	NA	MF422532
	F	C/B/F	MF422531, GQ214177*, GQ214175*
	G	C	MH118054, KC879491*, KC879488
	H1	C/D	MH144599, MH118066, MH118057
	H2		MH118069, MH118036, MH144590

* complete VP1 sequence.

Discussion

Genetic variation affects virus transmission, pathogenicity and epidemics (1, 5, 7, 8). HFMD is caused by various enteroviruses that are a RNA family with high genetic diversity (3–6). Proper subtyping/grouping of enteroviruses is crucial and useful for tracking the HFMD epidemic (2, 8). However, current classification and nomenclature of enteroviruses are confusing and incoherent, and lack proper and uniform criteria (9–19). In this study, we reclassified four dominant HFMD-related enteroviruses EVA71, CVA16, CVA6 and CVA10 on the basis of a combined criteria of phylogenic relationship and genetic distance, and proposed a classification of 7 subtypes A-G for EVA71, 3 subtypes/sub-subtypes A-C for CVA16 and CVA6, and 9 subtypes/sub-subtypes A-G, H1 and H2 for

CVA10. Furthermore, 4, 1 and 3 unclassified potential subtypes were suggested for EVA71, CVA16, and CVA6, respectively. The new nomenclature proposal is intended as a reference guide for investigators to properly and consistently define new enteroviruses.

A large number of molecular epidemiological investigations have suggested that EVA71, CVA16, CVA6 and CVA10 were the predominant enteroviruses responsible for HFMD in China and other countries (8). Compared with previous nomenclatures of the four enteroviruses, the current version are simple and clear. One of the major features between previous and current nomenclatures was that multiple subtypes in previous literatures were often merged into a single subtype according to the phylogeny- and distance-based criteria. For example, previous EVA71 subtypes C1-C3 and C5 were reassigned as subtype G, some strains of CVA6 subtypes A-B, E-G, and C-E were reassigned as subtypes A, C and unclassified 3, respectively, and CVA10 subtypes A/F, C-D, C/D/G, C/B/F and C/D were reassigned as A, B, C, F and H1, respectively. Another feature was that the same subtypes in previous literatures could be divided into multiple different subtypes. For example, previous EVA71 subtype C strains were re-divided into subtypes C, D, F and G, CVA6 subtype A strains were re-divided into subtypes A, and unclassified 1 and 2, and CVA10 subtype C strains were re-divided into subtypes B, C, F, G and H1. According the new nomenclatures of the four enteroviruses, the predominant circulating strains of the four enteroviruses were EVA71 subtype C, CVA16 subtype B, CVA6 subtype C and CVA10 subtypes C and H.

Three near full-length genomic sequences from epidemiologically unlinked individuals are preferred to define a new subtype of a virus (22). Because of the difficulty of obtaining full-length genomic sequences, near-complete, especially partial VP1 sequences were often used to define the subtypes of enteroviruses (9, 10, 20). In most scenarios, full-length genomic sequences and near-complete VP1 sequences generate consistent phylogeny of enteroviruses. In some cases, however, inconsistent topologies of certain subtypes were observed between the phylogenetic trees of full-length genomic sequences and near-complete VP1 sequences, which might be indicative of the presence of inter-subtype recombination events, which frequently occur during the evolution of enteroviruses (11, 23, 24). Most observed inter-subtype recombination events in enteroviruses occurred in the distant past, and the recombinants had evolved into well-defined subtypes during a long evolutionary history. This was especially true for EVA71 subtypes D and F and CVA10 subtype B, which were demonstrated to experience recombination during early evolution. However, more recent recombination events were rarely observed in enteroviruses, albeit they can generate some strains having distinct phylogenetic location and being difficult to be classified as certain subtype in the analyses of genomic and/or complete VP1 sequences. In this study, we identified two EVA71 recombinants (RF01_CG and RF02_CG), which originated from recombination between subtypes C and G.

On the other hand, inconsistent phylogenetic topology to full-length genomic sequences was often observed when too short VP1 sequences (less than 450 nt) were used. The vast majority of enterovirus VP1 sequences available in GenBank had length of less than 450 nt, and these short VP1 sequences were often used to define or determine subtypes (9, 25). Because the short VP1 sequences carried too few informative sites to define or distinguish a subtype, they often resulted in misclassification of enteroviruses, which explained why there were much inconsistency between the current and previous nomenclatures of four dominant enteroviruses. On basis of our results, near-complete VP1 sequences with length of more than 870 nt are recommend to determine the subtypes of enteroviruses.

A cut-off of 25% genetic divergence in the complete VP1 region was previously suggested to distinguish or define novel types of enteroviruses (4, 6). Subtypes and sub-subtypes are the distinct clades seen in the phylogenies of certain of enterovirus types. We found that the mean genetic distance in the complete VP1 region was 12.2–33.8% between subtypes (including unclassified subtypes), 12–15% between sub-subtypes, and less than 12% within subtype/sub-subtypes. Inter-subtype genetic distance higher than 25% were commonly observed between subtype A and other subtypes in CVA16 and CVA10, and subtype C and other subtypes in CVA6. These indicate that enteroviruses are more divergent than previously thought, implying that there might have some new potential subtypes not yet discovered.

There are two limitations in our current study. First, some newly defined subtypes had less than three representative full-length genomics sequences, which did not meet perfectly the rigorous nomenclature criterion to designate a new subtype/sub-subtype. In particular, 13 subtypes of the four dominant enteroviruses had only one available genomic sequence, albeit some of them were further supported by additional complete VP1 sequences. Therefore, these subtypes should be further confirmed in future as more genomic sequences have become available. Furthermore, there is an increasing need of obtaining genomic sequences to confirm these unclassified subtypes that lack supportive genomic sequences. Second, we only classified four dominant HFMD-related enteroviruses EVA71, CVA16, CVA6 and CVA10, whereas some other common enterovirus types (e.g. CVA4, CVA9, etc.) were not covered in this study. Besides the four HFMD-related enteroviruses, as well as CVB3 that was recently classified by us (20), all other enteroviruses need to be classified under the same criteria in the future.

Conclusion

We provided a new framework for the classification of enteroviruses on the basis of phylogeny- and distance-based criteria. Four dominant HFMD-related EVA71, CVA16, CVA6 and CVA10 were re-classified into 7(A-G), 3(A-C), 3(A-C), and 9(A-G, H1 and H2) subtypes/sub-subtypes, respectively. Furthermore, 2 recombinant forms RF01_CG and RF02_CG) of EVA71 and 8 unclassified subtypes of EVA71, CVA16 and CVA6 were identified. The clear and consistent genetic classification proposal for enteroviruses will be useful for epidemiological surveillance of HFMD, disease management, and vaccine development.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests

Funding

This research was funded by the National Science and Technology Major Project of China under Grant [2017ZX10103009-002, 2019YFC1200603 and 2018ZX10711001] and the TOTAL foundation.

Authors' contributions

Conceptualization, C.Z.; methodology, C.Z. and Y.D.; validation, C.Z., G.Y., X.J., J.W. and Y.D.; formal analysis, Y.D. and C.Z.; investigation, Y.D., Z.W. and S.L.; data curation, C.Z. and Y.D.; writing—original draft preparation, C.Z. and Y.D.; writing—review and editing, X.J.; visualization, Y.D. and C.Z. supervision, C.Z. and G.Y.; funding acquisition, C.Z. and S.L.. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We gratefully acknowledge Kai Liu, Yingying Ma, Lulu Zuo at IPS, CAS, Xuemin Fu at Max Planck Institute of Psychiatry, Munich, Germany, and Guangming Shang at Soochow University for their technical support.

References

1. Zaoutis T, Klein JD. Enterovirus infections. Pediatrics in review. 1998;19(6):183-91.
2. Lei X, Cui S, Zhao Z, Wang J. Etiology, pathogenesis, antivirals and vaccines of hand, foot, and mouth disease. National Science Review. 2015;2(3):268-84.
3. Hyypia T, Hovi T, Knowles NJ, Stanway G. Classification of enteroviruses based on molecular and biological properties. The Journal of general virology. 1997;78 (Pt 1):1-11.
4. Lukashev AN, Vakulenko YA, Turbabina NA, Deviatkin AA, Drexler JF. Molecular epidemiology and phylogenetics of human enteroviruses: Is there a forest behind the trees? Rev Med Virol. 2018;28(6):e2002.
5. Muehlenbachs A, Bhatnagar J, Zaki SR. Tissue tropism, pathology and pathogenesis of enterovirus infection. The Journal of pathology. 2015;235(2):217-28.
6. Lukashev AN, Vakulenko YA. Molecular evolution of types in non-polio enteroviruses. The Journal of general virology. 2017;98(12):2968-81.
7. Wang C, Zhou S, Xue W, Shen L, Huang W, Zhang Y, et al. Comprehensive virome analysis reveals the complexity and diversity of the viral spectrum in pediatric patients diagnosed with severe and mild hand-foot-and-mouth disease. Virology. 2018;518:116-25.
8. Fu X, Wan Z, Li Y, Hu Y, Jin X, Zhang C. National Epidemiology and Evolutionary History of Four Hand, Foot and Mouth Disease-Related Enteroviruses in China from 2008 to 2016. Virol Sin. 2020;35(1):21-33.
9. Oberste MS, Maher K, Kilpatrick DR, Flemister MR, Brown BA, Pallansch MA. Typing of human enteroviruses by partial sequencing of VP1. Journal of clinical microbiology. 1999;37(5):1288-93.
10. Oberste MS, Maher K, Kilpatrick DR, Pallansch MA. Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. Journal of virology. 1999;73(3):1941-8.
11. Zhang C, Zhu R, Yang Y, Chi Y, Yin J, Tang X, et al. Phylogenetic analysis of the major causative agents of hand, foot and mouth disease in Suzhou City, Jiangsu province, China, in 2012-2013. Emerg Microbes Infect. 2015;4(2):e12.
12. Xu M, Su L, Cao L, Zhong H, Dong N, Dong Z, et al. Genotypes of the Enterovirus Causing Hand Foot and Mouth Disease in Shanghai, China, 2012-2013. PLoS One. 2015;10(9):e0138514.
13. Weng Y, Chen W, He W, Huang M, Zhu Y, Yan Y. Serotyping and Genetic Characterization of Hand, Foot, and Mouth Disease (HFMD)-Associated Enteroviruses of No-EV71 and Non-CVA16 Circulating in Fujian, China, 2011-2015. Med Sci Monit. 2017;23:2508-18.
14. Perera D, Yusof MA, Podin Y, Ooi MH, Thao NT, Wong KK, et al. Molecular phylogeny of modern coxsackievirus A16. Arch Virol. 2007;152(6):1201-8.
15. He YQ, Chen L, Xu WB, Yang H, Wang HZ, Zong WP, et al. Emergence, circulation, and spatiotemporal phylogenetic analysis of coxsackievirus a6- and coxsackievirus a10-associated hand, foot, and mouth disease infections from 2008 to 2012 in Shenzhen, China. Journal of clinical microbiology.

2013;51(11):3560-6.
16. Ji T, Guo Y, Huang W, Shi Y, Xu Y, Tong W, et al. The emerging sub-genotype C2 of CoxsackievirusA10 Associated with Hand, Foot and Mouth Disease extensively circulating in mainland of China. Scientific reports. 2018;8(1):13357.
17. Song Y, Zhang Y, Ji T, Gu X, Yang Q, Zhu S, et al. Persistent circulation of Coxsackievirus A6 of genotype D3 in mainland of China between 2008 and 2015. Scientific reports. 2017;7(1):5491.
18. Saxena VK, Sane S, Nadkarni SS, Sharma DK, Deshpande JM. Genetic diversity of enterovirus A71, India. Emerging infectious diseases. 2015;21(1):123-6.
19. Chan YF, Sam IC, AbuBakar S. Phylogenetic designation of enterovirus 71 genotypes and subgenotypes using complete genome sequences. Infect Genet Evol. 2010;10(3):404-12.
20. Fu X, Mao L, Wan Z, Xu R, Ma Y, Shen L, et al. High proportion of coxsackievirus B3 genotype A in hand, foot and mouth disease in Zhenjiang, China, 2011-2016. Int J Infect Dis. 2019;87:1-7.
21. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Journal of virology. 1999;73(1):152-60.
22. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014;59(1):318-27.
23. Oberste MS, Penaranda S, Maher K, Pallansch MA. Complete genome sequences of all members of the species Human enterovirus A. The Journal of general virology. 2004;85(Pt 6):1597-607.
24. Simmonds P, Welch J. Frequency and dynamics of recombination within different species of human enteroviruses. Journal of virology. 2006;80(1):483-93.
25. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. Journal of clinical microbiology. 2006;44(8):2698-704.

Figures

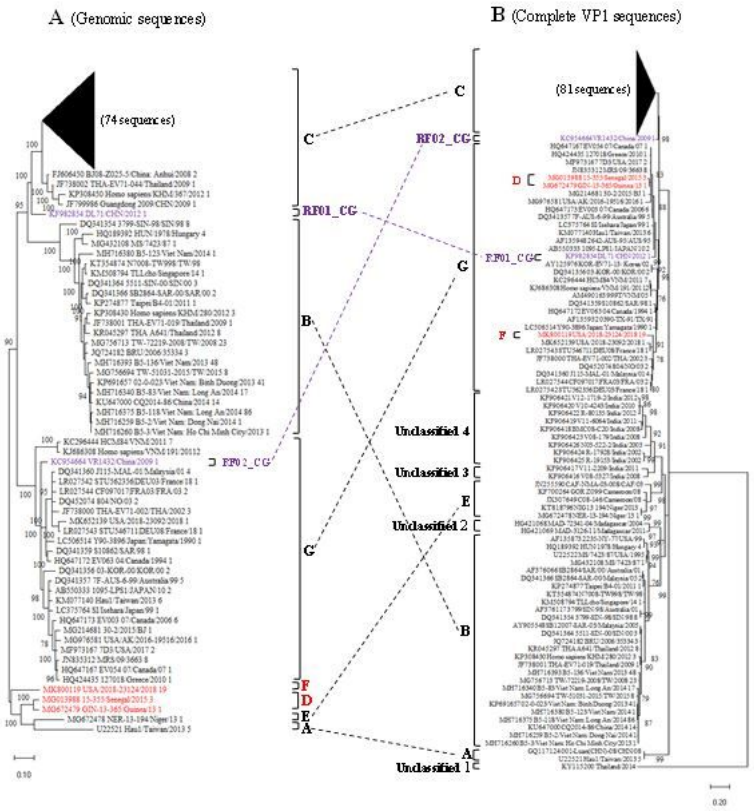

Figure 1

Genotype classification of EVA71 based on near full-length genomic (A) and near-complete VP1 (B) sequences. The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only the bootstrap values of >75% are shown at the nodes. Two subtypes experiencing early recombination events are highlighted in red font. Two recombinants are highlighted in purple font. The last number in the name of each strain indicates the number of completely identical genomic sequences. The last number “1” indicates that there was only one unique genomic sequence.

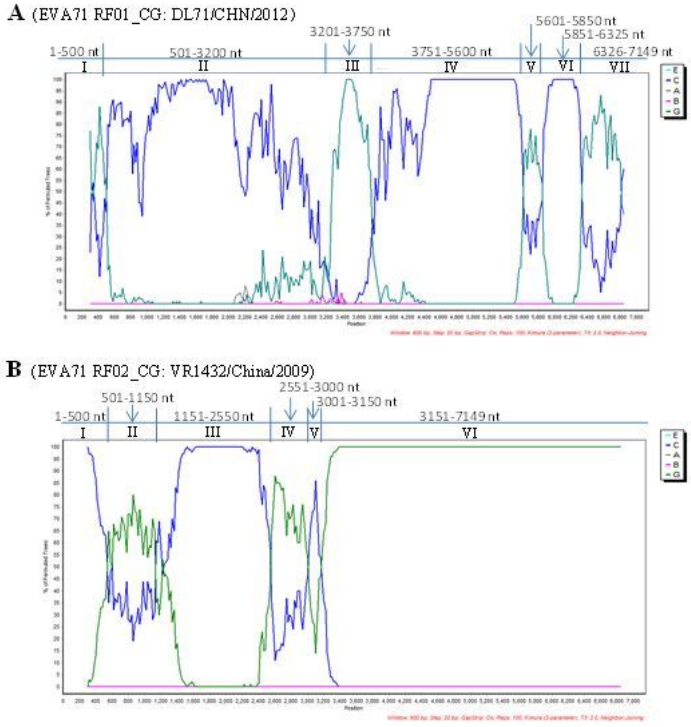

Figure 2

Bootscan analysis of two inter-subtype recombinants of EVA71. A: RF01_CG; B: RF02_CG. The reference strains used in the analyses were genotypes A (U22521), B (JF738001), C (JQ742002), D (MG672478) and G (HQ647172). The analyses were performed using a sliding 600-bp window with 20-bp steps.

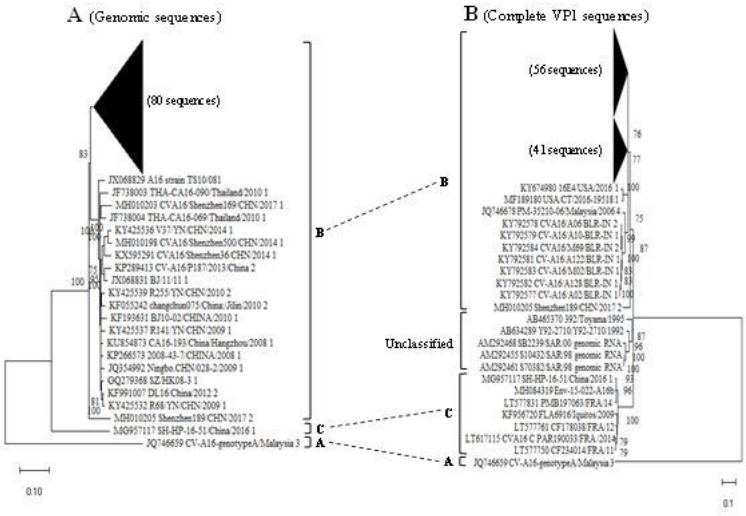

Figure 3

Genotype classification of CVA16 based on near full-length genomic (A) and near-complete VP1 (B) sequences. The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only the bootstrap values of >75% are shown at the nodes. The last number in the name of each strain indicates the number of completely identical genomic sequences. The last number “1” indicates that there was only one unique genomic sequence.

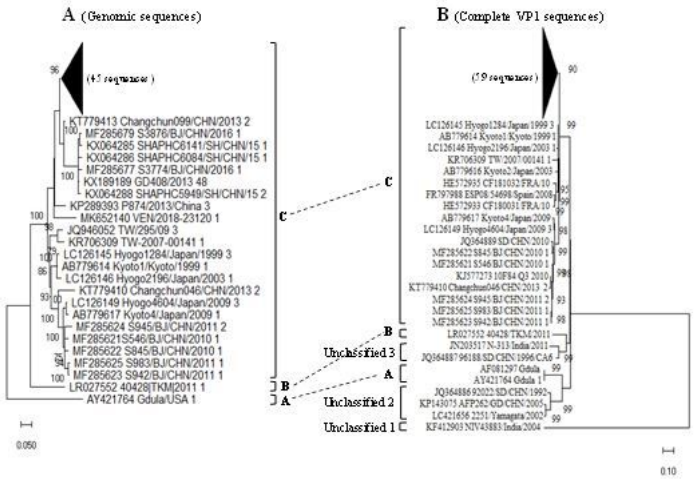

Figure 4

Genotype classification of CVA6 based on near full-length genomic (A) and near-complete VP1 (B) sequences. The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only the bootstrap values of >75% are shown at the nodes. The last number in the name of each strain indicates the number of completely identical genomic sequences. The last number “1” indicates that there was only one unique genomic sequence.

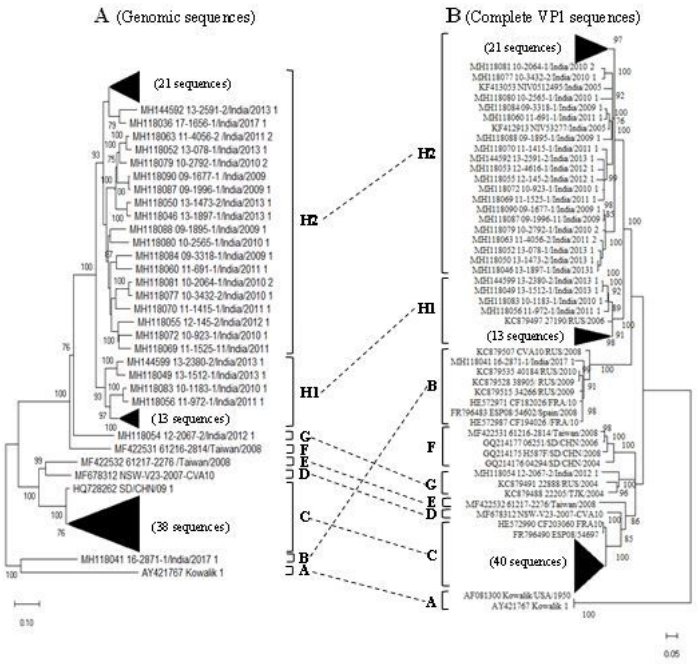

Figure 5

Genotype classification of CVA10 based on near full-length genomic (A) and near-complete VP1 (B) sequences. The stability of the nodes was assessed by bootstrap analysis with 1000 replications, and only the bootstrap values of >75% are shown at the nodes. The last number in the name of each strain indicates the number of completely identical genomic sequences. The last number “1” indicates that there was only one unique genomic sequence.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- idpsupplementarymaterials.docx