Recombination strategies and evolutionary dynamics of the *Human enterovirus* A global gene pool

Alexander N. Lukashev,1 Elena Yu. Shumilina,1 Ilya S. Belalov,1 Olga E. Ivanova,1 Tatiana P. Eremeeva,1 Vadim I. Reznik,2 O. E. Trotsenko,3 Jan Felix Drexler4 and Christian Drosten4

Correspondence
Alexander N. Lukashev
alexander_lukashev@hotmail.com

1Chumakov Institute of Poliomyelitis and Viral Encephalitides, Moscow, Russia
2Center of Hygiene and Epidemiology in Khabarovsk Region, Khabarovsk, Russia
3Khabarovsk Institute of Epidemiology and Microbiology, Khabarovsk, Russia
4Institute of Virology, University of Bonn Medical Centre, Bonn, Germany

We analysed natural recombination in 79 *Human enterovirus* A strains representing 13 serotypes by sequencing of VP1, 2C and 3D genome regions. The half-life of a non-recombinant tree node in coxsackieviruses 2, 4 and 10 was only 3.5 years, and never more than 9 years. All coxsackieviruses that differed by more than 7% of the nucleotide sequence in any genome region were recombinants relative to each other. Enterovirus 71 (EV71), on the contrary, displayed remarkable genetic stability. Three major EV71 clades were stable for 19–29 years, with a half-life of non-recombinant viruses between 13 and 18.5 years in different clades. Only five EV71 strains out of over 150 recently acquired non-structural genome regions from coxsackieviruses, while none of 80 contemporary coxsackieviruses had non-structural genes transferred from the three EV71 clades. In contrast to earlier observations, recombination between VP1 and 2C genome regions was not more frequent than between 2C and 3D regions.

Human enteroviruses are members of the genus *Enterovirus* within the family *Picornaviridae*. These small, non-enveloped RNA viruses have a single-stranded non-segmented genome of positive polarity of about 7500 nt. The genome encodes a single large polyprotein comprising four structural proteins, VP1–VP4 (collectively termed the structural genome region), and seven non-structural proteins, 2A, 2B, 2C, 3A, 3B, 3C and 3D (the non-structural genome region) (Racaniello, 2007).

Human enteroviruses comprise four species, *Human enterovirus* A–D (HEV-A–D). Enterovirus 71 (EV71) is the best studied non-polio enterovirus, and it was further classified into subtypes A, B0–B5, C1–C5 and D (Brown et al., 1999; McMinn, 2012). Enteroviruses are highly prevalent in the human population, especially in infants (Witsø et al., 2006). Infection with HEVs is usually subclinical, but can occasionally result in single cases and outbreaks of meningitis, sepsis-like infection and myocarditis. Infection with the three poliovirus serotypes (members of the HEV-C species) and EV71 can produce severe neurological lesions (Pallansch & Roos, 2007). Global prevalence of enteroviruses and high infection rates in infants greatly facilitate natural recombination. Recombination is apparently most prevalent in the HEV-B species (Simmonds & Welch, 2006). However, it is common also in HEV-A (Huang et al., 2009; Oberste et al., 2004; Simmonds & Welch, 2006) and HEV-C (Brown et al., 2003). As a result, enterovirus species exist as highly dynamic global gene pools (reviewed by Lukashev, 2005). It has been shown that the half-life of recombinant forms (combinations of discrete VP1 and 3D genes without further signs of recombination) in circulation may vary from 1.3 to 9.8 years between different HEV-B serotypes (McWilliam Leitch et al., 2010). Recombination in HEV-A was investigated in a number of studies, but these were confined by limited geographical coverage (Simmonds & Welch, 2006), sampling restricted to a few prototype strains (Oberste et al., 2004), or by analysing only the neurovirulent EV71, but no other contemporary HEV-A (Chen et al., 2010; Huang et al., 2008; Mirand et al., 2010; van der Sanden et al., 2011). The aim of this study was to determine recombination patterns in HEV-A on a whole species level with an extensive geographical and temporal coverage.

Seventy-eight HEV-A strains were isolated and identified according to a standard World Health Organization (WHO) protocol (WHO, 1992) in RD cell culture in the course of the WHO polio surveillance programme and enterovirus

---

GenBank accession numbers for identified sequences are KC879327–KC879556.

Two supplementary tables are available with the online version of this paper.
surveillance in Russia and New Independent States in 2001–2011 (Table S1, available in the online Supplementary Material). According to the national regulations, informed consent is not required for anonymized surveillance studies. Three genome regions were amplified by PCR, namely the complete VP1 and partial 2C and 3D. PCR primers are listed in Table S2. All viruses used for the study differed by at least 1% nucleotide sequence in their VP1 genome region; less divergent sequences were excluded upon preliminary analysis. GenBank accession numbers for sequences obtained here are KC879327–KC879556. All complete or nearly complete HEV-A sequences available in GenBank were aligned using CLUSTAL W (Thompson et al., 1994). Redundant sequences sharing more than 97% overall nucleotide sequence identity were then omitted, yielding 76 unique reference sequences. Sequence handling was performed with BioEdit v.7.0.5.2 software (Hall, 1999). Phylogenies were calculated using a maximum-likelihood tree algorithm with the Tajima–Nei substitution model in the MEGA 5.2 software package (Tamura et al., 2011).

In the VP1 genome region (nt 2439–3329 in the prototype EV71 strain BrCr, GenBank accession number U22521), phylogenetic grouping expectedly corresponded to serotypes (Fig. 1). A Bayesian likelihood-based algorithm implemented in BEAST version 1.7.4 (Drummond & Rambaut, 2007) was then used to date tree nodes that were conserved across all three genome regions. Bayesian molecular dating was performed separately for distinct serotypes, because analysis of the VP1 alignment that included multiple serotypes consistently yielded substitution rates incompatible with those reported in other studies and those resulting from analysis of individual serotypes (data not shown). The SRD06 substitution model optimized for coding sequences (Shapiro et al., 2006) was used with a lognormal relaxed clock setting, which was preferred over assumption of a strict clock upon Bayes factor testing (log10 Bayes factor of tree likelihood >10), while the exponential clock setting was not significantly superior to the lognormal clock. Each analysis was run over 10,000,000 generations, and trees were sampled every 1000 generations, resulting in 10,000 trees. Trees were annotated with TreeAnnotator v.1.4.8 using a burn-in of 2000 trees. Substitution rates in different serotypes (expressed as 10−3) were 4.05 (95% confidence interval 3.23–4.93) substitutions per site year−1 in EV71, 8.32 (5.35–11.6) in coxsackievirus A (CVA2), 5.53 (3.57–7.57) in CVA4 and 14.1 (4.01–24.2) in CVA10. The substitution rate for EV71 was remarkably similar to previously published estimates of 4.2–4.6 (Tee et al., 2010) and 3.7 (Mirand et al., 2010), but differed from the substitution rate of 7.2 reported in another work (McWilliam Leitch et al., 2012). Importantly, in the latter work, lower substitution rates were observed within individual EV71 subgenotypes (3.2–7.3), implying that the overall rate of 7.2 may be inaccurate. While there have been no published studies of substitution rates in other HEV-A types, the observed variation is compatible with the range of substitution rates in HEV-B serotypes (McWilliam Leitch et al., 2010). The time of the most recent common ancestor (tMRCA) of coxsackievirus serotypes (CVA2, 69 years; CVA4, 66 years; CVA10, 72 years) was lower than tMRCA of EV71 (91 years); therefore, EV71 may be a cause of an emerging infection, but according to the phylogenetic evidence, it cannot be deemed an emerging virus.

In the 2C (nt 4350–4766) and 3D (nt 5973–6602) genome regions, distribution of pairwise distances was relatively even, and did not allow assigning recombinant forms as was done earlier (McWilliam Leitch et al., 2012). Phylogenetic grouping generally did not correspond to the VP1 genome region (Fig. 1), implying common recombination. All coxsackieviruses except for type 16 were extensively shuffled in the non-structural genome regions. This kind of phylogenetic relationship is best exemplified by group I (Fig. 1) consisting of 23 viruses of 11 serotypes in the 2C genome region or group Ia that comprised 29 viruses of nine serotypes in the 3D genome region. The other extremity of recombination frequency was represented by EV71 and CVA16. In the 2C and 3D genome regions, most EV71 strains fell into three groups (groups II–IV, Table 1, Fig. 1), which corresponded to subgenotypes B4–B5, C1–C3 and C4, respectively. While grouping of all EV71 B subgenotypes was conserved throughout the genome in our study, it is known that recombination occurred within that group; therefore we regarded only the group comprising subgenotypes B4 and B5, which was also conserved in previous studies with a more extensive sampling of EV71 (McWilliam Leitch et al., 2012). There were few EV71 sequences with alternative 2C (four out of 48 sequences, two of B4 and two of C2 genotype) and 3D genome regions (five sequences, the four mentioned above and one EV71-C4). However, the three EV71 groups present in the 2C and 3D genome regions (a total of 51 and 50 sequences, respectively) did not contain a single sequence from another serotype. Within these EV71 groups, there was also no significant evidence of recombination (change of tree topology supported by bootstrap values over 70) between the three genome regions. Similarly, all 14 CVA16 isolates sequenced here formed a single phylogenetic group without signs of recombination between VP1, 2C and 3D genome regions (Fig. 1). These three groups, which were characterized by limited recombination, had a tMRCA of around two decades. Median ages of non-recombinant viruses from the root of their respective group (roughly equivalent to half-lives of non-recombinant viruses reported by McWilliam Leitch et al., 2012) were between 13 and 18.5 years (Table 1). These numbers make EV71 the least recombining enterovirus compared with echovirus 9 (E9), E30 and E11 that featured recombinant form half-lives of 1.3, 3.1 and 9.8 years, respectively (McWilliam Leitch et al., 2010). The difference from numbers reported for EV71 by McWilliam Leitch et al. (2012) (6 and 9 years for GtB and GtC, respectively) may be explained by the aforementioned uncertainty of the substitution rate and by limited sampling of GtB sequences (by either quantity or geographical coverage). The sampling of CVA16 was confined to Russian
Fig. 1. Phylogenetic trees of HEV-A in three genome regions. Strains were colour-coded according to their type. Scale bars indicate maximum likelihood distances. Numbers at tree nodes indicate bootstrap support values. Roman numbers indicate groups referred to in the text.
strains as opposed to the globally representative sampling of EV71. Therefore, it is not possible to conclude if this serotype is indeed less involved in recombination within HEV-A, especially as diverse recombinant forms of CVA16 were described in the study by McWilliam Leitch et al. (2012).

In accordance with earlier reports on genetic identity of novel HEV-A serotypes (Oberste et al., 2005), three isolates of EV76 and one of EV90 did not have signs of recombination with classical HEV-A serotypes. However, these viruses displayed multiple phylogenetic conflicts involving EV76, EV89, EV90 and EV91, supporting that these types might represent a subspecies of HEV-A.

All other coxsackievirus A types, except for CVA16, were highly involved into recombination within the global gene pool. Bayesian dating was done for three types with significant sampling, namely CVA2 (13 viruses), CVA4 (17 viruses) and CVA10 (13 viruses). The median age of tree nodes that were conserved across the three genome regions (reliably supported by bootstrap values above 70 in one genome region and intact, but not necessarily well supported, in another region) was just 3.6 years, and never more than 8.7 years, which is several-fold less than the age of non-recombinant EV71 groups that had tMRCA of 19–29 years. The median age of non-recombinant isolates was 3.5 years, and never exceeded 9 years.

To further illustrate two distinct recombination patterns in EV71, CVA16 and other coxsackieviruses, pairwise genetic distances in three genome regions were plotted separately within each of these two groups (Fig. 2).

In EV71 and CVA16, the pairwise distances in three genome regions correlated very well up to a genetic distance of about 0.14 (Fig. 2a, region α), indicating long-term circulation of viruses in the absence of recombination. There were discordant pairwise distances (low in VP1 and high in 2C and 3D, Fig. 2a, β region) that corresponded to the few EV71 strains that acquired an ‘alien’ 2C and 3D genome region (see above). There were no viruses with low pairwise distances in 2C/3D and divergent VP1 genome regions (Fig. 2a, γ region), which agreed with the absence of such viruses on phylogenetic trees.

The coxsackievirus group also included viruses without apparent signs of recombination (with concordant pairwise distances), but only with genetic distances up to approximately 0.07 (Fig. 2b, α region). Recombination events apparent as discordant genetic distances beginning from as low as 0.01 involved all genome regions (Fig. 2b, regions β, γ). In viruses that differed by over 0.07 of their RNA sequence in any genome region, pairwise sequence distances were largely discordant between all genome regions,

<table>
<thead>
<tr>
<th>Group</th>
<th>Serotypes</th>
<th>Number of members in VP1/2C/3D</th>
<th>Median non-recombinant virus age from node</th>
<th>Group tMRCA in VP1 (years)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>EV71 genotypes B4–B5</td>
<td>10/8/8</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>III</td>
<td>EV71 subtypes C1–C3</td>
<td>20/18/18</td>
<td>18.5</td>
<td>29</td>
</tr>
<tr>
<td>IV</td>
<td>EV71 subtype C4</td>
<td>25/25/24</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>–</td>
<td>CVA2, 4, 10 (7 conserved groups)</td>
<td>22/22/22</td>
<td>3.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>

*Time of tMRCA was calculated as node height minus age of the most recent isolate of a group.

---

**Table 1. Features of conserved HEV-A groups**

---

**Fig. 2.** Plot of pairwise distances among EV71 and CVA16 (a) and among other coxsackieviruses of species HEV-A (b) in three genome regions. Each point represents a pair of pairwise distances between the same strains in two genome regions that correspond to the symbol description. The uncorrected nucleotide sequence distance in the first genome region is plotted on the x-axis and that in the second region of the pair on the y-axis. Circled areas are described in the text.
indicating common shuffling of genome fragments by recombination (Fig. 2b, δ region). This observation corresponds well with the maximum age of conserved tree nodes for coxsackieviruses (8.7 years) multiplied by the estimated substitution rates \((5.5 \times 10^{-3} \text{ to } 1.41 \times 10^{-3} \text{ substitutions per site year}^{-1})\).

The number of conserved groups was the highest between the VP1 and 2C genome regions (14 groups), and about the same between the VP1 and 3D, and the 2C and 3D regions (seven and eight groups, respectively), while the distance between these genome regions is about equal (750 nt between VP1 and 2C and 870 nt between 2C and 3D). This observation contradicts the reports of an apparent recombination hot-spot on the border of structural and non-structural genes in enteroviruses and other picornaviruses (Benschop et al., 2010; Heath et al., 2006; Lukashev, 2005, 2010). It is likely that the apparent recombination hot-spot reported previously at the VP1–2A junction may only correspond to an artefact due to a stronger phylogenetic signal in the P1 genome region. This strong signal would result from the absence of recombination between capsid genes of different serotypes, higher variability of the structural genome region and, consequently, a better-resolved phylogeny. As a result, it is easier to detect a switch of phylogeny on the border of this region. The non-structural genes are arbitrarily shuffled between serotypes every few years, so the phylogenetic signal is quickly degraded and fewer recombination events can be detected with sufficient statistical support.

Before this study, EV71 comprised over 80% of all HEV-A complete genomic sequences in GenBank and was the sole serotype used in many studies. Recombination was implied by the emergence of several EV71 subtypes (McWilliam Leitch et al., 2012; Yoke-Fun & AbuBakar, 2006). The recombination partners in these recombination events remained unknown because there were no available sequences of contemporary HEV-A strains. Our study indicates that the recombination events leading to the emergence of EV71 subtypes involved members of the global HEV-A gene pool and not only EV71. This is evident because the three EV71 groups II–IV were evenly interspersed on the phylogenetic trees for the 2C and 3D genome regions and did not originate from a shared common ancestor. Surprisingly, EV71 subtypes almost ceased recombining with the rest of the HEV-A gene pool upon their emergence. Perfect correspondence of EV71 phylogenies in different genome regions was reported earlier (Bible et al., 2008; McWilliam Leitch et al., 2012; van der Sanden et al., 2011). This study highlights the degree of genetic conservation in EV71, which apparently recombines about five times less frequently than its peers in the HEV-A species CVA2, CVA4 and CVA10. Moreover, the high half-life of non-recombinant EV71 strains indicates that they are the most conserved types in terms of recombination frequency among enteroviruses that have been closely studied so far.

Several factors could explain the low apparent recombination in EV71. First, this could be due to reproductive isolation at the level of permissible/preferred cell types. However, EV71 and some other HEV-A members share the same putative cellular receptors PSGL-1 and SCARB2 (Nishimura & Shimizu, 2012). Also, use of different receptors by HEV-B serotypes (Racaniello, 2007) did not hamper promiscuous recombination (Lukashev et al., 2003; Simmonds & Welch, 2006), and poliovirus recombines freely with other HEV-C members that do not use the poliovirus receptor (Brown et al., 2003; Jegouic et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009). Another factor that could contribute to the low recombination rate in EV71 could be adaptation of non-structural genes to the EV71 capsid. Such adaptation of 2C and VP3 proteins was described in HEV-C species (Liu et al., 2009).

**Acknowledgements**

This study was supported by the Global Polio Eradication Initiative of the WHO through the WHO European Office and by the Government of the Russian Federation by the budgetary allocations for the implementation of activities under the Global Polio Eradication Initiative (Government act 1771-p of 14.10.2010). We would like to thank the staff of the following regional laboratories for technical assistance in collection of stool samples and virus isolation: Centres of Hygiene and Epidemiology in Moscow, Omsk region, Stavropol region; Regional Centre of Epidemiological Surveillance for Poliomyelitis, Pasteur Institute, St Petersburg; National Poliomyelitis Laboratories in Armenia, Georgia, Kazakhstan and Ukraine. The study was partially funded by the EU FP7 project EVA (contract number 228292) and by infrastructural support from the German Centre for Infection Research (DZIF).

**References**


