Infection, Genetics and Evolution 11 (2011) 276–289

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Repeated genomic transfers from echovirus 30 to echovirus 6 lineages indicate co-divergence between co-circulating populations of the two human enterovirus serotypes☆

J.-L. Bailly [a,*], A. Mirand [a,b], C. Henquell [b], C. Archimbaud [a,b], M. Chambon [a,b], C. Regagnon [b], F. Charbonné [a,b], H. Peigue-Lafeuille [a,b]

[a] Clermont Université, Université d'Auvergne, EA 3843, BP 10448, F-63000 Clermont-Ferrand, France
[b] CHU Clermont-Ferrand, Service Virologie, F-63003 Clermont-Ferrand, France

ARTICLE INFO

Article history:
Received 27 March 2010
Received in revised form 7 June 2010
Accepted 25 June 2010
Available online 6 July 2010

Keywords:
Aseptic meningitis
Evolutionary history
Picornavirus
Bayesian inference
Genetic recombination

ABSTRACT

Human echovirus types 6 (E-6) and 30 (E-30) cause seasonal epidemics of aseptic meningitis. These two enteroviruses are frequently observed in co-circulation, an epidemiological pattern that is prerequisite for the occurrence of dual infections, which can lead to recombination between co-infecting virus strains. Viral sequences were determined at loci 1D (VP1 capsid protein) and 3CD (non structural proteins) in 49 E-6 strains recovered in a single geographical region in France from 1999 to 2007, during the epidemiological survey of enterovirus infections. They were compared with previously recorded sequences of E-30 strains to investigate their evolutionary histories and possible recombination patterns. Phylogenetic analyses identified two distinct E-6 populations and different subpopulations. Assuming a relaxed molecular clock model and a Bayesian skyline demographic model in coalescent analyses with the BEAST program, the substitution rate in E-6 was estimated at 8.597×10^{-3} and 6.252×10^{-3} substitution/site/year for loci 1D and 3CD respectively. Consistent estimates of divergence times (t_{MRCA}) were obtained for loci 1D and 3CD indicating that two distinct E-6 populations originated in 1997 and 1999. Incongruent phylogenetic patterns inferred for the two loci were indicative of recombination events between the two populations. Phylogenies including the E-30 3CD sequences showed close genetic relationships between E-6 and discrete E-30 subpopulations. Recombination breakpoints were located with statistical significance in E-6 and E-30 genomes. Estimates of t_{MRCA} of phylogenetic recombinant clades indicated directional genetic transfers from E-30 to E-6 populations and their co-divergence over the time period studied.

1. Introduction

Echovirus types 6 (E-6) and 30 (E-30) are readily associated with acute infections in susceptible individuals and are recognized as leading causes of viral aseptic meningitis worldwide because they are regularly involved in sporadic infections and outbreaks within communities (Chomel et al., 2003; Thoelen et al., 2003; Khetsuriani et al., 2006; Richter et al., 2006; Roth et al., 2007; Cabrerizo et al., 2008; Ortner et al., 2009; Papa et al., 2009; Mao et al., 2010; Kapusinszky et al., 2010). These viruses represent two of nearly 100 types assigned to the human enteroviruses (HEVs; family *Picornaviridae*, genus *Enterovirus*). Types within HEVs are classified by sequence similarity into four species from HEV-A to HEV-D, and E-6 and E-30 are assigned to the HEV-B cluster (Stanway et al., 2005). The non enveloped virus particle of these viruses is composed of four capsid proteins (VP1–VP4) and includes a single-stranded, positive-sense RNA genome of about 7.5 kb in length (Racaniello, 2007). In addition to the capsid proteins, the genome encodes at least seven other non structural polypeptides whose genes are designated from 2A to 2C, and from 3A to 3D. The different HEV serotypes represent genetically defined clades that can be resolved through phylogenetic analysis of the 1D gene encoding the VP1 capsid protein (Oberste et al., 1999b). While the different serotypes share important structural and functional features, each of them appears to form a distinct monophyletic group including all circulating strains.

While the phylogenetic features of E-6 are unknown because of limited informative sequence data, the phylogeny of E-30 is

☆ This paper has been contributed through the 15th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology, Rotterdam, The Netherlands, September 7–11, 2009.

* Corresponding author at: Clermont Université, Université d'Auvergne, EA 3843, UFR Médecine, Laboratoire de virologie, 28 place Henri-Dunant, F-63000 Clermont-Ferrand, France. Tel.: +33 4 73 17 81 42; fax: +33 4 73 754 851.
E-mail address: j-luc.bailly@u-clermont1.fr (J.-L. Bailly).

1567-1348/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.06.019

characterized by distinct lineages among the virus populations that have been circulating over the last 50 years (Oberste et al., 1999a; Savolainen et al., 2001; Palacios et al., 2002). Molecular evolution of the 1D gene in E-30 populations is characterized by continual purifying selection against non synonymous substitutions and maintenance of amino acid polymorphism at only a few positions scattered in the VP1 protein (Bailly et al., 2009). Genetic recombination results in constant variation of the genome of circulating HEV strains (Santti et al., 1999; Oprisan et al., 2002; Lukashev et al., 2003; Lindberg et al., 2003; Oberste et al., 2004b; Simmonds, 2006; Bouslama et al., 2007). Recent investigations suggested that recombination is largely involved in the diversity of E-30 virus strains (Chen et al., 2007; Mirand et al., 2007; McWilliam Leitch et al., 2009). Homologous recombination is a recognized process of intermolecular RNA rearrangement in picornaviruses and was first discovered in poliovirus (Ledinko, 1963; Agol, 1997), an enterovirus assigned to the HEV-C species. Occurrence of outbreaks associated with recombinants of Sabin-derived polioviruses with wild-type polioviruses or with other genetically related serotypes have been reported (Kew et al., 2002; Yang et al., 2003; Shimizu et al., 2004; Adu et al., 2007; Rakoto-Andrianarivelo et al., 2007; Zhang et al., 2010), indicating a possible role of recombination in the diversification of strains within the HEV-C species (Brown et al., 2003; Jegouic et al., 2009). However, the key question of the real contribution of recombination to the evolution of an enterovirus population is still open. What is the biological relevance of the acquisition of various genome segments from other genetically related enterovirus types for any given enterovirus?

In this study, we enquire whether genetic recombination has consequences on the evolution of distinct populations of human enteroviruses detected in co-circulation over a short time period. In an earlier report, we showed that E-30 isolated from 1991 to 2005 in the same geographic region were distributed into five genetic clusters (noted from C0 to C4) by analyzing the gene sequences at locus 1D (Mirand et al., 2007). The genetic origin of clusters was associated with recombination because the topology of the phylogenetic tree estimated with the sequences determined at locus 3CD was different from the topology inferred for locus 1D. This suggested a polyclonal evolution of E-30 strains. While the incongruence between the phylogenetic patterns at two distant loci may arose from the mosaic structure of genomes, the discordant phylogenetic patterns can also be related to the fact that distinct lineages may have different evolutionary rates (Posada et al., 2002; Awadala, 2003). A comparative analysis of subgenomic segments covering the viral genome from the 1D gene to the 3CD sequence (# 4200 ntd), was performed with the bootscanning method (Salminen et al., 1995) for 13 strains selected among clusters C0–C4. The results suggested a mosaic structure of the E-30 genomes, in agreement with the recombination hypothesis, but the exact origin of the different genomic segments could not be established. This suggested that the sequence dataset was too limited and that substantial portions of the E-30 genomes had exogenous origins, i.e. they were derived from enterovirus strains of different types. To test this hypothesis, we compared the E-30 strains to the enterovirus strains of the other types detected in co-circulation during the epidemiological survey of aseptic meningitis (Chambon et al., 2001; Mirand et al., 2006, 2008). In this study, we present the findings of the comparative analysis between E-30 and E-6. The E-6 strains were isolated over a time period of 8 years (1999–2007), during which, in 2000, there was a large outbreak of aseptic meningitis. Our phylogenetic data provide evidence for short-term genomic exchanges that produce co-divergence patterns between co-circulating populations through repeated intertypic recombination events from E-30 to E-6 lineages.

2. Materials and methods

2.1. Virus strains

The 49 E-6 strains studied were recovered from inpatients admitted to two hospitals in France, the University Hospital of Clermont–Ferrand ($n = 47$) and the Trousseau Pediatric Hospital, Paris ($n = 2$) (Supplemental Table S1). The patients received a clinical diagnosis of aseptic enterovirus meningitis during the period 1999–2007. The 52 E-30 strains analyzed have been described elsewhere (Mirand et al., 2007).

2.2. Two-locus genotyping assay and determination of subgenomic sequences

The genotyping methods (RT-PCR assays and sequencing of 1D gene sequences) used for identifying the strains have been described previously (Mirand et al., 2006). The E-6 strains were either prospectively genotyped (strains isolated from 2005 to 2007; $n = 13$) or characterized retrospectively with the RT-PCR assay (strains recovered from 1999 to 2004; $n = 36$) for their inclusion in the study. The strains were also identified with a RT-PCR assay developed for characterizing the genomic sequence at locus 3CD, located 2229 nucleotides apart from the 1D locus in the E-6 genome. The 3CD RT-PCR assay was used according to a previously described method (Mirand et al., 2007) with a new set of oligonucleotide primers to obtain longer 3CD sequences. The primers had the following nucleotide sequences: HEVBS3C1, 5′GAC GGT GGG CCG TGT TGC CAC GIC AYG CIA ARC CYG GRC C3′ and HEVBR3D1, 5′GCT ATG AGA TGT CCG TCA AGC ATI ACI GGI ATY TTR CTC C3′ (R, A or G; Y, T or C; I, inosine). The 3CD sequences of the 52 E-30 strains were entirely determined with the new primers for their comparison with the E-6 sequences. A genomic segment of about 4000 nucleotides encompassing viral genes 1D, 2A, 2B, 2C, 3A, 3B, 3C and a 5′ portion of 3D, was determined in 14 selected E-6 strains using the above procedure, and the E-6 sequences were compared with the previously reported 13 E-30 sequences (Mirand et al., 2007).

All the sequences determined for the present study were deposited in GenBank, the Database of Japan and the EBI database under the accession numbers indicated in supplemental Table S1.

2.3. Data collection and compilation of nucleotide sequence data sets

The 1D gene sequences corresponding to the E-6 strains studied were compared with homologous sequences available in the international databases. GenBank entries corresponding to the 1D gene of an E-6 strain were retrieved from the National Centre for Biotechnology Information database (as of April 15, 2009). Only entries reporting a complete and dated nucleotide sequence were considered in the present study. The sequences are listed in supplemental table S2. Partial sequences were also used for phylogeographic analyses.

An alignment including sequences corresponding to the 3CD locus was constructed for a comparative analysis of the E-6 and E-30 strains. A number of sequences of three other enteroviruses available in GenBank were also included (supplemental table S2).

An alignment including subgenomic sequences determined in the E-6 strains, those of the E-30 strains, and other relevant sequences, was constructed following guiding principles (Higgins and Lemey, 2009). Briefly, an alignment of amino acid sequences was constructed with the MUSCLE program (default settings) and, after manual adjustments, the aligned amino acid sequences were back-translated to their original nucleotide sequences. All the sequence datasets were constructed with the BioEdit v.7.0.5 program (Hall, 1999).