Infection, Genetics and Evolution 9 (2009) 699–708

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Phylogeography of circulating populations of human echovirus 30 over 50 years: Nucleotide polymorphism and signature of purifying selection in the VP1 capsid protein gene

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

[a] Université d'Auvergne, Laboratoire de Virologie-EA3843, UFR Médecine, 28 place Henri-Dunant, F-63001 Clermont-Ferrand, France
[b] CHU Clermont-Ferrand, Laboratoire de Virologie, Centre de Biologie, F-63003 Clermont-Ferrand, France
[c] CHU Clermont-Ferrand, Laboratoire d'Hygiène hospitalière, Centre de Biologie, F-63003 Clermont-Ferrand, France

ARTICLE INFO

Article history:
Received 18 December 2007
Received in revised form 27 April 2008
Accepted 30 April 2008
Available online 28 May 2008

Keywords:
Picornavirus
Enterovirus
VP1 capsid protein
Purifying selection
Epidemics

ABSTRACT

A comprehensive set of 443 1D gene sequences (encoding the VP1 capsid protein) was analyzed to investigate the phylogenetic relationships and evolutionary patterns among strains of human echovirus 30 (E30; genus Enterovirus, family *Picornaviridae*) characterized over 50 years. Maximum-likelihood (ML) phylogenetic trees of complete and nonredundant 1D gene sequences (total length = 876 nucleotides) showed evidence of distinct lineages related to the isolation period of virus strains. Virus transportation was confirmed as a major epidemiological factor in the appearance of epidemics since recurrence of aseptic meningitis outbreaks in a given geographic area was associated with distinct E30 variants detected earlier in distant regions. Detection of the codon changes associated with E30 evolution was investigated with methods implemented in the Datamonkey web server. Evolution of the 1D gene was dominated by continual negative (purifying) selection against nonsynonymous substitutions at most codon sites, as determined by d*N*/d*S* ratio. Amino acid polymorphism was maintained at a limited number of sites (10/292) in the VP1 protein (within loops connecting β strands and C-terminus). Amino acid changes are allowed at these sites because they are likely exposed on the virion particle and nonsynonymous substitutions are observed in the corresponding codons because negative selection is relaxed.

1. Introduction

Echovirus 30 (E30) is one of 92 serotypes among the human enteroviruses (HEVs) belonging to the genus Enterovirus, in the family *Picornaviridae*. It is a major causative agent of aseptic meningitis, an epidemic disease of children and adults (Modlin, 2000; Peigue-Lafeuille et al., 2002). E30 infection is usually eradicated in a few days and, by analogy with the poliovirus infection, virus particles and virus infected cells are probably, cleared through a humoral response (Pallansch and Roos, 2007). However, the precise immune mechanisms involved in E30 infection have never been thoroughly investigated. Multiple outbreaks and nationwide epidemics have been reported worldwide over the last 15 years (see Bailly et al., 2000a, 2000b, 2002; CDC, 2003, 2006; Khetsuriani et al., 2006; Oberste et al., 1999; Savolainen et al., 2001), but it is not known whether the larger number of observations is the result of changes in medical practice, such as advances in molecular diagnostic procedures for detecting enteroviruses during infection of the central nervous system, or of epidemiological changes in the frequency of interindividual transmission.

Enteroviruses have a single-stranded, positive-sens RNA genome of about 7.5 kb in length. The genome contains a single open reading frame, which encodes the four capsid proteins VP1–VP4, within the P1 genomic region, and seven nonstructural proteins, regions P2 and P3 (Racaniello, 2007). Phylogenetic analysis of the 1D gene sequences showed that HEV serotypes are divided into four species (A–D) with E30 belonging to the HEV-B species (Stanway et al., 2005). The enterovirus capsids are made up of 60 copies of each of the four capsid proteins assembled into virions, which have an icosahedral structure, and only the VP1–VP3 proteins are exposed on the capsid surface (Racaniello, 2007). The structure of the E30 capsid has not been determined at the molecular level by X-ray crystallography and antigenic sites are

* Corresponding author. Tel.: +33 4 73 178 140; fax: +33 4 73 754 851.
E-mail address: j-luc.bailly@u-clermont1.fr (J.-L. Bailly).

1567-1348/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2008.04.009

not known. The antigenic variability of virus particles was determined in earlier E30 strains recovered in the 1960s but has not been investigated in strains circulating thereafter (Wenner et al., 1967).

Molecular epidemiology studies of E30 meningitis outbreaks show that point mutations give rise to substantial genetic diversity in the 1D gene encoding the VP1 capsid polypeptide (Yoshida et al., 1999; Bailly et al., 2002; Wang et al., 2002; Thoelen et al., 2003; Zhao et al., 2005; Chen et al., 2007; Mirand et al., 2007). However, no comprehensive phylogenetic investigation of the significance of genetic diversity patterns for E30 evolution has been performed. While molecular epidemiology investigations suggested that multiple virus variants exist among circulating strains of E30, which can be detected as distinct lineages in phylogenetic trees, each outbreak was usually associated with only one or a handful of virus variants. Genotypes bearing distinct amino acid variations possibly associated with different antigenic patterns have also been observed, but opinions differ on how they should be defined (Oberste et al., 1999; Savolainen et al., 2001; Palacios et al., 2002). Longitudinal molecular epidemiology studies of virus strains showed phylogenetic evidence of a rapid turnover of virus lineages. It was postulated that immunity of the general population could wipe out periodically occurring antigenic variants (Oberste et al., 1999; Savolainen et al., 2001; Palacios et al., 2002). In other studies, synonymous variations were reported as a main feature among recently circulating strains, which challenges the role of amino acid changes as a contributing factor in the epidemiology of E30 (Bailly et al., 2002; Thoelen et al., 2003).

To gain insight into the genetic polymorphism of E30 and into the factors that shape its diversity and phylogeography, we performed an extensive analysis of the evolutionary relationships of virus populations that have been circulating over the last 50 years. To achieve this, we assembled the largest dataset of viral isolates compiled to date for this enterovirus serotype, including samples taken from a wide range of geographical localities and over an extensive time-span. We elected to analyze a genomic segment that included the 1D gene encoding the VP1 capsid polypeptide because there was a large dataset available for this genomic sequence, which is widely used for enterovirus genotyping (Oberste et al., 2000). An additional interest of the 1D gene is the critical role of the VP1 protein in enterovirus immunity and infectivity: structural features of this protein are the primary antigens that induce protective immunity and others are involved in receptor binding (Minor, 1990; Mateu, 1995; Melnick, 1996; Racaniello, 2007).

We used different maximum likelihood methods because they are consistent with an analytical approach, to investigate phylogenetic relationships among 443 complete E30 1D gene sequences. We found phylogenetic evidence of eight lineages including virus strains closely related to the time period of their isolation over 50 years. We also obtained phylogenetic and statistically significant evidence that the VP1 protein is subject to extensive negative selection against amino acid changes and that positive selection might operate at only few distinct codon positions.

2. Materials and methods

2.1. E30 nucleotide sequences

More than 900 nucleotide sequences corresponding to the 1D gene of an E30 isolate were identified in the National Center for Biotechnology Information database and downloaded (as of October 15, 2007) into a local database including the E30 sequences determined in our laboratory ($n = 175$) that have been reported elsewhere (Bailly et al., 2000a, 2000b, 2002, Mirand et al., 2006, 2007, 2008). Sequences were excluded from the analysis if: (1) sequence length was lower than 30% of the complete 1D gene length (876 nucleotides encoding 292 amino acid residues), (2) multiple sequences were from the same individual, i.e. sequences obtained from different biological specimens (there was one exception, a group of highly divergent sequences from an immunodeficient patient that was conserved in the final alignment), (3) multiple sequences from different individuals were identical. A subset ($n = 420$) was formed by including only complete and nonredundant 1D gene sequences and was used for estimating phylogenetic and evolutionary features of E30.

Twenty three sequences were determined from virus isolates or cerebrospinal fluid specimens using methods described earlier (Mirand et al., 2006, 2008). Samples were obtained from different geographical regions in France ($n = 14$), Belgium ($n = 3$), Switzerland ($n = 2$), Tunisia ($n = 2$) and Bahrein ($n = 2$). All the datasets were constructed and edited with the BioEdit v.7.0.5 program (Hall, 1999). The sequences were deposited in the international sequence databases under the accession numbers AM946153–AM946175 and were included in the final dataset ($n = 443$) used for this study; the alignment is available on request.

2.2. Determination of the outgroup for the phylogenetic analysis of E30

A phylogenetic tree was reconstructed by the neighbor-joining (NJ; Saitou and Nei, 1987) method using the complete nucleotide sequences of the P1 genomic region, which includes the four capsid genes (1A–1D encoding capsid proteins VP1–VP4) for the 56 currently known serotypes that make up the HEV-B species. The maximum composite likelihood (MCL) method implemented in the MEGA v4.0 software (Tamura et al., 2007) was used to evaluate genetic distances, and 10,000 bootstrap replications were performed to estimate confidence values (P_B) of groups detected in the NJ tree (Felsenstein, 1985). The analysis showed that six serotypes were the most closely related to E30: echovirus 6 (E6), E21, E25, E29, enterovirus 73 (EV73) and EV83. A similarity analysis was then carried out to compare the 1D gene sequences of the six serotypes and that of E30, in order to determine a suitable outgroup (data not shown). The comparisons including E6, E25, EV73 and EV83 showed a lower nucleotide similarity at the 5′ and 3′ ends of the 1D gene and in several stretches scattered throughout the sequence. In contrast, comparisons performed with E21 and E29 showed a higher conservation of nucleotide similarity. E21 was chosen to infer the root of the E30 phylogenetic tree because the 1D gene sequences of the two serotypes were aligned unambiguously without indels.

2.3. Phylogenetic analysis with the maximum likelihood (ML) method

Phylogenetic relationships between sequences were inferred with an ML method using different computer programs. E30 phylogeny was first assessed using PAUP* v4.0b10 (Swofford, 2003) and substitution model settings determined from the hierarchical likelihood ratio test (hLRT) and the Akaike information criterion (AIC) implemented in the Modeltest v3.1 program (Posada and Crandall, 1998; Posada and Buckley, 2004). The model selected for phylogenetic analyses was TrN + I + G with hLRT ($p < 0.01$; $-\ln L = 25{,}769$) and GTR + I + G with AIC ($-\ln L = 25{,}718$). Base frequencies, transition/transversion ratio (R), the proportion of invariant sites (pINV) and the alpha parameter of the gamma rate distribution were assessed with the model selected, i.e. GTR + I + G. An initial NJ tree was used to perform the ML analysis and the tree