A Sabin 3-Derived Poliovirus Recombinant Contained a Sequence Homologous with Indigenous Human Enterovirus Species C in the Viral Polymerase Coding Region†

Minetaro Arita,1* Shuang-Li Zhu,2 Hiromu Yoshida,1 Tetsuo Yoneyama,1 Tatsuo Miyamura,1 and Hiroyuki Shimizu1

Department of Virology II, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama-shi, Tokyo 208-0011, Japan, and National Reference Laboratory of Poliomyelitis, Chinese Center for Disease Control and Prevention, Beijing, China2

Received 14 April 2005/Accepted 20 July 2005

Outbreaks of poliomyelitis caused by circulating vaccine-derived polioviruses (cVDPVs) have been reported in areas where indigenous wild polioviruses (PVs) were eliminated by vaccination. Most of these cVDPVs contained unidentified sequences in the nonstructural protein coding region which were considered to be derived from human enterovirus species C (HEV-C) by recombination. In this study, we report isolation of a Sabin 3-derived PV recombinant (Cambodia-02) from an acute flaccid paralysis (AFP) case in Cambodia in 2002. We attempted to identify the putative recombination counterpart of Cambodia-02 by sequence analysis of nonpolio enterovirus isolates from AFP cases in Cambodia from 1999 to 2003. Based on the previously estimated evolution rates of PVs, the recombination event resulting in Cambodia-02 was estimated to have occurred within 6 months after the administration of oral PV vaccine (99.3% nucleotide identity in VP1 region). The 2BC and the 3D\textsuperscript{pol} coding regions of Cambodia-02 were grouped into the genetic cluster of indigenous coxsackie A virus type 17 (CAV17) (the highest [87.1%] nucleotide identity) and the cluster of indigenous CAV13-CAV18 (the highest [94.9%] nucleotide identity) by the phylogenetic analysis of the HEV-C isolates in 2002, respectively. CAV13-CAV18 and CAV17 were the dominant HEV-C serotypes in 2002 but not in 2001 and in 2003. We found a putative recombination between CAV13-CAV18 and CAV17 in the 3D\textsuperscript{pol} coding region of a CAV17 isolate. These results suggested that a part of the 3D\textsuperscript{pol} coding region of PV3(Cambodia-02) was derived from a HEV-C strain genetically related to indigenous CAV13-CAV18 strains in 2002 in Cambodia.

Poliovirus (PV) is a small nonenveloped virus with a single-strand positive genomic RNA of about 7,500 nucleotides (nt) belonging to the family Picornaviridae, known as the causative agent of poliomyelitis. Currently, the global eradication program for poliomyelitis is continuing by utilizing both inactivated and live attenuated vaccines (44, 46). The endemicity of indigenous wild PVs was confirmed to be restricted to Afghanistan, Egypt, India, Niger, Nigeria, and Pakistan as of 2004 (http://www.polioeradication.org/progress.asp).

The Sabin strains (Sabin 1, 2, and 3) are attenuated PV strains and have been widely used as live oral PV vaccine (OPV) (44). Following the administration of OPV, the viruses infect the mucosal tissues and are commonly excreted for 3 to 7 weeks from immunocompetent individuals (1, 18) and occasionally for 10 to 22 years from immunodeficient patients (2, 25, 32; reviewed in reference 48). During the replication of the Sabin strains, revertants with increased virulence could emerge and cause vaccine-associated paralytic poliomyelitis in rare cases. The rate of vaccine-associated paralytic poliomyelitis has been estimated as one case per 520,000 doses associated with the first dose of OPV (35). The revertants have been isolated from healthy individuals and also from the environment (21, 52).

Recently, outbreaks of poliomyelitis caused by circulating vaccine-derived PV (cVDPV) have been reported in Egypt, Hispaniola, the Philippines, and Madagascar (6, 8, 10, 24, 51). Sequence analysis of the genomes of cVDPVs showed unidentified sequences in the nonstructural protein coding region. These sequences are considered to be derived from recombination with unidentified nonpolio enterovirus (NPEV) during the circulation of VDPVs for 1 to 10 years (6, 8, 10, 24, 49, 51). However, a highly evolved derivative of Sabin strains without recombination by an unidentified counterpart has been isolated from an acute flaccid paralysis (AFP) case after a long-term circulation (12). Therefore, the biological role of the recombination of cVDPVs with unidentified counterpart remains to be elucidated. At present, increased transmissibility of cVDPVs compared with that of the parental Sabin strains has been proposed as a result of the recombination (3, 16); however, no virological evidence has been provided so far.

Indigenous wild PVs have been eliminated in regions where cVDPVs have been reported (1991 in the Americas [42], 1993 in the Philippines [11], and 1998 in Madagascar [41]) except Egypt. Therefore, the field NPEVs genetically closely related to PV or highly mutated Sabin derivatives are considered the possible counterparts of the recombination. Among NPEVs, coxsackie A viruses (CAVs) belonging to human enterovirus species C (HEV-C) are the suspected origin of the recombination because of the higher similarity of the genomic se-

---

* Corresponding author. Mailing address: Department of Virology II, National Institute of Infectious Diseases, 4-7-1 Gakuen, Musashimurayama-shi, Tokyo 208-0011, Japan. Phone: 81-42-561-0771. Fax: 81-42-561-4729. E-mail: minetaro@nih.go.jp.
† Supplemental material for this article may be found at http://jvi.asm.org.
Isolation of a Sabin 3-derived Poliovirus Recombinant

Vol. 79, 2005

Isolation of a Sabin 3-derived poliovirus recombinant

12651

MATERIALS AND METHODS

Cells and viruses. RD cells (derived from human rhabdomyosarcoma), HEp-2c cells (derived from human larynx epidermoid carcinoma) and L20B cells (derived from mouse L tk− aprt− fibroblast) were cultured as monolayers in Eagle's minimum essential medium supplemented with 2% fetal calf serum (33, 50). RD, HEp-2c, and L20B cells were used for the virus isolation from fecal samples of AFP cases. Virus stocks were stored at −70°C.

Sequence analysis of the genomes of enterovirus isolates. Viral genomic RNA was isolated from the culture fluid of infected cells by using a High Pure viral RNA purification kit (Roche). DNA fragments used for the DNA sequencing were prepared by reverse transcription-PCR (RT-PCR) using the viral genomic RNA as the template by use of a Titan one-tube RT-PCR system (Roche). PCR products were purified by using a QIAquick PCR purification kit (QIAGEN). DNA sequencing was performed using a BigDye Terminator v3.0 cycle sequencing ready reaction kit (Applied Biosystems), and then sequences were analyzed by use of an ABI PRISM 3100 genetic analyzer (Applied Biosystems). The sequences of the 5′ end of the viral genomes were determined by the 5′ rapid amplification of cDNA ends method by using a 5′ rapid amplification of cDNA ends system, version 2.0 (Invitrogen), according to the manufacturer’s instructions. The sequence of the 3′ end of the viral genome was determined from an RT-PCR product obtained with UG16 primer (20) and EcoRI-3END− (Table 1). The percentage of the mutated synonymous sites among all synonymous sites (Ks) was calculated for the VP1 coding region as previously reported (2, 12, 17). Phylogenetic trees were constructed by the neighbor-joining method after bootstrapping 1,000 times (14, 45) using PHYLIP software (Joseph Felsenstein 1990, version 3.6c; University of Washington). The nucleotide substitutions among the isolates were estimated 1,000 times (14, 45) using PHYLIP software (Joseph Felsenstein 1990, version 3.6c; University of Washington). The nucleotide substitutions among the isolates were estimated by the Kimura-2 parameter method (26). The rate of transition-transversion was set at 2.0. Similarity plot analysis of HEV-C isolates was performed by using SimPlot (29).

Primers used for the sequence analysis are listed on Table 1. Primers UG16 and UC12 were used for the analysis of a part of the 3Dpol coding region (20). Primers EVP4 and OL68-1 were used for the analysis of the VP4 coding region (39, 43). Primers 2A2+ and 2C− were designed and used for the analysis of a part of the 2BC coding region. Primers 292 and 222 were used for the initial analysis of the VP1 coding region (37). Genomic sequences used for the phylogenetic analysis were as follows: 207 nt of the VP4 coding region (corresponding to nt 743 to 949 of the Sabin 3 genome), 337 nt of the 2BC coding region (corresponding to nt 3854 to 4190 of the Sabin 3 genome), and 352 nt of the 3Dpol coding region (corresponding to nt 6137 to 6488 of the Sabin 3 genome). Identification of NPEV isolates. A panel of horse antisera against commonly found NPEVs (RIVM, Bithoven, The Netherlands), which include echo and coxsackie B viruses, was used for the identification of HEV-B. Antisera against CAVs were purchased from the American Type Culture Collection. A total of 100 50% cell culture infectious doses of enterovirus isolates were incubated with 20 units of antisera for 2 h at 35°C, and then HEp-2c cell or RD cell suspensions in 10% fetal calf serum–minimum essential medium were added and incubated at 35.5°C (50). Inoculated cells were observed for cytopathic effect until 24 h after the complete appearance of cytopathic effect in the cells inoculated with the isolates in the absence of antisera.

Accession numbers of the nucleotide sequences. All the nucleotide sequences determined in this study were submitted to the DNA Data Bank of Japan (DDBJ). The GenBank/EMBL/DDBJ accession numbers of each sequence were as follows. The accession numbers of the VP4 coding region, the 2BC coding region, and the 3Dpol coding region of the NPEV isolates are AB206334 to AB206380, AB206709 to AB206757, and AB205529 to AB205546, respectively (see Fig. 2, 3, and 5 and the supplemental material). The accession numbers of the VP1 coding region of CAM1952, CAM2033, CAM2038, and CAM2083 are AB207264, AB207263, AB207265, and AB207266, respectively (Table 1). The accession numbers of the genomic sequences of Cambodia-02, CAM1900, CAM1972, CAM2069, and CAM2101 are AB205395, AB205397, AB205396, and AB205398, respectively.

RESULTS

Isolation of a type 3 PV recombinant from an AFP case. In 2002, type 3 PVs were isolated from three AFP cases in Cambodia. These PV isolates were initially characterized by sequencing of the VP1 coding region, and all the isolates were classified as OPV-like PVs according to the criteria of the World Health Organization (less than 1% nucleotide difference from the parental Sabin 3) (50). However, we found that one of these PV isolates (Cambodia-02) contained an unidentified sequence in the 3Dpol coding region which was apparently not related to those of the Sabin strains (Fig. 1). Further sequence analysis of the genome of Cambodia-02 showed that the 5′ part of the genome (from nt 1 to 3777), including the 5′ nontranslated region (5′NTR), the structural protein coding region, and a part of the 2Apro coding region, was derived from Sabin 3 followed by an unidentified sequence from the 2Apro coding region to the 3′ end of the genome (from nt 3778 to the 3′ end) (Fig. 1B and C). The nonstructural protein coding region of Cambodia-02 showed only low similarity with those of Sabin strains (Fig. 1B).
The nucleotide identity of the 5′ part of the Cambodia-02 genome to Sabin 3 was 99.5%. The $K_s$ value of Cambodia-02 calculated for the VP1 coding region was $1.35 \times 10^{-2}$ (with a standard error of $0.77 \times 10^{-2}$). Using evolution rates of PV observed for immunodeficiency cases ($2.85 \times 10^{-2}$ to $3.28 \times 10^{-2}$ synonymous substitutions per synonymous site per year) or for transmission of wild PV recombinants ($3.45 \times 10^{-2}$ synonymous substitutions per synonymous site per year) (2, 17, 28), we estimated that Cambodia-02 was isolated within 6 months after the administration of OPV. Cambodia-02 had reversions at the major attenuation determinants of Sabin 3 at nt 472 (U to C) and nt 2034, which resulted in an amino acid change of VP3 Phe91 to Ser (13, 31) (Fig. 1D). The Cambodia-02 genome contained multiple mutations in the structural protein coding region in addition to VP3 Phe91, as previously reported for temperature-resistant revertants of Sabin 3 (15, 31, 34).

**Isolation and identification of HEV-C from AFP cases in Cambodia.** We analyzed the genome of NPEV isolates from AFP cases around 2002 in Cambodia to identify the putative recombination counterpart of Cambodia-02. In 2002, we isolated NPEVs from 53 AFP cases (one was from a mixed case with PV) among a total of 155 AFP cases (Table 2). For the initial molecular typing of the isolates, we analyzed the VP4 coding region (nt 743 to 949 of the Sabin 3 genome; 207 nt) to classify the isolates into each genomic species (HEV-A, HEV-B, and HEV-C) (23). We found that 21 isolates were grouped into HEV-C by the phylogenetic tree analysis of the sequence of the VP4 coding region (data not shown). We identified the serotype of HEV-C isolates by a neutralization assay using type-specific antisera or by sequence analysis of the VP1 coding region. We could not discriminate CAV13 from CAV18 or CAV11 from CAV15 by the sequence analysis or by the neutralization assay, consistent with previous reports (4,
36). The deduced amino acid sequence of the VP1 protein of CAM2033 and CAM2038 showed a high nucleotide identity with those of CAV17 (94.1%) and CAV11-CAV15 (96.7%), respectively. We could not identify the serotype of CAM2083 from the deduced amino acid sequence of the VP1 protein, which showed low similarity with known enteroviruses. We observed the highest amino acid identity only with CAV24 (DN-19 strain) (74.1%) or with a CAV24 variant (73.1%). Consequently, HEV-C isolates in Cambodia in 2002 consisted of CAV1, CAV11-CAV15, CAV17, CAV13-CAV18, CAV20, CAV24, and an untypable HEV-C strain CAM2083 (Table 3).

Sequence analysis of HEV-C isolates in the 3Dpol coding region. We then analyzed the genomic sequence in the 3D pol coding region of the HEV-C isolates. The phylogenetic analysis of a part of the 3Dpol coding region (corresponding to a region of nt 6137 to 6488 of the Sabin 3 genome; 352 nt) showed that the isolates formed distinct genetic clusters from those of the prototype HEV-C strains, as observed for the sequence analysis of the VP4 coding region (Fig. 2). The phylogenetic analysis of the 3Dpol coding region failed to show a clear relationship between the serotypes of isolates and their genetic clusters. For CAM1974, CAM2083, and CAM2091, we could not obtain the corresponding DNA fragment by RT-PCR. In the phylogenetic analysis, a genetic cluster of indigenous CAV13-CAV18 strains was the closest to Cambodia-02. We found that a CAV13-CAV18 isolate (CAM2101) showed the highest nucleotide identity (94.0%) to Cambodia-02 among the HEV-C isolates. We further analyzed the nonstructural protein coding region of CAM2101 and found that CAM1900 showed a high identity to Cambodia-02 only in the 3Dpol coding region (94.9%) but not in other regions (Fig. 3A and C).

In the phylogenetic analysis of the 3Dpol coding region, we found that a CAV17 isolate, CAM2069, was located apart from other CAV17 isolates in a genetic cluster of indigenous CAV13-CAV18 strains. The sequence analysis of CAM2069 showed that CAM2069 exhibited a high (95.5 to 97.7%) nucleotide identity to CAM2101 (a CAV17 isolate) for the 2C and the 3AB coding regions. However, the 3CDpro coding region of CAM2069 showed high (93.3 to 95.5%) nucleotide identity to that of CAV13-CAV18, as observed in the phylogenetic analysis of the 3Dpol coding region (Fig. 2) and in the similarity plot analysis (Fig. 3C). This observation suggested that recombination between CAV13-CAV18 and CAV17 could occur at least in the 3CDpro coding region.
Sequence analysis of HEV-C isolates in the 2BC coding region. Next, we analyzed the sequence of another nonstructural protein coding region of HEV-C isolates, because the analysis in the 3Dpol coding region failed to identify the recombination counterpart. For this purpose, we designed a new primer set for RT-PCR and DNA sequencing, 2A2+ and 2C−, in the 2A sno coding region and in a cis-acting replication element in the 2C coding region (19, 40), respectively. By using this primer set, we analyzed a sequence of the 2BC coding region (corresponding to nt 3854 to 4190 of the Sabin 3 genome; 337 nt) for all the NPEV isolates (Fig. 4). We then analyzed the sequence of three isolates (CAM1920, CAM1936, and CAM2034) for which we could not analyze the sequence of the VP4 coding region. However, one isolate (CAM1952, a CAV24 strain) failed to give an RT-PCR product. In the phylogenetic analysis of the 2BC coding region, we observed a close relationship between the serotypes of isolates and the genetic clusters. In this phylogenetic analysis, Cambodia-02 was again located close to the genetic clusters of indigenous HEV-C but not to those of the HEV-C prototypes. CAM2101 (a CAV17 isolate) showed the highest nucleotide identity to Cambodia-02 in the 2BC coding region; however, the identity was not significantly high (86.9%).

HEV-C isolates from AFP cases in Cambodia from 1999 to 2003. To identify the recombination counterpart of Cambodia-02, we further analyzed the NPEV isolates from 1999 to 2003 in Cambodia. HEV-C was a dominant HEV species isolated from AFP cases in this period, suggesting that the prevalence

---

**FIG. 3.** Alignment of the genomes of Cambodian HEV-C isolates. The numbers in each region represent the percentages of nucleotide identity, and the numbers in parentheses represent the percentages of amino acid identity. The genomic regions that showed more than 90% amino acid identity are colored with light gray, and the genomic regions that showed more than 92% nucleotide identity are colored with dark gray. (A) Alignment of Cambodia-02 with a CAV13-CAV18 isolate (CAM1900). (B) Alignment of a CAV17 isolate (CAM2069) with CAM1900 (CAV13-CAV18) and CAM2101 (CAV17). (C) Multisequence analysis of HEV-C isolates and Cambodia-02 by similarity plot analysis calculated by SimPlot. CAM1900 was used as the reference. A window size of 200 bp with an increment of 20 bp was used. The locations of 3Cpro and 3Dpol coding regions are shown in the plot. (D) Alignment of a part of the genome of CAM1900 with that of CAM2069 around the putative recombination junction near the 3C pro coding region. The part representing unidentified sequence is colored with gray. NS; nonstructural protein, ND; not determined.
The dominant serotypes of HEV-C were different from year to year; in 2002, they were CAV17 and CAV13-CAV18. In the phylogenetic analysis of the 2BC coding region, Cambodia-02 was grouped into a cluster of the indigenous CAV15-CAV17-CAV20 isolates. However, the nucleotide identity was not significantly high, and the highest (90.1%) was found with a CAV17 isolate in 2000 (data not shown). Therefore, we could not identify the exact recombination counterpart among the HEV-C isolates examined. However, these results suggested that the recombination counterpart of Cambodia-02 was genetically closely related to the indigenous HEV-C strains in Cambodia.

**DISCUSSION**

A Sabin 3-derived PV recombinant (Cambodia-02) analyzed in this study was isolated from an AFP case in Cambodia in 2002. Cambodia-02 was classified as OPV-like PV by sequence analysis in the VPI coding region. In 2002, we isolated three type 3 PVs, including Cambodia-02, from AFP cases in Cambodia, but these PV isolates were genetically unrelated to each other (data not shown). Therefore, this evidence suggested that Cambodia-02 was isolated from a sporadic AFP case and did not result from a circulating strain.

Cambodia-02 was isolated within 6 months after the administration of OPV, as determined on the basis of the previous estimation of the evolution rate of PVs (2, 17). Thus, the unidentified sequence of Cambodia-02 should have retained the original genetic feature of the recombination counterpart. The last indigenous PV case in Cambodia was reported in 1997 (7); therefore, we examined HEV-C strains for the recombination counterpart. We examined HEV-C isolates from AFP cases in Cambodia; however, these isolates represented only a minor population of circulating HEV-C strains that would mostly result in asymptomatic infection. Therefore, through this strategy, we could expect to find some HEV-C strains that were only related to the recombination counterpart. We found that HEV-C was a dominant enterovirus species that could be isolated from the AFP cases in Cambodia (Table 2). The HEV-C isolates consisted of CAV1, CAV11-CAV15, CAV13-CAV18, CAV17, CAV20, CAV21, CAV24, and an untypable serotype represented by CAM2083 (Table 3). HEV-C strains have not been isolated as major NPEVs through the established enterovirus surveillance systems (5, 9) (http://idsc.nih.go.jp/las/promp/circle-g/meningi/menin.html [in Japanese]). However, recently, a high frequency of HEV-C isolation (~50% of the isolates) was reported in Madagascar, where type 2 cVDPVs emerged in 2002 (10, 41). Therefore, HEV-C might be a dominant HEV species among the circulating enteroviruses in tropical areas. However, the prevalence of HEV-C in other tropical areas remained to be further investigated.

We performed a comprehensive sequence analysis of HEV-C isolates in three different genomic regions, including the 2BC coding region and the 3D<sup>pol</sup> coding region (20, 39, 43) (Fig. 2 and 4). We designed a new primer set, 2A<sup>2+</sup> and 2C<sup>−</sup>, for the analysis of the 2BC coding region (Table 1). The 2A<sup>2+</sup> primer was designed in the 2A pro coding region, and the 2C<sup>−</sup> primer was designed in a cis-acting replication element of the enterovirus (19, 40). This primer set showed a broad spectrum of applicability for HEV-A, HEV-B, and HEV-C isolates, and sequence analysis using this primer set failed for only one isolate (CAM1952 [CAV24]) among 216 NPEV isolates. Recombination junctions of cVDPV have been identified in the 2AB coding region (6, 8, 10, 24). Therefore, with a wide spectrum of applicability for NPEV, this primer set would serve as a useful tool to identify HEV-C strains related to the recombination counterpart of cVDPV.

From the sequence analysis, we found that the nonstructural protein coding regions of Cambodia-02 were grouped into the genetic clusters of the indigenous CAV17 and CAV13-CAV18 strains in Cambodia, distinct from those of the HEV-C prototype strains (Fig. 2 and 4). We isolated CAV17 from 2001 to 2002 and CAV13-CAV18 from 2002 to 2003. Thus, both CAV17 and CAV13-CAV18 were highly prevalent and could be available as the recombination counterpart of Cambodia-02 in 2002. One of the CAV13-CAV18 isolates (CAM1900) showed the highest (94.9%) nucleotide identity to Cambo-
diation in the 3Dpol coding region (Fig. 4). The nucleotide identities in the 3Dpol coding region among isolates with the same HEV-C serotype isolated in 2002 were 95 to 96% (Fig. 3), and Cambodia-02 showed a nucleotide identity comparable to indigenous HEV-C isolates in 2002. This suggested that these HEV-C isolates had evolved independently before the putative epidemic in 2002. Interestingly, we found a putative CAV recombinant (CAM269) in the 3CDpol coding region, suggesting frequent interserotypic recombinant formation between CAV17 and CAV13-CAV18 in this region (Fig. 4). Actually, the 3CDpol coding region showed high similarities among CAV13-CAV18, CAV11-CAV15, CAV17, and CAV20 (4, 22). This was reminiscent of frequent recombination among HEV-B strains and also among human rhinoviruses where the 3Dpol coding regions of the same serotype were not monophyletic (27, 30, 38, 47). However, we could not find the exact recombination counterpart among the indigenous HEV-C isolates from AFP cases, although their nonstructural protein coding region was closely related to that of Cambodia-02.

In summary, we isolated a type 3 PV recombinant from an AFP case in Cambodia and suggested that the unidentified sequence of the recombinant was derived from an indigenous HEV-C strain in Cambodia. In addition to the poor population immunity, the high prevalence of HEV-C would be another critical factor for the emergence and/or the evolution of cVDPV. The biological roles of the recombination of cVDPV remain to be further studied.

ACKNOWLEDGMENTS

We thank Keith Feldon and Cambodian local and regional EPI staffs for their expert surveillance. We are grateful to Junko Wada for her excellent assistance. This work was supported by grants-in-aid for “Promotion of Polio Eradication” from the Ministry of Health, Labor and Welfare, Japan, and by a grant for health research from the regional office for the Western Pacific, World Health Organization. H.S. was supported in part by grants-in-aid for “Development of Expanded Programme on Immunization and Accelerating Measles Control in the Polio-free Era” from the Ministry of Health, Labor and Welfare, Japan.

REFERENCES