Genetic analysis of the VP4/VP2 coding region in human rhinovirus species C in patients with acute respiratory infection in Japan

Naoko Kiyota,1 Izumi Kushibuchi,2 Miho Kobayashi,3 Hiroyuki Tsukagoshi,3 Akihide Ryo,4 Koichi Nishimura,1 Asumi Hirata-Saito,2 Seiya Harada,1 Mika Arakawa,2 Kunihisa Kozawa,3 Masahiro Noda5 and Hirokazu Kimura5

1Kumamoto Prefectural Institute of Public Health and Environmental Sciences, Uto-shi, Kumamoto 869-0425, Japan
2Tochigi Prefectural Institute of Public Health, Utsunomiya-shi, Tochigi 329-1196, Japan
3Gunma Prefectural Institute of Public Health and Environmental Sciences, Maebashi-shi, Gunma 371-0052, Japan
4Department of Molecular Biodefence Research, Yokohama City University Graduate School of Medicine, Yokohama-shi, Kanagawa 236-0004, Japan
5Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan

Detailed genetic analysis was carried out of the VP4/VP2 coding region in human rhinovirus species C (HRV-C) strains detected in patients with acute respiratory infection in Japan. Phylogenetic trees were constructed by the neighbour-joining (NJ) and maximum-likelihood (ML) methods. The NJ phylogenetic tree assigned 11 genotypes to the present strains, whilst the ML tree showed that the strains diversified sometime in the early 1870s. Moreover, the pairwise distance among the present strains was relatively long, and the rate of molecular evolution of the coding region was rapid (3.07×10^{-3} substitutions per site per year). The results suggest that the present HRV-C strains have a wide genetic divergence and a unique evolutionary timescale.

INTRODUCTION

Human rhinoviruses (HRVs) belong to the genus Enterovirus (family Picornaviridae) and cause various acute respiratory infections (ARIs) such as the common cold, bronchiolitis and pneumonia (Turner & Couch, 2007). There are three species of HRVs – HRV-A, HRV-B and HRV-C – with HRV-C being the most recently discovered (Lamson et al., 2006; McErlean et al., 2007; Turner & Couch, 2007). Whilst HRV-A and -B are culturable using human fibroblasts, HRV-C cannot be isolated by conventional cell-culture methods. However, a recently developed organ-culture method now enables HRV-C to be propagated (Bochkov et al., 2011), and it has been found that HRV-C may, in fact, be prevalent, as well as HRV-A in Japan. HRV-C may be associated with various ARIs, such as upper respiratory infection (URI), bronchiolitis and pneumonia (Arakawa et al., 2012; Smuts et al., 2011; Watanabe et al., 2010), as well as with other diseases such as virus-induced asthma in many countries (Fujitsuka et al., 2011; Linsuwanon et al., 2009; Smuts et al., 2011).

HRV-C has been shown to have a wide genetic divergence (Arakawa et al., 2012; Wisdom et al., 2009). Indeed, our previous report suggested that the genomes of HRV-C detected in Japanese subjects with various ARIs showed >30% divergence based on sequences in the VP4/VP2 coding region, and these strains could be classified into many genotypes by phylogenetic analysis (Arakawa et al., 2012). Furthermore, studies of the VP4/VP2 coding region have allowed detailed phylogenetic analysis of the genotypes of HRV-A to -C (Arakawa et al., 2012; Cuevas et al., 2012; McIntyre et al., 2010; Simmonds et al., 2010). It appears, then, that this region could be useful for the detailed genetic analysis of HRV-C, although the evolution of the region is

Abbreviations: AdV, adenovirus; ARI, acute respiratory infection; CI, confidence interval; DR, different rate; EV, enterovirus; FEL, fixed effects likelihood; HBoV, human bocavirus; HMPV, human metapneumovirus; HPV, human parainfluenza virus; HRV, human rhinovirus; IFEL, internal fixed effects likelihood; LRI, lower respiratory infection; LRT, likelihood ratio test; ML, maximum-likelihood; NJ, neighbour joining; NPS, nasopharyngeal swab; REL, random effects likelihood; RSV, respiratory syncytial virus; SLAC, single likelihood ancestor counting; SR, single rate; SRDT, single rate dated tips; URI, upper respiratory infection.

The GenBank/EMBL/DDBJ accession numbers for the sequences of the HRV-C strains analysed in this study are AB731139–AB731157.
not fully understood. The neighbour-joining (NJ) method is frequently used in phylogenetic analysis to examine the molecular epidemiology of various viral genomes (Kimura, 1980; Saitou & Nei, 1987). This method is based on a cluster classification algorithm that allows cluster analysis and reveals the rate of viral evolution. Another technique, the maximum-likelihood (ML) method, enables the evolutionary timescale to be estimated (Felsenstein, 1981; Mizuta et al., 2011). In this study, using these two methods, we conducted a detailed genetic analysis of the VP4/VP2 coding region of HRV-C strains from patients with ARIs in Japan.

**METHODS**

**Samples and patients.** Nasopharyngeal swabs (NPSs; n = 1337) and eight tracheal aspiration samples were collected from patients aged 0–96 years (3.0 ± 2.8 years; median ± quartile deviation) with ARIs. Patients were diagnosed mainly with URI and lower respiratory infection (LRI; bronchitis, bronchiolitis and pneumonia). Samples were obtained by the local health authorities of Tochigi (central area of Japan: 607 NPSs and three tracheal aspiration samples collected) and Kumamoto (southern area: 730 NPSs and five tracheal aspiration samples collected) prefectures for the surveillance of viral diseases in Japan between April 2009 and December 2011. A distance of 1000 km separates each area. Informed consent was obtained from the subjects, or from the parents of underage subjects, for sample donation. Primarily, we used cell-culture methods to isolate various viruses (using VerO E6, RD-185, HEP-2 and Madin–Darby canine kidney cells). In addition, we attempted to isolate and/or detect respiratory bacteria such as Haemophilus influenzae, Legionella pneumophila, Neisseria species and Bordetella pertussis by culture and Streptococcus species using a kit (Bisno et al., 1997). We also detected Mycoplasma pneumoniae by PCR (Nadal et al., 2001). However, none of these bacteria was isolated or detected in any of the tested samples.

**RNA extraction, RT-PCR and sequencing.** Viral RNA was extracted from clinical samples using a QIAamp Viral RNA Mini kit (Qiagen). Reverse transcription was performed using ReverTra Ace reverse transcriptase (Toyobo) according to the manufacturer’s instructions. RT-PCR of the HRV VP4/VP2 coding region was carried out using the following primers: forward primer EvP4 (5′-CTACTTGTGGTGTCGGTT-3′) and reverse primer OLE8-1 (5′- GTGAAAATTTCACACACANCGC-3′). PCR was initiated at 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min, with a final extension at 72 °C for 10 min (Ishiko et al., 2002). The PCR product was purified with a Wizard SV Gel and PCR Clean-Up System (Promega). The purified products were sequenced using a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), using the two primer sets above. Sequence analysis was performed on an ABI 3500 Genetic Analyzer (Applied Biosystems). We attempted to detect other respiratory viruses such as influenza virus, human parainfluenza virus (HPIV), adenovirus (AdV), respiratory syncytial virus (RSV), human metapneumovirus (HMPV), enterovirus (EV) and human bocavirus (HBoV) as described previously (Allander et al., 2005; Echevarria et al., 1998; Ishiko et al., 2002; Matsuzaki et al., 2009; Nakauchi et al., 2011; Sullender et al., 1993; Xu et al., 2000; Zhang & Evans, 1991).

**Phylogenetic analysis with the NJ method.** For genotypic assignment of the HRV-strains – the 19 HRV-C strains analysed in this study and 28 reference strains (Simmonds et al., 2010) – phylogenetic analysis using the NJ method was performed as described previously (Arakawa et al., 2012). Evolutionary distances were estimated according to Kimura’s two-parameter method, and the reliability of the tree was estimated with 1000 bootstrap replications (Kimura, 1980; Saitou & Nei, 1987).

**Phylogenetic analysis, estimation of timescale, and datasetting of the ML method.** To construct the phylogenetic tree by the ML method, which is the best nucleotide substitution model, the general time reversible with gamma-distributed rates across sites (GTR + Γ) (Yang, 1994a, b) was selected using the KAKUSAN4 program v4.0. We used three phylogenetic models of sequence evolution. One is the different rate (DR) model (Felsenstein, 1981), the most general, which assumes that each branch of the unrooted phylogenetic tree has a different substitution rate (Rambaut, 2000). The DR model was used as a suitable general model against which to test the assumption of constant rates of the fit of the single rate (SR) and the single rate dated tips (SRDT) models (Goldman, 1993). Phylogenetic analysis of ML based on the DR model was constructed using RAXML BlackBox (http://phylobench.vital-it.ch/raxml-bb/index.php). The reliability of the phylogenetic hypothesis was assessed using bootstrap analysis of 100 ML iterations. The SR model assumes the same rate of evolution in all branches (i.e. a molecular clock), and the SRDT model is an SR model that relaxes the assumption of contemporaneous sequences and uses the date of isolation of each sequence to estimate the substitution rate (Rambaut, 2000). To estimate the rate of molecular evolution (and hence a timescale) for a phylogeny consisting of dated tips, phylogenetic analyses by ML based on the SR and SRDT models were performed using TipDate (http://mobyle.pasteur.fr/cgi-bin/portal.py?forms::tipdate) (Rambaut, 2000). The likelihood of the SR and SRDT models (with likelihood L0) was compared with that of the DR model (with likelihood L1) in a likelihood ratio test (LRT; Felsenstein, 1981) of the fit of the model. The test statistic is the difference in the log-likelihood (Δ) between the SR or SRDT model and the DR model. In the LRT statistic, twice the ratio of log-likelihoods (L0/L1) (2Δ*) is expected to be χ2 distributed with the degrees of freedom equal to those in their models. The DR model has 2n – 3 free parameters, the SR model has n – 1 and the SRDT model has n – 2 (a tree of n tips). In the LRT of the fit of the SR and SRDT models compared with the DR, if the SR model is rejected in favour of the DR model but the SRDT model is not, the SRDT model can be accepted as no worse a description of the evolution date than the DR model. A P value of ≥ 0.01 was considered statistically significant for the phylogenetic models. The resulting phylogenetic trees were described using TreeExplorer v2.12. The rates of nucleotide substitution, the estimated date of the root of the tree and the corresponding upper and lower 95% confidence intervals (CIs) were calculated under the SRDT model using TipDate. In general, it is difficult to apply the ML method to examine >50 strains (Yang, 2006). Taking these limitations into account, we screened gene sequences using the datasets to enable us to analyse the present HRV-C strains by the ML method. As a result, 19 HRV-C present strains and 13 HRV-C reference strains were applicable to the ML method in the present study.

**Calculation of pairwise distances.** To assign genotypes to each of the present strains, we used the 28 HRV-C reference strains (Arakawa et al., 2012; Simmonds et al., 2010). In addition, to assess interspecies frequency distributions of HRV-C, we calculated pairwise distances for all of the HRV-C strains, both present strains and reference strains, as described previously (Arakawa et al., 2012; Mizuta et al., 2010).

**Positive pressure analysis.** To evaluate the action of selective pressure on the VP4/VP2 coding region across all HRV-C strains, we estimated the rates of synonymous (dS) and non-synonymous (dN) changes at amino acid sites by conservative single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), internal fixed effects likelihood (IFEL) and the random effects likelihood (REL) method using DATAMONKEY (http://www.datamonkey.org) (Pond & Frost, 2012).
conducted genetic analysis of the VP4 deviation). These patients were diagnosed mainly with sequences of the Phylogenetic analysis of the nucleotide respectively. No viruses were detected in 629 samples. 70, 52, 28, 19, 24, 62, 7, 314 and 5 clinical samples, RSV, HMPV, EV and HBoV were isolated or detected in detected comprised influenza A (H1N1) pdm09 virus, cases), bronchitis (two cases), wheezing bronchitis (two cases), influenza A (H3N2) virus, influenza B virus, HPIV, AdV, RSV, HMPV, EV and HBoV were isolated or detected in 70, 52, 28, 19, 24, 62, 7, 314 and 5 clinical samples, respectively. No viruses were detected in 629 samples.

Phylogenetic analysis of the nucleotide sequences of the VP4/VP2 coding region in HRV-C by the NJ method

The phylogenetic tree (NJ method) based on the nucleotide sequences (390 nt) of the VP4/VP2 coding region is shown in Fig. 1. The 19 strains studied were genetically assigned to 11 types in the phylogenetic tree. The number of HRV-C strains in each type was as follows: pat10, two; HRV-C39, one; HRV-C43, two; pat14 or pat16, two; pat17, two; pat22, two; pat18, one; HRV-C40, three; HRV-C36, one; HRV-C37, two; and HRV-C46, one.

LRT of the fit of the models and timescale of evolution of the VP4/VP2 coding region in HRV-C by the ML method

Based on the SRDT model, we constructed a phylogenetic tree consisting of dated tips using the nucleotide sequences of the HRV-C VP4/VP2 coding region of the 19 strains and 13 reference strains by the ML method. As a result, the SR model was rejected as an adequate description of the evolution of the VP4/VP2 coding region (P<0.01). Therefore, the SRDT model was not significantly worse than the DR model, indicating that it adequately described the substitution process (P=0.0108). Through these processes, we obtained a phylogenetic tree by the ML method. The year of the first major division in this tree was estimated as 1872 (Fig. 2; 95% CI 1786–1905), which resulted in 11 types being genotypically assigned. In addition, the rate of molecular evolution was estimated from the tree as 3.07 × 10^{-3} substitutions per site per year (95% CI 1.83 × 10^{-3}–4.15 × 10^{-3}).

Analysis of pairwise distances and selective pressure of the VP4/VP2 coding region in HRV-C

The nucleotide and amino acid sequence identities of the 19 HRV-C strains were 68.2–99.5 and 77.7–100%, respectively. In addition, we calculated the intercluster distances of HRV-C from the distribution of the pairwise distances. Based on the nucleotide sequences, the pairwise distance was 0.253 ± 0.044 (mean ± 50%; Fig. 3) for the 19 strains and for 28 reference strains. Selection pressure analysis was performed on the 19 strains, and the SLAC method revealed a low mean dS/dN ratio of 0.058. No positively selected sites were detected in any position by any method, whilst over 100 negatively selected sites were found.

DISCUSSION

Using ML and NJ methods, we carried out phylogenetic analysis of the VP4/VP2 coding region of HRV-C from patients with ARIs in Japan between April 2009 and December 2011. We also calculated genetic distances (p-distance) and estimated the positively selected sites among the strains. The results revealed the following: (i) a phylogenetic tree by the ML method showed that the present HRV-C strains evolved approximately 150 years ago resulting in the 11 genotypes (Figs 1 and 2); (ii) the estimated evolutionary rate of the VP4/VP2 coding region, as well as for influenza virus subtype A, was rapid (3.07 × 10^{-3} substitutions per site per year); (iii) the p-distance of the present strains was relatively long (>0.25) and the nucleotide identity was 68.2–99.5%; (iv) no positively selected sites were found. Our results indicated

Table 1. Subjects and HRV-C strains in this study

<table>
<thead>
<tr>
<th>Clinical symptom</th>
<th>No. HRV-C strains*</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
<td>40 (14)</td>
</tr>
<tr>
<td>Bronchitis</td>
<td>18 (2)</td>
</tr>
<tr>
<td>Wheezing bronchitis</td>
<td>3 (2)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Total</td>
<td>63 (19)</td>
</tr>
</tbody>
</table>

*Numbers in parentheses show the number of HRV-C strains analysed in this study for each clinical condition.
Fig. 1. Phylogenetic tree of the VP4/VP2 coding region constructed by the NJ method. The phylogenetic tree is based on the nucleotide sequence of the VP4/VP2 coding region (390 nt) of the 19 strains analysed in this study and 28 reference strains. HRV-A and -B were used as outgroups. Distance was calculated according to Kimura’s two-parameter method and the tree was plotted using the NJ method, with labelling of branches showing at least 70% bootstrap support. Bar, nucleotide substitutions per site. Detailed procedures are described in the text.
that the present HRV-C strains evolved uniquely and rapidly over a period of approximately 150 years and diverged into multiple types.

Phylogenetic analysis by the NJ method is frequently used in molecular epidemiological studies of various viruses (Lin et al., 2001; Parveen et al., 2006; Vicente et al., 2011). The viral evolutionary rate and cluster classification can be obtained from phylogenetic trees using this method. Our report showed previously that the genotypic assignment of HRV-A, -B and -C viruses can be based on genetic divergence of the VP4/VP2 coding region (Arakawa et al., 2012). Furthermore, the ML method enabled us to analyse not only the timescale of evolution but also the rate of evolution of viral genes. In the present study, we applied the NJ and ML methods to analyse the phylogeny of the VP4/VP2 coding region in HRV-C. Using the NJ method, the present HRV-C strains were assigned to 11 genotypes (Fig. 1). In addition, we found that the present HRV-C strains evolved approximately 150 years ago, resulting in the formation of multiple genotypes. To our best knowledge, this paper is the first to report such observations in HRV-C.
Recent reports have shown that HRV-C is a major ARI agent. HRVs were thought previously to be associated mainly with the common cold, leading to mild respiratory manifestations (Gern, 2009; Turner & Couch, 2007). However, recent studies have suggested that HRVs are major agents of the induction of wheezing and exacerbation of asthma (Fujitsuka et al., 2011; Khadadah et al., 2010). In the present study, HRV-C strains were detected in patients with various ARIs such as URI and bronchitis. In addition, previous reports have suggested that ~30 genotypes with high genetic divergence are found in various HRV-C strains (McIntyre et al., 2010; Simmonds et al., 2010). The present strains were also classified into 11 genotypes with a relatively wide divergence, although only a small number of HRV-C strains were examined in the current study. Thus, these findings are in agreement with previous reports (Arakawa et al., 2012; Fujitsuka et al., 2011).

The evolutionary rate of viral genes differs across viruses, although it is not known why such differences occur (Mizuta, 2011). For example, the evolutionary rate of other virus genes such as the haemagglutinin gene in influenza virus subtype A (1.7 × 10⁻³ substitutions per site per year; Furuse et al., 2010; Graham et al., 2011) and attachment glycoprotein gene in RSV subgroup A (1.92 × 10⁻³ substitutions per site per year; Yoshida et al., 2012) is faster than that of virus genes such as the haemagglutinin–neuraminidase glycoprotein gene in HPIV type 1 (7.68 × 10⁻⁴ substitutions per site per year; Mizuta et al., 2011). A previous report estimated that the evolutionary rate of the VP4/VP2 coding region in HRV-C was 6.6 × 10⁻⁴ substitutions per site per year (Briese et al., 2008). In the present strains, the evolutionary rate of the VP4/VP2 coding region was faster (3.07 × 10⁻³ substitutions per site per year) than previous data for HRV-C. This may be due to differences in the regions analysed, the number of nucleotide analysed or the analytical tools used (Mizuta et al., 2011). In addition, previous reports have suggested that the rate of evolution is different for each virus (Sanjuán et al., 2010). Although the precise mechanisms are not known, it is possible that genome properties other than size, such as polarity or structure, may be associated with substitutions of the viral genome (Sanjuán et al., 2010). Taken together, the evolutionary rate of the VP4/VP2 coding region of the HRV-C strains detected in this study was as rapid as that of the haemagglutinin gene in influenza virus subtype A and the attachment glycoprotein gene in RSV.

No positively selected sites were found in the VP4/VP2 coding region in the present HRV-C strains, whilst many negative selected sites (>100 amino acids substitutions) were found. In general, positively selected sites are due to positive pressure in the host (Zhang et al., 2006). Previous reports have suggested that the VP2 and VP4 proteins may not be major antigens in HRV (Ledford et al., 2004). To escape positive pressure, positive selection may readily generate major epitopes, whilst negative selection plays an important role in maintaining the long-term stability of biological structures by removing deleterious mutations (Donker & Kirkwood, 2012; Yoshida et al., 2011). Thus, negative selection of the VP4/VP2-coding region in the present HRV-C strains may be responsible for maintaining the stability of biological structures of VP proteins by removing deleterious mutations.

A previous study suggested that there were no seasonal variations in HRV detection in Japan (Fujitsuka et al., 2011), whereas our later study did find such variations (Arakawa et al., 2012). No seasonal variations were found in HRV-C in the present study. These contradictory findings may be due to various epidemiological factors including the regions analysed and the virus species (Khor et al., 2012; Shek & Lee, 2003). In addition, we studied only a relatively small number of HRV-C strains. Thus, detailed studies of seasonal variations in HRV-C detection may be required.

In conclusion, the HRV-C strains analysed in this study were found to include multiple genotypes and may be associated with various ARIs such as URI, bronchitis and pneumonia in Japan. Evolution of the VP4/VP2-coding region in HRV-C is fast, with wide divergence. Larger
studies to analyse the molecular epidemiology of HRV-C may be needed.

**ACKNOWLEDGEMENTS**

This work was supported partly by a Grant-in-Aid from the Japan Society for the Promotion of Science and for Research on Emerging and Re-emerging Infectious Diseases from the Ministry of Health, Labour and Welfare, Japan.

**REFERENCES**


