Genetic analyses of the fusion protein genes of human parainfluenza virus types 1 and 3 among patients with acute respiratory infections in Eastern Japan from 2011 to 2015

Rika Tsutsui,1,2 Hiroyuki Tsukagoshi,3 Koo Nagasawa,4 Masaki Takahashi,5 Yuki Matsushima,6 Akihide Ryō,7 Makoto Kuroda,8 Hideki Takami2 and Hirokazu Kimura4,7,*

Abstract

Purpose. To genetically explore the fusion protein gene (F) in human parainfluenza virus type 1 (HPIV1) and type 3 (HPIV3) strains, we analysed them in patients with acute respiratory infections in Eastern Japan from 2011 to 2015.

Methodology. We constructed phylogenetic trees based on the HPIV and HPIV3 F gene using the maximum likelihood method and conducted P-distance and selective pressure analyses. We also predicted the linear epitopes of the protein in the prototype strains. Furthermore, we mapped the amino acid substitutions of the proteins.

Results. Nineteen strains of HPIV1 and 53 strains of HPIV3 were detected among the clinical acute respiratory infection cases. The phylogenetic trees indicated that the HPIV1 and HPIV3 strains were classified into clusters II and III and cluster C, respectively. The P-distance values of the HPIV1 and HPIV3 F genes were <0.03. Two positive selection sites were inferred in the HPIV1 (aa 8 and aa 10), and one positive selection site was inferred in the HPIV3 (aa 108), but over 10 negative selection sites were inferred. Four epitopes were predicted for the HPIV1 prototype strains, while five epitopes were predicted for the HPIV3 prototype strain. A positive selection site (aa 108) or the HPIV3 F protein was involved in the predicted epitope. Additionally, we found that an amino acid substitution (R73K) in the LC76627 HPIV3 strain presumably may affect the resistance to neutralization by antibodies.

Conclusion. The F gene of HPIV1 and HPIV3 was relatively well conserved in the eastern part of Japan during the investigation period.

INTRODUCTION

Human parainfluenza virus (HPIV) type 1 (HPIV1) and type 3 (HPIV3) belong to the genus *Respirovirus* and the family *Paramyxoviridae*. These viruses are responsible for a myriad of acute respiratory infections (ARIs), including the common cold, croup, bronchitis, bronchiolitis and pneumonia [1]. Previous epidemiological studies suggest that primary HPIV1 and HPIV3 infections have usually occurred in over 80% of children by the age of 5 years [2, 3]. Some infants with primary HPIV infection may develop severe clinical symptoms, including pneumonia accompanied by wheezing [4, 5]. Furthermore, HPIV infection, seasonal influenza and human respiratory syncytial virus (HRSV) infections may occur throughout life [4, 6, 7]. Although the epidemiology of HPIV is unclear, previous epidemiological research suggests that HPIV1 and HPIV3 are the dominant
Neuraminidase (HN) protein that causes primary infection in infants, rather than HPIV2 and HPIV4 [1, 6, 8].

The HPIV1 and HPIV3 genomes translate into seven and eight proteins, respectively [1, 8]. Of these, two structural proteins – the fusion (F) protein and the haemagglutinin–neuraminidase (HN) protein – have been identified as the major antigens [1, 9]. Indeed, it has been suggested that the HPIV F protein is pivotal to infection in the host cell [10–13]. The F protein may bind to host cells to facilitate infection, which also acts on syncytium formation between infected cells and F proteins of other viruses, such as HRSA, measles virus, mumps virus and human metapneumovirus [7, 14, 15].

Some molecular epidemiological studies suggest that HPIV1 can be further divided into three major genetic clusters or clades [16–18]. HPIV3 can also be classified into three distinct clusters (A, B and C) with the subdivision of cluster C further divided into five distinct subclusters (C1–C5) [19–23]. However, detailed genetic properties of the F gene in HPIV1 and HPIV3 strains detected in Japan are not known [24], although it is important to understand the characteristics of the F gene and its protein product in these viruses [25]. Therefore, we performed genetic analyses of the F gene in HPIV1 and HPIV3 in patients with ARIs in Eastern Japan (Aomori, Iwate and Gunma prefectures) between 2011 and 2015.

METHODS

Samples and patients

The study protocol was approved by the National Institute of Infectious Disease Ethics Committee (no. 495). Samples were obtained by the local health authorities of Aomori, Iwate and Gunma prefectures for the surveillance of viral diseases in Japan between April 2011 and March 2015. A total of 2069 pharyngeal or nasopharyngeal swab samples were collected from patients with ARIs, including upper respiratory illness, asthmatic bronchitis, bronchitis and pneumonia. Informed consent was obtained from subjects (or from the parents of underaged subjects) for sample donation. Either HPIV1 or HPIV3 was detected in 72 of the included patients (72/2069; 3.5 %).

RNA extraction, reverse transcription PCR (RT-PCR) and sequencing

Viral RNA was extracted from clinical samples using a QiaAmp viral RNA mini kit (Qiagen). RT-PCR was performed using a QIAIEN OneStep RT-PCR kit (Qiagen), using PCR primers for cDNA synthesis and amplification. The detection of HPIV1 and HPIV3 was carried out using multiplex PCR methods as previously described [26]. To analyse the full-length sequences of the HPIV1 and HPIV3 F genes with the primed walking method, some primer sets were newly designed by Primer Express version 1.5 software (Applied Biosystems) [27]. The full-length nucleotide sequence of the F gene spanned bases 5088–6754 (1688 nt) of the prototype Washington/1964 strain in HPIV1 and the 14702 strain, 5072–6691 (1620 nt) in HPIV3. The primer sequence data are shown in Table S1 (available in the online Supplementary Material). The PCR products were purified with a QIAquick PCR purification kit (Qiagen) and then sequenced with a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems), using newly designed primer sets. Sequence analysis was performed on an ABI 3500 genetic analyser (Applied Biosystems). The GenBank accession numbers of the nucleotide sequences obtained in the present study are LC076586–LC076603, LC076605, LC076622, LC076624–LC076630 and LC078993.

Phylogenetic analysis by the maximum likelihood (ML) method

Phylogenetic analysis of the nucleotide sequence of the HPIV1 and HPIV3 F genes was conducted using the Molecular Evolutionary Genetics Analysis (MEGA) software version 6 [28]. Evolutionary distances were estimated according to Kimura’s two-parameter method, and the phylogenetic tree was constructed with the ML method [29, 30]. We used Kakusan4 to select an appropriate substitution model, and GTR+G was selected [31]. The reliability of the tree was estimated with 1000 bootstrap replications using the GTR+G substitution model [32]. To construct the phylogenetic tree, we collected comprehensive data on the HPIV1 and HPIV3 F genes sequences from GenBank. The GenBank accession numbers of the strains were AB012132, AB736166, AF457102, EU326526, EU424062, FJ458542, JQ901971–JQ901980, JQ901983–JQ901997, Q901999–JQ092003, JQ902005, JQ902006, JQ902008, JQ902010, KF530196–KF530198, KF530203, KF530205, KF530211, KF530212, KF530217, KF530221, KF530225, KF530226, KF530229, KF530230, KF530232–KF530234, KF530236, KF530241–KF530243, KF530245, KF530249–KF530253, KF530256, KF530257, KF687307, KF687310–KF687315, F687318–KF687321, KF687323, KF687325–KF687328, KF687330, KF687332–KF687334, KF687336, KF687340, KF687344, KF687346–KF687349, KF687351–KF687354, KF687357, KF687358, KJ672527, KJ672530–KJ672533, KJ672535–KJ672539, KJ672541, KJ672542, KJ672545, KJ672547, KJ672549, KJ672553–KJ672555, KJ672559, KJ672560, KJ672568, KJ672574–KJ672576, KJ672579, KJ672582, KJ672584, KJ672586, KJ672588, KJ672591, KJ672593, KJ672595, KJ672596, KJ672601, KJ672604, KJ672605, KJ672607–KJ672610, KJ672612–KJ672616, KJ672618 and S82195.

Pairwise distance calculation

To assess the genetic distances of all present strains, we calculated pairwise distances (P-distance) for the HPIV1 or HPIV3 strains (both the current and reference strains), as previously described [16, 24].

Selective pressure analyses

To evaluate the action of selective pressure on the F gene across all HPIV1 or HPIV3 strains, we estimated the rates of non-synonymous (dN) and synonymous (dS) changes at amino acid sites by conservative single likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL) and internal
fixed-effects likelihood (IFEL) methods using DataMonkey (www.datamonkey.org) [33]. Positive/negative selection ($dN$/$dS$) was determined by $P \leq 0.05$ for SLAC, FEL and IFEL.

B-cell linear epitope analyses

To examine the relationship between the positive selection sites and epitopes, the linear B-cell epitopes of the HPIV1 and HPIV3 strains were predicted using Epitopia, BCPred, FBCPred, BepiPred and Antigenic as previously reported [34–38]. We used these tools under default conditions, with the exception of the epitope length. In the present study, the epitope lengths predicted by Antigenic and FBCPred were set at 10-mer amino acids, while the epitope length predicted by BCPred was set at 12-mer amino acids. We accepted the sites as the linear B-cell epitopes which were inferred by three or more methods and with more than 10-mer consecutive amino acids as previously reported [38].

Mapping of amino acid substitutions of the F protein in HPIV1 and HPIV3

To assess the relationships between amino acid substitutions and antibody reactivity against the F protein, we mapped the amino acid substitutions of the F protein of the present strains, as previously described [39]. The models were constructed using MODELLER v9.16 using the Washington 1964 strain and the Wash/47885/57 strain as templates for HPIV1 and HPIV3, respectively [40]. Homology modelling for HPIV1 and HPIV3 was based on the crystal structure of 1ZTM (Protein Data Bank accession number for the HPIV3 F protein), and the energy of each constructed model was minimized using Swiss-PDBdb Viewer v4.1 [41]. Then, the models were evaluated by Ramachandran plots produced using the RAMPAge server [42]. Finally, the models were coloured using Chimera v1.10.2, and the amino acid substitutions corresponding to the prototype strains were mapped on the predicted models [43]. Furthermore, we also mapped the neutralization-related amino acids for the F protein in HPIV3 as suggested by a previous report [44].

RESULTS

Detection of HPIV1 and HPIV3 in this study

We collected 2069 pharyngeal or nasopharyngeal swab samples from patients with ARIs and detected 19 strains of HPIV1 (0.92 %) and 53 strains of HPIV3 (2.6 %) using multiplex PCR methods. These HPIV1- or HPIV3-positive patients were variably diagnosed with upper respiratory illness, lower respiratory illness, bronchitis, bronchiolitis or pneumonia. HPIV1 strains were typically detected in patients aged 21.8±36.0 years (range, 1 month to 88 years), and HPIV3 strains were typically detected in patients aged 2.2±2.4 years (range, 1 month to 13 years). No differences were seen in the male-to-female ratios.

Phylogenetic analysis, nucleotide identities and P-distance in HPIV1 or HPIV3 strains

To assess the identified HPIV1 and HPIV3 strains phylogenetically, we constructed the phylogenetic trees according to the F gene nucleotide sequences of HPIV1 and HPIV3 using the ML method (Fig. 1a, b). The identified HPIV1 strains could be classified into two clusters (clusters II and III), and HPIV3 strains were classified into cluster C. In addition, HPIV3 was further classified into distinct four subclusters (C1, C2, C3 and C5). The nucleotide sequence identities were 93.8 % to 100 % and 93.2 % to 100 %, respectively. In addition, we calculated the P-distance among the HPIV1 and HPIV3 strains to be 0.018±0.011 and 0.026±0.018, respectively (Fig. 2a, b).

Positive/negative selection sites in the F gene of HPIV1 or HPIV3

We analysed the positive and negative selection sites in the strains using SLAC, FEL and IFEL methods and calculated the $dN$/$dS$ values. The details of these analyses are presented in Tables 1 and 2. For HPIV1, we found one positive selection site at aa 5 by the FEL method and one positive selection site at aa 8 by the IFEL method. For HPIV3, we found one positive selection site at aa 108 by the IFEL method. Many negative selection sites were detected by all methods for both strains. More than 10 and 60 negative selection sites were estimated in HPIV1 and HPIV3, respectively.

Linear B-cell epitopes analyses

The predicted linear epitopes are shown in Table 3 and Figs 3 and 4. As shown in Table 3, in the HPIV1 and HPIV3 strains, four and five epitopes were predicted, respectively. Among them, a common sequence of seven amino acids (YICPxDP) was found in both of the strains. The common predicted epitope motif was located on the head of the F protein (Figs 3 and 4), which corresponded to the positions at aa 341–347 in the F protein of the HPIV1 prototype strain and at aa 338–344 of the HPIV3 prototype strain, respectively. Moreover, a predicted epitope of HPIV3 was involved with the positive selection sites (aa 108), while the HPIV3 epitopes did not correspond to the neutralization reactive sites of HPIV3 (aa 73, aa 396–398). The epitopes in the HPIV1 F protein were not involved with the positive selection sites.

Amino acid substitutions of the F protein in the predicted models

The predicted structural models of the HPIV1 and HPIV3 F proteins are shown in Figs 3 and 4, respectively. The HPIV1 model shows aa 22–97 and aa 142–484 corresponding to the F protein of the HPIV1 prototype strains (Washington 1964 strain), and the model for HPIV3 shows aa 25–97 and aa 145–487 corresponding to the protein of the HPIV3 prototype strains (Wash/47885/57 strains). Ramachandran plots showed that over 95 % of the models’ residues were in favoured regions. The sum of the amino acid substitutions in each cluster on the surface of the F protein is indicated in red. The other HPIV1 strains
Fig. 1. Phylogenetic trees of the F gene for HPIV1 (a) and HPIV3 (b) constructed by the ML method. The GenBank accession numbers of the reference strains are indicated in parentheses.
in clusters II and III had 11 amino acid substitutions and 29 amino acid substitutions corresponding to the HPIV1 prototype strain, respectively (Table S2). Among them, 5 or 11 amino acid substitutions were estimated on the surface of the \( F \) protein for the HPIV1 clusters II and III, respectively (Fig. 3). Moreover, our HPIV3 strains had 34 amino acid substitutions corresponding to the HPIV3 prototype strain (Table S3). In addition, there were nine amino acid substitutions on the surface of the protein. Of these, R73K could affect resistance to the neutralization by monoclonal antibodies and was estimated to be within the HPIV3 cluster C strains [44].

**DISCUSSION**

HPIV3 and HPIV1 were classified into some lineages; for example, Mizuta et al. [16, 24] have shown that HPIV3 isolated in Yamagata prefecture, Japan, could be classified into three lineages according to the \( HN \) gene analysis. They also reported that HPIV1 isolated in the same areas could be classified into two clusters according to the \( HN \) gene [16]. However, to the best of our knowledge, there are no reports regarding the genetic analysis of the \( F \) gene in HPIV1 and HPIV3 detected in Japan. In this study, we showed that HPIV1 and HPIV3 detected in the Eastern part of Japan could be classified into three and two lineages, respectively, by \( F \) gene analysis. Our results also indicated that both strains were well conserved. To compare the phylogenetics of the \( F \) genes of HPIV1 and HPIV3 with those detected in Asia, America, Africa and Europe, we tried to collect comprehensive data on the HPIV1 and HPIV3 \( F \) gene sequences from GenBank (full length). Both the strains in this study and the foreign strains were closely located in the phylogenetic trees. Furthermore, the genetic distances (\( P \)-distances) of these strains were relatively short. Thus, HPIV1 and HPIV3 \( F \) genes might have similar genetic properties in these strains, although we were unable to collect adequate numbers of \( F \) gene sequences for the virus strains. To better understand the genetic properties of the \( F \) genes in these strains, larger studies are needed.

Next, we inferred the positive and negative selection sites in the present strains. A few positive selection sites were inferred in the \( F \) proteins of both HPIV1 and HPIV3. The positive selection sites of the \( F \) protein in HPIV1 were located in the N-terminus of the protein (Lys5Glu, Gly5Glu or Asp5Glu and Ile8Leu or Phe8Leu). Previous research has suggested that N-terminal amino acid residues

**Table 1.** Positive selection sites in the HPIV1 and HPIV3 \( F \) genes

<table>
<thead>
<tr>
<th>Virus</th>
<th>aa position</th>
<th>Change</th>
<th>SLAC</th>
<th>FEL</th>
<th>IFEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPIV1</td>
<td>5Glu</td>
<td>Lys, Gly, Asp</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8Leu</td>
<td>Ile, Phe</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPIV3</td>
<td>108Glu</td>
<td>Lys, Gly</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>108Lys</td>
<td>Glu, Gly</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance level for SLAC, FEL and IFEL is \( P<0.05 \).

**Table 2.** Negative selection sites in the HPIV1 and HPIV3 \( F \) genes

<table>
<thead>
<tr>
<th>Virus</th>
<th>SLAC</th>
<th>FEL</th>
<th>IFEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPIV1</td>
<td>13</td>
<td>46</td>
<td>11</td>
</tr>
<tr>
<td>HPIV3</td>
<td>74</td>
<td>124</td>
<td>69</td>
</tr>
</tbody>
</table>

Significance level for SLAC, FEL, and IFEL is \( P<0.05 \).
1 to 19 of the HPIV1 F protein act as the signalling peptide and that amino acid substitutions are common in these regions [8]. Thus, it is plausible that the amino acid substitutions in the present strains may cause alteration of the function of the F protein as a signalling peptide [1]. In contrast, the positive selection sites of the F protein in HPIV3 were located at amino acid residue 108 (Lys108Glu or Gly108Glu and Glu108Lys or Gly108Lys), although this region consisted of a portion of the fusion peptide [8]. In recent years, it has been suggested that a mutation in the F protein may affect the growth rate and membrane fusion activity of HPIV [45–47]. Thus, these amino acid substitutions may reflect functional changes of the HPIV3 F protein, although we did not examine the molecular activities of the protein in vitro.

When viral antigen receives immune pressure by the host, the virus may try to escape recognition using a variety of evasion mechanisms [48]. Thus, it may be important to examine the relationships between the positive selection sites and epitopes. Considering these circumstances, we predicted the linear epitopes in the HPIV F proteins as previously described [38]. As shown in Tables 1 and 3, a linear epitope was found to be involved in a positive selection site (aa 108) within the HPIV3 F protein. Moreover, this site

<table>
<thead>
<tr>
<th>Virus</th>
<th>Prototype strain (accession no.)</th>
<th>Position</th>
<th>Predicted epitopes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPIV1</td>
<td>Washington/1964 (AF457102)</td>
<td>101–111</td>
<td>DTVTVNNDPQT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>341–350</td>
<td>YICPRDPTQL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>448–457</td>
<td>VGPAVSIRPV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>546–555</td>
<td>RNPYMGNNSN</td>
</tr>
<tr>
<td></td>
<td>Wash/47885/1957 (S82915)</td>
<td>57–66</td>
<td>IEDNSCGDQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>98–109</td>
<td>ESNENTDPRTRK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>337–346</td>
<td>SYICPSDPGF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>370–379</td>
<td>SDIVPRYAFV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>525–534</td>
<td>NRVQNDKPY</td>
</tr>
</tbody>
</table>

Table 3. Predicted linear B-cell epitopes in the HPIV1 and HPIV3 fusion proteins

Fig. 3. Position of amino acid substitution sites mapped to the predicted models of the fusion (F) protein in the Washington 1964 strain (the HPIV1 prototype strain). These models were constructed by homology modelling using 1ZTM (Protein Data Bank accession number for the uncleaved HPIV3 F protein). Chains A, B and C are indicated by white grey, light grey and black, respectively. Amino acid substitutions in the protein within each cluster are shown in red. The substituted amino acids in chain A are also shown in the figure. The predicted linear epitopes are presented in cyan.
In conclusion, we conducted a genetic analysis of the F protein of HPIV1 and HPIV3, similar to HPIV3. Overall, our data were based on relatively small numbers of each strain, making such assessments difficult. Thus, this site may not be associated with neutralizing sites. However, it remains unknown as to how such sites are involved in the function of the epitope and evasion of antibody detection initiated by the host defence mechanisms [49]. In contrast, this site was incompatible with the linear epitopes and neutralization sites of mouse monoclonal antibodies. This may be due to differences in epitope recognition by the immune system between mice and humans [50]. Further studies regarding these relationships may be required.

To assess the relationships between amino acid substitutions and antibody reactivity against F proteins, we also modelled the F protein. This modelling showed that some amino acid substitutions could be predicted. Of these, R73K, which could be a neutralization-related amino acid, was detected in only one HPIV3 strain (LC76627) [44]. Additional surveillance is needed to elucidate the prevalence and clinical relevance of the mutation. In contrast, we could not determine the amino acids pertaining to neutralization-related sites in the HPIV1 F protein, since there is a lack of data pertaining to the neutralization sites of the F protein of HPIV1, similar to HPIV3. Overall, our data were based on relatively small numbers of each strain, making such assessment difficult.

In conclusion, we conducted a genetic analysis of the F genes of HPIV1 and HPIV3 from samples taken in Eastern Japan between 2011 and 2015. Despite this research, it remains important to continue to accumulate additional data about both strains, and we recommend larger molecular epidemiology studies in the future. Indeed, although HPIV is a major cause of ARIs, there have been too few genetic studies with regard to the most common strains, HPIV1 and HPIV3, and there is a need to remedy this situation in the future.
24. Almajhdi FN.


