Emerging Genotypes of Human Respiratory Syncytial Virus Subgroup A among Patients in Japan

Yugo Shobugawa,1* Reiko Saito,1 Yasuko Sano,2 Hassan Zaraket,1 Yasushi Suzuki,1 Akihiko Kumaki,1 Isolde Dapat,1 Taeko Oguma,1 Masahiro Yamaguchi,1 and Hiroshi Suzuki1

Division of Public Health, Department of Infectious Control and International Medicine, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan,1 and Sano Clinic, Niigata, Japan2

Received 20 January 2009/Returned for modification 5 May 2009/Accepted 17 June 2009

Human respiratory syncytial virus (HRSV) is a common etiological agent of acute lower respiratory tract disease in infants. We report the molecular epidemiology of HRSV in Niigata, Japan, over six successive seasons (from 2001 to 2007) and the emerging genotypes of HRSV subgroup A (HRSV-A) strains. A total of 488 HRSV samples were obtained from 1,103 screened cases in a pediatric clinic in Niigata. According to the phylogenetic analysis, among the PCR-positive samples, 338 HRSV-A strains clustered into the previously reported genotypes GA5 and GA7 and two novel genotypes, NA1 and NA2, which were genetically close to GA2 strains. One hundred fifty HRSV-B strains clustered into three genotypes, namely, GB3, SAB3, and BA, which has a 60-nucleotide insertion in the second hypervariable region of the G protein. The NA1 strains emerged first, in the 2004–2005 season, and subsequently, the NA2 strain emerged in the 2005–2006 season. Both strains caused large epidemics in the 2005–2006 and 2006–2007 seasons. The average age of children who were infected with NA2 strains was significantly higher than that of those infected with GA5 and the frequency of reinfection by NA2 was the highest among all genotypes, suggesting that this genotype possessed new antigenicity for evading past host immunity. This is the first paper to show a possible correlation between an emerging genotype, NA2, and large outbreaks of HRSV in Japan. Continuing studies to follow up the genetic changes and to clarify the mechanism of reinfection in HRSV are important steps to understand HRSV infections.

Human respiratory syncytial virus (HRSV) is the most common cause of serious acute lower respiratory tract disease among infants and young children, and is found mainly in late fall, winter, and spring in temperate zones of the world (6). Some 50% to 70% of infants experience infection in the first year of life, and virtually all are infected by 2 years of age (3). In a population-based birth cohort study, 1.1% of the cohort were admitted to the hospital within 12 months of birth with HRSV-induced bronchiolitis (14). The consequences of HRSV infection in children with underlying conditions, such as prematurity, cardiac and pulmonary disease, or immunosuppression, may include prolonged substantial illness and even death (18, 38). Reinfections are very common throughout life (7, 10, 11, 13).

HRSV belongs to the family Paramyxoviridae and has a nonsegmented, negative-sense RNA genome of approximately 15,200 nucleotides (3). HRSV has been classified into antigenic subgroups A and B (HRSV-A and HRSV-B, respectively), initially, on the basis of the reactivity of the virus with monoclonal antibodies directed against the attachment glycoprotein (G protein) (1, 5, 15, 21) and now by genetic analyses (9, 19, 31–33).

The G protein is the most variable HRSV protein, with two hypervariable regions. Its C-terminal region (the second hypervariable region) accounts for strain-specific epitopes (2–4, 9, 16, 27, 29, 31). The molecular epidemiology and evolutionary patterns of the G protein have provided important information about the epidemiological features of HRSV. Some studies showed that several different genotypes cocirculated and some predominated in a community every year (23, 27). However, the relationship between strain diversity and the clinical and epidemiological features of HRSV has yet to be elucidated in detail.

The subgroups have been subdivided further, into genotypes, by genetic analyses. HRSV-A is divided into seven genotypes (GA1 to -7) and HRSV-B into four genotypes (GB1 to -4) (3, 8). An additional HRSV-A genotype, SAA1, has been proposed, as well as the new HRSV-B genotypes SAB1 to -3 (37). Another HRSV-B genotype includes the Buenos Aires (BA) type strain, which has a 60-nucleotide insertion in the second hypervariable region of the G protein and has been reported in Buenos Aires in 1999, as well as in other areas of the world (31, 35, 40). BA strains are further subdivided into six clusters (BA-I to BA-VI) (36).

National surveillance of HRSV infection based on weekly reports from sentinel pediatric clinics throughout Japan began in November 2003 under the Infectious Disease Control Law. The surveillance indicated that the number of HRSV-infected patients increased dramatically for unknown reasons in the 2005–2006 and 2006–2007 seasons. This study reports large HRSV-A epidemics in the 2005–2006 and 2006–2007 seasons in Niigata, Japan, and provides a possible link to high-level circulation in the same seasons nationwide. It also discusses the relationship between antigenic variation and clinical features of HRSV in regard to molecular epidemiology and the evolutionary patterns of the G protein.
PCR and nucleotide sequencing, viral RNA was extracted from 100-
/H9262 cycle sequencing kit (Applied Biosystems, Foster, CA) according to the manu-
/inghamshire, United Kingdom), labeled with a BigDye terminator (version 3.1)
/MicroSpin S-300 HR column PCR purification kit (Amersham Bioscience, Buck-
/verse transcriptase (Invitrogen Corp. Carlsbad, CA) (31).

Materials and Methods

HRSV surveillance system in Japan. National surveillance of HRSV infection
/tpc287.html). Each HRSV case is defined on the basis of clinical respiratory
/symptoms, such as rhinorrhea, wheeze, and cough, and laboratory diagnosis, such
/as virus isolation, serologic test, or rapid antigen test, is carried out at the
/outpatient clinic or on admission to the hospital. The number of patients
/diagnosed with HRSV infection is reported on a weekly basis from approximately
/0,000 sentinel pediatric clinics/hospitals throughout Japan. These sentinel sites
/forward clinical data to approximately 60 prefectural or municipal public health
/institutes, and the data generated are electronically reported to the Infectious
/Disease Surveillance Center in the National Institute of Infectious Diseases (Tokyo).

Study population and clinical samples. This study was conducted for six
/seasonal seasons, from November 2001 through June 2007, at one pediatric
/outpatient clinic in Niigata City, Japan. Niigata City is the prefectural capital of
/Niigata Prefecture, and its total population is approximately 0.8 million. Children
/under 5 years of age with acute lower respiratory tract illness, such as wheezing,
cough, rhinorrhea, and fever, were enrolled in the study. We obtained written
/informed consent from their guardians and nasopharyngeal aspirates or nasal
/swabs from the patients. Basic clinical data, such as sex, age, date of clinic visit,
/and date of onset, were recorded by a pediatrician at the clinic. This study was
/approved by the Medical Faculty Ethics Committee of Niigata University.

Clinical specimens were kept at 4°C in the clinic; were transported to the
/Department of Public Health, Graduate School of Medical and Dental Sci-
/ences, Niigata University, within 5 days of sampling; and were then kept
/frozen at −80°C for further examination.

RT-PCR and nucleotide sequencing. Briefly, for reverse transcription (RT)-
/PCR and nucleotide sequencing, viral RNA was extracted from 100-μl samples
/of the nasopharyngeal aspirates or swabs by using an Extranen II kit (Kainos,
/Tokyo, Japan) according to the manufacturer’s instructions. RT to create cDNA
/was performed using random primers and Moloney murine leukemia virus reverse
/transcriptase (Invitrogen Corp. Carlsbad, CA) (31).

In accordance with a method previously published (31), first and heminested
/PCRs targeting a 270-nucleotide segment of the G protein gene’s second hyper-
/variable region were performed. Amplified PCR products were purified with a
/MicroSpin S-300 HR column PCR purification kit (Amersham Bioscience, Buck-
/inghamshire, United Kingdom), labeled with a BigDye terminator (version 3.1)
/cycle sequencing kit (Applied Biosystems, Foster, CA) according to the manu-
/facturer’s instructions, and then analyzed on an ABI 3100 automatic DNA
/sequencer.

Phylogenetic analysis. Multiple sequence alignments of G genes for HRSV-A
/and -B strains were compiled with BioEdit, version 7.0.9.0 (12). Phylogenetic
/trees of the G protein gene’s second hypervariable region were generated by
/using the neighbor-joining method with MEGA, version 4 (34). Bootstrap prob-
/abilities for 1,000 iterations were calculated to evaluate confidence estimates. All
/sequences for both subgroup A and B viruses were included in the phylogenetic
/analysis. In the phylogeny, strains from Rochester, NY; Winnipeg, Manitoba,
/Canada; Houston, TX.; St. Louis, MO; Soweto, South Africa; Birmingham, AL;
/West Virginia; and Buenos Aires, Argentina, were included for comparison.

Pairwise nucleotide distances (p distances), that is, the number of pairwise
/nucleotide differences divided by the total number of nucleotides in the se-
/quenced segment, between representative paired sequences from each genotype
/of HRSV-A were calculated using MEGA, version 4 (34).

Patient characteristics by genotypes. Clinical data on patients, such as sex and
/average age, were compared among seasons and infecting-strain genotypes. Sta-
tistical analysis for comparison of mean values in more than three groups was
/ performed by using Scheffe’s test. Comparison of the proportions was accom-
/plished with 2-by-multiple tables. Statistical significance was concluded if the
/P value was less than 0.05.

Deduced amino acids and N-glycosylation site analysis. The deduced amino
/acids sequences at the second hypervariable region of HRSV-A and HRSV-B
/strains were compared to those of the prototype A2 strain, spanning 270 nucle-
/otides, and the BA type strain (BA4128/99B), with 330 nucleotides, respectively
/(16, 35). Potential N-glycosylation sites were predicted if the encoded amino
/acids were NXT, where X is not a proline (16, 25, 28).

Nucleotide sequence accession numbers. The GenBank accession numbers of
/the nucleotide sequences obtained in the present study are AB470478 to
/AB470482.

Results

National surveillance of HRSV cases in Japan. National surveillance for HRSV in Japan started in November 2003, and
/the disease was most prevalent and clearly active during au-
tumn and early winter (Fig. 1). Large outbreaks of HRSV
/of the outbreak was bigger in the latter season. In general, the

FIG. 1. Comparison of monthly number of HRSV cases between a hospital-based survey in Niigata and national surveillance in Japan. The monthly distribution of numbers of screened cases in hospital-based surveillance of HRSV in Niigata City is shown by the stacked bar graph; closed boxes indicate numbers of PCR-positive samples, and open boxes denote PCR-negative samples. The numbers of cases in national surveillance of HRSV in Japan are shown by the line graph.
seasonality and the size of the epidemic in our survey matched the national surveillance data.

Study population and demographic details. Among 1,103 screened cases, a total of 488 HRSV strains were positive by PCR for the six successive seasons from November 2001 through June 2007 in Niigata City, Japan (Table 1). All HRSV-positive samples were sequenced and further analyzed for genotype. The yearly distribution of positive cases varied, depending on the season, from 33 to 154, with the highest positivity rate being observed in the 2006–2007 season (Table 1). The average age of patients in the 2006–2007 season (1.8 years old) was significantly higher than in the 2001–2002 (0.8 years old, P < 0.001), 2002–2003 (1.1 ± 0.8 years old, P < 0.005), and 2003–2004 (1.2 ± 0.8 years old, P < 0.05) seasons, respectively. On the other hand, the average ages of HRSV-negative infants were not significantly different during the study period.

Molecular epidemiology and emergence of new genotype of HRSV in Niigata City. Phylogenetic analysis revealed that 338 samples were HRSV-A and 150 were HRSV-B (Table 2). The HRSV-A strains clustered into four genotypes, namely, GA5, GA7, and the two new genotypes NA1 and NA2, with bootstrap values of 70% to 100% (Fig. 2a). The BA strains, first reported in Buenos Aires in 1999, have a 60-nucleotide insertion in the second hypervariable region of the G protein (35).

The predominant genotype changed during the study period. In HRSV-A strains, genotype GA5 circulated intermittently from the 2001–2002 season through the 2003–2004 season, shifting to NA1 and NA2 from the 2004–2005 season to the 2006–2007 season (Table 2). Among the HRSV-B strains, BA viruses were predominant from the 2002–2003 to the 2006–2007 season. BA viruses in this study were further classified into three groups, namely, BA-II, BA-IV, and BA-V (36). In the 2002–2003 season, 1 BA-II and 40 BA-V strains were detected, and in the subsequent seasons, from the 2003–2004 through the 2006–2007 season, only the BA-IV genotype was found, with the numbers of isolates being 3, 5, 33, and 16 in the 2003–2004, 2004–2005, 2005–2006, and 2006–2007 seasons, respectively.

Age distribution of children with HRSV infections by genotype. A significant difference in average ages was observed between patients infected with genotypes GA5 and NA2 (1.0 ± 0.8 years old versus 1.7 ± 1.1 years old, P < 0.001) (Table 3).
in HRSV-A infection cases. On the other hand, the average age of patients with each genotype among HRSV-B infection cases and between patients infected with HRSV-B and HRSV-A strains as a whole were not statistically different.

The proportion of children more than 2 years old infected with NA2 was significantly higher than the proportion infected with GA5 (37.6% versus 15.7%, $P < 0.001$). Among patients infected with other genotypes, the proportions were not statistically different (Table 3).

Amino acid analysis and N-glycosylation sites. Four putative N-glycosylation sites were identified on the second variable region of the G protein among the HRSV-A strains. The first site, amino acid (aa) position 237, was conserved among the GA5 and NA2 strains but disappeared in the NA1 strain. The second site, aa 250, was present in the genotype GA5 strain only; the third site, aa 251, was identified in the prototype A2 and the NA1 strains as having amino acid coding NTT, and the NA2 strain as having NIT. The fourth site, aa 294, was conserved among all the three genotypes. Other amino acid mutations in comparison to the prototype were observed at residues 226, 256, 262, 274, 290, and 297 (Fig. 3a).

Three N-glycosylation sites were identified among the group
antigeneic variation between and within the subgroups may contribute to reinfections by evading the host immune responses. Almost all homologous HRSV reinfections observed in this study were caused by HRSV-A viruses. This is in line with the findings of previous studies (20, 41) which suggested that HRSV-A may not elicit a complete and/or long-lasting subgroup-specific immune response compared to that elicited by HRSV-B. Interestingly, reinfection caused by NA2 occurred more frequently than reinfections by other genotypes, although it only circulated during two seasons. Thus, our observations led us to surmise that a weaker neutralizing immune response against NA2 was a cause for a higher rate of reinfection with NA2 strains.

However, it was unclear why NA1 was replaced by NA2 and why it did not become the dominant genotype in the following epidemic. Genetically, NA1 strains were very close to NA2 strains, as the p distance between the NA1 and NA2 strains was 0.041 to 0.070. However, several amino acid changes and different N-glycosylation sites in the second hypervariable region on the G protein might have contributed to a change in antigenicity between the genotypes (3, 24). Yet, this remains to be confirmed by serological studies.

After the appearance of the BA virus, which has a duplication of 60 nucleotides, in the 2002–2003 season (31), BA viruses were in circulation Niigata consistently thereafter. Trento et al. analyzed HRSV-B collected in Buenos Aires and classified the isolates into clusters BA-I to BA-IV by analyzing the entire G gene sequence (36). Also, they compared a worldwide collection of BA strains by partial sequence analysis of the last 330 nucleotides and reported six clusters, BA-I to BA-VI. In our analysis, we found that three clusters, BA-II, -IV, and -V, circulated during the study period. In particular, BA-V was unique to Niigata, and it circulated only in the 2002–2003 season. A specific amino acid change of P231Q in the second hypervariable region was observed, which can lead to aa position 230 being a potential N-glycosylation site. After the 2002–2003 season, BA-V disappeared and BA-IV became dominant. It eventually became the most diverse genetic cluster, as can be seen from the results of this study and previous reports from

### Table 3. Age comparison among groups divided by genotypes

<table>
<thead>
<tr>
<th>Virus and genotype</th>
<th>No. of cases</th>
<th>Avg age (yr) ± SD</th>
<th>Age distribution [% of patients]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0–1 yr</td>
</tr>
<tr>
<td>HRSV-A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA1</td>
<td>71</td>
<td>1.4 ± 0.9</td>
<td>26 (36.6)</td>
</tr>
<tr>
<td>NA2</td>
<td>141</td>
<td>1.7 ± 1.1</td>
<td>35 (24.8)</td>
</tr>
<tr>
<td>GA5</td>
<td>121</td>
<td>1.0 ± 0.8</td>
<td>71 (58.7)</td>
</tr>
<tr>
<td>GA7</td>
<td>5</td>
<td>1.3 ± 1.2</td>
<td>3 (60.0)</td>
</tr>
<tr>
<td>HRSV-B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>143</td>
<td>1.4 ± 1.0</td>
<td>58 (40.6)</td>
</tr>
<tr>
<td>GB3</td>
<td>5</td>
<td>1.4 ± 1.1</td>
<td>3 (60.0)</td>
</tr>
<tr>
<td>SAB3</td>
<td>2</td>
<td>0.7 ± 0.3</td>
<td>1 (50.0)</td>
</tr>
</tbody>
</table>

a For HRSV-A strains, a statistically significant difference in average age was shown between patients with genotypes NA2 and GA5 (p < 0.001), and a statistically significant higher proportion of patients with genotype NA2 than with GA5 that were >2 years old (p < 0.001) was shown.

b For HRSV-B strains, no statistically significant age differences were found between patients with different genotypes.

B strains (Fig. 3b). The first site was identified at aa 230 in the BA-V strains only, while the other two sites, at aa 296 and aa 310, were conserved among almost all the strains at the C-terminal end of the G protein gene. Other amino acid mutations in comparison to the prototype were observed at residues 218, 247, 270, and 287 (Fig. 3b).

**Reinfection.** Reinfection was defined as an infection that occurred after a month-long period from the first infection without any HRSV clinical manifestation of acute respiratory infection. Over the study period, 39 (8.0%) of 488 patients were documented with HRSV reinfections (Table 4). Among them, eight (20.5%) were homologous reinfections, four (10.3%) being GA5, three (7.7%) NA2, and one (2.6%) BA. In our study, 20 of 39 reinfections were caused by NA2, the highest frequency of all the genotypes. Of these, the first infection was due to homologous NA2 in three cases, to NA1 and BA strains in eight cases each, and to a GA5 strain in one case.

**DISCUSSION**

Our epidemiological study indicates that large epidemics of HRSV occurred in Niigata in the 2005–2006 and 2006–2007 seasons, which coincided with the Japanese national surveillance data. It is well known that large influenza outbreaks occurred due to the emergence of new serotypes or subtypes of the influenza virus (30, 39), but this phenomenon has not been reported for HRSV. In Niigata, the BA strain emerged as a predominant genotype in the 2002–2003 season and continued to circulate during six successive seasons until the 2006–2007 season but did not cause large outbreaks thereafter, although a variety of genotypes were detected within the BA strains (31).

In addition, two novel genotypes, NA1 and NA2, which were GA2 variants of HRSV-A, were detected. The two genotypes were very close to each other but distinguished from GA2 in the phylogenetic tree and pairwise nucleotide distances. Both genotypes caused large epidemics in the 2005–2006 season, with NA1 and NA2 cocirculating, and in the 2006–2007 season, with NA2 predominating. In recent years, GA5 and GA2 were the most common genotypes among HRSV-A (17, 25, 41), but the new GA2 variants NA1 and NA2 have not been reported in Japan before. Thus, it was speculated that the large epidemics occurred due to the emergence of NA1 and NA2, because most of the children did not have immunity. We also detected NA2 strains from pediatric samples in Lebanon and Vietnam in 2008 (unpublished data), suggesting that this strain might be circulating globally. However, there are no reports of large outbreaks caused by the GA2 variants in other areas, so further studies are warranted to confirm our findings.

In general, most individuals are infected with HRSV in the first year of life and few escape infection by this virus during infancy or early childhood. That is why the age of children infected with HRSV tended to be younger. However, the results of our study indicated that the average age of the HRSV cases during the large outbreak in the 2006–2007 season was significantly older than in the other seasons. Of note, the ages of NA2-infected children were significantly older than the ages of children infected with other genotypes. This could be due to a low level of past immunity against the newly emerged NA2 among most infants and children.
FIG. 3. Deduced amino acid alignments of the second variable region of the G protein gene from HRSV-A (a) and HRSV-B (b) Niigata strains. Alignments are shown relative to the sequences of prototype strain A2 (GenBank accession number M11486) (16) and genotype BA strain BA4128/99B (GenBank accession number AY333364) (35). The amino acid numbering corresponds to strain A2 G protein positions 220 to 298 for the HRSV-A viruses and to strain BA4128/99B G protein positions 219 to 315 for the HRSV-B viruses. Identical residues are indicated by dots. Stop codons are indicated by asterisks. Potential N-glycosylation sites (NXT, where X is not a proline) are indicated by gray shading. In panel a, second N-glycosylation sites (aa 250) are indicated by rectangles and third sites (aa 251) by dotted rectangles. In panel b, the two copies of the duplicated 20-amino-acid region in HRSV-B viruses are indicated by rectangles.
other countries (36). The reason for the successful circulation of this subgenotype for years remains to be determined.

In conclusion, we found novel emerging genotypes of HRSV-A, namely, NA1 and NA2, in Niigata, Japan. The analysis of both clinical and molecular biological information showed a possible correlation between the emergence of genotype NA2 and higher chances for reinfection, which may eventually lead to large outbreaks of HRSV in Japan. Furthermore, we would like to emphasize the importance of long-term molecular epidemiological surveys for early detection of newly emerging genotypes and the benefit of studies to elucidate the genetic and antigenic mechanisms underlying reinfection for better understanding of HRSV infections.

ACKNOWLEDGMENTS

We thank the staff at Sano Pediatric Clinic who assisted with collection of the clinical information and samples; Naohito Tanabe and Asami Sasaki and the staff in the Department of Public Health, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan, for help with the study design and statistical analysis; Akemi Watanabe for her excellent technical assistance; Akinori Miyashita and Ryozo Kuwano in the Center for Bioresources, Brain Research Institute, Niigata University, for utilization of the DNA sequencer; and Miwako Saikusa in Yokohama Municipal Laboratory for valuable advice and providing positive controls to initiate this study.

REFERENCES


