Original Article

Genotyping and Phylogenetic Analysis of the Major Genes in Respiratory Syncytial Virus Isolated from Infants with Bronchiolitis

Kazuko Goto-Sugai, Hiroyuki Tsukagoshi, Katsumi Mizuta, Shunji Matsuda, Masahiro Noda, Toshiyuki Sugai, Yoshihiro Saito, Nobuhiko Okabe, Masato Tashiro, Kunihisa Kozawa, Ryota Tanaka, Yukio Morita, Atsuyoshi Nishina, and Hirokazu Kimura

National Hospital Organization Yokohama Medical Center, Kanagawa 245-8575;
Gunma Prefectural Institute of Public Health and Environmental Sciences, Gunma 371-0052;
Yamagata Prefectural Institute of Public Health, Yamagata 990-0031;
National Hospital Organization Ehime National Hospital, Ehime 791-0281;
Influenza Virus Research Center, National Institute of Infectious Diseases, Tokyo 208-0011;
Department of Pediatrics, Jikei University School of Medicine, Tokyo 105-8461;
Department of Surgery, Institute of Medical Sciences, Kyorin University, Tokyo 181-8611; and
Yamagata Prefectural Yonezawa Women’s Junior College, Yamagata 992-0025, Japan

(Received April 16, 2010. Accepted August 23, 2010)

SUMMARY: We performed the genotyping and phylogenetic analysis of respiratory syncytial virus (RSV) isolated from 17 infants with bronchiolitis in Kanagawa Prefecture, Japan, in 2005 and 2006. The major genes in these samples (attachment [G] glycoprotein gene, fusion [F] protein gene, and nucleoprotein [N] gene) were sequenced and analyzed genetically. Phylogenetic analysis of these genes revealed that 7 and 10 strains could be classified into subgroups A and B, respectively. Phylogenetic analysis of the G gene revealed that the subgroup A and B strains were unique genotypes GA2 and BA, respectively. Moreover, the amino acid sequences for these genotypes suggested a relatively high frequency of amino acid substitutions in the G and F proteins in these strains, whereas the N protein was highly homologous. These results suggest that RSV genotypes GA2 and BA may be associated with bronchiolitis in the cases studied here.

INTRODUCTION

Respiratory syncytial virus (RSV; genus Paramyxoviridae) causes acute respiratory infection (ARI), which is the most common disease in children around the world (1,2). RSV can also cause severe obstructive pulmonary disease such as bronchiolitis and bronchopneumonia (3–5). However, despite the fact that RSV is a common cause of bronchiolitis or pneumonia, particularly in infants (aged under 1 year), the underlying mechanisms leading to such conditions remain unknown (6). RSV infections can be life-threatening in these cases (2,7), and recent studies have suggested that RSV infections might trigger or exacerbate airway hyperresponsiveness, including asthma (8,9). Indeed, it is generally recognized by pediatricians that virus-induced asthma caused by RSV or other viruses can result in refractory airway disease (10).

The RSV genome encodes around 10 proteins, the major antigens of which are attachment (G) glycoprotein and fusion (F) glycoprotein (2). As nucleocapsid (N) protein is also an essential structural protein of antigenic replicative intermediate RNA (2), these major proteins might mediate infection of the host cell and RNA replication.

Molecular epidemiological studies focusing on the analysis of G gene have suggested that RSV is classified into subgroups A and B (11). On the basis of this classification, it has been suggested that a specific genotype, GA3, might be associated with a significantly greater severity of illness (12), although other reports have suggested that the severity of illness is linked to the amount of RSV in nasopharyngeal aspirate rather than the viral subgroup (13,14). Thus, the association between a specific virus type and the severity of RSV infections, including bronchiolitis, has not been accurately addressed. In light of this, we performed genetic analysis of the major genes (G, F, and N genes) of RSV isolated from infants (aged under 1 year) with RSV-related bronchiolitis.

MATERIALS AND METHODS

Patients: Each of the 17 patients (mean age ± standard deviation [SD], 4.7 ± 3.4 months) examined for this study exhibited fever and wheeze on initial examination. Chest X-rays of all patients showed the typical hyperinflation and reticulogranular pattern indicative of bronchiolitis. None of the patients had a history of contact with each other. All patients were admitted to the Fujisawa City Hospital or the National Hospital Organization Yokohama Medical Center and treated

---

*Corresponding author: Mailing address: Gunma Prefectural Institute of Public Health and Environmental Sciences, 378 Kamioi-machi, Maebashi-shi, Gunma 371-0052, Japan. Tel: +81-27-232-4881, Fax: +81-234-8438, E-mail: tsuka-hiro@gref.gunma.jp

1These authors contributed equally to this study.
with infusion, oxygen, and beta2-agonist or epinephrine nebulization. Patient data are summarized in Table 1.

**Virus isolation:** HEP-2 (ATCC CCL-23), HEL (WI-38 strain, ATCC CCL-75), RD (ATCC CCL-136), and MDCK cells (ATCC CCL-34) were obtained from the American Tissue Culture Collection (Rockville, Md., USA) and grown in Dulbecco's modified Eagle minimal medium (Invitrogen, Carlsbad, Calif., USA) containing 10% fetal bovine serum at 37°C in a humidified atmosphere of 5% CO₂ for cell culture and viral isolation (15). Throat swabs from patients were inoculated into these cell lines grown in 24-well microplates (Corning, Corning, N.Y., USA) and incubated at 33°C (for isolation of influenza virus) or 36°C in a humidified atmosphere of 5% CO₂ (15). Viruses were isolated as described previously (15). No significant presence of any pathogen other than RSV was detected in any of the patients.

**Reverse transcription-polymerase chain reaction (RT-PCR):** Viral suspensions (isolates) were centrifuged at 3,000 × g at 4°C for 30 min for RNA extraction, RT-PCR, and sequence analysis. The supernatants (culture supernatants) were used for RT-PCR and sequence analysis, as described previously (16–19). Briefly, RSV RNA was extracted from 140 μL of the culture supernatants using a QIAamp Viral RNA Mini Kit (Qiagen, Valencia, Calif., USA). The resulting RNA solution was treated with DNase I (Takara, Tokyo, Japan), and the reverse transcription reaction mixture was incubated with random hexamer primers at 42°C for 75 min, followed by incubation at 99°C for 5 min, and then amplified by thermal cycling.

The G gene of RSV was the target for the external and seminested PCRs. External PCR was carried out with primers ABG490 and F164. The forward primer, ABG490 (5'-ATGATTTWCAATTTGAGTTGTC-3'), corresponded to bases 497–519 of the G protein gene of strain A2, and bases 491–513 of the G protein gene of strain 18537. The reverse primer, F164 (5'-GGTATGACACTGGTTATACCAACC-3'), corresponded to bases 164–186 of the F protein gene of strain 18537 (with one mismatch with the G protein gene of strain A2), and has previously been used to amplify the G protein genes of both groups (16). Amplification was carried out at 94°C for 1 min, followed by 35 cycles of 94°C for 40 s, 50°C for 45 s, and 72°C for 45 s, with a final extension at 72°C for 10 min. One microliter of diluted external PCR product was used for the seminested PCR. Subgroup A-specific primer AG655 (5'-GATC YCAAACCTCAACCCAC-3'), corresponding to bases 655–674 of the G protein gene of strain A2 for subgroup A, and subgroup B-specific primer BG517 (5'-TTGGTT CCCTGTAGTATATGTG-3'), corresponding to bases 517–538 of the G protein gene of strain 18537 for subgroup B, were used as forward primers, and F164 was used as the reverse primer for the seminested PCR (17). Amplification was carried out at 94°C for 1 min, followed by 40 cycles of 94°C for 40 s, 54°C for 45 s, and 72°C for 45 s, with a final extension at 72°C for 10 min. The nested amplicons were 450/585 bp and 645 bp for group A/B and BA viruses, respectively. The primers for F gene were designed as follows: the anterior half of the F gene (nucleotides 3–1,068 of the N terminal region) was amplified at 94°C for 1 min, followed by 35 cycles of 94°C for 60 s, 50°C for 60 s, and 72°C for 60 s, with a final extension at 72°C for 10 min, using the RSV-U primer (5'-GGCAATAAACATGGAGTTG-3') and the RSV-4R primer (5'-AAAGGAATGACTGATCTG-3') (18). The primers for N gene were as follows: RSVN3 (5'-GCC GCA GCT ACT ACA AGC AG-3'), corresponding to bases 426–451 of the N gene of strain A2, and RSVN5 (5'-TCG GGC GGG ACC TGG ACC TC-3'), corresponding to bases 748–773 of the N gene of strain A2 (19). Amplification was carried out at 94°C for 1 min, followed by 35 cycles at 94°C for 40 s, 68°C for 30 s, and 72°C for 45 s, ending with extension for an additional 10 min at 72°C. Purification of the DNA fragments and determination of the nucleotide sequences was performed as described previously (20).
Phylogenetic and deduced amino acid sequence analysis: Phylogenetic analysis based on the nucleotide sequences of the \( G \), \( F \), and \( N \) genes was performed as described previously (11,16). The nucleotide sequence of \( G \) gene spanned bases 673–912 (240 bp) of prototype strain A2 (GenBank accession no. M11486) (11). For group B viruses, the sequence corresponded to bases 670–963 (294 bp) of a genotype BA strain from Argentina (strain BA4128/99B; GenBank accession no. AY333364) (11). The \( F \) gene is located within the F2 subunit and the 550 nucleotides of \( F \) gene (nucleotides 23–572 of the N-terminal region) were found to be distinct (21,22). The nucleotide sequences of the partial \( N \) gene of RSV (positions 452–747: 296 bp) were analyzed as described previously (19) using the CLUSTAL W program on the DNA Data Bank of Japan (DDBJ) homepage (http://www.ddbj.nig.ac.jp/Welcome-j.html) and TreeExplorer (version 2.12) (http://evolgen.biol.metro-u.ac.jp/TE/). Evolutionary distances were estimated using Kimura’s two-parameter method, and phylogenetic trees were constructed using the neighboring (NJ) method (20,23). The reliability of the tree was estimated using 1,000 bootstrap replications and intercluster frequency distributions were calculated using pairwise genetic distances for all strains, as described previously (24).

Prototype strains were used to compare the deduced amino acid sequences of these genes encoding the proteins, as described previously (11,16). Thus, strains A2 (for subgroup A, GenBank accession no. M11486) and BA4128/99B (for subgroup B, GenBank accession no. AY333364) (11) were used to compare G protein sequences, whereas strains A2 (for subgroup A, GenBank accession no. M11486) and 18537 (for subgroup B, GenBank accession no. D00334) (16) were used to compare F protein. Finally, strains A2 (for subgroup A, GenBank accession no. M11486) and 18537 (for subgroup B, GenBank accession no. D00736) were used to compare N protein sequences.

RESULTS

Phylogenetic and amino acid analysis of \( G \) gene/G protein: The phylogenetic tree based on the nucleotide sequences of the \( G \) gene is shown in Fig. 1. Seven and 10 of the 17 RSV strains were classified to subgroups A and

![Phylogenetic tree based on G gene sequences of various RSV strains.](image)
Subgroup A

G protein as position 220

301

A2 strain
RSV/Kanagawa, JPN/Sept 05/Ka-3
K 153K K K 136K K 153K K K

RSV/Kanagawa, JPN/Nov 05/Ka-9
K 153K K K 153K K 153K K K

RSV/Kanagawa, JPN/Oct 05/Ka-10
K 153K K K 153K K 153K K K

RSV/Kanagawa, JPN/Jan 06/Ka-12
K 153K K K 153K K 153K K K

Fig. 2. Amino acid alignments of G protein gene from subgroup A (A) and subgroup B (B) strains. Alignments are shown relative to the sequences of prototype strain A2 and prototype strain BA4128/99B strain. Amino acids shown correspond to strain A2 G protein, positions 220 to 298 for the group A viruses, or to strain BA4128/99B G protein, positions 219 to 315 for the group B viruses. Identical residues are indicated by dashes.

B, respectively. Moreover, all strains belonging to subgroup A were located in genotype GA2, and all strains belonging to subgroup B were genotyped as BA. The homology between A2 (prototype strain) and the present GA2 type strains was 85.8–88.3% at the nucleotide level and 73.1–78.2% at the amino acid level, whereas the homology between the prototype of genotype BA (BA4128/99B strain) and the present genotype BA strains was 94.6–98.0% at the nucleotide level and 90.8–95.9% at the amino acid level. The mean (SD) intercluster distances in this tree were 0.049 ± 0.025 and 0.033 ± 0.018 between the present subgroup A and B strains, respectively. The amino acid sequences of subgroup A and B strains were compared to the prototype A2 and BA4128/99B strains, respectively (Fig. 2). Relatively frequent amino acid substitutions were seen in the present strains belonging to both subgroups. Moreover, when G protein amino acid sequences were compared for the present genotype GA2 strains and Long strain (a representative prototype subgroup A strain), the positively selected changes Pro226Leu, Ser269Thr, Pro289Ser, and Pro290Leu were seen in all the genotype GA2 strains (25).

Phylogenetic and amino acid analysis of F gene/F protein: The phylogenetic tree based on the nucleotide sequences of the F gene is shown in Fig. 3. Similarly to the G gene, 7 and 10 strains were classified into subgroups A and B, respectively. The homology between A2 strain and the present strains was 90.2–94.1% at the nucleotide level and 84.2–92.4% at the amino acid level, whereas the homology between the prototype of subgroup B (18537 strain) and the present strains was 94.7–97.3% at the nucleotide level and 88.0–98.4% at the amino acid level. The mean (SD) intercluster distances between the present subgroup A and B strains in this tree were 0.062 ± 0.026 and 0.031 ± 0.016, respectively. The amino acid sequences of subgroup A and B strains were compared to those of A2 and 18537 strains, respectively (Fig. 4), and clusters of amino acid substitutions of F protein were found in the present strains belonging to both subgroup A and B.

Phylogenetic and amino acid analysis of N gene/N protein: As was the case with the G and F genes, 7 and 10 strains were classified into subgroups A and B, respectively (Fig. 5). The homology of N gene between the A2 strain and the present strains was 96.6–97.0% at the nucleotide level and 98.0–99.0% at the amino acid level, whereas the homology between the prototype of subgroup B (18537) strain and the present strains was 97.6–98.0% at the nucleotide level and 99.0–100% at the amino acid level. The mean (SD) intercluster distances in this tree were 0.010 ± 0.010 and 0.001 ± 0.001 between the present subgroup A and B strains, respectively. In addition, one amino acid substitution (at nt 1783, from histidine [His] to tyrosine [Tyr]; amino acid [aa] 216) was found in the present subgroup A strains with respect to the A2 strain (representative subgroup A strain); no substitution was found upon comparing our subgroup B strain with the 18537 strain (representative subgroup B strain). Thus, the present study appears to suggest that the high degree of homology for the N gene in RSV may be associated with bronchiolitis in infants.

**DISCUSSION**

Phylogenetic analysis of the G, F, and N genes from the 17 RSV isolates obtained from Japanese infants with bronchiolitis revealed that 7 and 10 strains could be classified into subgroups A and B, respectively. The phylogenetic tree based on the G gene sequences showed that all strains belonging to subgroups A and B could be classified into types GA2 and BA, respectively. There was greater nucleotide divergence between the G and F genes in the present subgroup A strains than in the subgroup B strains. In addition, the amino acid sequences of the G and F genes indicated that substitutions were relatively frequent with respect to the prototype strains. In contrast, the amino acid sequences of the N gene were highly conserved compared with the prototype.
strains. In the present cases, RSV strains with a relatively wide genetic diversity as regards the G and F genes might be associated with bronchiolitis.

Previous studies have suggested that G and F proteins are the major antigens of RSV and are linked to neutralization, antigenicity, and virulence (16,18,21). Thus, the phylogenetic analysis of RSV G gene by Martinello et al. showed that the GA3 genotype is associated with greater severity of illness in, for example, bronchiolitis and pneumonia (12). Similarly, correlations between certain strains and/or genotypes of RSV and slight differences in disease severity have been described by Hall et al. and Walsh et al. (26,27). In recent years, some genotypes, such as GA2, GA5, and BA, have been reported to be prevalent throughout the world (11,28–30). However, to gain a better understanding of the molecular epidemiology of RSV, it may be important to determine which RSV genotypes circulate in different areas. In the present study, two RSV genotypes (GA2 and BA) were found to be associated with bronchiolitis, in accordance with previous reports. In addition, the severity of illness caused by subgroup A isolates did not differ from that caused by subgroup B isolates in the present study. Indeed, previous reports have suggested that the severity of the illness is not linked to subgroups or genotypes but to the quantity of RSV in nasopharyngeal aspirate (13,31), and a very recent study suggested that the ectodomain of G protein was subject to strong positive selection, with 29 positively selected amino acid sites in RSV subgroup A (25). A comparison of the amino acid sequences of G protein between genotype GA2 and Long strain (a prototype subgroup A strain) found some positively selected sites (Pro226Leu, Ser269Thr, Pro289Ser, and Pro290Leu), whereas Pro226Leu defined genotype GA3 (25). In addition, some of these sites (226 and 290) are known epitopes in RSV subgroup A. Thus, variations of the amino acids at these sites may play a key role in severe respiratory infections such as bronchiolitis. The G protein is a major target (together with the F glycoprotein) of the RSV human immune response. Antigenic and genetic variations have been shown to occur more frequently in G protein than in F protein, thus suggesting a remarkable genetic flexibility in the G protein sequences.

---

**Fig. 3.** Phylogenetic tree based on F gene sequences of various RSV strains. Evolutionary distance was calculated using Kimura’s two-parameter method, and the tree was plotted using the NJ method. Numbers at main branches indicate the bootstrap values of the clusters supported by that branch. GenBank accession numbers of the strains are AB500666 to AB500682.
of RSV. Such a high level of genetic variation might be associated with the fact that RSV plays an important role in facilitating infection in RSV. Riccetto et al. have also demonstrated that the severity of illness of RSV infection in infants is associated with other factors such as body weight and prematurity (32). The association between RSV infection and severity of illness is not yet well understood, thus suggesting the need for additional studies.

Recent studies have shown that RSV antigenic epitopes are insufficiently recognized by innate TLR4-expressing immune cells, which results in low antibody avidity for protective epitopes (33). Awomoyi et al., for example, have suggested that TLR4 polymorphism is linked to symptomatic RSV (34), therefore, both the antigenicity of RSV and the host immune conditions, such as innate immunity, may play important roles in the pathophysiology of severe respiratory infections such as bronchiolitis and pneumonia.

RSV is a major causative agent of severe respiratory infections such as bronchiolitis and pneumonia in infants. The results of the present and other studies suggest that bronchiolitis-associated RSV might show wide genetic diversity, although additional molecular epidemiologic studies are warranted to better understand such RSV infections.

Acknowledgments This work was supported in part by Research on Emerging and Re-emerging Infectious Diseases, Labour and Welfare Programs from the Ministry of Health, Labour and Welfare, Japan (No. 24–19211301).

Conflict of interest None to declare.

REFERENCES

Fig. 5. Phylogenetic tree based on N gene sequences of various RSV strains. Evolutionary distance was calculated using Kimura’s two-parameter method, and the tree was plotted using the NJ method. Numbers at main branches indicate the bootstrap values of the clusters supported by that branch. GenBank accession numbers of the strains are AB370030 to AB370046.


