Genetic Diversity and Molecular Epidemiology of the G Protein of Subgroups A and B of Respiratory Syncytial Viruses Isolated over 9 Consecutive Epidemics in Korea

Eun Hwa Choi¹,a and Hoan Jong Lee¹,2

To study genetic variation and molecular epidemiology of the G protein of respiratory syncytial virus (RSV), 253 strains from a children’s hospital in Korea over 9 consecutive epidemics were analyzed. Restriction analysis of the entire G protein gene demonstrated 24 genotypes among 188 subgroup A and 6 among 65 subgroup B isolates. Two to 4 dominant genotypes of subgroup A cocirculated, and different genotypes predominated in each epidemic. Predominant genotypes were replaced with new genotypes during consecutive epidemics. One of 2 dominant genotypes among subgroup B predominated alternately or concurrently. Phylogenetic analysis revealed that there were multiple lineages, with clustering related to their location and time of isolation among strains from Korea and worldwide. Geographic and temporal distinction have been shown more clearly for subgroup B than subgroup A. These results suggest that the G protein of RSV is continuously evolving, with a distinct pattern presumably due to immune selection in a localized region over time.

Respiratory syncytial virus (RSV) is the major viral cause of lower respiratory tract infections among infants and young children [1]. It has also been recognized as an important etiologic agent that causes pneumonia in elderly patients. Annual epidemics occur during the winter months in temperate regions. In Korea, RSV has contributed to ~60% of all viral lower respiratory tract infections in children’s hospital [2].

RSV has 2 distinct characteristics: it can (1) infect infants in the presence of maternal antibody and (2) cause repeat infections throughout the child’s life [3, 4]. The ability to cause reinfections could be due either to an inadequate immune response or to variability of the virus. The fact that natural infection with RSV provides only limited protective immunity might to some extent be explained by the extensive genetic heterogeneity of RSV isolates. The G protein shows the largest antigenic and genetic differences between 2 antigenic subgroups of RSV and is one of the targets for the neutralizing and protective antibody response [5–8].

Received 1 November 1999; revised 14 February 2000; electronically published 15 May 2000.


Financial support: Seoul National University Hospital Research Fund (04-96-020).

¹Present affiliation: Immunocompromised Host Section, Pediatric Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

²Present affiliation: Immunocompromised Host Section, Pediatric Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.

Reprints or correspondence: Dr. Hoan Jong Lee, Dept. of Pediatrics, Seoul National University Children’s Hospital, 28 Yonjung-dong, Chongno-gu, Seoul 110-744, Korea (hoanlee@plaza.snu.ac.kr).

The Journal of Infectious Diseases 2000; 181:1547–56
© 2000 by the Infectious Diseases Society of America. All rights reserved. 0022-1899/2000/18105-0004$02.00

Materials and Methods

Viruses. Two hundred eighty-four RSV strains were obtained, by use of HEp-2 cell monolayers, from children with lower respiratory tract infections at the Department of Pediatrics, Seoul National University Children’s Hospital, over 9 consecutive epidemics from November 1990 through February 1999. The viruses had been kept frozen at −70°C. A total of 257 strains were suc-
cessfully recovered in HEp-2 cells. The number of strains included in the study represented 90% of the isolates during the study period. Genetic variability was initially analyzed by restriction digestion for 253 strains. For 4 of the strains, the G protein gene could not be amplified, and these strains were excluded from further study. Thirty-five strains were chosen for sequencing of the entire G protein gene after genotypes were defined by restriction analysis.

**RNA extraction.** The frozen stock viruses were propagated in HEp-2 cells. When an extensive cytopathic effect was evident, the viral RNA was extracted from infected cells lysate by use of a kit (RNaid Plus; Bio 101, La Jolla, CA) according to the manufacturer’s directions.

**Reverse transcription (RT)—polymerase chain reaction (PCR) analysis.** Viral RNA prepared from infected cell lysate was used as a template for cDNA synthesis. cDNA was made in a 20-µL reaction volume containing 8.2 µL of viral RNA (1–2 µg/µL), 200 U of avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim, Indianapolis), 20 µL of RNase inhibitor, and 20 pmol of primer F164 (GTTATGACACTGG TATACCAACC) [18] in 1 mM dNTPs, 10 mM Tris-HCl, 50 mM KCl, and 2.5 mM MgCl2. RT-PCR was done with primers F164, G10 (GCAACACTGTCC AAAAAACAG, for subgroup A) and G32 (GCAACCATGTCC AACAACACAG, for subgroup B) [18]. PCR was done in a 100-µL mixture containing 10 µL of 10× PCR buffer (100 mM Tris [pH 8.3], 500 mM KCl, and 15 mM MgCl2), 1 µL of a 25-mM dNTP mixture, 200 ng of each primer, 10 µL of cDNA, and 0.5 µL (2 U) of Taq DNA polymerase (Boehringer Mannheim). The amplification was done for 35 cycles, each of them composed of 1 min at 95°C, 1 min at 55°C, and 1 min at 72°C, and was followed by 7 min of extension at 72°C. The cDNA synthesis and PCR were done following a strict protocol with precautions to prevent contamination. Ten microliters of RT-PCR product was separated on a 0.7% agarose gel and visualized under UV light (302 nm) after staining with ethidium bromide.

**Restriction-enzyme digestion.** The PCR products were analyzed for their genetic diversity of G protein according to digestion patterns by restriction enzymes. Six restriction enzymes, *AluI, Hinfl, MboI, PstI, RsaI, and TaqI*, were used for both subgroup A and B strains. Restriction patterns were analyzed by electrophoresis in 3% agarose gels with ethidium bromide.

**Nucleotide sequencing.** To more accurately define the extent of genetic variability within the subgroups, nucleotide sequences of the whole G protein gene were determined for selected isolates (24 strains of subgroup A and 11 strains of subgroup B). The strains were selected randomly, with an attempt to include 1 or 2 strains of each genotype from each epidemic period. The PCR products were purified by agarose gel electrophoresis, followed by DNA extraction with a kit (GeneClean II; Bio 101). Nucleotide sequencing reactions were done with the Dye Terminator Cycle Sequencing Kit and ABI 373A DNA sequencer (PE Biosystems, Foster City, CA). Sequencing primers for subgroup A were as follows: G10 (sense), G248 (antisense; CTTGATCTGGGTGTTGATC), G267 (sense; GATGCAACAAGCCAGATCAAG), AG546 (sense; CCCTGACAGATATCGAC), AG801 (antisense; GAGTTGGAGGAATTC), and F164 [18, 21, 22]. Primer sequences were provided by W. M. Sullender (University of Alabama at Birmingham) before publication. Nucleotide sequences were confirmed with duplicate reactions that used sense and antisense primers. Nucleotide and deduced amino acid sequences were analyzed with Sequence Analysis version 1.2.1 software (PE Biosystems). The nucleotide sequences determined from the cDNAs of the mRNAs of the G protein genes of 24 subgroup A and 11 subgroup B isolates were submitted to GenBank (see Appendix).

**Phylogenetic analysis.** Phylogenetic analysis was done by the maximum parsimony method with PAUP version 3.1.1 [23], to evaluate geographic and temporal relatedness. A single parsimonious tree, rooted at the midpoint of the greatest patristic distance, was generated. Subgroup A and B strains were analyzed separately. For subgroup A RSV, phylogenetic trees were compared between Korean strains and strains from Madrid (Spain), Montevideo (Uruguay), and Birmingham (United Kingdom) from 1989 through 1994, on the basis of sequences of the entire G protein gene through GenBank [10, 11, 24]. Phylogenetic trees for subgroup B strains were analyzed on the basis of 132 nucleotides of the C-terminal variable region, because it is very difficult to interpret phylogenetic relationships among the limited number of entire G protein gene sequences available from GenBank. Eleven Korean strains were compared with prototype virus, 4 old strains, and 11 isolates from Alabama from epidemics during 1993–1996 [17]. Maximum parsimony trees were evaluated statistically by calculation of bootstrap probabilities for 1000 iterations.

**Results**

**Subgroup distribution.** Strains were divided into 2 antigenic subgroups, A and B, according to the reactivity with monoclonal antibodies developed by the authors. Of the 253 isolates investigated in this study, 188 belonged to subgroup A and 65 to subgroup B, showing a predominance of subgroup A strains. Subgroup A viruses were predominant during 7 epidemic periods, whereas subgroup B viruses dominated during the other 2 periods (1991/1992 and 1998/1999). During 2 epidemics, viruses of only 1 subgroup were isolated (subgroup A in the 1991/1992 season and subgroup B in the 1993/1994 season; tables 1 and 2).

**Genotype distribution.** The various restriction enzyme patterns observed for subgroup A and B are shown in figure 1. For a particular restriction enzyme, a lower-case letter was assigned to each restriction pattern; the letters assigned within each restriction enzyme designation are independent of the letters assigned to others. Similarly, the letters assigned to enzyme patterns for subgroup A are independent of the letters assigned to subgroup B. Each genotype was defined by a combination of the digestion patterns by 6 different restriction enzymes for subgroups A and B separately. A number was assigned to each genotype after the prefix GP-A (for subgroup A) or GP-B (for subgroup B) in decreasing order of frequency (e.g., GP-A1, GP-A2).

The restriction analysis revealed that genetic diversity was greater in subgroup A than subgroup B. Twenty-four genotypes for subgroup A and 6 types for subgroup B were identified.
Table 1. Distribution of various genotypes of G protein gene of respiratory syncytial virus among 188 strains of subgroup A isolated at Seoul National University Children’s Hospital over 9 consecutive epidemics, 1990–1999.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GP-A1</td>
<td>53</td>
<td>4 (19)</td>
<td>15 (50)</td>
<td>14 (45)</td>
<td>4 (17)</td>
<td>2 (15)</td>
<td>1 (4)</td>
<td>3 (11)</td>
<td>10 (70)</td>
<td>16 (103)</td>
</tr>
<tr>
<td>GP-A2</td>
<td>33</td>
<td>6 (29)</td>
<td>6 (20)</td>
<td>1 (4)</td>
<td>4 (13)</td>
<td>3 (23)</td>
<td>1 (4)</td>
<td>7 (58)</td>
<td>7 (56)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>GP-A3</td>
<td>21</td>
<td>6 (24)</td>
<td>11 (48)</td>
<td>6 (20)</td>
<td>1 (4)</td>
<td>3 (12)</td>
<td>1 (4)</td>
<td>2 (15)</td>
<td>2 (15)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>GP-A4</td>
<td>16</td>
<td>7 (33)</td>
<td>8 (24)</td>
<td>1 (4)</td>
<td>4 (13)</td>
<td>3 (23)</td>
<td>1 (4)</td>
<td>7 (58)</td>
<td>1 (6)</td>
<td>1 (6)</td>
</tr>
<tr>
<td>GP-A5</td>
<td>16</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A6</td>
<td>11</td>
<td>2 (7)</td>
<td>8 (26)</td>
<td>6 (20)</td>
<td>1 (4)</td>
<td>5 (19)</td>
<td>4 (15)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A7</td>
<td>9</td>
<td>6 (18)</td>
<td>2 (7)</td>
<td>1 (4)</td>
<td>4 (15)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A8</td>
<td>4</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A9</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A10</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A11</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A12</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A13</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A14</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A15</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A16</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A17</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A18</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A19</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A20</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A21</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A22</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A23</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>GP-A24</td>
<td>1</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Total</td>
<td>188</td>
<td>21 (11.2)</td>
<td>33 (17.4)</td>
<td>30 (15.9)</td>
<td>22 (11.7)</td>
<td>19 (10.1)</td>
<td>14 (7.5)</td>
<td>27 (14.3)</td>
<td>9 (5.3)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE. Data are no. of isolates (% among subgroup A strains in each epidemic period).

The annual distribution for 9 representative genotypes among subgroup A is shown in figure 2A. Three to 9 genotypes were identified during each epidemic period, and 2–4 predominant genotypes accounting for >70% of subgroup A isolates for the respective year circulated concurrently. The dominant genotype(s) differed in each epidemic, and the predominant genotype(s) were replaced with new predominant genotype(s) during consecutive epidemics.

Of particular interest is that these predominant genotypes tended to prevail in 1–6 (mostly 2–4) consecutive epidemics after appearance and then disappeared (figure 2A). GP-A1, the most frequently isolated genotype during the first 7 years, was the predominant genotype during the 1991/1992, 1992/1993, and 1994/1995 epidemics, constituting 41%–50% of the isolates of subgroup A in each epidemic. Then the proportion of the GP-A1 type declined to <20% over the next 2 epidemics, and this type was not detected in the 1997/1998 and 1998/1999 seasons. The GP-A2 genotype was the only one observed in every epidemic in which subgroup A strains were isolated, although the proportions were relatively low, except in the 1998/1999 epidemic. However, sequence and phylogenetic analysis revealed that GP-A2 strains isolated during the 1990/1991, 1991/1992, and 1992/1993 seasons may be a different lineage from those isolated during the 1995/1996, 1997/1998, and 1998/1999 seasons (see below). The GP-A4 type was predominant in the 1990/1991 epidemic and then declined during the subsequent 2 epidemics and was not detected for the latter 6 consecutive epidemics. It is likely that this type has been the predominant lineage involved in epidemic(s) before the 1990/1991 epidemic.

Strains belonging to GP-A3 and GP-A6 genotypes were detected for the first time in the 1994/1995 and 1992/1993 epidemics, respectively, when they made up 19% and 7% of isolates. The proportion of these 2 types in each epidemic of the following years increased to 48% and 26%, respectively, and then declined and disappeared in the subsequent years. GP-A5, GP-A8, and GP-A9 genotypes, which were undetected until the 1996/1997 or 1997/1998 epidemic, predominated in the 1997/1998 epidemic and declined or disappeared in the following year.

Genotypes of subgroup B were quite stable, compared with those of subgroup A. Among the subgroup B strains, 2 genotypes, GP-B1 and GP-B2, were dominant throughout the study period and made up 66% and 25%, respectively, of total subgroup B isolates during the study period (table 2). GP-B1 and GP-B2 genotypes circulated alternately or concurrently in each epidemic (figure 2B).

Nucleotide and amino acid sequence analysis. Among the strains whose G genes were sequenced, there were no identical sequences. Viruses belonging to the same genotype had an overall amino acid difference of <1% (range, 0.1%–0.9%), except in
the case of GP-A2 (see below), whereas strains with different genotypes differed by 2%–7%.

All genetic changes were base substitutions, and no deletion, insertion, or frameshift mutations were identified. Pairwise comparisons of the full sequence of the G protein gene showed that nucleotide sequence homology of subgroup A was 91%–93% between prototype strains (A2, Long) and Korean isolates and 93%–99% among Korean isolates. Nucleotide sequence homology of subgroup B was 91%–96% between prototype strains (CH18537, 8/60) and Korean isolates and 95%–99% among Korean isolates. Deduced amino acid sequence homology of subgroup A was 85%–90% between prototype strains and Korean isolates and 88%–99% among Korean isolates, and those of subgroup B were 87%–90% and 92%–99%, respectively.

Seven of 24 Korean subgroup A strains encoded 297 amino acids, deficient 1 amino acid at the 3' termini, whereas the rest, as well as the prototype strains, encoded 298 amino acids (figure 3A). All Korean subgroup B strains encoded 295 amino acids, adding 3 amino acids at the 3' termini of prototype strains (figure 3B). The amino acid difference within the C-terminal third among Korean isolates was found to be as great as 23% and 22% for subgroup A and B isolates, respectively. Compared with the prototype strains (A2 and CH18537), Korean subgroup A and B strains differed up to 29% and 27%, respectively (figure 3A, 3B).

An average of 52% of the entire nucleotide differences resulted in amino acid coding changes. However, considering the 3' terminal region shown in figure 3, percentages of nucleotide changes resulting in amino acid changes were as great as 65%.

As noted above, 2 genetically distinct types may be included in the genotype GP-A2, which was isolated throughout the study period. G genes of 7 strains with the GP-A2 genotype were sequenced. The number of amino acids of 3 strains isolated in the 1990/1991 or 1992/1993 season (91-142, 92-528, and 93-06) was 297 (arbitrarily termed “early GP-A2”), whereas that of 4 strains isolated in the 1995/1996, 1996/1997, or 1998/1999 season (95-107, 96-308, 98-523, and 98-579) was 298 (arbitrarily termed “late GP-A2”). In addition, pairwise differences between the early and late isolates were as great as 1.9%–5.8%, whereas the differences in nucleotides within each group were <1% (0.2%–0.9%). This suggestion is supported by the phylogenetic tree, which has shown that these 2 groups constitute distinct separate lineages (figure 4A). Thus, it is possible that strains belonging to the GP-A2 genotype that was circulating since the 1995/1996 season may be genetically distinct from the strains of GP-A2 isolated by the 1992/1993 season, although they were not distinguished by restriction analysis with enzymes used in this study.

**Phylogenetic analysis.** Both subgroups A and B, belonging to different lineages, correlated well with the genotypes we described (figure 4). Phylogenetic trees of subgroup A based on the entire G sequence showed multiple lineages and 3 broad clusters (figure 4A). The prototype strains belonged to 2 separate clusters in which some of the strains from Montevideo and Birmingham were included. All isolates from Seoul and Madrid appeared in the cluster distinct from those that included prototypes. Korean strains could be grouped into 4 broad clusters. Strains from other geographic area were partly overlapped with genotypes GP-A1, GP-A3, and GP-A19 of Korean strains. However, most strains tended to cluster by geographic areas as subdivided by detailed sublineages with significant bootstrap probability. None of the strains from other geographic areas included here were related to GP-A4 and GP-A22 genotypes of Korean isolates. Geographic clustering was more evident when Korean strains were separately compared with those from each of the above 3 areas (trees not shown).

In the phylogenetic tree of subgroup B strains based on the 132 nucleotides of the C-terminal region, distinct separation was evident between isolates from Alabama and Korea, although they were isolated during temporally close epidemics (figure 4B). Temporal distinction was also observed between old strains and recent strains from Seoul and Alabama.

**Discussion**

RSV is divided into subgroups A and B and further into antigenic variants within each subgroup by reactivity with a panel of monoclonal antibodies developed against both subgroups of RSV. The greatest differences between 2 RSV subgroups are seen in the G protein [25, 26]. Earlier studies have focused on demonstration of antigenic or genetic variation within the same subgroup of RSV for a limited number of

| Table 2. Distribution of various genotypes of G protein gene of respiratory syncytial virus among 66 strains of subgroup B isolated at Seoul National University Children’s Hospital over 9 consecutive epidemics, 1990–1999. |
|----------|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| GP-B1    | 43    | 1 (100)   | 9 (90)    | 9 (100)   | 4 (57)    | 10 (100)  | 10 (77)   |           |           |           |
| GP-B2    | 16    | 10 (77)   |           |           | 2 (100)   | 3 (43)    | 1 (8)     |           |           |           |
| GP-B3    | 3     | 1         |           |           |           |           |           |           |           |           |
| GP-B4    | 1     | 1         |           |           |           |           |           |           |           |           |
| GP-B5    | 1     | 1         |           |           |           |           |           |           |           |           |
| GP-B6    | 1     |           |           |           |           |           |           |           |           |           |
| Total    | 65    | 13        | 0         | 1         | 10        | 9         | 2         | 7         | 10        | 13        |

**NOTE.** Data are no. of isolates (% among subgroup B strains in each epidemic period).
strains that were temporally and geographically diverse in their origin. It was also demonstrated that both subgroups A and B and their antigenic or genetic variants cocirculate in a single epidemic. Only recently, a few studies looked at quantitative changes in genetic variants of RSV strains isolated during consecutive epidemics in the same region.

Cane et al. [12] examined genetic variability among RSV strains isolated over 5 consecutive epidemics in Birmingham. They found that multiple lineages were circulating in a single epidemic, different lineages predominated in each epidemic, and some lineages of subgroup A strains appeared to increase in numbers over several years and then decline, possibly indicating a buildup of resistance in the community to a particular genotype [12, 27]. Coggins et al. [17] observed similar findings by restriction-fragment analysis of the G gene and nucleotide sequence determination of a variable region of the G gene of RSV subgroup A isolated over 3 successive epidemics. Peret et al. [28] also demonstrated that a shift of the predominant ge-

Figure 1. Restriction-fragment analysis of amplified cDNAs of the G protein gene of subgroup A (A) and subgroup B (B) respiratory syncytial virus. Polymerase chain reaction (PCR) products were digested with AluI, HincII, MboI, PstI, RsaI, and TaqI and were separated by 3% agarose gel electrophoresis. PstI digestion for subgroup B is not shown, because it did not cleave subgroup B products. Lanes M, PCR molecular-weight markers; sizes (top to bottom), 1000, 750, 525, 500, 400, 300, 200, 100, and 50 bp, respectively.
Figure 2. Distribution of representative genotypes of the G protein gene of respiratory syncytial virus among 188 strains of subgroup A (A) and 66 strains of subgroup B (B) isolated at Seoul National University Children’s Hospital over 9 consecutive epidemics, 1990-1999. Genotypes with >3 isolates are shown. The vertical axis represents percentages of each genotype in each subgroup. Genotypes are indicated at the right side of the figure. *Nucleotide sequence data and phylogenetic comparison suggest that GP-A2 genotype strains isolated by the 1992/1993 season might be genetically distinct from those isolated since the 1995/1996 season (see text).

notype or subtype occurred each year in the phylogenetic analysis of the variable region of the G gene over 5 successive epidemics.

Our data are generally consistent with the above studies, in that multiple genotypes cocirculated in each epidemic and a shift occurred in the predominant genotype(s) of the G gene of subgroup A viruses over successive epidemics. However, this study, spanning 9 years, shows more clearly the genetic displacement of RSV subgroup A, with the disappearance of several genotypes and appearance of new genotypes of the G protein within the same community over 9 consecutive epidemics. Many of the genotypes with a few isolates were identified during only 1 epidemic period, but most of the predominant genotypes were isolated in more than 2 consecutive epidemics. They tended to increase in numbers over several years and decline.

Compared with the genotypic displacement of subgroup A over successive epidemics, subgroup B viruses were relatively stable. In this study, 2 dominant genotypes circulated alter-
Figure 3. Alignment and comparison of deduced amino acid sequences of C-terminal third of the G protein gene of respiratory syncytial virus subgroup A (A) and subgroup B (B) strains isolated in Korea. Sequences of the G protein gene of A2 and CH18537 are shown: for other viruses, only differences relative to A2 and CH18537 are indicated. Genotypes are indicated at the left of the strain no. *Termination codon.

nately or concurrently throughout the study period. In the other studies mentioned above, the data for subgroup B strains were too limited to look at the trend among subgroup B strains.

The G protein is involved in the neutralizing and protective antibody response. The fact that a high percentage of nucleotide changes resulted in amino acid changes suggests that there may be a selective advantage to G protein changes [15, 19]. However, it is unclear that genetic change is subject to antigenic difference [13, 22, 29]. Although it is not clear whether the shift in genotypes of G gene can explain repeated infections, consistent displacement of genotypes over consecutive epidemics suggests that disappearance of a particular genotype may be caused in part by increased levels of herd immunity to that genotype.

Sequence analysis of numerous RSV isolates, mostly subgroup A, has shown extensive variability of the G glycoprotein [13, 15, 19, 21]. Viruses from antigenic subgroup A belonged to different lineages that correlate with previously identified genotypes [13, 16]. As noted above, viruses from different lineages were isolated during the same epidemic and in the same place.

The genotypes we described are based on digestion patterns by restriction enzymes, and sequences were analyzed for a limited number of strains. Therefore, there are limitations in the interpretation of the results. The degree of differences in the sequences between genotypes may be quite variable, as seen in the phylogenetic tree of figure 4; some genotypes are widely separated, whereas some are closely related (e.g., in figure 4A, GP-A5, GP-A8, and GP-A9 and GP-A2 and GP-A11; in figure 4B, GP-B2 and GP-B3 and GP-B1 and GP-B6). However, among the viruses whose G protein genes were sequenced, differences in the sequences of amino acids belonging to the same genotype were !1% (0.1%–0.9%), except in the case of GP-A2, whereas those among different genotypes were 2%–7%. This fact suggests that genotypes we described reflect the sequences quite well except for GP-A2. In case of the GP-A2 genotype, the sequence differences among 3 isolates before the 1992/1993 season and among 4 isolates after the 1995/1996 season were both !1%, whereas the differences between isolates before the 1992/1993 season and after the 1995/1996 season were 1.9%–5.8%. This is compatible with differences seen among or within other genotypes and suggests that GP-A2 strains that were circulating since the 1995/1996 season may be genetically distinct from the strains of GP-A2 isolated by the 1992/1993 season.

Phylogenetic analysis based on nucleotide or amino acid sequences revealed that RSV subgroup A evolves with multiple
lineages. There is insufficient data about the evolutionary pattern of subgroup B because, in general, subgroup B strains are isolated less frequently than are subgroup A strains. However, phylogenetic trees derived from several subgroup B strains also show multiple evolutionary lineages. A recent report observed the correlation of the G protein length and the position of viruses in the phylogenetic tree [20].

Although both subgroups of RSV exhibit multiple evolutionary lineages, our data suggests that there may be difference in evolutionary patterns between subgroups A and B shown by yearly genotype distribution and phylogenetic comparison. Subgroup A RSVs may have greater ability to diverge against selective pressure, which may explain greater epidemics of subgroup A than subgroup B [2, 12, 30].

There have been conflicting data regarding whether evolution of RSV is related temporally or geographically. Restriction mapping and partial sequencing of selected segments of the SH, N, and G genes revealed that similar viruses are present simultaneously in widely separated countries [11, 31]. Generally, it is understood that RSV genotypes may have a worldwide distribution, and viruses isolated in distant places and in slightly different years may be more closely related than viruses isolated in the same place on 2 consecutive days [32, 33]. Analysis of the sequences of the variable region of the G gene of RSV isolated over 38 years in the different parts of the world showed that clustering was more closely related to their date of isolation than to the location of isolation [19]. These findings suggest that the evolutionary pattern of subgroup A correlated with temporal rather than geographic variation.

On the other hand, the G protein gene sequences of strains isolated recently from the Gambia and Cuba were distinct from those of contemporary isolates in the developed world [34, 35].
In this study, subgroup A viruses with similar lineages were found throughout Korea, Spain, Uruguay, the United Kingdom, and the United States during the same or different epidemic periods. However, Korean strains were distinct, in that Korean isolates for 9 years were not related to prototype strains and some genotypes of subgroup A among Korean strains were not related to any strains from other geographic areas. These findings were more evident when Korean strains were separately compared with those from other individual regions—Madrid, Montevideo, or Birmingham—and, on the basis of the C-terminal region, with those from Alabama [17] (trees not shown).

Strains of subgroup B could be distinguished related to their geographic region and time of isolation as shown in the phylogenetic trees based on the C-terminal region.

Considering the above findings together, it is possible that there may be global spread of the strains, but, afterward, viruses may evolve with a distinct pattern determined by viral strain, host, and local characteristics.

In summary, we have described the evident genetic replacement of the G protein of subgroup A strains that were isolated in the same region through nine successive epidemics. We observed that there are distinct lineages of the G protein among both subgroups A and B between geographic regions. These findings may suggest the positive immune selection in the same region.

Acknowledgments

We thank W. M. Sullender and W. B. Coggins (University of Alabama at Birmingham) for providing the oligonucleotide sequences for the polymerase chain reaction analysis, K. J. Song (Korea University College of Medicine) for assistance with analysis of phylogenetic trees, and K. H. Park (Clinical Research Institute, Seoul National University Hospital) for assistance with nucleotide sequence determination.

Appendix

GenBank Accession Numbers

Accession numbers for 24 subgroup A strains are as follows: SEL/03/91087, AF193304; SEL/04/91142, AF193305; SEL/05/91242, AF193306; SEL/10/91399, AF193307; SEL/01/92011, AF193308; SEL/09/92308, AF193309; SEL/10/92359, AF193310; SEL/11/92415, AF193311; SEL/12/92528, AF193312; SEL/01/93006, AF193313; SEL/02/93057, AF193314; SEL/11/94118, AF193315; SEL/12/94191, AF193316; SEL/01/95026, AF193317; SEL/11/95103, AF193318; SEL/12/95118, AF193319; SEL/12/95107, AF193320; SEL/05/96113, AF193321; SEL/08/96308, AF193322; SEL/12/97385, AF193323; SEL/12/97434, AF193324; SEL/02/98072, AF193325; SEL/09/98523, AF193326; and SEL/10/98579, AF193327. Accession numbers for 11 subgroup B strains are as follows: SEL/01/91001, AF193328; SEL/03/91080, AF193329; SEL/03/91088, AF193330; SEL/03/91110, AF193331; SEL/04/91141, AF193332; SEL/10/93366, AF193333; SEL/12/93418, AF193334; SEL/03/94055, AF193335; SEL/12/94212, AF193336; SEL/02/95043, AF193337; and SEL/11/95100, AF193338.

References

9. Johnson PR, Collins PL. The 1B (NS2), 1C (NS), and N proteins of human respiratory syncytial virus (RSV) of antigenic subgroups A and B: sequence conservation and divergence within RSV genomic RNA. J Gen Virol 1989;70:1539–47.
19. Cane PA, Pringle CR. Evolution of subgroup A respiratory syncytial virus:


