Natural History of Human Respiratory Syncytial Virus Inferred from Phylogenetic Analysis of the Attachment (G) Glycoprotein with a 60-Nucleotide Duplication

Alfonsina Trento,† Mariana Viegas,† Mónica Galiano, Cristina Videla, Guadalupe Carballal, Alicia S. Mistchenko, and José A. Melero*

Unidad de Biología Viral, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain; Laboratorio de Virología, Hospital de Niños Dr. Ricardo Gutiérrez, Gallo 1330, C1425EFD Buenos Aires, Argentina; and Laboratorio de Virología, Centro de Educación Médica e Investigaciones Clínicas, CEMIC, Hospital Universitario, Av. Galván 4102, C1431FWO Buenos Aires, Argentina

Received 17 August 2005/Accepted 20 September 2005

A total of 47 clinical samples were identified during an active surveillance program of respiratory infections in Buenos Aires (BA) (1999 to 2004) that contained sequences of human respiratory syncytial virus (HRSV) with a 60-nucleotide duplication in the attachment (G) protein gene. This duplication was analogous to that previously described for other three viruses also isolated in Buenos Aires in 1999 (A. Trento et al., J. Gen. Virol. 84:3115–3120, 2003). Phylogenetic analysis indicated that BA sequences with that duplication shared a common ancestor (dated about 1998) with other HRSV G sequences reported worldwide after 1999. The duplicated nucleotide sequence was an exact copy of the preceding 60 nucleotides in early viruses, but both copies of the duplicated segment accumulated nucleotide substitutions in more recent viruses at a rate apparently higher than in other regions of the G protein gene. The evolution of the viruses with the duplicated G segment apparently followed the overall evolutionary pattern previously described for HRSV, and this genotype has replaced other prevailing antigenic group B genotypes in Buenos Aires and other places. Thus, the duplicated segment represents a natural tag that can be used to track the dissemination and evolution of HRSV in an unprecedented setting. We have taken advantage of this situation to reexamine the molecular epidemiology of HRSV and to explore the natural history of this important human pathogen.

Human respiratory syncytial virus (HRSV)—the prototype of the Pneumovirus genus within the Paramyxoviridae family—is a major cause of lower respiratory tract infections in children (19, 20, 21) and in vulnerable adults (12, 14, 15). HRSV infections are clustered in epidemics that occur during the winter months in temperate countries or during the rainy season in tropical regions (reviewed in reference 9). A characteristic feature of HRSV is that moderate levels of antibody do not provide lasting protection. Thus, reinfections are common throughout life, although prior infections reduce the severity of the disease (19).

The viral genome, a negative-sense single-stranded RNA molecule, encodes at least 11 distinct proteins, two of which (G and F) are the major surface glycoproteins anchored in the viral membrane (for a review, see reference 9). The G protein is the attachment protein that mediates virus binding to cells (26). The F glycoprotein mediates fusion of the viral and cell membranes for virus entry into the cell and fusion of the infected cell membrane with that of adjacent cells to promote syncytia formation (53).

The G protein is a type II glycoprotein that shares neither sequence nor structural features with the attachment proteins (HN or H) of other paramyxoviruses (54). It is produced in two different forms in the infected cell: (i) a membrane-bound form (Gm) of 292 to 319 amino acids (depending on the viral strain) that is incorporated into virions and (ii) a soluble form (Gs), lacking the signal and membrane anchor region, that is secreted from infected cells. Gs is formed by alternative translation initiation from a second in-frame AUG codon in the G protein open reading frame (M48) (38). Gs remains monomeric, whereas Gm forms oligomers, probably tetramers (13).

Variability of HRSV isolates was first demonstrated at the antigenic level in a neutralization test performed with hyperimmune serum (7). Different panels of monoclonal antibodies were later used to classify HRSV isolates into two antigenic groups, groups A and B (1, 31), which correlated with genetically distinct viruses (10). Further studies of HRSV evolution by sequence analysis have focused mainly in the G protein gene for two reasons: (i) the G protein is the viral gene product with the highest antigenic and genetic variability between HRSV isolates from either the same or different antigenic groups, and (ii) the G protein is one of the targets for the neutralizing and protective antibody response (for a review, see reference 29).

Accumulated sequence data has provided insights into the pattern of HRSV evolution, which resembles that of influenza B viruses (28, 40, 56), with the following characteristics. (i) Epidemics are commonly produced by viruses of the two antigenic groups, classified into more than one genotype. At a local level, the predominant genotype(s) of one epidemic is frequently replaced by other genotypes in consecutive years (4). (ii) Most HRSV genotypes are distributed worldwide. In addition, viruses isolated in distant places and different years...
may be more closely related than viruses isolated in the same place and during the same epidemic (17). (iii) Within the same genotype, there is a tendency for accumulation of sequence changes with time in the G protein gene (5, 17). These changes are concentrated in two hypervariable regions that flank a highly conserved central region of the G protein ectodomain, which includes a cluster of four cysteines. (iv) Whereas synonymous nucleotide changes are distributed uniformly throughout the G protein gene, nonsynonymous changes accumulate preferentially in the two variable regions of the G protein, lending support to the idea of a positive selection for some of those changes (29, 55, 57, 58).

Recently, three antigenic group B viruses were reported (49) that were isolated in Buenos Aires (BA) in 1999 and had 60 nucleotides duplicated in the G protein gene. Subsequently, G protein sequences with the same duplication were found in more recent clinical samples from distantly related places (25, 32, 42, 43, 57, 58) (sequences from the Quebec samples are available in GenBank, accession number AY927401 [18]). We now report 47 additional samples, isolated in Buenos Aires between 1999 and 2004, that contain HRSV G sequences with the duplicated segment. Phylogenetic analysis provided strong evidence that all G protein sequences with the 60-nucleotide duplication derived from a common ancestor, dated about 1998. We have taken advantage of this 60-nucleotide natural tag to reexamine HRSV evolution during propagation in its natural host.

MATERIALS AND METHODS

Clinical specimen. Nasopharyngeal aspirates were taken in phosphate-buff ered saline from children hospitalized in the “Hospital de niños Dr. Ricardo Gutierrez” of Buenos Aires during six consecutive years (1999 to 2004) (51, 52). Rapid detection of viral antigens was performed by indirect immunofluorescence with specific antibodies, and the samples were then stored at −80°C until further analysis. Samples positive for HRSV were classified into groups A and B by multiplex nested reverse transcriptase PCR (RT-PCR) as described previously (44). Virus nomenclature used in this study uses a letter code representing the isolate site of isolation (e.g., BA represents Buenos Aires) followed by the isolate number, dated about 1998. We have taken advantage of this 60-nucleotide natural tag to reexamine HRSV evolution during propagation in its natural host.

RNA extraction, DNA amplification, and sequencing of group B G sequences. One microliter of glycogen (Sigma) was added to samples prior to RNA extraction. Total RNA was extracted directly from 300 µl of frozen clinical specimen by the guanidinium isothiocyanate method followed by precipitation with isopropanol in 0.75 M ammonium acetate (6). The dried RNA pellet was dissolved in 90 µl of 0.1% diethyl pyrocarbonate-treated water.

Seven microliters of total RNA was used as a template for RT-PCR amplification with a commercial kit (QIAGEN). Oligonucleotide F164 (5′-GTTATGA GATGATTACCATTTTGAAGTGTTCA-3′; nucleotides 141 to 164 of the F protein gene; negative sense) was used as the primer of the reverse transcription reaction that was carried out at 50°C for 30 min followed by 15 min at 95°C to inactivate RT and to activate HotStart Taq DNA polymerase. Oligonucleotides F164 and OG1-21 (5′-GGGGCAATGCAACCATGTCC-3′; nucleotides 1 to 21 of the G protein gene from the CH18537 strain; positive sense) were used as primers for amplification of the full-length G gene. Alternatively, F164 and OG1-21 (5′-GATGATTACCATTTTGAAGTGTTCA-3′; nucleotides 496 to 519 of the G protein gene; positive sense) were used to amplify the C-terminal half of the G protein gene, nonsynonymous changes accumulate preferentially in the two variable regions of the G protein, lending support to the idea of a positive selection for some of those changes (29, 55, 57, 58).

RNA amplification products were then purified using the commercial kit Wizard PCR preps DNA purification system (Promega) and subjected to forward and reverse cycle sequencing with the BigDye terminator sequencing kit (Applied Biosystems). Besides the PCR primer sets, an additional negative sense primer, CH253 (5′-GG TTTTATATGTGTGAACT-3′; nucleotides 253 to 233 of the G protein gene), was used to complete the G nucleotide sequence.

Phylogenetic analysis and genotype distribution. Maximum-likelihood phylogenetic trees were obtained using PAUP+ package version 4.0b10 (47). The best-fitting nucleotide substitution model was determined from the data. A total of 31 partial sequences (270 nucleotides long) of HRSV group B viruses isolated worldwide, and previously assigned to specific genotypes, were retrieved from GenBank for comparative studies. Similarly, 37 sequences (330 nucleotides long) with a 60-nucleotide duplication in the G protein gene were downloaded from the same database.

Nucleotide and amino acid sequence analysis. Sequences were aligned with ClustalX version 1.81 (48) and manually edited with BioEdit version 7.0.1. The nucleotide and amino acid pair-wise distance values (p distances) were calculated using MEGA version 3.0 (24). The inferred amino acid sequences (represented using universal code) and polymorphisms were analyzed with DNAp 4.0 software (41).

Evolutionary rate and date of most recent common ancestor. These parameters were calculated by plotting the isolation date of HRSV samples versus the genetic distance of the corresponding sequences, calculated as the number of nucleotide substitutions from the putative node of divergence (“root”). The best-fitting line was determined by regression analysis. The root sequence was set to maximize the R² value of the regression analysis. The slope of the regression line was used to calculate the evolutionary rate, and the absissa crossing point was taken as the date of the most recent common ancestor for all the sequences under analysis. The evolutionary rate and the date of the most recent common ancestor were also calculated by the maximum-likelihood method using the TipDate software package version 1.2 (37). An initial tree was constructed under the TN+F model without assuming a molecular clock, and the root of the tree was set at the sequence that maximized the likelihood of the SRDT model in the TipDate software.

Nucleotide sequence accession numbers. The sequences reported here were deposited in the GenBank database under accession numbers DQ227363 to DQ227408.

RESULTS

Clinical samples. About 17,000 samples of nasopharyngeal aspirates were tested by indirect immunofluorescence for respiratory viruses during an active surveillance program of respiratory infections in a public hospital of Buenos Aires from 1999 to 2004 (51, 52). A total of 353 of those samples were positive for HRSV and were classified in antigenic group A or B by multiplex nested RT-PCR of the G protein gene, as described by Stockton et al. (44). As found in other studies done in Argentina (16, 52), group A sequences were detected more frequently than group B sequences throughout the study period, although group B samples were more abundant than group A sequences in 1999 and 2002 (Table 1). HRSV group
B infections were not detected in the year 2000 (52).

The amplified DNA products of 47 samples, out of 121 group B amplicons, migrated more slowly in agarose gels than the DNA of other group B samples but comigrated with the amplified DNA of three previously reported viruses (49) with a 60-nucleotide duplication in the G protein gene (not shown). The number of samples that yielded a slow-migrating amplified DNA product (presumably as result of a 60-nucleotide duplication that was later confirmed by DNA sequencing) represented 14.3% of all group B samples in 1999 (Table 1). The same percentage was seen in 2001 (although with a lower number of samples), and that value increased to 51.1% in 2002, 50.0% in 2003, and 60.0% in 2004. Thus, the percentage of samples that yielded a slow-migrating DNA band increased considerably during the six epidemics analyzed in this study.

Phylogenetic relationship of BA sequences with the 60-nucleotide duplication. Nucleotide sequences of the entire G protein gene could be obtained for 35 samples that yielded the slow-migrating DNA band after RT-PCR amplification (Table 1). In other cases, only partial sequences could be obtained due to availability of the clinical samples and the sample from 2001 yielded no readable sequence after several attempts. The complete G sequences were subjected to phylogenetic analysis. The three previously reported sequences with the 60-nucleotide duplication, from BA viruses isolated in 1999 (49), were also included in this analysis. All new sequences exhibited the duplicated segment of 60 nucleotides, starting after residue 792 of the G protein gene (numbers refer to the sequence of the group B strain CH18539) (23). No other insertions or deletions were observed to compensate for the extra 20 amino acids added to the G protein by the duplicated sequence. Interestingly, whereas the duplicated segment in the 1999 samples was an exact copy of the preceding 60 nucleotides, nucleotide changes accumulated in the two copies of the duplication in more-recent viruses (see below).

The maximum-likelihood method was used to construct a phylogenetic tree of complete BA sequences with the 60-nucleotide duplication (Fig. 1). Some of the samples (BA/3833/99 and BA/4128/99; BA/1606/02 and BA/1208/02; BA/619/02 and BA/164/02; BA/1562/02 and BA/495/02) yielded identical sequences, but in most cases unique sequences were obtained from each amplified DNA. The sequences could be classified into at least three branches (BA-I, BA-II, and BA-III), with significant (>70) bootstrap values. The sequences in BA-I included all the samples from 1999 (7 samples) and 12 samples from 2002. The remaining 10 sequences from 2002 samples were clustered in BA-II and BA-III with other sequences from 2003 and 2004 samples. Two sequences from 2004 were grouped in a separate branch (BA-IV) with a low bootstrap value; however, the bootstrap increased to significant values when further sequences were included in the analysis (Fig. 2).

Overall identity values of the G sequences represented in Fig. 1 were 98.7% at the nucleotide level and 97.6% at the amino acid level, reflecting higher sequence variation at the protein than at the RNA level, as observed generally for HRSV isolates (27, 39, 45, 50). Pair-wise comparison of those sequences indicated that transitions were more frequent than transversions (Ti/Tv = 3.97).

Comparison of BA sequences with those of other HRSV samples with the 60-nucleotide duplication isolated worldwide. After the initial description of HRSV group B viruses with the 60-nucleotide duplication in the G protein gene (49), other laboratories have reported partial G sequences from clinical samples with a duplicated 60-nucleotide segment. The viruses present in all those samples were classified within antigenic group B of HRSV on the basis of either group-specific RT-PCR tests or nucleotide sequencing. The different reports included samples isolated in Sapporo in 2000 (32) and 2002 (25), in Niigata in 2003 (42), in Kenya in 2003 (43), in Quebec in 2001 and 2002 (18), and in Belgium from 1999 to 2004 (58). Notably, all these samples were isolated more recently than those of the first reported group B viruses with the 60-nucleotide duplication (June and August 1999) (49). In all cases, the borders of the duplicated segment were the same. As noted with BA samples, some degree of sequence divergence was observed in most cases within the two copies of the duplicated segment, except for very early samples. Although all these data suggested that viruses with a duplicated G segment shared a common ancestor and originated from a single duplication event, the occurrence of several independent duplication events could not be formally excluded.

To test whether or not viruses with the 60-nucleotide duplication have common ancestry, a phylogenetic analysis was
FIG. 2. Phylogenetic tree of HRSV group B constructed by the maximum-likelihood method. A total of 46 partial nucleotide sequences of the G protein gene (nucleotides 652 to 982) with the 60-nucleotide duplication from samples isolated in BA were aligned using the CLUSTAL software. Of these, only 30 were unique sequences (some unique BA sequences of Fig. 1 were identical in the partial segment used here) that were compared with the same segment of unique sequences with the duplication reported from Quebec (QUE), Niigata (NG), Sapporo (SAP), Kenya (Ken), and Belgium (BE). These samples are marked with color-coded diamonds in the figure, according to countries. Partial sequences (nucleotides 652 to 922) of reference group B genotypes GB1, GB2, GB3, GB4, SAB1, SAB2, SAB3, URU1, and URU2 and the reference strain CH18537 were included in the analysis for comparison. The length of horizontal lines is proportional to the genetic distance between viruses. The bar represents one 0.02 nucleotide substitution per site, and the tree was unrooted. The numbers at the internal nodes indicate the number of bootstrap probabilities (500 replicates) (shown only for values greater than 70). Names of viruses refer to the place/year of isolation, when known. Genotypes are indicated at right by brackets. Within BA genotype, branches BA-I to BA-VI are shown also by brackets. The number of BA sequences identical to those shown in the figure is indicated between parentheses at right of the sample name.
made of this type of sequence from samples isolated worldwide. In addition, 31 sequences of different HRSV group B genotypes without the duplicated segment were included in the analysis for comparison. Some of the later sequences were obtained from samples isolated in Buenos Aires, or in proximal places (Montevideo, Uruguay) (2, 16), and were contemporary to the BA sequences with the 60-nucleotide duplication. Since most group B sequences reported to date included only C-terminal partial sequences of the G protein gene, sequences of the last 270 nucleotides (or 330 nucleotides for sequences with the duplication) were used to generate the phylogenetic tree of Fig. 2.

All sequences with the 60-nucleotide duplication (labeled with colored diamonds in Fig. 2) were clustered in a new genotype (BA), separated from those identified previously among HRSV group B isolates. Conversely, none of the prototype sequences of previously reported group B genotypes was clustered with the sequences of the new BA genotype. These results lend strong support to the hypothesis of a common ancestor for all sequences with the duplicated G segment.

When the first or second copy of the duplicated sequences was excluded from the phylogenetic analysis (to rule out possible disturbances introduced in the analysis by the duplication), the overall structure of the phylogenetic tree of Fig. 2 remained unchanged (not shown). In addition, although the BA genotype has a low bootstrap value in Fig. 2 (due to the use of short partial sequences), this parameter increased significantly (>70) when complete G sequences, only available for a subset of the samples, were used in the analysis (not shown).

Two new branches (BA-V and BA-VI) were identified within the BA genotype of Fig. 2 that had not been recognized previously in the tree of Fig. 1. BA-V included only three sequences from samples isolated during an HRSV outbreak in Niigata in 2003, and BA-VI contained almost exclusively sequences from samples isolated in Belgium, except one from Quebec and another from Sapporo. Notably, the BA-III branch shown in Fig. 2 contained only sequences from samples isolated in Buenos Aires from 2002 to 2004. The most heterogeneous branch, regarding the place and date of the samples, was BA-IV. This branch included the sample BE/12445/99, which was reclassified into BA-I when a tree was made with the sequence of the entire G protein ectodomain, as available for the Belgian samples (not shown). In contrast, clustering of the sequences from the Argentinian samples included in Fig. 2 remained essentially unchanged when the complete G sequences were considered (see Fig. 1).

Sequence variation within the 60-nucleotide duplication. The alignment of partial nucleotide sequences containing the duplicated segment from samples clustered within the BA genotype of Fig. 2 is shown in Fig. 3. As mentioned before, the duplication had two identical copies only in some very early samples. In all other cases, a limited number of nucleotide substitutions had accumulated in either copy of the duplicated segment, except in the second copy of the BA-I samples. The transition T754C was found in the sequences of all branches, except BA-I. Other nucleotide substitutions were restricted only to certain branches or were specific of either single sequences or limited sequence subsets. Overall, an accumulation of nucleotide changes with time was observed in the sequences of Fig. 3.

The alignment of partial amino acid sequences including the duplicated G segment is also shown in Fig. 3. In general, variations at the amino acid level reflected the changes observed in the nucleotide sequences. For instance, the amino acid substitution S247P was found in the sequences of all branches, except BA-I. Other amino acid changes were branch specific or restricted to certain sequences. The use of alternative stop codons (not shown in the figure), as reported for other group B viruses (27), generally determined the protein length. However, other unusual changes, outside the duplicated segment, also contributed to G protein length polymorphism. For instance, three sequences (BA/1161/02, BA/4862/03, and BA/5140/03) had three As inserted after residue 712, adding an extra Lys in position 234 of the G protein. In addition, the sequence BA/100/04 had six nucleotides deleted after residue 489. Insertions and deletions of other triples have been found in the Belgian sequences included in Fig. 3 (58), and in one of these samples (BE/13159/02) a premature stop codon shortened the G polypeptide to 282 amino acids. Despite these changes, an overall correlation between the classification of sequences in the different branches and protein length could be seen (Fig. 3).

It is worth mentioning that the residue Thr229 (outside the duplicated segment) that is conserved in all group B sequences reported to date was changed to Ile (T229I) in all sequences of the BA genotype (except in BA/733/02, for which the change was T229F). This observation also supports the notion of a common ancestor for viruses with the 60-nucleotide duplication.

To compare the variability of the duplicated segment with respect to other regions of the G protein gene, the nucleotide sequences of the BA samples shown in Fig. 1 were aligned using the CLUSTAL software. Then, sequence changes were calculated using nonoverlapping windows of 60 nucleotides, framed to match the duplicated segment (Fig. 4A). Accordingly, this analysis could only be done with complete BA sequences. The percentages of nucleotide changes differed considerably between different regions of the G protein gene, following the general pattern observed in previous studies (29). Two regions, encoding the transmembrane domain (nucleotides 132 to 191) and the cysteine cluster segment of the G protein ectodomain (nucleotides 492 to 551), were conserved in all BA sequences. In contrast, high levels of sequence variation were observed in other parts of the G protein gene. Notably, the first copy of the duplicated segment (nucleotides 732 to 791; black bar, Fig. 4A) showed a significantly higher level of nucleotide variation than other parts of the G protein gene, including the second copy of the duplicated segment (nucleotides 792 to 851).

Date of the most recent common ancestor and rate of evolution of sequences with the 60-nucleotide duplication. To estimate the date of the most recent common ancestor for BA viruses with the duplicated segment, the phylogenetic branch length (p distance) calculated in Fig. 1 was plotted against the year of isolation of each sample (Fig. 5). A straight line could be fitted to the actual data by linear regression analysis. The root sequence was set to maximize the $R^2$ value of the regression analysis. The origin of the common ancestor was dated (by extrapolating to abscissas) between 1998 and 1999, i.e., shortly
FIG. 3. Partial sequences of the G protein gene with the 60-nucleotide duplication. The sequences of samples with the 60-nucleotide duplication used to construct the tree of Fig. 2 were aligned using the CLUSTAL software. Nucleotide sequences are shown at left of the figure, and the translated amino acid sequences are shown at right, with the branch name in the middle. The two copies of the duplicated sequences are framed by rectangles. Identical residues are indicated by dots. The name of samples and the color code are the same as described for Fig. 2. The protein length is shown at the right of each partial amino acid sequence.
before the date of the first viral isolates with the 60-nucleotide duplication (June to August 1999).

The rate of evolution of the G protein gene calculated from the slope of the regression line was $3.0 \times 10^{-3}$ nucleotide substitutions/site/year, with a good correlation between branch length and sampling time ($R^2 = 9.173$). The values calculated by the linear regression analysis correlated well with those obtained by the alternative maximum-likelihood method under a single-rate model (SRDT). In this case, the date of the most common recent ancestor for BA sequences was 1999 (95% confidence interval, 1997 to 1999) and the rate of evolution was $2.5 \times 10^{-3}$ nucleotide substitutions/site/year (95% confidence interval, 1.7 to $3.5 \times 10^{-3}$ nucleotide substitutions/site/year).

The same type of plotting was made with the partial sequences of all samples isolated worldwide that contained the duplicated segment (shown in Fig. 2). The high dispersion of data, due to the limited number of nucleotides used in the analysis, prevented reaching statistically significant values (not shown). Nevertheless, the date of the most recent common ancestor (1998) and the rate of evolution ($3.0 \times 10^{-3}$ nucleotide substitutions/site/year) estimated by regression analysis of all sequences with the 60-nucleotide duplication were very similar to those calculated with the complete BA sequences.

**DISCUSSION**

It has been postulated that a relatively stable secondary structure of the viral RNA, formed transiently during the replication process, could account for the generation of the duplicated segment in the HRSV G protein gene (49). Although this is probably a rare genetic event—which extends the list of unusual sequence changes observed in the G protein gene of HRSV isolates (29, 58)—it might have occurred more than once during HRSV evolution. The fact that the borders of the duplicated segment were identical in all the sequences with the 60-nucleotide duplication (Fig. 3) might simply reflect structural or functional constraints for that change in other parts of the G protein gene. However, the observation that all HRSV G sequences with the 60-nucleotide duplication were clustered in the same genotype (BA) of Fig. 2—that excluded all other group B sequences without the duplication—provided strong evidence for a common ancestry and offered us an unprecedented setting to reevaluate the epidemiological features of HRSV. Since the pattern of evolution of HRSV sequences with the duplicated segment was not different from that observed generally for group A and B viruses (for reviews, see references 3 and 46), Fig. 2 provides an exceptional example of the natural history of HRSV during six consecutive years.

At a local level, Fig. 1 reflects the circulation of BA viruses during the six consecutive years of this study. Whereas viruses from the 1999 epidemic were all closely related to the putative ancestor (BA-I branch) and contained an exact copy of the duplicated segment, viruses circulating in later epidemics were grouped into four separate branches (BA-I, BA-II, BA-III, and BA-VI). This result exemplifies the replacement of a locally predominant HRSV genotype in successive epidemics, as reported previously for group A and B viruses without the 60-nucleotide duplication (4, 8, 34, 35). It is worth noting that...
BA-I viruses were apparently extinct after 2002 in Buenos Aires (Fig. 1) and in other places (Fig. 2). It may be speculated that incorporation of further nucleotide substitutions in the G protein gene led to enhancement of viral fitness and replacement of the original viruses. Conspicuously, the transition T754C was found in the sequences of all BA branches, except BA-I. In addition, the second copy of the duplicated segment contained changes in the sequences of branches other than BA-I.

Figures 2 and 3 reflect the global dissemination of BA viruses since they originated about 1998 (Fig. 5). As mentioned before, viruses of the BA-I branch circulated extensively in Buenos Aires from June to August of 1999. In December of the same year, the first BA-I sequence with an exact copy of the duplicated segment from a non-Argentinean sample was found in Belgium (58). Thus, it is clear that BA-I viruses crossed the Atlantic (in either direction) in a period of few months. In later years, viruses that had accumulated limited number of changes in the duplicated segment were circulating in very distant places, providing direct evidence for the rapid dissemination of HRSV worldwide and the generation of genetic variability.

The locations where BA viruses have circulated as well as the corresponding genotypes and years of circulation are illustrated in the map of Fig. 6. Other places where circulation of BA viruses has also been reported but where sequences of those viruses are still unavailable are also indicated in the map. Obviously, the routes followed by the BA viruses to reach the different places are not known but may have imposed certain bottleneck effects, with an impact in HRSV evolution, which contributed to the exclusive dissemination of certain branches in locally restricted areas. For instance, BA-III viruses were found only in Buenos Aires and BA-V viruses were found only in Niigata (Fig. 2).

BA-IV viruses were not found in Buenos Aires before 2004. Since these viruses were closely related to those circulating in other places (particularly Quebec) during previous years, and shared with them an unusual protein length, it is likely that they were imported from some of those places. Consequently, two types of viruses with the duplicated G segment circulated in Buenos Aires in 2004: (i) those of the BA-III branch that probably evolved from previous local viruses, since sequences of this branch have not been reported from other places, and (ii) those of the BA-IV branch that were probably imported. This complex situation may confound the conclusions reached in other phylogenetic studies, as highlighted by Cane (3), when the origin of a common ancestor is uncertain and emphasizes the weight of the conclusions that may be reached from this study.

The high mutation rate of HRSV, as for RNA viruses in general (11), probably contributed to generate the genetic variation observed among sequences of the BA genotype. However, the rapid accumulation of sequence changes with time in the G protein of these viruses (Fig. 5) is indicative of selective forces acting to accelerate the genetic drift. Notably, the evolutionary rate calculated for BA sequences (2.5 to 3.0 × 10^{-3} nucleotide substitutions/site/year; Fig. 5) is significantly higher than that reported for other HRSV sequences of both antigenic groups: 1.83 × 10^{-3} for group A and 1.95 × 10^{-3} for group B (22, 57, 58). It is worth mentioning that immunological positive selection at certain sites of the HRSV G protein ectodomain has been claimed to operate during HRSV evolution (29, 55, 57, 58). Interestingly, some of the positively se-
lected sites reported for group B viruses are located in the duplicated G segment or in nearby sequences (58). This may have contributed to the accelerated accumulation of mutations in the duplicated segment compared with the results seen with other segments of the G protein gene (Fig. 4A and 4B).

Although G protein gene sequences with duplications or deletions of 1 to 2 triplets have been reported previously (references 2, 17, 29, 30, and 58 and this study), they were found sporadically in certain epidemics and lost in later years. Thus, it is remarkable that viruses with the largest duplication described so far (60 nucleotides) have been circulating for at least 6 years and have replaced other prevailing genotypes in certain places. For instance, sequences with the 60-nucleotide duplication account for 50 to 60% of all group B viruses circulating in Buenos Aires in recent years (Table 1). Similarly, viruses with sequences of the BA genotype have predominated among group B viruses in recent epidemics in Belgium (58) and in Japan (42). The selective advantage of BA viruses over other genotypes is unknown, but, as argued above, the immunologically naive condition of the human population with respect to these new viruses may have contributed to their rapid dissemination worldwide. Further studies will determine whether or not BA viruses reach extinction in future years, as has been found with other predominant HRV genotypes. In this respect, surveillance programs such as the one implemented for this study may be of instrumental importance in tracing the fate of HRV with the duplicated G segment. Knowledge gained from this type of study may have an impact not only on our understanding of HRV evolution but also on other areas, such as vaccine development.

In summary, as exemplified in this study, the 60-nucleotide duplication tag found frequently in the G protein gene of group B HRV isolates offers an unique opportunity to study the natural history of this virus and the selecting forces that drive evolution of this important human pathogen.

ACKNOWLEDGMENTS

We are most grateful to Pat Cane for critical reading of the manuscript and further comments. We acknowledge Patricia Rivero, Beatriz Ebekian, and Carmen Ricarte for excellent technical support, the personnel of the Genomic core facility (ISCHI, Majadahonda, Madrid, Spain) for running the sequence samples, and our colleagues who deposited in GenBank some of the sequences used in this study.

This work was supported in part by grants from Ministerio de Educación y Ciencia, Spain (SAF2003-08250), Instituto de Salud Carlos III, Spain (03/ESP30-1).

A.S.M. is a member of “Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC),” G.C. is a member of “Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET),” and M.V. is a Fellow of “CONICET.”

REFERENCES