Genetic Variability and Molecular Evolution of the Human Respiratory Syncytial Virus Subgroup B Attachment G Protein

Kalina T. Zlateva, Philippe Lemey, Elien Moës,† Anne-Mieke Vandamme, and Marc Van Ranst*

Laboratory of Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, University of Leuven, Leuven, Belgium

Human respiratory syncytial virus (HRSV) is the most important cause of acute respiratory disease in infants. Two major subgroups (A and B) have been identified based on antigenic differences in the attachment G protein. Antigenic variation between and within the subgroups may contribute to reinfections with these viruses by evading the host immune responses. To investigate the circulation patterns and mechanisms by which HRSV-B viruses evolve, we analyzed the G protein genetic variability of subgroup B sequences isolated over a 45-year period, including 196 Belgian strains obtained over 22 epidemic seasons (1982 to 2004). Our study revealed that the HRSV-B evolutionary rate (1.95×10^{-3} nucleotide substitutions/site/year) is similar to that previously estimated for HRSV-A (1.83×10^{-3} nucleotide substitutions/site/year). However, natural HRSV-B isolates appear to accommodate more drastic changes in their attachment G proteins. The most recent common ancestor of the currently circulating subgroup B strains was estimated to date back to around the year 1949. The divergence between the two major subgroups was calculated to have occurred approximately 350 years ago. Furthermore, we have identified 12 positively selected sites in the G protein ectodomain, suggesting that immune-driven selective pressure operates in certain codon positions. HRSV-A and -B strains have similar phylodynamic patterns: both subgroups are characterized by global spatiotemporal strain dynamics, where the high infectiousness of HRSV permits the rapid geographic spread of novel strain variants.

Human respiratory syncytial virus (HRSV) is the most common causative agent of serious respiratory tract infections in infants and young children (6, 16, 20, 25). HRSV is also increasingly recognized as an important agent of disease in the elderly (11), in immunocompromised patients (10, 12) and even in the general adult population (9, 21, 39).

HRSV is an enveloped virus with a nonsegmented negative-sense RNA genome of approximately 15,200 nucleotides and belongs to the genus *Pneumovirus* (6). Two major antigenic groups, HRSV-A and -B, have been identified on the basis of reactions with monoclonal antibodies (1, 38). Only antibodies directed against the G and F surface glycoproteins have been shown to be neutralizing in vitro or protective in vivo (7, 57, 64). Nucleotide sequencing has revealed that the attachment G protein is the most divergent between and within the two subgroups (27). The G protein is a type II integral protein of 289 to 299 amino acids, depending on the viral strain (13, 32, 55). The mature 90-kDa form of the G protein is obtained through extensive glycosylation of a 32-kDa polypeptide precursor by the addition of N-linked sugars to Asn residues and O-linked sugars to Ser and Thr residues (67, 68). The attachment G protein can be divided into an intracellular domain, a transmembrane domain, and a large ectodomain. The latter is comprised of two hypervariable mucin-like regions separated by a highly conserved putative receptor-binding site (4, 27, 55).

HRSV infections occur continually throughout life, which is indicative for partial cross-immunity against different strains (18, 23, 69). Children initially infected with an HRSV-A strain appear more likely to experience reinfection with an HRSV-B strain (37, 65). These findings suggest that antigenic variability of the G protein between and within the two major HRSV antigenic groups may facilitate reinfection by evading the pre-existing host immune response (37, 54).

Most reports on the genetic variability and evolution of the attachment G protein are based on HRSV-A strains (2, 13, 52, 75), and there is little information available about the genetic diversity and molecular evolution of HRSV-B strains (29, 32, 55, 59). This can be explained by the fact that HRSV-A strains are generally more often isolated during HRSV outbreaks (1, 19, 24, 36). The predominance of HRSV-A over HRSV-B viruses has been attributed to the higher variability among the HRSV-A strains (5, 46, 49, 61). Differences in the extent of intragenetic diversity in the G protein between the two HRSV subgroups may reflect differences in the evolutionary patterns of these viruses. We have previously reported 13 amino acid sites to be under positive selection in the G attachment protein of HRSV-A, suggesting that evolution under immune selective pressure occurs in certain codon positions (75). A detailed understanding of the evolutionary mechanisms employed by HRSV to evade the host immune responses will be beneficial in predicting possible changes in virulence and will be pivotal in vaccine development programs.

---

* Corresponding author. Mailing address: Laboratory of Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, University of Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium. Phone: 32 16 347908. Fax: 32 16 347900. E-mail: marc.vanranst@uz.kuleuven.ac.be.
† Present address: School of Biology, University of St. Andrews, St. Andrews, United Kingdom.

In the current study we investigated the genetic diversity, circulation patterns, and mode of evolution of HRSV-B viruses. We have therefore sequenced 724 bp of the attachment G protein from 196 subgroup B HRSV strains obtained over 22 epidemic seasons (1982 to 2004) in Belgium.

MATERIALS AND METHODS

Viral isolates. Nasopharyngeal secretions were collected during 1982 to 2004 from patients between 10 days and 73 years of age (mean age, 13 months) who were admitted with a lower respiratory tract illness to the Gasthuisberg University Hospital in Leuven, Belgium. Samples were included in this study when they reacted positively with an HRSV antigen test (TestPack RSV; Abbott Laboratories, Abbott Park, IL) or exhibited a typical HRSV cytopathic effect in cell cultures (HeLa and HEp-2). HRSV-positive nasopharyngeal aspirates and HRSV-infected cells were stored at −80°C. The nomenclature adopted for the Belgian isolates indicates the place of isolation (Belgium, BE), followed by a laboratory isolate number and the epidemic season (e.g., BE/14610/03-04).

PCR primer sequences. HRSV-B-specific oligonucleotide primers were designed based on available GenBank subgroup B sequences of the attachment G protein and the intergenic G-F region. The forward primer BGF located in the G gene (5′-GCAGCCATAATATTCATCATCTCT-3′) and the reverse primer BGR located in the intergenic G-F region (5′-TGCCCCAGRTTTAATTTCGT TC-3′) correspond to nucleotides (nt) 4858 to 4881 and 5637 to 5658, respectively, in strain WV/B1/85 (GenBank accession number AF013254). The expected size of the PCR product was 0.8 kb.

One-step RT-PCR. Total RNA was extracted from 140 μl of HRSV-positive sample using a QIAmp viral RNA mini kit (QIAGEN, Westburg, The Netherlands) and was dissolved in 60 μl of elution buffer. The reverse transcription (RT)-PCR assay was performed with a OneStep RT-PCR Kit (QIAGEN) in a 50-μl volume containing 30 pmol of forward and reverse primer and 10 μl of the extracted RNA. Thermocycling was conducted on a GeneAmp PCR system 9600 thermal cycler (Applied Biosystems, Foster City, Calif.) programmed as follows: 55°C for 30 min for reverse transcription and 94°C for 15 min for DNA-polymerase activation; 40 cycles of 94°C for 30 s, 63°C for 1 min, and 72°C for 1 min, followed by a final extension step at 72°C for 10 min. The amplified products were subjected to 6% polyacrylamide gel electrophoresis. The 0.8-kb amplicons were purified using a QIAquick PCR Purification Kit (QIAGEN).

DNA sequencing. The purified PCR products were sequenced in forward and reverse direction using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). Besides the PCR primer set, an additional forward primer (5′-AGAGACCCAAAAACACYAGCCAA-3′) and reverse primer (5′-ACAGGGAACGAAGTTGAACACTTCA-3′), both located in the G protein gene, were used to obtain the full-length sequence of the amplified products. Sequencing analysis was performed on an ABI PRISM 3100 DNA sequencer (Applied Biosystems).

Sequence analysis. The chromatogram sequences were inspected with Chromas 2.2 (Technelysium, Helensvale, Australia), and contigs were prepared using SeqMan II (DNASTAR, Madison, Wis.). Multiple sequence alignment of the HRSV-B G gene sequences included in this study was performed using CLUSTAL X, version 1.83 (58), and manually edited in the GeneDoc version 2.6.002 alignment editor (41). Identical sequences and amino acid variability were estimated using the DAMBE (data analysis in molecular biology) program, version 4.2.13 (71). O glycosylation sites were determined using the program NetOglyc version 3.1 (22, 28).

Phylogenetic analysis. The appropriate nucleotide substitution model was determined using hierarchical likelihood ratio testing (47). Phylogenetic reconstruction was performed using the PAUP (phylogenetic analysis using parsimony) package version 4b10 (56). Model parameters were optimized on an initial neighbor-joining tree. Using the estimated parameters and the neighbor-joining tree as a starting tree, a heuristic maximum likelihood (ML) search was performed, evaluating tree topologies generated by both nearest neighbor-interchange and tree bisection-reconnection.

Evolutionary rate analysis. An exploratory root-to-tip linear regression was performed in Path-O-Gen, kindly provided by Andrew Rambaut (University of Oxford). This program takes a phylogenetic tree as input and performs a linear regression between the genetic distance from the root and the sampling time for each strain. The root of the tree was assigned on the long branch separating the early samples from the others. A more accurate estimate for the evolutionary rate and the most recent common ancestor (MRCA) was obtained using the single rate dated tips (SRDT) model, a statistically well founded likelihood method (48). The ML tree was rooted at the position that yielded the highest

likelihood under the SRDT model. For the amino acid data set including both HRSV-A and -B sequences, the phylogeny was obtained by connecting the ML trees of both groups at the positions that yielded the highest likelihood under the SRDT model in their separate analyses. The likelihood ratio test comparing the SRDT model with the different rates model was used to test the molecular clock. These analyses were performed using the TipDate program (48) for nucleotides and the PAML (phylogenetic analysis by maximum likelihood) program package version 3.13 (74) for amino acids.

Selective pressure analysis. Positively selected sites were identified under probabilistic models of codon substitution that allow for variable nonsynonymous/synonymous substitution rate ratios (dN/dS, or ω) among sites (42, 72, 73). The likelihood ratio test was used to determine whether allowing for sites with ω of >1 significantly improved the fit of the model to the data. If the dN/dS ratio for any site class was above 1, the Bayes theorem was used to calculate the posterior probability that each site, given its data, belonged to such a site class. Three different model comparisons were performed: the neutral model M1, which assumes a proportion p_0 of negative (conserved) sites with ω_0 of 0 and a proportion $p_1 = 1 - p_0$ of neutral sites with ω_1 of 1, was tested against the selection model M2, allowing for an extra class of sites under diversifying selection with frequency $p_2 = 1 - p_0 - p_1$ and with ω_2 of >1 estimated from the data. The selection model M2 does not allow for sites with $0 < \omega < 1$. Model M0 allowing for one ω for all sites was tested against model M3 with a discrete distribution of three site classes (with proportions p_0, p_1, p_2 and $\omega_0, \omega_1, \omega_2$). M7, incorporating a beta distribution with parameters p and q to account for variable ω_0 in the interval (0 to 1) among neutral or negatively selected sites, was tested against model M8, to which an extra component allowing for positively selected sites is added with proportion p_2 and ω_2. All calculations were performed using the CODEML program from the PAML package version 3.13 (74).

Nucleotide sequence accession numbers. The nucleotide sequences from the Belgian isolates were deposited in the GenBank database under accession numbers AY751084 to AY751282.

RESULTS

Genetic diversity in the HRSV-B G protein. Genetic variability was determined by nucleotide sequencing of a 724-bp fragment in the G gene (nt 177 to 900 in reference strain WV/B1/85; GenBank AF013254) of 196 HRSV-B strains isolated in Belgium over 22 epidemic seasons (1982 to 2004). No subgroup B strains were obtained from the epidemic seasons 1983–1984 and 1994–1995. The nucleotide sequences from the Belgian isolates were compared to 15 HRSV-B sequences available in the GenBank database (Table 1).

Among the 196 Belgian strains, there were 124 strains with a unique sequence in the G gene, with 32 of the unique sequences being repeated one or more times among the remaining 72 strains (Fig. 1). An absolutely conserved 23-amino-acid region (amino acids 165 to 187 according to strain WV/B1/85) was recognized among all HRSV-B G protein sequences (Fig. 2A).

There were significant differences in the length of the deduced G protein sequences among the HRSV-B strains. The predicted complete G proteins of the Belgian isolates were of 13 different amino acid lengths (254, 278, 282, 292, 293, 295, 296, 297, 299, 312, 315, 317, and 319 amino acids) (Fig. 2). Differences in the G protein lengths were caused by the usage of alternative termination codons; by the presence of in-frame duplications, deletions and insertions; by one frameshift mutation; and by two premature stop codons. Three alternative stop codons were observed at nucleotide positions 877, 886, and 898 (referring to strain WV/B1/85), leading to G proteins of, respectively, 292, 295, and 299 amino acids (Fig. 2A). A six-nucleotide in-frame deletion after residue 474 was identified in 12 isolates. Due to the usage of the second and third termination codons, two strains with this deletion had G pro-

TABLE 1. HRSV-B GenBank sequences used in this study

| Strain	Accession no.	Isolation place	Year of isolation	Reference
SW/8/60	M73545	Sweden	1960	55
CH/18537/62	M17213	District of Columbia	1962	27
WV/4843/80-81	M73540	West Virginia	1980	55
WV/10010/83	M73541	West Virginia	1983	55
WV/B1/85	AF013254	West Virginia	1985	30
WV/15291/85	M73542	West Virginia	1985	55
NM/1355/89	M73543	New Mexico	1989	55
MON/15/90	AY333361	Montevideo (Uruguay)	1990	59
NY/CH10b/90-91	AF065250	New York	1990	46
NY/CH9b/92-93	AF065251	New York	1992	46
NY/CH18b/92-93	AF065252	New York	1992	46
NY/CH53b/92-93	AF065253	New York	1993	46
Sap/4/00-01	AB117522	Sapporo (Japan)	2000	40
BA/3833/99	AY333362	Buenos Aires (Argentina)	1999	59
BA/3859/99	AY333363	Buenos Aires (Argentina)	1999	59

tein lengths of 293 amino acids (BE/1732/82-83, which is identical to BE/1733/82-83) and another two strains (BE/3455/86-87 and BE/946/86-87) were 297 amino acids long (Fig. 2B). One isolate (BE/450/82-83) with the same deletion also had a single nucleotide gap at nucleotide position 702, resulting in amino acid alterations after residue 233 and G protein length of 254 amino acids (Fig. 2C). During the period 1999 to 2004, we have identified 47 isolates with a 60-nt duplication of the region 721 to 780 according to strain WV/B1/85 (Table 2 and Fig. 2D to F). The same three alternative termination codons were observed among the strains with this duplication leading to protein lengths of 312, 315, and 319 amino acids (Fig. 2D). Seven isolates with the 60-nt duplication had protein lengths of 317 amino acids, due to the six-nucleotide deletion described above and the usage of the third termination codon (Fig. 2E). As a result of a single nucleotide mutation at position 862 (CAG→TAG) (referring to strain BA/3833/99), one strain with the 60-nt duplication (BE/13159/02-03) had a protein length of 282 amino acids (Fig. 2F). In addition to the G protein length polymorphism observed among the strains with the 60-nt duplication, we recognized several nonsynonymous amino acid substitutions in the duplicated region indicated on Fig. 3. A three-nucleotide insertion (GAA) after residue 678 was identified in five isolates from 1995–1996 (Fig. 2G and H), and two other strains from the epidemic seasons 1989–1990 and 1990–1991 had a three-nucleotide (AAA) insertion after residue 702 (Fig. 2I). All isolates with the three-nucleotide insertions with the exception of strain BE/1334/95-96 used the second termination codon and had G proteins of 296 amino acids (Fig. 2G and I). Because of the three-nucleotide GAA insertion and the stop codon-inducing nucleotide substitutions at positions 835 and 837 (CAA→TAG), isolate BE/1334/95-96 had a protein length of 278 amino acids (Fig. 2H).

O-glycosylated sites. Using the program NetOglyc version 3.1 (22, 28), we predicted 105 serine and threonine residues to be potentially O glycosylated in 16.67% to 100% of the HRSV-B viruses, with a best general score predictor (G score) ranging between 0.50 and 0.77 (Fig. 4). Among these potentially O-glycosylated residues, 22 have also been predicted to be the most likely to contain O-linked sugars in HRSV-A strain A2 (6). The amino acid positions of the 22 serine and threonine residues referring to strain WV/B1/85 are 100, 113,

117, 119, 137, 138, 139, 141, 147, 199, 203, 220, 228, 232, 236, 254, 270, 271, 276, 283, 284, and 287.

Phylogenetic clustering and circulation patterns of the Belgian HRSV-B strains. A total of 124 unique Belgian HRSV-B sequences and 15 GenBank-derived HRSV-B sequences were included in the phylogenetic analysis (Fig. 1). The Belgian HRSV-B strains clustered into 13 main lineages designated as genotypes GB1 to GB13. Seven Belgian sequences (BE/456/86-87, BE/2756/85-86, BE/11712/93-94, BE/7374/92-93, BE/10495/93-94, BE/23/90-91, and BE/12015/96-97) were not assigned to any lineage. Since there are only a few published subgroup HRSV-B G gene sequences covering the 724-bp region analyzed in this study, it was difficult to present our data in the context of previous classification schemes (45, 46, 60, 61). We could designate two lineages as GB1 and GB2, based on close clustering with previously assigned HRSV-B viruses (45, 60, 61). Most Belgian strains isolated during the same epidemic season clustered closely together in the phylogenetic tree (e.g., GB4, GB5, and GB9). All isolates possessing the 60-nt duplication grouped together in lineage GB13. Lineages GB1, GB2, GB3, GB5, GB6, GB9, and GB12 were supported with bootstrap values of 84 to 100%, based on 1,000 replicates.

HRSV-B viruses showed complex circulation patterns with up to six lineages cocirculating during the same epidemic season (Table 2). Certain genotypes were found throughout more than five epidemic seasons (GB1, GB3, and GB12), while genotypes GB5, with the exception of one strain isolated in 1987/1988, and GB9 were detected during one season. Genotype GB1 was the most prevalent genotype in Belgium during the 80s, isolated over eight epidemic seasons (1982–1983 to 1989–1990), with the exceptions of the 1983–1984 season, during which no HRSV-B strains were identified, and the epidemic season 1984–1985 in which GB5 was the dominating lineage observed. Most HRSV-B strains demonstrated a gradual build-up and then replacement of the predominant genotypes with new ones. For example, GB1 and GB4 (1989/1990) were replaced by GB3 (1990/1991), and lineages GB3 and GB12 (2000/2001) were replaced by GB13 (2001/2002). The predominant genotypes tended to prevail for one to two consecutive epidemics before being replaced by another genotype. Lineage GB13 was first detected in Belgium in 1999–2000, and these strains had a predicted G protein length of 315 amino acids.
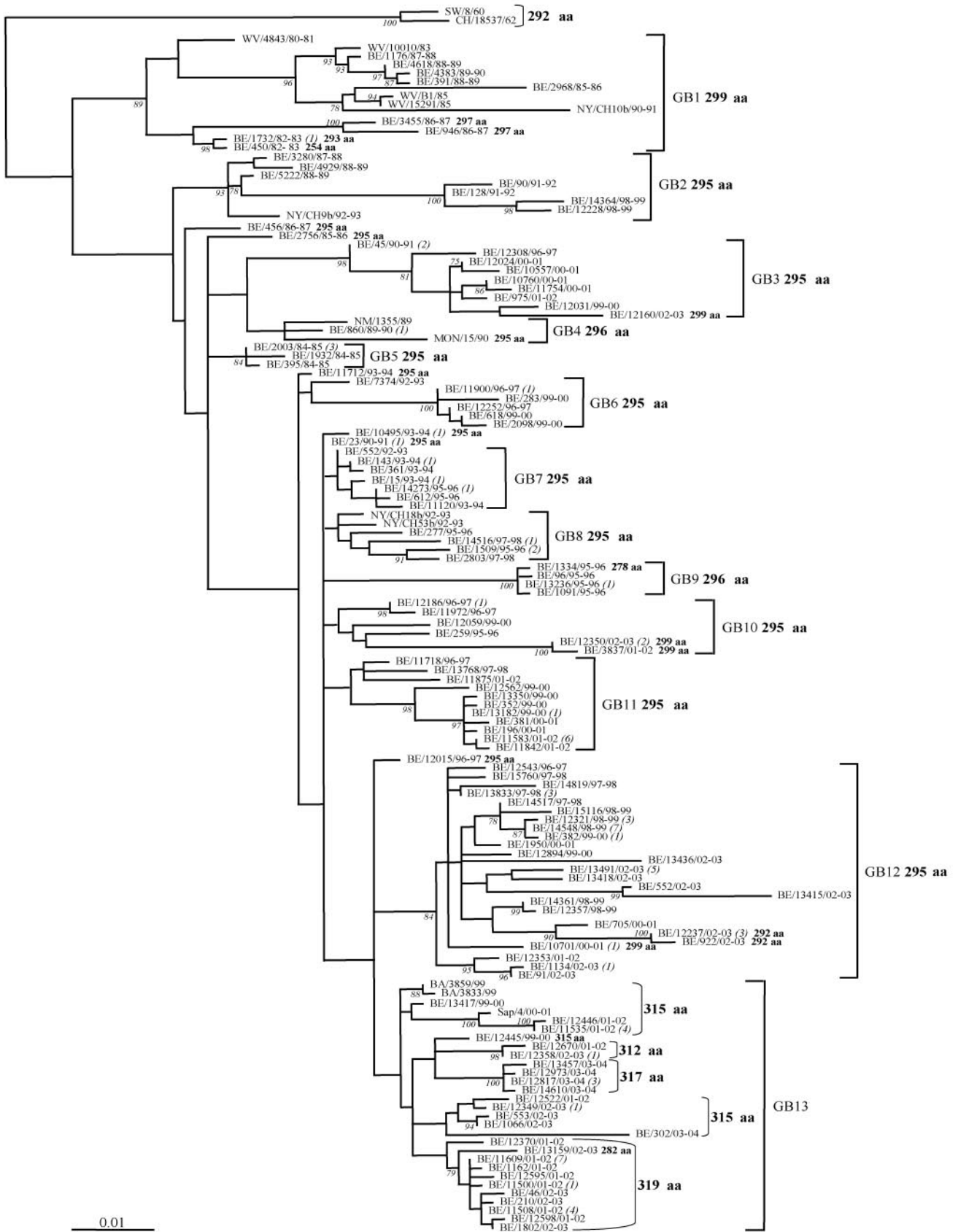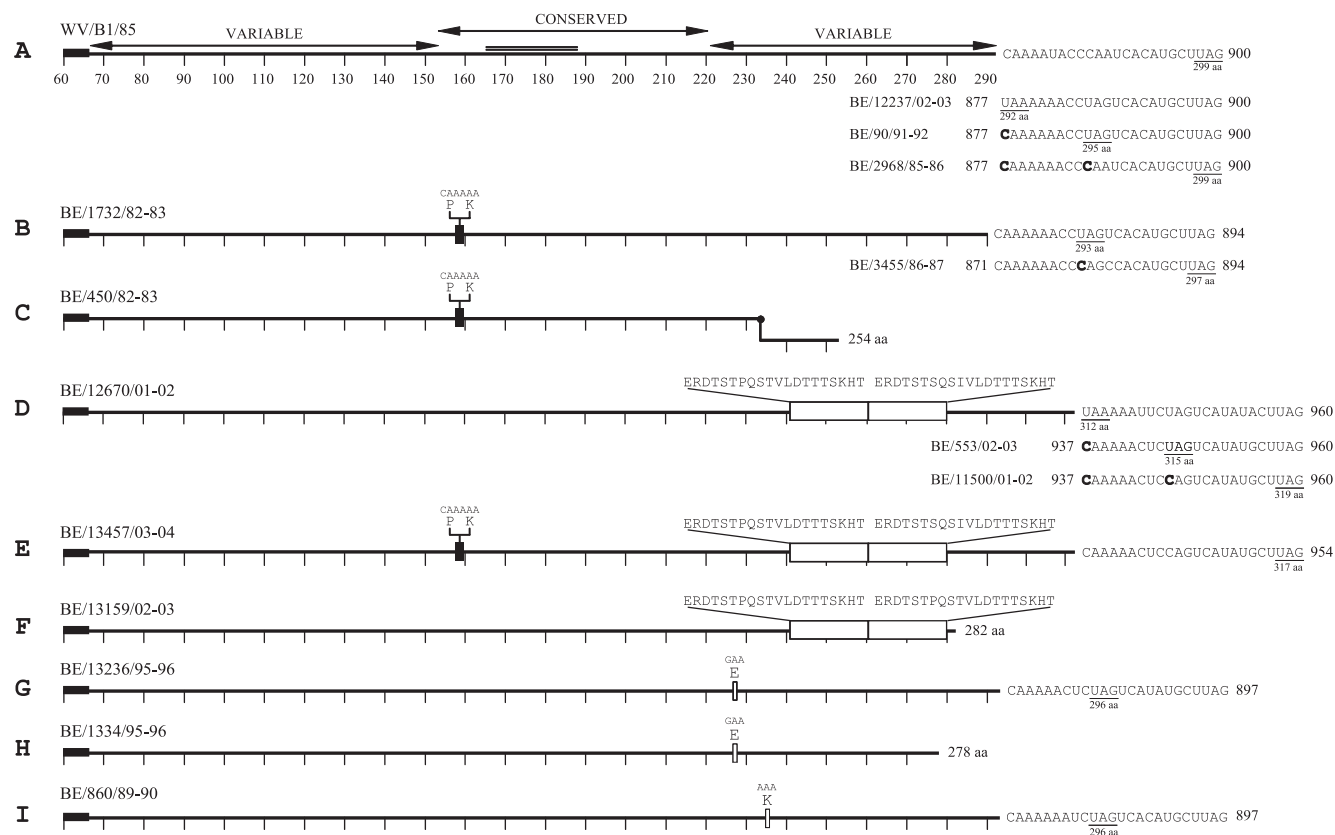
FIG. 2. Schematic diagrams presenting the predicted G protein length variability among the HRSV-B Belgian isolates. The linear presentations are given according to the reference strain WV/B1/85 with the two variable regions and the central conserved region (residues 153 to 221) indicated above the top diagram. The absolutely conserved amino acid region (amino acid positions 164 to 187) among the HRSV-B isolates is indicated with a double line (A). The underlined sequences present the three alternative termination codons described previously (32), and the predicted full protein lengths of Belgian representative sequences are given. The six-nucleotide deletion is indicated with a filled box (B, C, and E), the 60-nt duplication is indicated with a large open and divided box (D to F), and the three-nucleotide insertions are indicated with a small open box (G to I). The position of the frame-shifting in the frameshift mutant is indicated (●) in panel C. The two premature terminated sequences with their respective protein lengths are presented in panels F and H.

GB13 became the predominating lineage during 2001–2002 with the cocirculation of isolates of three different protein lengths (312, 315, and 319 amino acids) (Fig. 3). After being replaced by GB12 in the 2002–2003 season, GB13 became again the prevailing lineage during the following 2003–2004 epidemic season, and strains with protein lengths of 317 amino acids were dominating.

Analysis of selective pressure. The average nonsynonymous/synonymous substitution rate ratio (*dN/dS*) in the HRSV-B G protein ranged from 0.44 to 0.70 among all codon substitution models, suggesting that a nonsynonymous mutation has about 44 to 70% as much chance as a synonymous mutation of becoming fixed in the population. The simpler one-ratio (M0),

neutral (M1), and beta distribution (M7) models were rejected when compared with the discrete (M3), the selection (M2), and the beta and ω (M8) models. Although model M2 (selection) fits the data set better than model M1 (neutral), it does not suggest presence of positively selected sites ($p_2 = 0.62$; $\omega_2 = 0.17$). Model M3 suggested that approximately 8% ($p_2 = 0.08$) of the amino acid sites in the HRSV-B G-protein are under positive selection ($\omega_2 = 2.47$). Parameter estimates under model M8 (beta and ω) suggested that about 11% ($p_2 = 0.11$) of the sites are under diversifying selection with ω_2 of 2.16. In conclusion, models M3 and M8 provide consistent evidence for the presence of a small proportion of positively selected sites in the HRSV-B G attachment protein. Models

FIG. 1. HRSV-B G protein maximum likelihood phylogenetic tree. The nucleotide sequences of the G protein gene of the Belgian isolates (BE) were compared with those from Buenos Aires (BA), District of Columbia (CH), Montevideo (MON), New Mexico (NM), New York (NY), Sweden (SW), and West Virginia (WV). The reference strain SW/8/60 was used as outgroup sequence in the tree. Numbers at internal nodes represent bootstrap percentages, as determined for 1,000 iterations by the neighbor-joining method. Only bootstrap values greater than 75% are shown. The italicized numbers in brackets at the terminal nodes correspond to the number of identical sequences. The genetic clusters obtained in the analysis are indicated by the square brackets and assigned as GB1 to GB13. The predicted G protein lengths are designated in boldface; exceptions are indicated next to the corresponding sequence.

TABLE 2. Epidemic season distribution of HRSV-B genotypes GB1 to GB13

Genotype	Number of HRSV-B strains isolated from each epidemic season[a]																					
	1980–1981	1982–1983	1983–1984	1984–1985	1985–1986	1986–1987	1987–1988	1988–1989	1989–1990	1990–1991	1991–1992	1992–1993	1993–1994	1995–1996	1996–1997	1997–1998	1998–1999	1999–2000	2000–2001	2001–2002	2002–2003	2003–2004
GB1	(1)	3	(1)	(1)	1(1)	2	1	2	1	(1)												
GB2							1	2				(1)										
GB3											2						2					
GB4							1	(1)	5 (1)	3					1			1	4	1	1	
GB5			1							1												
GB6															3			3				
GB7													6	3		3						
GB8												1		4								
GB9												(2)		5								
GB10														1	3			1		1		
GB11															1	1		5	2	9	3	
GB12															1	7	10	7	4	1	19	
GB13																	(2)	2	(1)	25	12	8

[a] The numbers of non-Belgian GenBank HRSV-B strains that were included in this study are indicated in parentheses.

M3 and M8 gave similar lists of positively selected sites, although the posterior probabilities varied somewhat among the two models. Model M3 identified 12 positively selected sites with posterior probabilities P ($\omega > 1$) of >0.5, while model M8 located 22 sites to be under positive selection, including the 12 sites found by M3. In this study, we have only presented the posterior probabilities for positively selected sites identified under model M3.

Four positively selected sites were located in the first hypervariable region (amino acid positions 98, 99, 142, and 152 referring to strain WV/B1/85) and eight positively selected sites were located in the carboxy-terminal third of the ectodomain (amino acid positions 219, 223, 237, 247, 251, 257, 258, and 259) (Fig. 5). Six amino acids (positions 219, 223, 247, 251, 257, and 258) were identified to be under positive selection above the 90% level. Amino acids 99 and 237 were above the 80% level, amino acids 259 and 98 were above the 70% and 60% level, respectively, and amino acids 142 and 152 were above the 50% level.

Evolutionary rate. The correlation between phylogenetic root-to-tip divergence and time of sampling of HRSV-B strains is displayed in a regression plot (Fig. 6). The rate of substitution as estimated from this linear regression analysis was 2.36×10^{-3} nucleotide substitutions/site/year. The MRCA of HSRV-B was estimated to date back to 1955. Although linear regression gives a good correlation ($R^2 = 0.7$) between the sampling and genetic distance of the viral strains, the data points cannot be considered independent because of the phylogenetic structure. Using a genealogy-based maximum likelihood method, taking into account the phylogenetic dependency, the molecular clock assumption was significantly rejected ($P < 0.01$). Nevertheless, a previous simulation study has shown that non-clock rates should be reliable indicators of the average substitution rate, providing that data sets are relatively large and informative, and the inclusion of isolation dates into a single-rate model improves its likelihood (26), which is the case in our study. Under the SRDT-model, the evolutionary rate for HRSV-B viruses was calculated as 1.95×10^{-3} nucleotide substitutions/site/year (95% confidence interval [CI], 1.15×10^{-3} to 2.34×10^{-3}) and the MRCA dates back to 1949 (95% CI, 1938 to 1955). The upper CIs of the maximum likelihood estimates of the evolutionary rate and the MRCA are highly similar to the linear regression estimates. The maximum-likelihood estimate of the MRCA for HRSV-A and -B, based on the amino acid sequences of our previous study (75) and this one, was dated back to 1648 (95% CI, 1544 to 1752).

DISCUSSION

Genetic diversity among the HRSV-B strains. The genetic diversity observed among the G protein of the HRSV-B strains was significantly higher than the diversity observed in the Belgian HRSV-A strains, in which only amino acid substitutions were identified (75). The predicted G proteins of the subgroup B isolates were of 13 different amino acid lengths. The mechanisms responsible for this variability include amino acid substitutions, insertions, deletions, duplications, changes in the stop codon usage, and one frameshift mutation. Of particular interest is the great genetic diversity observed within the novel
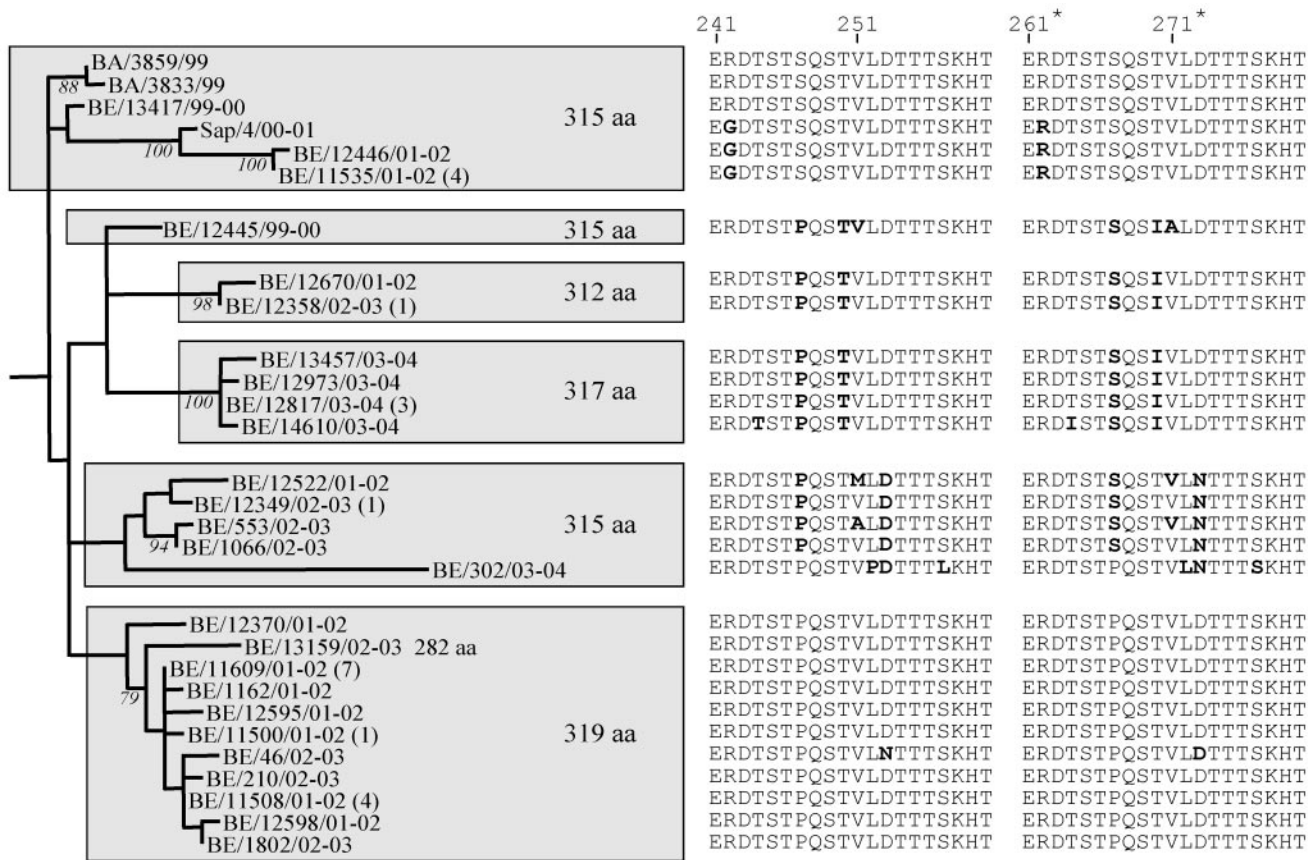
FIG. 3. Phylogenetic tree of the GB13 lineage containing HRSV-B strains with the 60-nt duplicated region isolated from Belgium (BE), Buenos Aires (BA), and Sapporo (Sap). Bootstrap values above 75% for 1,000 iterations are shown at the branch. The amino acid sequence of the duplicated region is shown next to each isolate with nonsynonymous substitutions presented in boldface.

GB13 lineage. We have identified GB13 isolates of five different protein lengths (282, 312, 315, 317, and 319 amino acids) due to six-nucleotide in-frame deletions, one premature stop codon, and the usage of three alternative termination codons (Fig. 2D to F and 3). Changes in stop codon usage have been associated with important antigenetic variations in certain HRSV-A escape mutants that were selected in vitro with monoclonal antibodies (15, 33, 50, 51). Therefore, we suggest that mutations leading to alternative termination codons and premature stop codons might give a significant advantage to HRSV-B viruses to persist in the community by escaping the preexisting host immune response.

In this study, we describe for the first time 12 isolates with a six-nucleotide (amino acid positions 159 to 160) in-frame deletion in a region of the attachment G protein (amino acid residues 153 to 221) that was previously considered to be conserved among HRSV-B viruses (Fig. 2B, C, and E) (55). It is worth mentioning that five of these strains were isolated during the 80s, and seven strains were from the 2003–2004 epidemic season and possessed also the 60-nt duplicated region. We speculate that these deletions originated by two separate mutational events, one in the early 80s and the other during the last epidemic season (2003–2004). In comparison to HRSV-A, it appears that HRSV-B viruses can accommodate more drastic changes in their G proteins and that these changes are not restricted to the carboxy-terminal third of the ectodomain.

The six-nucleotide deletions, the three-nucleotide insertions, and the single nucleotide gap identified in the HRSV-B isolates were found next to clusters of adenosine residues. Previous studies suggested that error-prone regions in the HRSV-A G gene appear to be concentrated next to adenosines in runs of six or seven residues (3, 15). These mutations seem likely to have occurred after a polymerase "stuttering" error with reiterative copying of the same template due to polymerase pausing and shifting on the nascent mRNA chain (62, 63). The HRSV-B frameshift mutant strain (BE/450/82-83) had a predicted protein length that was 45 amino acids shorter than the WV/B1/85 reference strain, and the 23 carboxy-terminal residues were altered (Fig. 2C). To our knowledge, this is the first report of a natural HRSV-B isolate with such extensive changes in the G protein.

Circulation patterns of HRSV-B isolates. The phylogenetic tree revealed complex circulation patterns with multiple lineages cocirculating in a single epidemic and a shift of the prevailing genotype occurring every 1 or 2 years (Table 2 and Fig. 1). HRSV-B strains have a worldwide distribution, with viruses isolated in distant places during the same season being closely related (e.g., BE/2968/85-86 and WV/B1/85; NM/1355/89, BE/860/89-90, and MON/15/90). The circulation of multi-
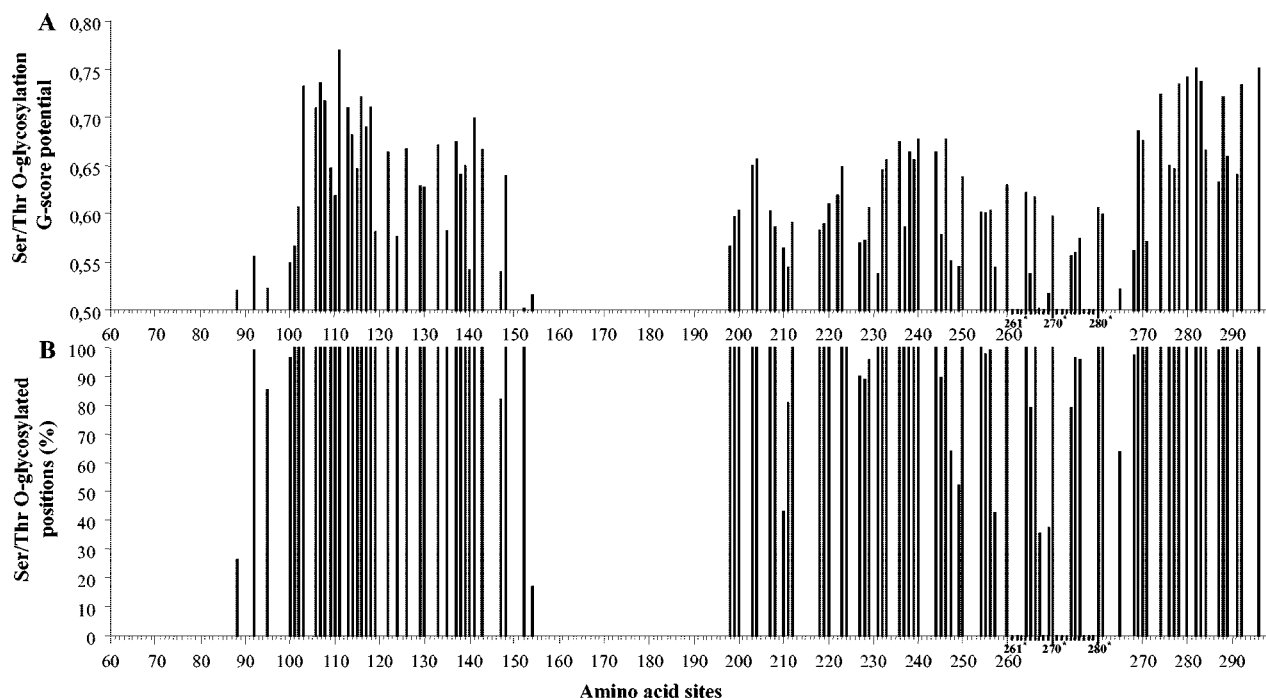
FIG. 4. Predicted Ser/Thr O-glycosylated residues in the HRSV-B viruses presented according to reference strain WV/B1/85. (A) Best general score potentials (G score) of the predicted O-glycosylated Ser/Thr sites. (B) Percentages of the HRSV-B sequences in which a particular Ser/Thr is predicted to be O glycosylated. The numbers with a superscript asterisk and the scale lines in bold indicate the amino acid positions in the duplicated 60-nt region identified in some HRSV-B strains.

ple identical sequences obtained during the same outbreak implies that certain strains can take advantage of their slight immunological differences to evade existing immunity and infect more hosts. Dominant genotypes appeared to show a gradual build-up and then replacement with new ones, with the disappearance of some lineages and the appearance of others. For example, lineages GB1 and GB5 were circulating in Belgium during the 80s, while GB6, GB7, GB8, and GB9 were found only during the 90s (Table 2). The isolation of particular lineages over more than five epidemic seasons (GB1, GB3, and GB12) indicates that certain HRSV-B genotypes can remain present for relatively long periods of time (Table 2).

All Belgian strains with the 60-nt duplication clustered together in a separate lineage, GB13, implying that this novel genotype has originated from a unique mutational event (Fig. 1 and 3). In June and August 1999, three strains with a 60-nt duplication in the carboxy-terminal third of the G gene were isolated in Buenos Aires (59). Similar strains (BE/13417/99-00 and BE/12445/99-00) were detected in Belgium for the first time in December 1999 (Fig. 2). This novel lineage became the predominant genotype in Belgium during the epidemic seasons 2001–2002 and 2003–2004. Strains with the 60-nt duplication were also recently reported in Japan and Kenya (40, 53). These findings indicate that this novel strain is now emerging globally.

Phylodynamic patterns are influenced primarily by natural selection that arises from cross-immunity and secondarily by neutral epidemiological processes such as spatial population separation. The survival of phylogenetic lineages depends, therefore, on the prevailing epidemiological and immunological forces (17). HRSV-B viruses show complex phylogenetic patterns with cocirculation of multiple lineages and replacement of the prevailing genotype, illustrating the global spatiotemporal strain dynamics of these viruses.

Positive selection in the HRSV-B G attachment protein. In many proteins, a high proportion of amino acids can be largely invariable due to strong functional constraints, and therefore adaptive evolution is likely to affect only certain amino acids (8, 31). As was observed in our previous study on HRSV-A, the average ratios of nonsynonymous to synonymous nucleotide substitutions in the HRSV-B G protein identified under all models were not indicative for positive selection because these values did not surpass the threshold of an ω of 1. Under the discrete model (M3), we were able to detect 12 codon sites in the HRSV-B G protein ectodomain to be under positive selection with posterior probabilities above 0.5 (Fig. 5). Although there were no positively selected sites found in common between HRSV-A and HRSV-B, the distribution of these codon positions in the G protein ectodomain was similar (75). Four positively selected sites were identified in the first hypervariable region for both subgroups (sites 98, 99, 142, and 152 for HRSV-B). In the second hypervariable region, nine positively selected sites were detected in HRSV-A viruses and eight sites were detected in HRSV-B viruses (sites 219, 223, 237, 247, 251, 257, 258, and 259) (Fig. 5). Previous studies have shown that HRSV-A strain-specific epitopes are preferentially located in the carboxy-terminal third of the G protein (33, 50), suggestive for immunodominance of this hypervariable domain. The clustering of positively selected sites in the carboxy-terminal third of HRSV-B G protein implies that this region contains important antigenic determinants. The differences in the positively
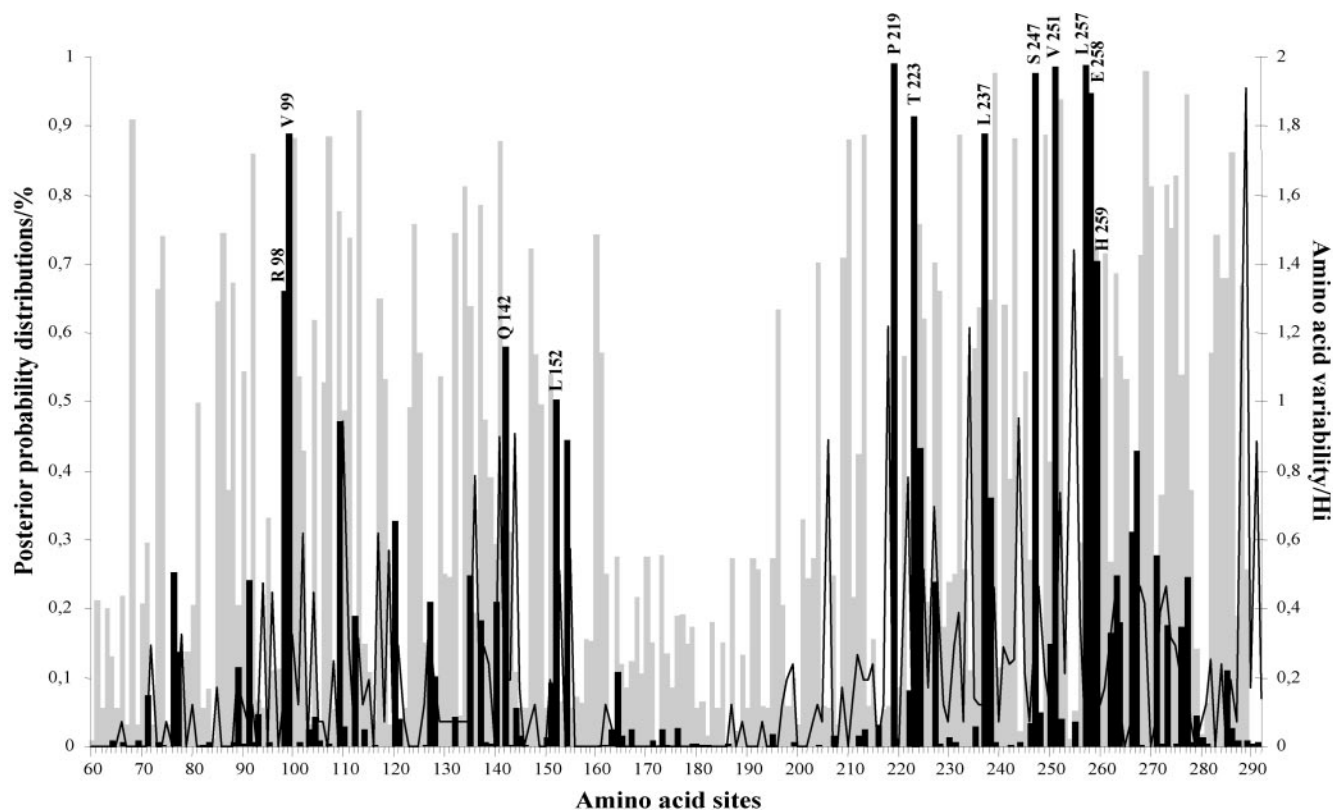
FIG. 5. Posterior probabilities of sites for different classes along the G protein ectodomain region under the discrete model M3. This model assumes three classes of sites in the gene: positive sites (black bars), neutral sites (open graph), and negative sites (gray bars). Positively selected amino acid sites with posterior probabilities above 50% are indicated according to strain WV/B1/85. Amino acid variability, measured by entropy (H_i), is plotted to the second y axis.

selected codon positions between HRSV-A and -B strains could be relevant for a different host immune response induced by the two subgroups. A stronger immune response against HRSV-B might lead to a lower rate of reinfection with HRSV-B strains. Alternatively, the immune response against HRSV-A might offer more cross-protection against HRSV-B infection than vice versa (35). These hypotheses might offer an explanation for the observation that in most epidemic seasons, HRSV-A strains are more often isolated than HRSV-B strains.

Modulation of the number and distribution patterns of potential N- and O-linked glycosylation sites can influence the expression of certain epitopes by either masking or contributing to recognition by carbohydrate specific antibodies (34, 43, 44). In this study, we have located 105 serine and threonine residues to be potentially O glycosylated in the ectodomain of the HRSV-B G attachment protein, including 22 of the 25 sites that have been predicted to be the most likely to contain O-linked sugars in the HRSV-A strain A2 (Fig. 4) (6, 22, 66). The serine and threonine residues at positively selected sites 152, 219, 223, and 237 were predicted to be O glycosylated in 100% of the strains, and the serine residues at positively selected sites 247 and 257 were predicted to be O glycosylated in 64% and 43% of the strains, respectively. Furthermore, we have identified 10 Ser and Thr residue positions to be potentially O glycosylated in the 60-nt duplicated region (Fig. 4, 264* to 267*, 269*, 270*, 274* to 276*, and 280*). Since O-linked

sugars have been shown to be important for the functional properties of the G protein, amino acid substitutions at these positions could give a selective advantage to the virus to avoid neutralization by preexisting antibodies. Further analysis is necessary to see if amino acid substitutions in the HRSV-B positively selected sites can lead to antibody escape mutants. Understanding the functional importance of these positively selected amino acid positions could help to predict possible changes in virulence as well as responses to certain strains in future vaccination programs.

Rate of evolution of HRSV-B strains. The evolutionary rate estimate of the G protein of HRSV-B (1.95×10^{-3} nucleotide substitutions/site/year; 95% CI, 1.15×10^{-3} to 2.34×10^{-3}) is very similar to that identified for HRSV-A (1.83×10^{-3} nucleotide substitutions/site/year; 95% CI, 1.44×10^{-3} to 2.26×10^{-3}) in our previous study (75). Therefore, the MRCA estimate of the currently circulating HRSV-B strains (1949; 95% CI, 1938 to 1955) is almost identical to the MRCA of the HRSV-A strains (1944; 95% CI, 1937 to 1950). Our results indicate that in the human host, HRSV-B has an equally high potential as HRSV-A for accumulation of genetic diversity. Our calculations date the MRCA of HRSV-A and -B viruses back to 1648 (95% CI, 1544 to 1752), implying that the divergence between the two subgroups occurred approximately 350 years ago.

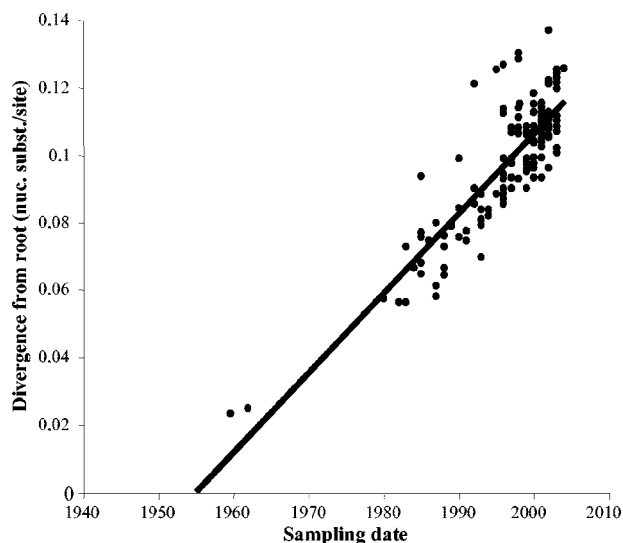Although HRSV-B strains are isolated less frequently than

FIG. 6. Linear root-to-tip regression plot presenting the correlation between the branch lengths and the sampling dates of the HRSV-B isolates included in this study.

HRSV-A strains, our study shows that these two subgroups have very similar circulation patterns and evolutionary rates. Natural HRSV-B isolates appear to accommodate more drastic changes in their attachment G proteins such as 60-nt duplications, 6-nt deletions, premature stop codons, and frameshift mutations. In our study, 12 codon positions in the HRSV-B attachment G protein seem to be under adaptive evolution. Three positively selected sites (sites 152, 223, and 257) had also been described previously (70), and we report an additional nine novel positively selected sites (sites 98, 99, 142, 219, 237, 247, 251, 258, and 259). The world population growth, widespread urbanization, and the development of modern transportation have probably aided sustained transmission and global dissemination of HRSV viruses. We speculate that amino acid substitutions in certain epitope-related codon positions in the G attachment protein of HRSV-A and -B strains have enabled these viruses to escape the preexisting immunity and reinfect the population. It has been reported previously that antibody responses after infections with HRSV-A were more cross-reactive than were the responses that followed primary infection by HRSV-B (35). Therefore, the observed differences in the positively selected sites between HRSV-A and HRSV-B might reflect differences in the immunological responses that they elicit. Understanding the significance of possible amino acid changes in the positively selected positions identified in both subgroups could give useful insights for future vaccine development studies.

REFERENCES

1. **Anderson, L. J., J. C. Hierholzer, C. Tsou, R. M. Hendry, B. F. Fernie, Y. Stone, and K. McIntosh.** 1985. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. J. Infect. Dis. **151:**626–633.
2. **Cane, P. A., and C. R. Pringle.** 1995. Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein. J. Virol. **69:**2918–2925.
3. **Cane, P. A., D. A. Matthews, and C. R. Pringle.** 1993. Frequent polymerase errors observed in a restricted area of clones derived from the attachment (G) protein gene of respiratory syncytial virus. J. Virol. **67:**1090–1093.
4. **Cane, P. A., D. A. Matthews, and C. R. Pringle.** 1991. Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. J. Gen. Virol. **72:**2091–2096.
5. **Coggins, W. B., E. J. Lefkowitz, and W. M. Sullender.** 1998. Genetic variability among group A and group B respiratory syncytial viruses in a children's hospital. J. Clin. Microbiol. **36:**3552–3557.
6. **Collins, P. L., R. M. Chanock, and B. R. Murphy.** 2001. Respiratory syncytial virus, p. 1443–1485. In D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology, 4th ed. Lippincott Williams and Wilkins, Philadelphia, Pa.
7. **Connors, M., P. L. Collins, C. Y. Firestone, and B. R. Murphy.** 1991. Respiratory syncytial virus (RSV) F, G, M2 (22K), and N proteins each induce resistance to RSV challenge, but resistance induced by M2 and N proteins is relatively short-lived. J. Virol. **65:**1634–1637.
8. **Crandall, K. A., C. R. Kelsey, H. Imamichi, H. C. Lane, and N. P. Salzman.** 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol. Biol. Evol. **16:**372–382.
9. **Dowell, S. F., L. J. Anderson, H. E. Gary, Jr., D. D. Erdman, J. F. Plouffe, T. M. File, Jr., B. J. Marston, and R. F. Breiman.** 1996. Respiratory syncytial virus is an important cause of community-acquired lower respiratory infection among hospitalized adults. J. Infect. Dis. **174:**456–462.
10. **Englund, J. A., L. J. Anderson, and F. S. Rhame.** 1991. Nosocomial transmission of respiratory syncytial virus in immunocompromised adults. J. Clin. Microbiol. **29:**115–119.
11. **Falsey, A. R., and E. E. Walsh.** 2000. Respiratory syncytial virus infection in adults. Clin. Microbiol. Rev. **13:**371–384.
12. **Fleming, D. M., and K. W. Cross.** 1993. Respiratory syncytial virus or influenza? Lancet **342:**1507–1510.
13. **Garcia, O., M. Martin, J. Dopazo, J. Arbiza, S. Frabasile, J. Russi, M. Hortal, P. Perez-Brena, I. Martinez, B. Garcia-Barreno, and J. A. Melero.** 1994. Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. J. Virol. **68:**5448–5459.
14. **Garcia-Barreno, B., T. Delgado, and J. A. Melero.** 1994. Oligo(A) sequences of human respiratory syncytial virus G protein gene: assessment of their genetic stability in frameshift mutants. J. Virol. **68:**5460–5468.
15. **Garcia-Barreno, B., A. Portela, T. Delgado, J. A. Lopez, and J. A. Melero.** 1990. Frame-shift mutations as a novel mechanism for the generation of neutralization resistant mutants of human respiratory syncytial virus. EMBO J. **9:**4181–4187.
16. **Glezen, P., and F. W. Denny.** 1973. Epidemiology of acute lower respiratory disease in children. N. Engl. J. Med. **288:**498–505.
17. **Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes.** 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science **303:**327–332.
18. **Hall, C. B., E. E. Walsh, C. E. Long, and K. C. Schnabel.** 1991. Immunity to and frequency of reinfection with respiratory syncytial virus. J. Infect. Dis. **163:**693–698.
19. **Hall, C. B., E. E. Walsh, K. C. Schnabel, C. E. Long, K. M. McConnochie, S. W. Hildreth, and L. J. Anderson.** 1990. Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children. J. Infect. Dis. **162:**1283–1290.
20. **Hall, C. B., A. E. Kopelman, R. G. Douglas, Jr., J. M. Geiman, and M. P. Meagher.** 1979. Neonatal respiratory syncytial virus infection. N. Engl. J. Med. **300:**393–396.
21. **Han, L. L., J. P. Alexander, and L. J. Anderson.** 1999. Respiratory syncytial virus pneumonia among the elderly: an assessment of disease burden. J. Infect. Dis. **179:**25–30.
22. **Hansen, J. E., O. Lund, N. Tolstrup, A. A. Gooley, K. L. Williams, and S. Brunak.** 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj. J. **15:**115–130.
23. **Henderson, F. W., A. M. Collier, W. A. Clyde, Jr., and F. W. Denny.** 1979. Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children. N. Engl. J. Med. **300:**530–534.
24. **Hendry, R. M., L. T. Pierik, and K. McIntosh.** 1989. Prevalence of respira-

tory syncytial virus subgroups over six consecutive outbreaks: 1981–1987. J. Infect. Dis. **160:**185–190.

25. Holberg, C. J., A. L. Wright, F. D. Martinez, C. G. Ray, L. M. Taussig, and M. D. Lebowitz. 1991. Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. Am. J. Epidemiol. **133:** 1135–1151.

26. Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. **54:**156–165.

27. Johnson, P. R., M. K. Spriggs, R. A. Olmsted, and P. L. Collins. 1987. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. Proc. Natl. Acad. Sci. USA **84:**5625–5629.

28. Julenius, K., A. Molgaard, R. Gupta, and S. Brunak. 2005. Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology **15:**153–164.

29. Kamasaki, H., H. Tsutsumi, K. Seki, and S. Chiba. 2001. Genetic variability of respiratory syncytial virus subgroup B strain isolated during the last 20 years from the same region in Japan: existence of time-dependent linear genetic drifts. Arch. Virol. **146:**457–466.

30. Karron, R. A., D. A. Buonagurio, A. F. Georgiu, S. S. Whitehead, J. E. Adamus, M. L. Clements-Mann, D. O. Harris, V. B. Randolph, S. A. Udem, B. R. Murphy, and M. S. Sidhu. 1997. Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. Proc. Natl. Acad. Sci. USA **94:**13961–13966.

31. Li, W. H. 1997. Molecular evolution. Sinauer Assoc., Sunderland, MA.

32. Martinez, I., O. Valdes, A. Delfraro, J. Arbiza, J. Russi, and J. A. Melero. 1999. Evolutionary pattern of the G glycoprotein of human respiratory syncytial viruses from antigenic group B: the use of alternative termination codons and lineage diversification. J. Gen. Virol. **80:**125–130.

33. Martinez, I., J. Dopazo, and J. A. Melero. 1997. Antigenic structure of the human respiratory syncytial virus G glycoprotein and relevance of hypermutation events for the generation of antigenic variants. J. Gen. Virol. **78:**2419–2429.

34. Melero, J. A., B. Garcia-Barreno, I. Martinez, C. R. Pringle, and P. A. Cane. 1997. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. J. Gen. Virol. **78:**2411–2418.

35. Muelenaer, P. M., F. W. Henderson, V. G. Hemming, E. E. Walsh, L. J. Anderson, G. A. Prince, and B. R. Murphy. 1991. Group-specific serum antibody responses in children with primary and recurrent respiratory syncytial virus infections. J. Infect. Dis. **164:**15–21.

36. Mufson, M. A., R. B. Belshe, C. Orvell, and E. Norrby. 1988. Respiratory syncytial virus epidemics: variable dominance of subgroups A and B strains among children, 1981–1986. J. Infect. Dis. **157:**143–148.

37. Mufson, M. A., R. B. Belshe, C. Orvell, and E. Norrby. 1987. Subgroup characteristics of respiratory syncytial virus strains recovered from children with two consecutive infections. J. Clin. Microbiol. **25:**1535–1539.

38. Mufson, M. A., C. Orvell, B. Rafnar, and E. Norrby. 1985. Two distinct subtypes of human respiratory syncytial virus. J. Gen. Virol. **66:**2111–2124.

39. Murry, A. R., and S. F. Dowell. 1997. Respiratory syncytial virus: not just for kids. Hosp. Pract. **32:**87–88.

40. Nagai, K., H. Kamasaki, Y. Kuroiwa, L. Okita, and H. Tsutsumi. 2004. Nosocomial outbreak of respiratory syncytial virus subgroup B variants with the 60 nucleotides-duplicated G protein gene. J. Med. Virol. **74:**161–165.

41. Nicholas, K. B., H. B. Nicholas, and D. W. Deerfield. 1997. GeneDoc: analysis and visualization of genetic variation. Embnet News **4:**14.

42. Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:**929–936.

43. Palomo, C., P. A. Cane, and J. A. Melero. 2000. Evaluation of the antibody specificities of human convalescent-phase sera against the attachment (G) protein of human respiratory syncytial virus: influence of strain variation and carbohydrate side chains. J. Med. Virol. **60:**468–474.

44. Palomo, C., B. Garcia-Barreno, C. Penas, and J. A. Melero. 1991. The G protein of human respiratory syncytial virus: significance of carbohydrate side-chains and the C-terminal end to its antigenicity. J. Gen. Virol. **72:**669–675.

45. Peret, T. C., C. B. Hall, G. W. Hammond, P. A. Piedra, G. A. Storch, W. M. Sullender, C. Tsou, and L. J. Anderson. 2000. Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. J. Infect. Dis. **181:**1891–1896.

46. Peret, T. C., C. B. Hall, K. C. Schnabel, J. A. Golub, and L. J. Anderson. 1998. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. J. Gen. Virol. **79:**2221–2229.

47. Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. **50:**580–601.

48. Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics **16:**395–399.

49. Roca, A., M. P. Loscertales, L. Quinto, P. Perez-Brena, N. Vaz, P. L. Alonso, and J. C. Saiz. 2001. Genetic variability among group A and B respiratory syncytial viruses in Mozambique: identification of a new cluster of group B isolates. J. Gen. Virol. **82:**103–111.

50. Rueda, P., C. Palomo, B. Garcia-Barreno, and J. A. Melero. 1995. The three C-terminal residues of human respiratory syncytial virus G glycoprotein (Long strain) are essential for integrity of multiple epitopes distinguishable by antiidiotypic antibodies. Viral Immunol. **8:**37–46.

51. Rueda, P., T. Delgado, A. Portela, J. A. Melero, and B. Garcia-Barreno. 1991. Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies. J. Virol. **65:**3374–3378.

52. Sanz, M. C., O. M. Kew, and L. J. Anderson. 1994. Genetic heterogeneity of the attachment glycoprotein G among group A respiratory syncytial viruses. Virus Res. **33:**203–217.

53. Scott, P. D., R. Ochola, M. Ngama, E. A. Okiro, D. J. Nokes, G. F. Medley, and P. A. Cane. 2004. Molecular epidemiology of respiratory syncytial virus in Kilifi district, Kenya. J. Med. Virol. **74:**344–354.

54. Sullender, W. M., M. A. Mufson, G. A. Prince, L. J. Anderson, and G. W. Wertz. 1998. Antigenic and genetic diversity among the attachment proteins of group A respiratory syncytial viruses that have caused repeat infections in children. J. Infect. Dis. **178:**925–932.

55. Sullender, W. M., M. A. Mufson, L. J. Anderson, and G. W. Wertz. 1991. Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. J. Virol. **65:**5425–5434.

56. Swofford, D. L. 1998. PAUP* 4.0: phylogenetic analysis using parsimony. Sinauer Assoc., Sunderland, MA.

57. Taylor, G., E. J. Stott, M. Bew, B. F. Fernie, P. J. Cote, A. P. Collins, M. Hughes, and J. Jebbett. 1984. Monoclonal antibodies protect against respiratory syncytial virus infection in mice. Immunology **52:**137–142.

58. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.

59. Trento, A., M. Galiano, C. Videla, G. Carballal, B. Garcia-Barreno, J. A. Melero, and C. Palomo. 2003. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. J. Gen. Virol. **84:**3115–3120.

60. Venter, M., M. Collinson, and B. D. Schoub. 2002. Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: comparison of viruses and genotypes responsible for different disease manifestations. J. Med. Virol. **68:**452–461.

61. Venter, M., S. A. Madhi, C. T. Tiemessen, and B. D. Schoub. 2001. Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new subgroup A and B genotypes. J. Gen. Virol. **82:**2117–2124.

62. Vidal, S., J. Curran, and D. Kolakofsky. 1990. A stuttering model for paramyxovirus P mRNA editing. EMBO J. **9:**2017–2022.

63. Vidal, S., J. Curran, and D. Kolakofsky. 1990. Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity. J. Virol. **64:**239–246.

64. Walsh, E. E., J. J. Schlesinger, and M. W. Brandriss. 1984. Protection from respiratory syncytial virus infection in cotton rats by passive transfer of monoclonal antibodies. Infect. Immun. **43:**756–758.

65. Waris, M. 1991. Pattern of respiratory syncytial virus epidemics in Finland: two-year cycles with alternating prevalence of groups A and B. J. Infect. Dis. **163:**464–469.

66. Wathen, M. W., P. A. Aeed, and A. P. Elhammer. 1991. Characterization of oligosaccharide structures on a chimeric respiratory syncytial virus protein expressed in insect cell line Sf9. Biochemistry **30:**2863–2868.

67. Wertz, G. W., M. Krieger, and L. A. Ball. 1989. Structure and cell surface maturation of the attachment glycoprotein of human respiratory syncytial virus in a cell line deficient in O glycosylation. J. Virol. **63:**4767–4776.

68. Wertz, G. W., P. L. Collins, Y. Huang, C. Gruber, S. Levine, and L. A. Ball. 1985. Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. Proc. Natl. Acad. Sci. USA **82:**4075–4079.

69. Wilson, S. D., K. Roberts, K. Hammond, J. G. Ayres, and P. A. Cane. 2000. Estimation of incidence of respiratory syncytial virus infection in schoolchildren using salivary antibodies. J. Med. Virol. **61:**81–84.

70. Woelk, C. H., and E. C. Holmes. 2001. Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). J. Mol. Evol. **52:**182–192.

71. Xia, X., and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. **92:**371–373.

72. Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15:**496–503.

73. Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431–449.

74. Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:**555–556.

75. Zlateva, K. T., P. Lemey, A.-M. Vandamme, and M. Van Ranst. 2004. Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup A: positively selected sites in the attachment G glycoprotein. J. Virol. **78:**4675–4683.