

Molecular Evolution and Circulation Patterns of Human Respiratory Syncytial Virus Subgroup A: Positively Selected Sites in the Attachment G Glycoprotein

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

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

Received 28 June 2003/Accepted 16 January 2004

Human respiratory syncytial virus (HRSV) is the most common etiological agent of acute lower respiratory tract disease in infants and can cause repeated infections throughout life. In this study, we have analyzed nucleotide sequences encompassing 629 bp at the carboxy terminus of the G glycoprotein gene for HRSV subgroup A strains isolated over 47 years, including 112 Belgian strains isolated over 19 consecutive years (1984 to 2002). By using a maximum likelihood method, we have tested the presence of diversifying selection and identified 13 positively selected sites with a posterior probability above 0.5. The sites under positive selection correspond to sites of O glycosylation or to amino acids that were previously described as monoclonal antibody-induced in vitro escape mutants. Our findings suggest that the evolution of subgroup A HRSV G glycoprotein is driven by immune pressure operating in certain codon positions located mainly in the second hypervariable region of the ectodomain. Phylogenetic analysis revealed the prolonged cocirculation of two subgroup A lineages among the Belgian population and the possible extinction of three other lineages. The evolutionary rate of HRSV subgroup A isolates was estimated to be 1.83×10^{-3} nucleotide substitutions/site/year, projecting the most recent common ancestor back to the early 1940s.

Human respiratory syncytial virus (HRSV) is an enveloped virus with a nonsegmented negative-strand RNA genome (family *Paramyxoviridae*, genus *Pneumovirus*). HRSV is the major cause of acute lower respiratory tract infections in infants and young children. It is also recognized as an important pathogen in immunocompromised patients and the elderly (12, 58, 65). Two major subgroups of HRSV (HRSV-A and HRSV-B) have been described based on antigenic and sequencing studies (2, 6, 9, 25, 34). An unusual feature of HRSV is that repeated infections with both subgroups can occur, despite the development of mucosal and systemic immune responses (18, 19, 20). Although HRSV-A and -B strains can cocirculate, subgroup A is dominant in more epidemics (1, 21, 23, 33). Several genotypes can cocirculate in a single epidemic season, and different genotypes can predominate in consecutive epidemics (3, 7, 21, 42). In addition, viruses isolated in geographically distant places and in different years may be more closely related than viruses isolated in the same place during the same season (5, 14, 32).

The attachment G glycoprotein is a type II integral membrane protein (56, 64) that shows the highest degree of diversity both between and within the two HRSV subgroups (25, 54). The G glycoprotein is one of the main antigens responsible for inducing a neutralizing immune response (11). It is highly glycosylated, and the majority of the potential glycosylation

sites are clustered in two mucin-like variable regions of the ectodomain separated by a highly conserved 13-amino-acid motif, which is considered to be the putative site for virus attachment to the cell receptor (8, 25, 54).

Three types of epitopes have been identified in the G molecule: (i) conserved epitopes, shared by subgroups A and B (2, 34), (ii) subgroup-specific epitopes, and (iii) strain-specific or variable epitopes (2, 16). The strain-specific epitopes have been mapped within the hypervariable C-terminal third of the G glycoprotein ectodomain, and there is evidence of progressive accumulation of genetic and antigenic changes in this region (6, 14, 50).

The epidemiology and the possible mechanisms by which HRSV can evade the immune response are still not completely elucidated. The significance of genetic drift and nonprogressive random variation in the molecular evolution of HRSV is a question that remains to be answered. No studies have been conducted on sequence data over more than 10 consecutive HRSV epidemics, and most reports are based on sequencing of only a limited region of the G glycoprotein (41, 45, 62). In order to investigate the tempo and the mode by which HRSV-A evolves, we sequenced 629 bp of the G glycoprotein ectodomain from 112 subgroup A strains obtained during 19 consecutive years in Belgium (1984 to 2002). In the present work, we confirm five previously described amino acid sites and report eight novel sites in the G glycoprotein to be under positive selection. These results suggest that HRSV-A strains are evolving under selective pressure operating in certain codon positions of the genome. Furthermore, we have determined the most recent common ancestor (MRCA) of HRSV-A to date back to the early 1940s.

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

MATERIALS AND METHODS

Viral isolates. Nasopharyngeal samples were collected from infants and young children between 2 weeks and 5 years old (mean age, 11 months) who were hospitalized with respiratory distress symptoms in the pediatric ward of the Gasthuisberg University Hospital in Leuven, Belgium, during the period from 1984 to 2002. Samples that reacted positive with an HRSV antigen test (TESTPACK RSV, Abbott Laboratories, Abbott Park, Ill.) and exhibited a typical HRSV cytopathic effect in cell cultures (HeLa and HEp-2) were included in this study. HRSV-infected cells were harvested and 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/1150/99-00).

PCR primer sequences. The primers for reverse transcription and PCR were based on the published sequences of the attachment G glycoprotein and the fusion F protein genes of HRSV-A (55). The forward primer G267 (5'-GATG CAACAAGCCAGATCAAG-3') corresponds to bases 247 to 267 in the G glycoprotein mRNA of the A2 strain (GenBank accession number M11486) and has previously been demonstrated to be subgroup specific (55). The reverse primer F164 (5'-GTTATCACACTGGTATACCAACC-3') was complementary to bases 164 to 186 in the F protein mRNA sequence (accession number M11486). The expected size of the PCR product was 0.9 kb.

RNA extraction. The viral RNA to be used as a template for cDNA synthesis was extracted from 140 μl of the supernatant of each cell-cultured sample by using the QIAmp Viral RNA mini kit (QIAGEN, Westburg, The Netherlands), and the extracted viral RNA was dissolved in 60 μl of elution buffer.

cDNA synthesis. An aliquot of 10 μl of the extracted viral RNA was first denatured at 80°C for 2 min and then added to the reaction mixture, consisting of 2 μl of $10\times$ PCR buffer II (Applied Biosystems, Foster City, Calif.), 3 mM MgCl_2 , 1 mM each deoxynucleoside triphosphate, 20 pmol of primer F164, 20 U of RNase inhibitor (Applied Biosystems), and 50 U of murine leukemia virus reverse transcriptase (Applied Biosystems) in a total volume of 20 μl . The reaction was initiated at room temperature for 10 min and further incubated at 42°C for 40 min and then heated at 95°C for 5 min.

PCR. For the PCR, 5 μl of the cDNA product was added in a final volume of 50 μl containing 2.5 mM MgCl_2 (pH 9.0), 15 pmol of primers F164 and G267, 200 μM deoxynucleotide triphosphates, and 1.5 U of AmpTaq (Perkin Elmer). The thermocycling procedure consisted of 40 cycles (denaturation at 95°C for 35 s, primer annealing at 65°C for 45 s, and extension at 72°C for 45 s), followed by a final extension step at 72°C for 10 min, in a GeneAmp PCR system 9700 thermal cycler (Perkin Elmer). The 0.9-kb PCR products were subjected to a 6% polyacrylamide gel electrophoresis (9- μl PCR product), stained with ethidium bromide, and visualized under UV light. The 0.9-kb amplicons were purified by using a QIAquick PCR purification kit (QIAGEN).

DNA sequencing. The purified PCR products were cycle sequenced in forward and reverse directions by using an ABI PRISM BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems). Besides the PCR primer set, an additional forward primer (5'-ACATCCGAGTACCTATCACAA-3') and reverse primer (5'-GCTTGGTGGTGGTTTCTTTC-3'), both located in the G glycoprotein gene (nucleotides [nt] 859 to 879 and 599 to 619, respectively; GenBank accession number M11486), were used to obtain the 864-bp nucleotide sequence. Sequencing analysis was performed on an ABI PRISM 3100 genetic analyzer (Applied Biosystems).

Nucleotide sequence analysis. The chromatogram sequences were inspected with Chromas 2.2 (Technelysium, Helensvale, Australia), and contigs were prepared with SeqMan II (DNASTAR, Madison, Wis.).

Phylogenetic analysis. Multiple sequence alignments of 629 nt of the C-terminal end of the G glycoprotein gene were prepared by using CLUSTAL W (59) and manually edited in the GeneDoc version 2.6.002 alignment editor (35). Amino acid variability was calculated by using the data analysis in molecular biology (DAMBE) program version 4.0.36 (67). O glycosylation sites were determined by using the program NetOglyc version 2.0 (22). The appropriate nucleotide substitution model was determined by using hierarchical likelihood ratio testing (43). Phylogenetic reconstruction was performed by using the phylogenetic analysis using parsimony (PAUP) package version 4b10 (57). Model parameters were optimized on an initial neighbor-joining tree. By using the estimated parameters and the neighbor-joining tree as a starting tree, a heuristic maximum likelihood 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 with VirusRates 1.0, kindly provided by Andrew Rambaut. This method reconstructs a neighbor-joining tree and performs a linear regression between the time of sampling of each tip and the genetic distance from the root.

The root of the tree was chosen to maximize the R^2 value of the regression. Confidence intervals (CI) for the evolutionary rate and the time to the MRCA were calculated by bootstrapping the sequence data. A more accurate estimate for the evolutionary rate and the MRCA was obtained under the maximum likelihood framework (44). Given a genealogy, the likelihood is calculated under the constraint that the positions of the tips are proportional to their sampling dates, a model called the single rate dated tips (SRDT) model. Here, the evolutionary rate is an additional parameter used to scale the times of the internal nodes into units of the expected number of substitutions per site. The root was picked so as to maximize the likelihood under the SRDT model. The likelihood ratio test can be used to test the molecular clock by comparing the likelihood under the SRDT with the likelihood under a model where each branch is allowed to have its own rate, termed the different rates model.

Selective pressure analysis. Positively selected sites were identified under probabilistic models of codon substitution that allow for variable nonsynonymous/synonymous substitution rate ratios (ω) among sites (36). The likelihood ratio test was used to determine whether allowing for sites with a ω of >1 significantly improves the fit of the model to the data. If the ω ratio for any site class is above 1, we used the Bayes theorem to calculate the posterior probability that each site, given its data, is from such a site class. Three different model comparisons were performed. The neutral model M1, which incorporates a class of negatively selected (with proportion p_0 , giving ω_0) as well as neutral sites (p_1 , giving ω_1), was tested against the selection model M2, allowing for an extra class of sites under diversifying selection (p_2 , giving ω_2); model M0 with a single class of sites was tested against model M3 with a discrete distribution of three classes (with proportions p_0 , p_1 , and p_2 , giving ω_0 , ω_1 , and ω_2 equal to each with a specific ω ratio; M7, incorporating a beta distribution (with parameters p and q) to account for variable ω among neutral or negatively selected sites, was tested against model M8, to which an extra component allowing for positively selected sites was added (p_2 , giving ω_2). All calculations were performed using the CODEML program from the phylogenetic analysis by maximum likelihood (PAML) package version 4b10 (69).

Nucleotide sequence accession numbers. The nucleotide sequences from the Belgian isolates were deposited in the GenBank database under accession numbers AY343549 to AY343660.

RESULTS

Nucleotide sequence analysis. Genetic variability was determined by nucleotide sequencing of an 864-bp fragment in the G glycoprotein gene and the F protein gene of 112 HRSV-A strains isolated during 19 years from 1984 to 2002. The nucleotide sequences of a 629-bp fragment of the G glycoprotein gene (nt 284 to 912 in reference strain A2; GenBank accession number M11486) from the Belgian isolates were compared to 48 sequences published in GenBank (Table 1). Among the 112 Belgian strains, there were 81 unique sequences and 18 groups of identical sequences ($n = 49$, with a representative from each group in the phylogenetic tree). Fourteen groups with identical sequences were isolated during the same epidemic season; three groups with identical sequences were isolated during two consecutive epidemics, and one group contained isolates from three consecutive epidemic seasons. One isolate from Belgium (BE/1591/89-90) was found to be identical to one isolate from Madrid (MAD/4/90). Another strain from Madrid (MAD/4/91) was identical to three Belgian isolates (BE/6374/91-92, BE/6274/91-92, and BE/64/91-92) from the 1991-1992 season but contained an in-frame insertion of 6 nt after nt 593 (relative to the A2 reference strain; GenBank accession number M11486). No subgroup A strains were obtained from the epidemic season 1990-1991. All genetic changes in the Belgian isolates were base substitutions, and no deletions, insertions, or introductions of premature stop codons were identified.

Phylogenetic clustering of HRSV-A sequences. Eighty-one unique Belgian HRSV-A sequences and 48 GenBank-derived HRSV-A sequences were included in the phylogenetic analysis

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

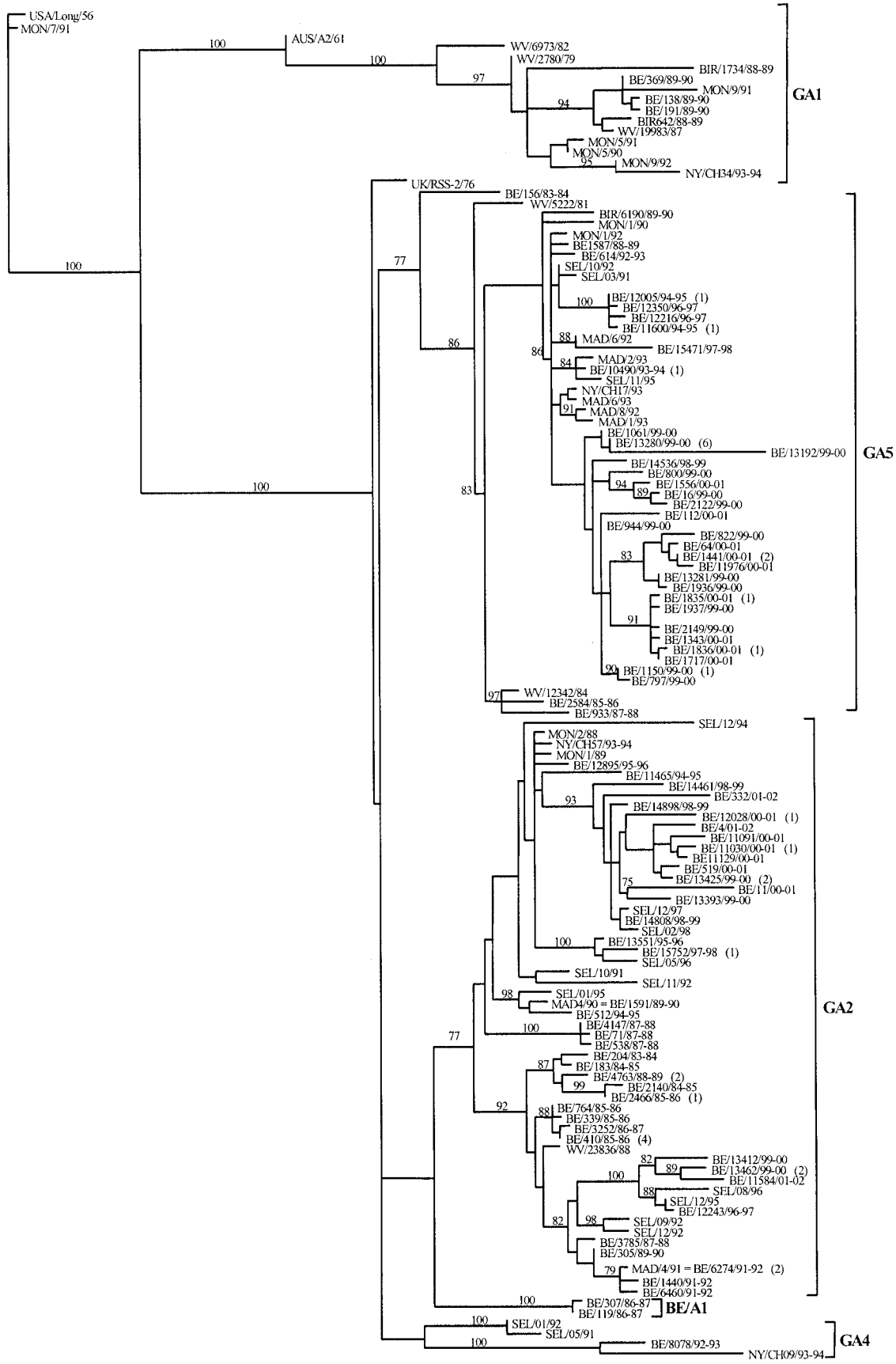
Strain	Accession no.	Country of isolation	Yr of isolation	Reference
USA/Long/56	M17212	U.S.A.	1956	25
AUS/A2/61	M11486	Australia	1961	64
BIR/642/89	X73354	U.K.	1989	8
BIR/1734/89	X73350	U.K.	1989	8
BIR/6190/89	X73352	U.K.	1989	8
MAD/4/90	Z33416	Spain	1990	14
MAD/4/91	Z33420	Spain	1991	14
MAD/6/92	Z33418	Spain	1992	14
MAD/8/92	Z33419	Spain	1992	14
MAD/1/93	Z33414	Spain	1993	14
MAD/2/93	Z33493	Spain	1993	14
MAD/6/93	Z33410	Spain	1993	14
MON/2/88	Z33424	Uruguay	1988	14
MON/1/89	Z33422	Uruguay	1989	14
MON/1/90	Z33494	Uruguay	1990	14
MON/5/90	Z33427	Uruguay	1990	14
MON/5/91	Z33428	Uruguay	1991	14
MON/7/91	Z33429	Uruguay	1991	14
MON/9/91	Z33431	Uruguay	1991	14
MON/1/92	Z33423	Uruguay	1992	14
MON/9/92	Z33432	Uruguay	1992	14
NY/CH09/93	AF065254	U.S.A.	1993	42
NY/CH17/93	AF065255	U.S.A.	1993	42
NY/CH57/94	AF065258	U.S.A.	1994	42
NY/CH34/94	AF065257	U.S.A.	1994	42
UK/RSS-2/76	U39662	U.K.	1976	60
SE/03/91	AF193304	Korea	1991	10
SE/05/91	AF193306	Korea	1991	10
SE/10/91	AF193307	Korea	1992	10
SE/01/92	AF193308	Korea	1992	10
SE/09/92	AF193309	Korea	1992	10
SE/10/92	AF193310	Korea	1992	10
SE/11/92	AF193311	Korea	1992	10
SE/12/92	AF193312	Korea	1992	10
SE/12/94	AF193316	Korea	1994	10
SE/01/95	AF193317	Korea	1995	10
SE/11/95	AF193318	Korea	1995	10
SE/12/95	AF193319	Korea	1995	10
SE/05/96	AF193321	Korea	1996	10
SE/08/96	AF193322	Korea	1996	10
SE/12/97	AF193323	Korea	1997	10
SE/02/98	AF193325	Korea	1998	10
WV/2780/79	AF065405	U.S.A.	1979	53
WV/5222/81	AF065406	U.S.A.	1981	53
WV/6973/82	AF065407	U.S.A.	1982	53
WV/12342/84	AF065409	U.S.A.	1984	53
WV/19983/87	AF065408	U.S.A.	1987	53
WV/23836/88	AF065410	U.S.A.	1988	53

(Fig. 1). The HRSV-A sequences clustered into five main lineages with bootstrap values of 77 to 100% (GA1, GA2, GA4, GA5, and BE/A1), with the exception of lineage GA4, which was supported by a lower bootstrap value (Fig. 1). Subgroup A lineages were designated GA1, GA2, GA4, and GA5 by phylogenetic comparisons with sequences that had previously been assigned to specific genotypes (42, 62). One reference strain (MAD/4/91) that was assigned to genotype GA3 in the study of Peret et al. (42) is assigned to genotype GA2 in this study. Lineages GA2 and GA5 comprise most of the Belgian isolates. Both GA2 and GA5 strains have been cocirculating in Belgium during 10 of the 19 epidemic seasons (1983-1984, 1985-1986, 1987-1988, 1988-1989, 1994-1995, 1996-1997, 1997-1998, 1998-1999, 1999-2000, and 2000-2001). No GA5 strains were found

in 6 of 19 epidemic seasons (1984-1985, 1986-1987, 1989-1990, 1990-1991, 1991-1992, and 1995-1996), and no GA2 strains were isolated in 3 of the 19 epidemic seasons (1990-1991, 1992-1993, and 1993-1994). Lineages GA1 and GA4 contain only a limited number of isolates, with only three Belgian isolates from the 1989-1990 season in lineage GA1 and one Belgian strain from the 1992-1993 season in lineage GA4. Two Belgian strains (BE/307/86-87 and BE/119/86-87) isolated in the epidemic season 1986-1987 grouped into a distinct cluster (lineage BE/A1). The phylogenetic tree (Fig. 1) shows that some strains isolated during the same epidemic season in Belgium (BE/12243/96-97 and BE/12216/96-97; BE/14808/98-99 and BE/14536/98-99) can show a greater divergence than isolates from two different geographical locations and different sampling years (BE/14808/98-99 and SEL/12/97; BE/12243/96-97 and SEL/12/95; BE/15752/97-98 and SEL/05/96; BE/10490/93-94, MAD/2/93, and SEL/11/95).

Analysis of selective pressure. The average nonsynonymous/synonymous substitution rate ratio in the HRSV-A G glycoprotein ranges from 0.5212 to 0.9304 among all model comparisons (Table 2), suggesting that a nonsynonymous mutation has about 52 to 93% as much chance as a synonymous mutation of being fixed in the population. The average acceptance rate of <1 indicates that, on average, purifying selection dominates the evolution of the G glycoprotein. Nevertheless, models that allow for positively selected sites (M2 [selection], M3 [discrete], and M8 [beta and ω]) all provide a significantly better fit to the data as evaluated by likelihood ratio test than do their counterpart models that only account for neutral and negatively selected sites (M1 [neutral], M0 [one ratio], and M7 [beta], respectively). Therefore, they all suggest the presence of positively selected sites with a proportion ranging from 7 to 10% (Table 2). Thirteen sites in the G glycoprotein gene had a posterior probability greater than 0.5. Four positively selected sites are located in the first hypervariable region (amino acid positions 97, 111, 117, and 132 referring to the Long strain), and nine positively selected sites are located in the C-terminal third of the ectodomain (amino acid positions 215, 225, 226, 247, 262, 274, 276, 290, and 297) (Fig. 2). Five amino acid positions (amino acid positions 117, 132, 225, 226, 262, and 290) were identified as being under positive selection at the 90% confidence level, four amino acid positions (amino acid positions 97, 215, 247, and 274) were above the 80% level, one positively selected site (site 297) was above the 70% level, and two amino acid positions (amino acid positions 111 and 276) were above the 60% level. By using the program NetOglyc (version 2.0), we predicted that all serine residues in positively selected sites 117 and 290 are O glycosylated with a high likelihood (for site 117, potential of 0.8857 to 0.9996 and threshold of 0.5571 to 0.6523; for site 290, potential of 0.9995 to 0.9944 and threshold of 0.5254 to 0.5483). By using the same program, we also predicted the threonine residues at positively selected site 225 to be O glycosylated (potential of 0.6118 to 0.7974 and threshold of 0.4830 to 0.5225).

Evolutionary rate. The correlation between phylogenetic branch length and the time of sampling of the viral strains is displayed in the root-to-tip regression plot (Fig. 3). The rate of substitution as estimated from this linear regression analysis was 1.47×10^{-3} nucleotide substitutions/site/year (95% CI, 1.11×10^{-3} to 2.18×10^{-3}). The MRCA of HRSV-A was



0.05

TABLE 2. Likelihood values and parameter estimates for the selection analysis of the HRSV G glycoprotein gene

Model	N ^a	ln L ^b	Parameter estimates	Avg ω
M0, one ratio	1	-5,370	$\omega = 0.5776$	0.5776
M1, neutral	1	-5,331	$p_0 = 0.25092$ ($p_1 = 0.74908$)	0.7491
M2, selection	3	-5,312	$p_0 = 0.24348, p_1 = 0.68638$ ($p_2 = 0.07014$), $\omega_2 = 3.47907$	0.9304
M3, discrete	5	-5,288	$p_0 = 0.49603, p_1 = 0.41734$ ($p_2 = 0.08663$), $\omega = 0.15010, \omega_1 = 0.84962, \omega_2 = 2.51499$	0.6469
M7, beta	2	-5,305	$p = 0.42523, q = 0.39058$	0.5212
M8, beta and ω	4	5,288	$p_0 = 0.89921, p = 0.65369, q = 0.80186$ ($p_2 = 0.10079$), $\omega = 2.37207$	0.6427

^a N, number of parameters in respective model.

^b ln L, log likelihood.

estimated to date back to 1940 (95% CI, 1,921 to 1,948). Although there appears to be a good correlation between branch length and time of sampling ($R^2 = 0.7565$), the linear regression is too conservative because the sequence data share a phylogenetic history and are therefore not independent. By using a genealogy-based maximum likelihood method, taking into account the phylogenetic dependency, the molecular clock assumption was significantly rejected ($P < 0.01$). Nevertheless, with another algorithm (SRDT), the estimates of the evolutionary rate and the MRCA are highly similar to the linear regression estimates. Under the SRDT model, the evolutionary rate was calculated as 1.83×10^{-3} nucleotide substitutions/site/year (95% CI, 1.44×10^{-3} to 2.26×10^{-3}), and the MRCA dates back to 1944 (95% CI, 1,937 to 1,950).

DISCUSSION

Circulation patterns of HRSV-A strains. This is the first study that has included HRSV strains isolated from Belgium. Phylogenetic analysis in this study confirmed previous observations that strains isolated in the same place during the same epidemic season (e.g., BE/12243/96-97 and BE/12216/96-97) may be more distantly related than isolates from two different geographical locations and different sampling years (e.g., BE/14808/98-99 and SEL/12/97) (6, 9, 10). We have also identified one isolate from Madrid (MAD/4/90) as being identical in the analyzed region to one Belgian strain (BE/1591/89-90), implying that certain outbreak strains may have the potential to spread over different countries. The presence of multiple identical sequences among the Belgian isolates suggests that certain strains are predominating in a given epidemic season. In addition, we have found one Belgian isolate from the epidemic season 1985-1986 to be identical to an isolate from 1987-1988, suggesting that HRSV-A strains can remain stable for more than one epidemic season.

Most of the Belgian strains belong to the two main lineages GA2 and GA5 (Fig. 1), and strains from these two genotypes have

been cocirculating among the Belgian population for at least 19 epidemic seasons. Lineages GA1, GA4, and BE/A1 contain only a limited number of Belgian isolates (Fig. 1). Previous studies have reported the disappearance of GA1 (designated genotype SHL5) in Birmingham and Vienna, after it had been the predominant genotype in the 1988-1989 epidemic season (7, 29). All three Belgian isolates from this lineage are from the epidemic season 1989-1990, implying rapid spread and predominant circulation of GA1 strains in Europe during the period 1988-1990. A recent study in Argentina (13) has identified GA1 strains isolated in 1996. These results indicate that the epidemiology of HRSV-A is likely determined by local factors, such as immune resistance in the community to certain genotypes. It has been suggested that herd immunity, or high levels of maternal immunity to a particular epidemic strain, may not only reduce the disease severity caused by that genotype but also may restrict the circulation of that genotype in subsequent epidemics (3). We therefore speculate that elevated levels of herd immunity to GA1, GA4, and BE/A1 may be a reason for reducing their effective population size, leading to the disappearance of these lineages in Belgium.

The circulation pattern of HRSV-A resembles that of the influenza B virus in that multiple lineages can cocirculate for extended periods of time, with new variants replacing older ones within each lineage, due to antigenic drift (26, 28, 46, 47).

Positive selection in the G glycoprotein ectodomain. Although we observe a high average ratio of nonsynonymous to synonymous nucleotide substitutions in the G glycoprotein gene ($\omega = 0.6469$), this value does not surpass the threshold of $\omega > 1$, and it is therefore not indicative of positive selection (68). Since the average ω is usually not sensitive enough to detect Darwinian selection at the molecular level, we made use of codon substitution models to detect sites under positive selection. By using these methods, we have identified 13 codon sites in the two hypervariable regions of the HRSV-A G glycoprotein ectodomain as being under positive selection pressure with posterior probabilities above 0.5 (Fig. 2). This indicates that, although a high proportion of amino acids can be

FIG. 1. Phylogenetic tree of HRSV-A strains as computed with paup4b10. The nucleotide sequences of the G glycoprotein gene of the Belgian isolates were compared with those from Madrid (MAD), Montevideo (MON), Seoul (SEL), West Virginia (WV), and Birmingham (BIR) with the original nomenclature RSB89-642 (BIR642/88-89), RSB89-1734 (BIR1734/88-89), and RSB89-6190 (BIR6190/89-90) and compared with those from New York (NY) with original nomenclature CH17 (NY/CH17/93), CH09 (NY/CH09/93), CH57 (NY/CH57/94), CH34 (NY/CH34/94), and the reference strains Long (USA/Long/56), A2 (AUS/A2/61), and RSS-2 (UK/RSS-2/76). The Long strain was used as the outgroup sequence in the tree. The numbers at the internal nodes represent the number of bootstrap probabilities, as determined for 1,000 iterations by the neighbor-joining method. Only bootstrap values greater than 77% are shown. The italicized numbers in brackets at the terminal nodes correspond to the numbers of identical sequences. The genetic clusters obtained in the analysis are indicated by the square brackets and the designations GA1, GA2, GA4, GA5, and BE/A1.

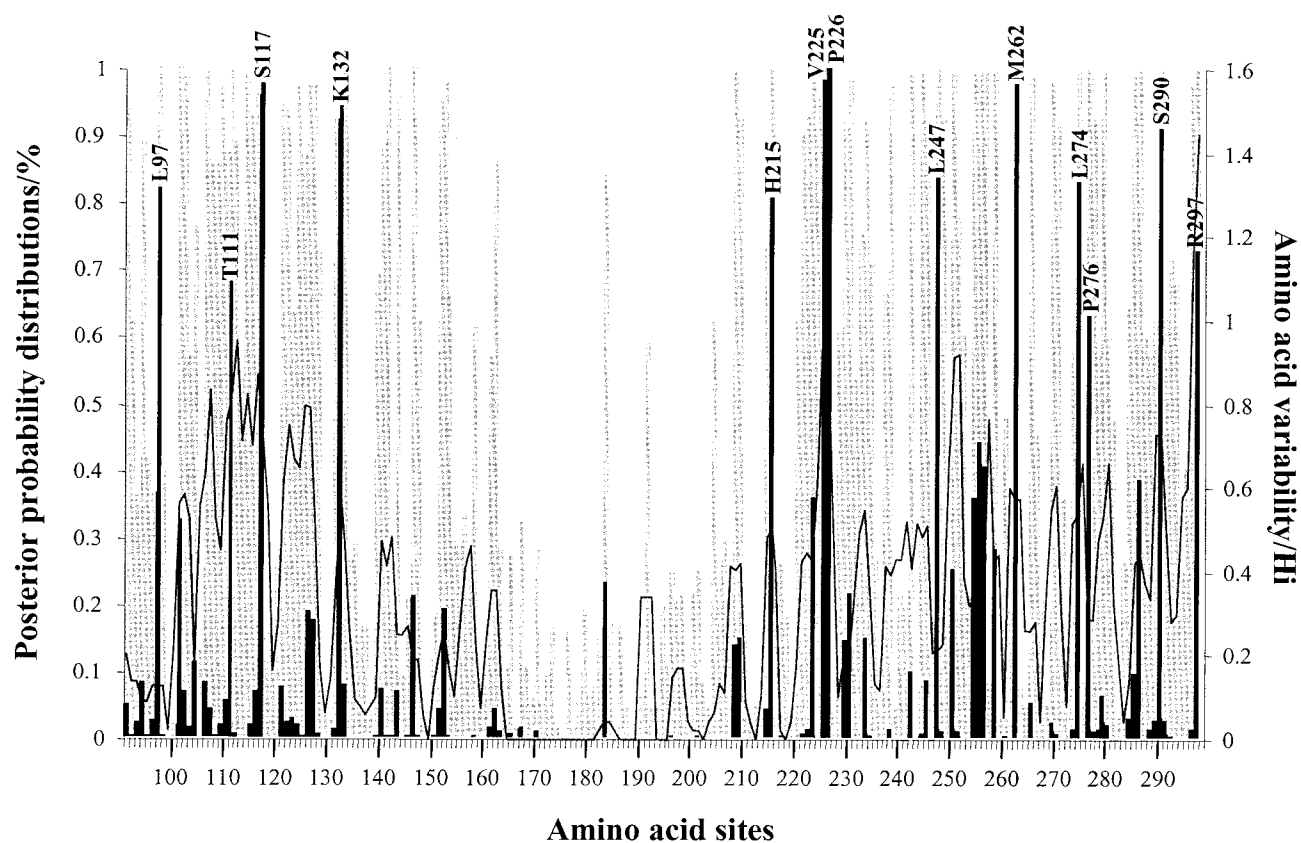


FIG. 2. Posterior probabilities of site classes along the G protein ectodomain region under the discrete model M3. This model assumes three classes of sites in the gene: positive sites (■), neutral sites (□), and negative sites (▣). Positively selected amino acid sites with posterior probabilities above 50% are indicated according to the Long strain. Amino acid variability, measured by entropy (H_i), is plotted to the second y axis.

largely invariable due to structural and functional constraints (52), adaptive evolution may occur at certain sites of the genome. Several linear epitopes have been detected in the G glycoprotein of HRSV-A by analyzing the reaction of human convalescent-phase sera with peptides based on the amino acid sequence of the ectodomain (4, 37) and by sequencing escape mutants selected with individual monoclonal antibodies (MAbs). There is a strong association between the positively selected sites identified in this study and the mapped neutralizing epitopes. Amino acid replacements in 7 of the 13 positively selected sites (sites 97, 215, 225, 226, 274, 290, and 297) have been described previously in escape mutants selected with specific MAbs. Positively selected site 97 in the first hypervariable region of the ectodomain corresponds to one of the amino acid substitutions (Leu to Arg) found in an escape mutant of the HRSV Long strain, which is resistant to the MAb c793 that recognizes a conserved (i.e., reacting with both HRSV-A and -B) G glycoprotein epitope (34, 49). Three positively selected sites in the C-terminal third of the ectodomain (sites 215, 225, and 226) are in a region that is thought to include important determinants for the induction of neutralizing antibodies (15, 38). A His-for-Pro substitution at positively selected site 215 contributed to the loss of reactivity to MAb 63G of HRSV-A strain RSB89-1734 (6). Positively selected site 225 is one of the changed sites in an escape mutant of the HRSV Long strain,

selected with group-specific MAb L9, which can neutralize both HRSV-A and -B strains (63). Leu-for-Pro changes in positively selected sites 226, 274, and 290 have been described for isolates that have lost both conserved and group-specific epitopes (sites 226 and 290) (31) and strain-specific epitopes (site 274) of the G glycoprotein (6, 49). The Arg-for-His amino acid substitution at positively selected position 297 has been shown to influence the integrity of multiple overlapping strain-specific epitopes (48). Virus infectivity has been shown to be sensitive to the limited removal of N-linked or O-linked oligosaccharides (27). Thus, sequence changes may influence the location of carbohydrate side chains, which are important determinants of the G glycoprotein antigenic structure (4, 17, 39, 40). We predicted that serine residues at positively selected sites 117 and 290 are O glycosylated with a high potential. Threonine residues at positively selected site 225 were also predicted to be O glycosylated. It is possible that strains that have lost O glycosylation can escape the immune system by losing recognition of a carbohydrate epitope.

Less is known about the effects of amino acid replacements in 5 of the 13 positively selected sites (sites 111, 117, 132, 247, 262, and 276). Positions 111, 117, and 132 are within an area (codons 82 to 153) presumed to be an antigenic region in the *bovine respiratory syncytial virus* (BRSV) G protein (61). Furthermore, at positively selected site 117, we predicted serine

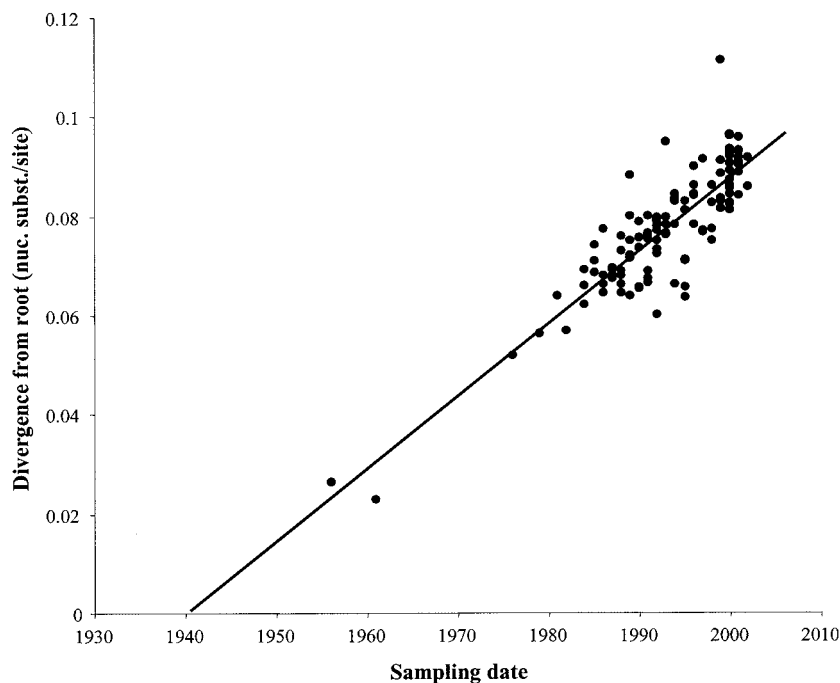


FIG. 3. Linear root-to-tip regression plot.

residues to be O glycosylated. Immunization with synthetic peptides that encompass the region that contains the positively selected site 132 (amino acid residues 124 to 203) has been shown to confer complete resistance to HRSV replication in mice (51), suggesting that this region plays an important role in protection. We were not able to find any involvement of positively selected sites 111, 247, 262, and 276 either in escape mutants or in O glycosylation. Further analysis of the immune responses to peptides with selected mutations in these amino acid positions will help to elucidate their functional importance in the G glycoprotein.

The uneven distribution of the positively selected sites identified along the G glycoprotein ectodomain of HRSV-A suggests that new antigenic variants can be selected by immune pressure operating mainly in the carboxy-terminal third of the G glycoprotein. Adaptive evolution in codon positions that are part of conserved and group-specific epitopes could lead to strains with an increased virulence because of a diminished serological protection of the antibodies obtained in previous HRSV infections.

Rate of evolution of HRSV-A strains. Our evolutionary rate analysis encompassed the largest available data set of HRSV-A sequences of 629-bp lengths isolated over 47 years, including newly sequenced strains obtained over 19 consecutive years. Sampling in time results in heterochronous data, allowing us to calculate an evolutionary rate and to estimate the time of the MRCA of the known up-to-date HRSV-A sequences. The estimated rate of accumulation of nucleotide changes in the G glycoprotein gene of HRSV (1.83×10^{-3} nucleotide substitutions/site/year) is comparable to that observed in other rapidly evolving RNA viruses, like the hemagglutinin gene of influenza B viruses, and is highly similar to a previous estimate of 1.6×10^{-3} nucleotide substitutions/site/year (24, 28). Although the

HRSV-A data set has only four sequences before 1981 (USA/Long/56, AUS/A2/61, UK/RSS-2/76, and WV/2780/79), the linear regression analysis shows that they fit well on the regression line (Fig. 3), implying a roughly constant rate of evolution during the period from 1956 to 1980. Furthermore, the linear regression analysis reveals a strong correlation between the accumulation of genetic divergence and the isolation date of the sequences included in this study. Therefore we suggest that the MCRA of what is considered now as HRSV-A dates back to the early 1940s.

A recent study on BRSV has shown the existence of stronger positive-selection pressure on the BRSV G protein in countries where BRSV vaccination had been more widely used, indicating evolution due to immunological pressure (61). Our results provide evidence that the mode of evolution of HRSV-A is likely influenced by selection of new antigenic variants due to positive selection operating at certain codon positions, located mainly in the second hypervariable region of the G glycoprotein ectodomain. This supports the findings of Woelk and Holmes (66), confirming five of the six positively selected sites (sites 117, 215, 225, 226, and 297) that they have reported for HRSV-A. We have identified eight novel codon sites as being under adaptive evolution (amino acid positions 97, 111, 132, 247, 262, 274, 276, and 290 according to the Long reference strain). A recent study has demonstrated a greater severity of illness associated with a particular HRSV-A clade by using sequence data from the G glycoprotein gene (30). Therefore, evaluation of sequence variation can be useful for elucidating the biological function of certain amino acids, as well as for identifying genetic factors associated with greater virulence. Furthermore, monitoring sequence changes in positions that are under positive selection may provide potentially

useful information for identifying future dominant epidemic strains.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Union Shipping and Trade Company Ltd., Sofia, Bulgaria, and by a doctoral scholarship in the framework of the Central and Eastern European Initiatives from the University of Leuven.

REFERENCES

- Akerlind, B., and E. Norrby. 1986. Occurrence of respiratory syncytial virus subtypes A and B strains in Sweden. *J. Med. Virol.* **19**:241–247.
- 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.
- Cane, P. A. 2001. Molecular epidemiology of respiratory syncytial virus. *Rev. Med. Virol.* **11**:103–116.
- Cane, P. A. 1997. Analysis of linear epitopes recognised by the primary human antibody response to a variable region of the attachment (G) protein of respiratory syncytial virus. *J. Med. Virol.* **51**:297–304.
- Cane, P. A., and C. R. Pringle. 1995. Molecular epidemiology of respiratory syncytial virus: a review of the use of reverse transcription-polymerase chain reaction in the analysis of genetic variability. *Electrophoresis* **16**:329–333.
- 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.
- Cane, P. A., D. A. Matthews, and C. R. Pringle. 1994. Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. *J. Clin. Microbiol.* **32**:1–4.
- 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.
- Cane, P. A., and C. R. Pringle. 1991. Respiratory syncytial virus heterogeneity during an epidemic: analysis by limited nucleotide sequencing (SH gene) and restriction mapping (N gene). *J. Gen. Virol.* **72**:349–357.
- Choi, E. H., and H. J. Lee. 2000. Genetic diversity and molecular epidemiology of the G protein of subgroups A and B of respiratory syncytial viruses isolated over 9 consecutive epidemics in Korea. *J. Infect. Dis.* **181**:1547–1556.
- Collins, P. L., K. McIntosh, and R. M. Chanock. 1996. Respiratory syncytial virus, p 1313–1351. *In* B. N. Fields, D. M. Knipe, P. M. Howley, et al. (ed.), *Fields virology*. Lippincott-Raven, Philadelphia, Pa.
- Falsey, A. R., and E. E. Walsh. 2000. Respiratory syncytial virus infection in adults. *Clin. Microbiol. Rev.* **13**:371–384.
- Frabasile, S., A. Delfraro, L. Facal, C. Videla, M. Galiano, M. J. de Sierra, D. Ruchansky, N. Viturra, M. Berois, G. Carballal, J. Russi, and J. Arbiza. 2003. Antigenic and genetic variability of human respiratory syncytial viruses (group A) isolated in Uruguay and Argentina: 1993–2001. *J. Med. Virol.* **71**:305–312.
- Garcia, O., M. Martin, J. Dopazo, J. Arbiza, S. Frabasile, J. Russi, M. Hortal, P. Perez-Brena, I. Martinez, and B. Garcia-Barreno. 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.
- 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.
- Garcia-Barreno, B., C. Palomo, C. Penas, T. Delgado, P. Perez-Brena, and J. A. Melero. 1989. Marked differences in the antigenic structure of human respiratory syncytial virus F and G glycoproteins. *J. Virol.* **63**:925–932.
- Garcia-Beato, R., I. Martinez, C. Franci, F. X. Real, B. Garcia-Barreno, and J. A. Melero. 1996. Host cell effect upon glycosylation and antigenicity of human respiratory syncytial virus G glycoprotein. *Virology* **221**:301–309.
- Glezen, W. P., L. H. Taber, A. L. Frank, and J. A. Kasel. 1986. Risk of primary infection and reinfection with respiratory syncytial virus. *Am. J. Dis. Child.* **140**:543–546.
- Glezen, W. P., A. Paredes, J. E. Allison, L. H. Taber, and A. L. Frank. 1981. Risk of respiratory syncytial virus infection for infants from low-income families in relationship to age, sex, ethnic group, and maternal antibody level. *J. Pediatr.* **98**:708–715.
- 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.
- Hall, C. B., E. E. Walsh, K. C. Schnabel, C. E. Long, K. M. McConochie, 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.
- 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. *Glycoconjugate J.* **15**:115–130.
- Hendry, R. M., L. T. Pierik, and K. McIntosh. 1989. Prevalence of respiratory syncytial virus subgroups over six consecutive outbreaks: 1981–1987. *J. Infect. Dis.* **160**:185–190.
- 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.
- 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.
- Kanegae, Y., S. Sugita, A. Endo, M. Ishida, S. Senya, K. Osako, K. Nerome, and A. Oya. 1990. Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. *J. Virol.* **64**:2860–2865.
- Lambert, D. M. 1988. Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins. *Virology* **164**:458–466.
- Lindstrom, S. E., Y. Hiromoto, H. Nishimura, T. Saito, R. Nerome, and K. Nerome. 1999. Comparative analysis of evolutionary mechanisms of the hemagglutinin and three internal protein genes of influenza B virus: multiple cocirculating lineages and frequent reassortment of the NP, M, and NS genes. *J. Virol.* **73**:4413–4426.
- Lukic-Grlic, A., P. A. Cane, A. Bace, C. R. Pringle, G. Mlinaric-Galinovic, and T. Popow-Kraupp. 1998. Antigenic and genomic diversity of central European respiratory syncytial virus strains. *Arch. Virol.* **143**:1441–1447.
- Martinello, R. A., M. D. Chen, C. Weibel, and J. S. Kahn. 2002. Correlation between respiratory syncytial virus genotype and severity of illness. *J. Infect. Dis.* **186**:839–842.
- 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.
- 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.
- 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.
- 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.
- Nicholas, K. B., Jr., H. B. Nicholas, and D. W. Deerfield II. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBnet News* **4**:1–4.
- 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.
- Norrby, E., M. A. Mufson, H. Alexander, R. A. Houghton, and R. A. Lerner. 1987. Site-directed serology with synthetic peptides representing the large glycoprotein G of respiratory syncytial virus. *Proc. Natl. Acad. Sci. USA* **84**:6572–6576.
- Olmsted, R. A., B. R. Murphy, L. A. Lawrence, N. Elango, B. Moss, and P. L. Collins. 1989. Processing, surface expression, and immunogenicity of carboxy-terminally truncated mutants of G protein of human respiratory syncytial virus. *J. Virol.* **63**:411–420.
- 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.
- 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.
- 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.
- 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.
- Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* **50**:580–601.
- Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**:395–399.
- 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.
- Rota, P. A., M. L. Hemphill, T. Whistler, H. L. Regnery, and A. P. Kendal.

1992. Antigenic and genetic characterization of the haemagglutinins of recent cocirculating strains of influenza B virus. *J. Gen. Virol.* **73**:2737–2742.
47. **Rota, P. A., T. R. Wallis, M. W. Harmon, J. S. Rota, A. P. Kendal, and K. Nerome.** 1990. Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology* **17**:59–68.
48. **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.
49. **Rueda, P., B. Garcia-Barreno, and J. A. Melero.** 1994. Loss of conserved cysteine residues in the attachment (G) glycoprotein of two human respiratory syncytial virus escape mutants that contain multiple A-G substitutions (hypermutations). *Virology* **198**:653–662.
50. **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.
51. **Simard, C., F. Nadon, C. Seguin, and M. Trudel.** 1995. Evidence that the amino acid region 124–203 of glycoprotein G from the respiratory syncytial virus (RSV) constitutes a major part of the polypeptide domain that is involved in the protection against RSV infection. *Antivir. Res.* **28**:303–315.
52. **Simmonds, P., and D. B. Smith.** 1999. Structural constraints on RNA virus evolution. *J. Virol.* **73**:5787–5794.
53. **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.
54. **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.
55. **Sullender, W. M., and G. W. Wertz.** 1991. Synthetic oligonucleotide probes differentiate respiratory syncytial virus subgroups in a nucleic acid hybridization assay. *J. Clin. Microbiol.* **29**:1255–1257.
56. **Sullender, W. M., and G. W. Wertz.** 1991. The unusual attachment glycoprotein of the respiratory syncytial viruses: structure, maturation, and role in immunity, p. 383–406. *In* D. Kingbury (ed.), *The paramyxoviruses*. Plenum Press, New York, N.Y.
57. **Swofford, D. L.** 1998. PAUP* 4.0—phylogenetic analysis using parsimony. Sinauer Associates, Sunderland, Mass.
58. **Taylor, G. S., I. B. Vipond, and E. O. Caul.** 2001. Molecular epidemiology of outbreak of respiratory syncytial virus within bone marrow transplantation unit. *J. Clin. Microbiol.* **39**:801–803.
59. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
60. **Tolley, K. P., A. C. Marriott, A. Simpson, D. J. Plows, D. A. Matthews, S. J. Longhurst, J. E. Evans, J. L. Johnson, P. A. Cane, V. B. Randolph, A. J. Easton, and C. R. Pringle.** 1996. Identification of mutations contributing to the reduced virulence of a modified strain of respiratory syncytial virus. *Vaccine* **14**:1637–1646.
61. **Valarcher, J. F., F. Schelcher, and H. Bourhy.** 2000. Evolution of bovine respiratory syncytial virus. *J. Virol.* **74**:10714–10728.
62. **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.
63. **Walsh, E. E., A. R. Falsey, and W. M. Sullender.** 1998. Monoclonal antibody neutralization escape mutants of respiratory syncytial virus with unique alterations in the attachment (G) protein. *J. Gen. Virol.* **79**:479–487.
64. **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.
65. **Whimbey, E., and S. Ghosh.** 2000. Respiratory syncytial virus infections in immunocompromised adults. *Curr. Clin. Top. Infect. Dis.* **20**:232–255.
66. **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.
67. **Xia, X., and Z. Xie.** 2001. DAMBE: data analysis in molecular biology and evolution. *J. Hered.* **92**:371–373.
68. **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.
69. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.