J Mol Evol (2001) 52:182–192
DOI: 10.1007/s002390010147

JOURNAL OF MOLECULAR EVOLUTION

Variable Immune-Driven Natural Selection in the Attachment (G) Glycoprotein of Respiratory Syncytial Virus (RSV)

Christopher H. Woelk, Edward C. Holmes

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received: 15 August 2000 / Accepted: 2 November 2000

Abstract. A maximum-likelihood analysis of selection pressures acting on the attachment (G) glycoprotein gene of respiratory syncytial virus (RSV) from humans (HRSV) and bovines (BRSV) is presented. Six positively selected sites were identified in both group A and group B of HRSV, although only one site was common between them, while no positively selected sites were detected in BRSV. All positively selected sites were located within the ectodomain of the G protein and showed some association with positions of immunoglobulin (Ig) epitopes and sites of O-glycosylation. These results suggest that immune (antibody)-driven natural selection is an important determinant of RSV evolution and that this selection pressure differs among strains. The passage histories of RSV strains were also shown to affect the distribution of positively selected sites, particularly in HRSV B, and should be considered whenever retrospective analysis of adaptive evolution is undertaken.

Key words: Respiratory syncytial virus — Positive selection — Maximum likelihood — Glycosylation — Passage history — Immune selection

Introduction

Respiratory syncytial virus (RSV) is a negative-stranded RNA virus (genus *Pneumovirus,* family *Paramyxoviridae*) that induces a respiratory tract disease commonly seen in infants and young children (Collins et al. 1996). Both human (HRSV) and bovine (BRSV) forms exist, with the latter causing similar symptoms in cattle to those observed in humans (Baker et al. 1986; Collins and McIntosh 1995). Two major groups of HRSV, designated HRSV A and HRSV B, have been identified by antigenic and phylogenetic analyses and differ by up to 45% in amino acid sequence (Johnson et al. 1987; Mufson et al. 1985), whereas no such distinction is present among BRSV strains.

Key to understanding the evolution and epidemiology of RSV, and hence central to our efforts to control the virus, is determining the role played by antigenic variation. For example, a variety of antibody escape mutants has been characterized in the viral G glycoprotein (reviewed by Melero et al. 1997) and antigenic differences between HRSV A and HRSV B, as well as smaller-scale antigenic variation within groups, may facilitate reinfections with the virus (Sullender et al. 1998). More pertinently, there is as yet no safe and effective vaccine to protect against RSV infection, and although a BRSV vaccine exists, it does not inhibit the circulation of the virus in cattle herds (Schrijver et al. 1998; Van der Poel et al. 1993).

In both HRSV and BRSV the viral G glycoprotein, responsible for attachment to the host cell, is the major target of the immune response (García et al. 1994). The protein varies in length from 257 to 299 amino acids and the mature 90-kDa form is obtained through extensive carbohydrate modification of a 32-kDa precursor polypeptide by the addition of N-linked carbohydrates to Asn residues and O-linked carbohydrates to Ser and Thr residues (Wertz et al. 1989). The protein can be divided into

Correspondence to: Dr. Edward C. Holmes; *e-mail:* Edward.Holmes@zoo.ox.ac.uk

an intracellular domain, a transmembrane domain, and a large ectodomain, the latter of which in turn is comprised of two highly glycosylated mucin-like variable regions that sandwich a conserved region between amino acid 148 and amino acid 197 (Langedijk et al. 1996). This conserved region includes a cluster of four cysteines (amino acids 173, 176, 182, and 186) and its moderate hydrophobicity suggests the presence of a receptor binding domain (Johnson et al. 1987).

Previous studies have suggested that certain regions of the G protein may be under immune selection. For example, it has been proposed that RSV modulates glycosylation sites to abolish or mask epitopes (Melero et al. 1997; Palomo 2000). Point mutations might also change epitopes (García et al. 1994; Martínez et al. 1997), as could hypermutations (Martínez et al. 1997; Rueda et al. 1994), frameshifts (García-Barreno et al. 1990), and alternative termination codons (Martínez et al. 1999; Rueda et al. 1995). Simple pairwise comparisons of the numbers of nonsynonymous (d_N) and synonymous (d_S) substitutions per site, either in the entire G gene or the variable regions considered alone, have also provided evidence for positive selection (Cane et al. 1991; García et al. 1994; Larsen et al. 1998; Stine et al. 1997, Melero, 1997; Sullender et al. 1991, 1998). Moreover, it has also been suggested that immunological pressures may differ among BRSV, HRSV A, and HRSV B because of underlying differences in the variability of the central hydrophobic region (Prozzi et al. 1997).

In this paper we have employed more sophisticated maximum-likelihood analyses of d_N/d_S to determine, on a site-specific basis, the selection pressures acting on the G protein of HRSV and BRSV and how these pressures might differ among viruses. Furthermore, since some RSV strains were originally obtained over 40 years ago and may have been extensively manipulated in cell culture since this time, we also investigated whether their inclusion affected the analysis of selection pressures.

Materials and Methods

Complete and nearly complete RSV G protein gene sequences were obtained from GenBank (Table 1). Sequences were divided into BRSV, HRSV A and HRSV B data sets and each was aligned using CLUSTALW (Thompson et al. 1994). All alignments are available from the authors on request. Even after identical sequences were removed, the BRSV and HRSV A data sets contained too many sequences for an efficient maximum-likelihood (ML) analysis of selection pressures so that further pruning was needed. All sequences with greater than 99% similarity to each other in the data set were therefore deleted. Yang (1998) proposed that the results of selection analysis will show little change when sequences with a high level of similarity are pruned from the data set. In all, 12 sequences were removed from the BRSV data set, leaving 18 sequences, and 18 HRSV A sequences were removed, to form a new data set of 26. Those sequences deleted are indicated in Table 1. Due to its small size, the HRSV B data set did not need pruning and contains 12 sequences. A duplicated codon that led to an insertion of a Thy amino acid in the C-terminal portion of HRSV B strain NM1355 was also removed. A further partial HRSV A data set was constructed by concatenating sequence data from amino acid 90 to amino acid 132 and from amino acid 262 to amino acid 298. This alignment was constructed from the complete HRSV A sequences described above and a previously published data set of partial G protein sequences (Cane et al. 1995). Sequences that had greater than 99% similarity were again pruned from the data set and thus MON/9/92 and 12 sequences from the Cane et al. (1995) data set were excluded. Those partial sequences retained and subjected to selection analysis are presented in Table 2.

To investigate the effects of passage history on positive selection, a further three data sets were compiled which mirrored the complete data sets just described except that strains isolated prior to 1980, and so possibly passaged many times, were excluded. Although the 1980 date is somewhat arbitrary, this analysis will at least show whether adaptation to cell culture might have affected our results. Consequently, BRSV strains A51908, FS-1, 375, 127, Snook, 220-69, NMK7, and Dorset were removed to form the BRSV′ data set, the Long, A2, and RSS-2 strains were removed to form the HRSV A′ data set, and the 18537, 8/60, and 9320 strains were removed to form the HRSV B′ data set.

ML trees of each of the seven data sets were estimated using the PAUP* package (Swofford, 1998). The HKY85 model of nucleotide substitution was used with values for both the transition/transversion ratio and the shape parameter (α) of a gamma distribution of rate variation among sites (with eight categories) estimated during tree reconstruction. These parameter values are given in Tables 3a and 3b (trees not shown; available at http://evolve.zoo.ox.ac.uk). For a review of the phylogenetic relationships of HRSV see Sullender (2000) and Elvander et al. (1998) for BRSV. To determine how well different groupings were supported, a bootstrap analysis was undertaken using 1000 replicate neighbor-joining trees with input distances estimated under the ML substitution model.

A ML approach was also used to examine selection pressures. This analysis utilizes a series of models of codon substitution (Yang et al. 2000). All models require a phylogenetic tree (described above) with the instantaneous transition/transversion rate ratio (K) and base frequencies at the three codon positions given. The models either estimate or fix ω parameters, equivalent to d_N/d_S ratios. The simplest model, M0, calculates a single ω parameter for all sites. The M1 model accounts for neutral evolution by assuming that a proportion of sites (p_0) is conserved with ω_0 fixed at 0, and that a proportion of sites (p_1) is neutral with ω_1 set to 1. The M2 model attempts to account for positive selection with the addition of a third category of sites (p_2) for which ω_2 is estimated from the data. The M3 model provides a more sensitive test for positive selection by using a discrete distribution to estimate the ω ratio for a predetermined number of classes (in this case three). The M7 and M8 models both use a discrete beta distribution (with 10 categories) to model different ω ratios among sites, although the M8 model can estimate a further class of sites with a $\omega > 1$. Models that are nested can be compared directly using a likelihood ratio test (LRT) (Nielsen and Yang 1998). In particular, as both M0 and M1 are nested with M2 and M3, these comparisons can be used to detect the presence of positive selection and hence reject neutral evolution (providing $\omega > 1$). Models M0 and M1 are not nested. M7 and M8 are also nested and thereby provide a further test for positive selection. Empirical Bayes theory can also be used to calculate the probability that a site is from a particular site class (i.e., conserved, neutral, or positively selected).

By making use of the free ratio (FR) model it is also possible to detect individual branches that have undergone positive selection (Yang 1998). Instead of analyzing sites, the FR model estimates the d_N/d_S ratio for the entire gene along each branch and can be compared to the M0 model, which assumes the same d_N/d_S ratio for the whole tree. The methods and models described were implemented using the CODEML program of the PAML package version 3.0prior (Yang 1997).

The NetOglyc program (version 2.0) was used to determine the extent of O-glycosylation at sites along the G protein (Hansen et al. 1998).

Table 1. RSV strains used in this study[a]

Virus	Accession No.	Strain	Country of isolation	Year of isolation	Reference
BRSV	**L08410**	**375**	**U.S.A.**	**1971**	**Stine et al. (1997)**
BRSV	L08411	88-294	U.S.A.	NK	Stine et al. (1997)
BRSV	L08412	87-14594	U.S.A.	NK	Stine et al. (1997)
BRSV	**L08413**	**234-489**	**U.S.A.**	**NK**	**Stine et al. (1997)**
BRSV	L08414	236-652	U.S.A.	NK	Stine et al. (1997)
BRSV	**L08415**	**1344R**	**U.S.A.**	**NK**	**Stine et al. (1997)**
BRSV	**L08416**	**1540R**	**U.S.A.**	**NK**	**Stine et al. (1997)**
BRSV	L08417	NDKS-7	U.S.A.	NK	Stine et al. (1997)
BRSV	L10925	375	U.S.A.	1971	Stine et al. (1997)
BRSV	M58307	391-2	U.S.A.	1984–1985[b]	Lerch et al. (1990)
BRSV	From paper	A51908	U.S.A.	1975	Mallipeddi & Samal (1993)
BRSV	From paper	FS-1	U.S.A.	1971	Mallipeddi & Samal (1993)
BRSV	**From paper**	**VC464**	**U.S.A.**	**1972**	**Mallipeddi & Samal (1993)**
BRSV	U24713	NMK7	Japan	1968–1969[b]	Prozzi et al. (1997)
BRSV	**U24714**	**MVR553**	**Belgium**	**1983**	**Prozzi et al. (1997)**
BRSV	U24715	Dorset	U.K.	before 1971	Prozzi et al. (1997)
BRSV	U24716	85-1330	U.S.A.	1985	Prozzi et al. (1997)
BRSV	**U33539**	**Lelystad**	**Netherlands**	**1983–1984[b]**	**Prozzi et al. (1997)**
BRSV	**U57823**	**BovX**	**Sweden**	**1967**	**Prozzi et al. (1997)**
BRSV	**U92098**	**8307027**	**Denmark**	**1983**	**Larsen et al. (1998)**
BRSV	**U92100**	**88LU195**	**Denmark**	**1988**	**Larsen et al. (1998)**
BRSV	U92101	9304899	Denmark	1993	Larsen et al. (1998)
BRSV	U92102	9314893	Denmark	1993	Larsen et al. (1998)
BRSV	**U92103**	**9402020**	**Denmark**	**1994**	**Larsen et al. (1998)**
BRSV	U92104	9402022	Denmark	1994	Larsen et al. (1998)
BRSV	Y08716	127	U.K.	1973	Furze et al. (1997)
BRSV	Y08717	WBH	Holland	1986	Furze et al. (1997)
BRSV	**Y08718**	**4642**	**U.K.**	**1976**	**Furze et al. (1997)**
BRSV	Y08719	Snook	U.K.	1976	Furze et al. (1997)
BRSV	Y08720	220-69	Belgium	1969	Furze et al. (1997)
HRSV A	AFO65254	CH09	U.S.A.	1990–1995[b]	Peret et al. (1998)
HRSV A	**AFO65255**	**CH17**	**U.S.A.**	**1990–1995[b]**	**Peret et al. (1998)**
HRSV A	**AFO65256**	**CH28**	**U.S.A.**	**1990–1995[b]**	**Peret et al. (1998)**
HRSV A	AFO65257	CH34	U.S.A.	1990–1995[b]	Peret et al. (1998)
HRSV A	**AFO65258**	**CH57**	**U.S.A.**	**1990–1995[b]**	**Peret et al. (1998)**
HRSV A	**AFO65405**	**WV2780**	**U.S.A.**	**1979**	**Sullender et al. (1998)**
HRSV A	AFO65406	WV5222	U.S.A.	1981	Sullender et al. (1998)
HRSV A	AFO65407	WV6973	U.S.A.	1982	Sullender et al. (1998)
HRSV A	**AFO65408**	**WV19983**	**U.S.A.**	**1987**	**Sullender et al. (1998)**
HRSV A	AFO65409	WV12342	U.S.A.	1984	Sullender et al. (1998)
HRSV A	AFO65410	WV23836	U.S.A.	1988	Sullender et al. (1998)
HRSV A	M11486	A2	Australia	1961	Wertz et al. (1985)
HRSV A	M17212	Long	U.S.A.	1956	Johnson et al. (1987)
HRSV A	U39662	RSS-2	U.K.	1976	Tolley et al. (1996)
HRSV A	X73350	RSB89-642[c]	U.K.	1989	Cane et al. (1991)
HRSV A	X73351	RSB89-5857	U.K.	1989	Cane et al. (1991)
HRSV A	X73352	RSB89-6190	U.K.	1989	Cane et al. (1991)
HRSV A	X73353	RSB89-6256	U.K.	1989	Cane et al. (1991)
HRSV A	X73354	RSB89-1734[c]	U.K.	1989	Cane et al. (1991)
HRSV A	X73355	RSB89-6614	U.K.	1989	Cane et al. (1991)
HRSV A	Z33410	MAD/6/93	Spain	1993	García et al. (1994)
HRSV A	Z33411	MAD/3/92	Spain	1992	García et al. (1994)
HRSV A	Z33412	MAD/1/89	Spain	1989	García et al. (1994)
HRSV A	**Z33414**	**MAD/1/93**	**Spain**	**1993**	**García et al. (1994)**
HRSV A	**Z33415**	**MAD/3/89**	**Spain**	**1989**	**García et al. (1994)**
HRSV A	Z33416	MAD/4/90	Spain	1990	García et al. (1994)
HRSV A	**Z33417**	**MAD/5/92**	**Spain**	**1992**	**García et al. (1994)**
HRSV A	**Z33418**	**MAD/6/92**	**Spain**	**1992**	**García et al. (1994)**
HRSV A	**Z33419**	**MAD/8/92**	**Spain**	**1992**	**García et al. (1994)**
HRSV A	**Z33420**	**MAD/4/91**	**Spain**	**1991**	**García et al. (1994)**
HRSV A	Z33421	MON/1/87	Uruguay	1987	García et al. (1994)
HRSV A	Z33422	MON/1/89	Uruguay	1989	García et al. (1994)

Table 1. Continued

Virus	Accession No.	Strain	Country of isolation	Year of isolation	Reference
HRSV A	**Z33423**	**MON/1/92**	**Uruguay**	**1992**	**García et al. (1994)**
HRSV A	**Z33424**	**MON/2/88**	**Uruguay**	**1988**	**García et al. (1994)**
HRSV A	**Z33425**	**MON/3/88**	**Uruguay**	**1988**	**García et al. (1994)**
HRSV A	Z33426	MON/4/90	Uruguay	1990	García et al. (1994)
HRSV A	Z33427	MON/5/90	Uruguay	1990	García et al. (1994)
HRSV A	**Z33428**	**MON/5/91**	**Uruguay**	**1991**	**García et al. (1994)**
HRSV A	**Z33429**	**MON/7/91**	**Uruguay**	**1991**	**García et al. (1994)**
HRSV A	**Z33430**	**MON/8/92**	**Uruguay**	**1992**	**García et al. (1994)**
HRSV A	Z33431	MON/9/91	Uruguay	1991	García et al. (1994)
HRSV A	Z33432	MON/9/92	Uruguay	1992	García et al. (1994)
HRSV A	**Z33493**	**MAD/2/93**	**Spain**	**1993**	**García et al. (1994)**
HRSV A	Z33494	MON/1/90	Uruguay	1990	García et al. (1994)
HRSV B	AFO13254	B1	U.S.A.	1985	Karron et al. (1997)
HRSV B	AFO65250	CH10b	U.S.A.	1990–1994[b]	Peret et al. (1998)
HRSV B	AFO65251	CH93-9b	U.S.A.	1990–1995[b]	Peret et al. (1998)
HRSV B	AFO65252	CH93-18b	U.S.A.	1990–1995[b]	Peret et al. (1998)
HRSV B	AFO65253	CH93-53b	U.S.A.	1990–1995[b]	Peret et al. (1998)
HRSV B	M17213	18537	U.S.A.	1962	Johnson et al. (1987)
HRSV B	M55633	8/60	Sweden	1960	Sullender et al. (1991)
HRSV B	M73540	WV4843	U.S.A.	1980	Sullender et al. (1991)
HRSV B	M73541	WV10010	U.S.A.	1983	Sullender et al. (1991)
HRSV B	M73542	WV15291	U.S.A.	1985	Sullender et al. (1991)
HRSV B	M63543	NM1355	U.S.A.	1989	Sullender et al. (1991)
HRSV B	M73544	9320	U.S.A.	1977	Sullender et al. (1991)

[a] Sequences in bold type indicate those deleted from the selection analysis. NK, not known.
[b] Strain isolated between the years given.
[c] GenBank designations for the sequences of RSB89-642 and RSB89-1734 have been transposed.

Results

Analysis of Positive Selection at the Lineage Level

Parameter estimates resulting from the implementation of models of codon evolution are given in Tables 4a and 4b, and a summary of LRTs performed between nested models is given in Tables 5a and 5b. These results indicate that the FR model did not outperform the M0 model in any analysis, suggesting that a single ω ratio for the entire tree is an adequate explanation of the data. Consequently, no further analysis of evolution on individual branches was undertaken.

Analysis of Positive Selection at the Codon Site Level

BRSV

The average ω ratio estimated over all sites ranged from 0.431 to 0.458 among models, excluding the worst-fitting M1 model. The M2 model fitted the data significantly better than both M0 and M1 (Table 5a) but did not support the action of positive selection since the majority of sites (72.7%) appear to be moderately conserved ($\omega_2 = 0.252$), with the remainder falling into the neutral class. The M3 model could not reject either M2 or M0 and parameter estimates also suggested that the majority of sites ($p_0 + p_1 \approx 74.6\%$) are moderately conserved (ω

Table 2. Partial RSV strains from Cane et al. (1995) used in this study

Strain	County of isolation	Year of isolation
RSV630068	U.S.A.	1963
RSG1545	U.K.	1965
NC2825	U.K.	1966
RSSw23891	Sweden	1970
NC8870	U.K.	1971
RSG1503	U.K.	1972
RSSw36129	Sweden	1974
NC8453	U.K.	1975
RSG4988	U.K.	1976
RSG3881/77	U.K.	1977
RSG3423/77	U.K.	1977
B1/938	U.S.A.	1979
B2/958	U.S.A.	1979
B7/1393	U.S.A.	1979
B9/1407	U.S.A.	1979
B16/1687	U.S.A.	1979
H1414	U.K.	1980
RSA5	U.K.	1982
RSA9	U.K.	1983
RSA32	U.K.	1984
NC2782	U.K.	1985
RSB87-9	U.K.	1985
RSB90-8336	U.K.	1990
RS03148	Norway	1991
RSF50676	Finland	1992
RSB93-0458	U.K.	1993
RSC6079	U.K.	1993

Table 3a. Maximum-likelihood parameters for RSV G gene phylogenetic trees[a]

Virus	Taxa	Codons	*ln L*	T_S/T_V	α
BRSV	18	257	−2611.348	4.742	1.433
HRSV A	26	297	−3362.437	4.030	0.813
HRSV A partial	52	60	−1780.488	4.112	1.459
HRSV B	12	299	−2128.365	3.457	0.482

[a] *ln L;* log likelihood; T_S/T_V, observed transition/transversion ratio; α, shape parameter of a gamma distribution of among-site rate variation.

Table 3b. Maximum-likelihood parameters for RSV G gene phylogenetic trees after removal of strains isolated prior to 1980[a]

Virus	Taxa	Codons	*ln L*	T_S/T_V	α
BRSV′	10	257	−1902.139	5.851	1.066
HRSV A′	23	297	−3124.122	4.127	0.725
HRSV B′	9	299	−1777.939	3.013	0.400

[a] *ln L;* log likelihood; T_S/T_V, observed transition/transversion ratio; α, shape parameter of a gamma distribution of among-site rate variation.

= 0.260), with the remainder essentially neutral (ω_2 = 1.039). Finally, the M8 model did not perform significantly better than M7. Together, these models failed to detect positive selection in the G gene of BRSV.

HRSV A

The average ω ratio ranges from 0.468 to 0.635 when the complete G gene of HRSV A was analysed. Table 5a shows that the M0 model is significantly rejected by models that allow for positive selection (M2 and M3). In turn, M1 and M2 can be rejected in favor of M3, which incorporates slightly more sites into its positively selected class (p_2 = 8.8%) than M2, but with less positive selection pressure (ω_2 = 2.433). The majority of sites (p_0 = 58.6%) are weakly conserved (ω_0 = 0.532), while 32.7% of sites are strongly conserved (ω_1 = 0.001). The M7 model is rejected in favor of M8, which indicates that 7.1% of sites are undergoing positive selection with a ω_1 ratio of 2.571. The ω estimates for the positively selected class under M3 and M8 are remarkably similar. Posterior probabilities identified only site 226 to be under positive selection at the 90% level for the M2 model. In contrast, the M3 model identified six positively selected sites: positions 117 and 208 at the 90% level, positions 225 and 297 at the 95% level, and positions 215 and 226 at the 99% level (Fig. 1). The M8 model was able to identify only three of these sites above the 90% level—positions 215, 226, and 297.

Parameter estimates for the selection analysis of the partial HRSV A data set generally mirrored those of the complete G gene (Table 5a) and led to identical LRT results (Table 5b). Although the M3 model estimates a positively selected class of similar size (p_1 = 7.8%) and under similar selection pressure (ω_1 = 2.518) as before, the majority of sites are no longer weakly conserved but essentially neutral (ω_0 = 0.990), and the remainder (p_2 = 43.6%) are moderately conserved (ω_2 = 0.224). The partial alignment covers amino acid regions in which positively selected sites 117 and 297 were identified in the analysis of the complete gene. Analysis of this partial data set confirmed the presence of positive selection at site 297 but not at site 117.

HRSV B

In the case of HRSV B, the M2, M3, and M8 models which allow positive selection significantly outperform those that do not (Table 5a) and their average ω ratio ranges from 0.493 to 0.619 (Table 4a). M2 estimates that approximately 1.9% of sites are under strong positive selection (ω_2 = 8.627). Although M3 was not significantly favored over M2, it too suggested that a small proportion of sites (0.6%) is under strong selection (ω_2 = 10.726). M3 also identified another class of positively selected sites with a ω_0 ratio of 3.592, which accounted for 6.7% of sites, although the majority of sites were moderately conserved (ω_1 = 0.269). The M8 model has a significantly better likelihood than M7 and estimates that 6.0% of sites are positively selected (ω_1 = 4.671).

The M2 model identified only position 223 to be under strong positive selection, with a posterior probability of 0.991. M3 and M8 produced identical results in identifying six sites to be under positive selection: positions 223 and 257 at the 99% level, positions 152, 227 and 295 at the 95% level, and position 232 at the 90% level (Fig. 1). With respect to M3, site 223 was assigned to the highly positively selected class (ω_2 = 10.726), whereas the other sites all belong to the weaker positively selected class (ω_0 = 3.592).

Analysis of Passaging Effects

To assess the possible effects of passage history on selection analysis the complete BRSV, HRSV A and HRSV B data sets were reanalyzed with strains isolated before 1980 removed. In BRSV none of the models allowing positive selection could significantly reject those that do not (Table 5b), confirming that there is little evidence for adaptive evolution in the G protein of this virus. In the case of HRSV A, there was still evidence for positive selection after cultured strains were removed. However, dramatic changes were seen in the parameter estimates of M3, in which most sites are now strongly rather than weakly conserved and there is an increase in the ω estimate associated with the positively selected class (Table 4b). This was also true of M2 and M8. Furthermore, the M8 model now identified only two positively selected sites (226 and 215) above the 90% level, while M3 could identify only site 226.

The most dramatic results were obtained with HRSV B such that selection analysis was no longer able to corroborate the position of positively selected sites since

Table 4a. Likelihood and parameter estimates for the selection analysis of the RSV G gene[a]

Virus	Model	P	$ln\ L$	d_N/d_S	Parameter estimate(s)
BRSV	FR	33	−2602.934	NS	NS
	M0	1	−2619.596	0.431	$\omega = 0.431$
	M1	1	−2622.330	0.631	$p_0 = 0.369$, $p_1 = 0.631$
	M2	3	−2616.153	0.456	$p_0 = 0.000$, $p_1 = 0.273$, $p_2 = 0.727$, $\omega_2 = 0.252$
	M3	5[b]	−2616.150	0.458	$p_0 = 0.369$, $p_1 = 0.377$, $p_2 = 0.254$, $\omega_0 = 0.260$, $\omega_1 = 0.260$, $\omega_2 = 1.039$
	M7	2	−2616.414	0.450	$p = 0.864$, $q = 1.056$
	M8	4	−2616.152	0.458	$p = 70.543$, $q = 198.970$, $p_0 = 0.748$, $p_1 = 0.252$, $\omega_1 = 1.043$
HRSV A	FR	49	−3387.166	NS	NS
	M0	1	−3411.652	0.468	$\omega = 0.468$
	M1	1	−3384.150	0.550	$p_0 = 0.450$, $p_1 = 0.550$
	M2	3	−3381.312	0.635	$p_0 = 0.443$, $p_1 = 0.527$, $p_2 = 0.029$, $\omega_2 = 3.701$
	M3	5[b]	−3376.751	0.525	$p_0 = 0.586$, $p_1 = 0.327$, $p_2 = 0.088$, $\omega_0 = 0.532$, $\omega_1 = 0.001$, $\omega_2 = 2.433$
	M7	2	−3382.137	0.470	$p = 0.175$, $q = 0.197$
	M8	4	−3377.129	0.527	$p = 0.481$, $q = 0.814$, $p_0 = 0.929$, $p_1 = 0.071$, $\omega_1 = 2.571$
HRSV A partial	FR	86	−1779.732	NS	NS
	M0	1	−1829.290	0.684	$\omega = 0.684$
	M1	1	−1823.663	0.845	$p_0 = 0.155$, $p_1 = 0.845$
	M2	3	−1820.688	1.022	$p_0 = 0.148$, $p_1 = 0.777$, $p_2 = 0.075$, $\omega_2 = 3.275$
	M3	5[b]	−1813.730	0.775	$p_0 = 0.486$, $p_1 = 0.078$, $p_2 = 0.436$, $\omega_0 = 0.990$, $\omega_1 = 2.518$, $\omega_2 = 0.224$
	M7	2	−1817.502	0.639	$p = 0.603$, $q = 0.341$
	M8	4	−1814.333	0.762	$p = 0.960$, $q = 0.771$, $p_0 = 0.865$, $p_1 = 0.135$, $\omega_1 = 2.093$
HRSV B	FR	21	−2095.112	NS	NS
	M0	1	−2109.732	0.436	$\omega = 0.436$
	M1	1	−2099.143	0.460	$p_0 = 0.540$, $p_1 = 0.460$
	M2	3	−2094.540	0.619	$p_0 = 0.525$, $p_1 = 0.456$, $p_2 = 0.019$, $\omega_2 = 8.627$
	M3	5[b]	−2092.109	0.550	$p_0 = 0.067$, $p_1 = 0.928$, $p_2 = 0.006$, $\omega_0 = 3.592$, $\omega_1 = 0.269$, $\omega_2 = 10.726$
	M7	2	−2098.910	0.433	$p = 0.012$, $q = 0.014$
	M8	4	−2093.694	0.493	$p = 122.029$, $q = 272.736$, $p_0 = 0.940$, $p_1 = 0.060$, $\omega_1 = 4.671$

[a] P, number of parameters in model; d_N/d_S, average ω ratio. NS: For simplicity, the ω ratios in these cases are not shown (available from authors on request).
[b] The M3 model was run with $K = 3$ so the number of parameters in the model is $2K - 1 = 5$.

Table 4b. Likelihood and parameter estimates for the selection analysis of the RSV G gene after the removal of strains isolated prior to 1980[a]

Virus	Model	P	$ln\ L$	d_N/d_S	Parameter estimate(s)
BRSV′	FR	17	−1880.042	NS	NS
	M0	1	−1884.217	0.402	$\omega = 0.402$
	M1	1	−1886.780	0.550	$p_0 = 0.450$, $p_1 = 0.550$
	M2	3	−1883.859	0.410	$p_0 = 0.194$, $p_1 = 0.000$, $p_2 = 0.806$, $\omega_2 = 0.508$
	M3	5[b]	−1883.859	0.410	$p_0 = 0.396$, $p_1 = 0.409$, $p_2 = 0.195$, $\omega_0 = 0.509$, $\omega_1 = 0.508$, $\omega_2 = 0.001$
	M7	2	−1883.886	0.410	$p = 1.658$, $q = 2.377$
	M8	4	−1883.859	0.410	$p = 0.005$, $q = 2.697$, $p_0 = 0.194$, $p_1 = 0.806$, $\omega_1 = 0.508$
HRSV A′	FR	42	−3151.403	NS	NS
	M0	1	−3168.150	0.465	$\omega = 0.465$
	M1	1	−3142.083	0.537	$p_0 = 0.463$, $p_1 = 0.537$
	M2	3	−3138.293	0.635	$p_0 = 0.455$, $p_1 = 0.519$, $p_2 = 0.026$, $\omega_2 = 4.449$
	M3	5[b]	−3134.613	0.535	$p_0 = 0.052$, $p_1 = 0.634$, $p_2 = 0.314$, $\omega_0 = 3.083$, $\omega_1 = 0.133$, $\omega_2 = 0.926$
	M7	2	−3140.294	0.463	$p = 0.161$, $q = 0.187$
	M8	4	−3134.636	0.533	$p = 0.453$, $q = 0.779$, $p_0 = 0.932$, $p_1 = 0.068$, $\omega_1 = 2.803$
HRSV B′	FR	15	−1749.655	NS	NS
	M0	1	−1761.375	0.384	$\omega = 0.384$
	M1	1	−1752.860	0.370	$p_0 = 0.630$, $p_1 = 0.370$
	M2	3	−1751.089	0.523	$p_0 = 0.633$, $p_1 = 0.341$, $p_2 = 0.025$, $\omega_2 = 7.193$
	M3	5[b]	−1750.383	0.513	$p_0 = 0.044$, $p_1 = 0.900$, $p_2 = 0.056$, $\omega_0 = 0.224$, $\omega_1 = 0.224$, $\omega_2 = 5.358$
	M7	2	−1752.978	0.400	$p = 0.006$, $q = 0.010$
	M8	4	−1750.383	0.513	$p = 57.625$, $q = 198.616$, $p_0 = 0.944$, $p_1 = 0.056$, $\omega_1 = 5.363$

[a] P, number of parameters in respective model; d_N/d_S, average ω ratio. NS: For simplicity, the ω ratios in these cases are not shown (available from authors on request).
[b] The M3 model was run with $K = 3$ so the number of parameters in the model is $2K - 1 = 5$.

Table 5a. Likelihood ratio tests (LRT) between models of codon evolution for the RSV G gene[a]

		BRSV		HRSV A		HRSV A partial		HRSV B	
LRT	df[b]	χ^2	*p* value	χ^2	*p* value	χ^2	*p* value	χ^2	*p* value
M0 vs. FR	—[c]	33.323	0.403	48.971	0.434	99.117	0.140	29.239	0.083
M0 vs. M2	2	6.887	**0.032**	60.680	**0.000**	17.204	**0.000**	30.384	**0.000**
M1 vs. M2	2	12.354	**0.002**	5.677	0.059	5.949	0.051	9.206	**0.010**
M0 vs. M3	4	6.892	0.142	69.801	**0.000**	31.121	**0.000**	35.246	**0.000**
M1 vs. M3	4	12.359	**0.015**	14.798	**0.005**	19.866	**0.001**	14.068	**0.007**
M2 vs. M3	2	0.005	0.997	9.122	**0.010**	13.917	**0.001**	4.862	0.088
M7 vs. M8	2	0.524	0.769	10.015	**0.007**	6.337	**0.042**	10.432	**0.005**

[a] See Table 5b, footnote a.
[b] Degrees of freedom between the respective models.
[c] df is 32 for BRSV, 48 for HRSV A, 85 for HRSV A partial, and 20 for HRSV B.

Table 5b. Likelihood ratio tests (LRT) between models of codon evolution for the RSV G gene after removal of strains isolated prior to 1980[a]

		BRSV′		HRSV A′		HRSV B′	
LRT	df[b]	χ^2	*p* value	χ^2	*p* value	χ^2	*p* value
M0 vs. FR	—[c]	8.349	0.938	33.494	0.791	23.440	0.053
M0 vs. M2	2	0.715	0.699	59.714	**0.000**	20.571	**0.000**
M1 vs. M2	2	5.841	0.054	7.580	**0.023**	3.542	0.170
M0 vs. M3	4	0.715	0.949	67.073	**0.000**	21.984	**0.000**
M1 vs. M3	4	5.841	0.211	14.940	**0.005**	4.955	0.292
M2 vs. M3	2	0.000	1.000	7.360	**0.025**	1.413	0.493
M7 vs. M8	2	0.054	0.973	11.317	**0.003**	5.190	0.075

[a] Likelihood ratio tests are performed by taking twice the difference in log likelihood between two models and comparing the value obtained with a χ^2 distribution (degrees of freedom equal to the difference in the number of parameters between the models). *p* values in boldface indicate comparisons where the null hypothesis can be rejected in favor of the alternative hypothesis (i.e., the model on the left is rejected in favor of the one on the right).
[b] Degrees of freedom between the respective models.
[c] df is 16 for BRSV, 41 for HRSV A, and 14 for HRSV B.

models accounting for positive selection could no longer unanimously reject those that do not (Table 5b). As with HRSV A, the removal of heavily passaged strains also led to major changes in parameter estimates for the M3 model and to variations in the ω ratios associated with the positively selected classes (Table 4b).

Discussion

Although our analysis provided no evidence for positive selection on specific lineages of either BRSV or HRSV, we did find a small number of codon sites in the G protein under positive selection in the human form of the virus. Interestingly, only one selected site was found in common between HRSV A and HRSV B, and this was at position 226 (homologous to position 227 in HRSV B) (Fig. 1), suggesting that these viral groups induce slightly different immune responses. The relevance of site 226 in either group is unknown but clearly merits further study. It is also noteworthy that all the positively selected sites we identified were located in the ectodomain of the G protein, while none were found in the receptor binding domain. It is possible that the latter region may lack variability because of structural–functional constraints or because it is not immunodominant (Walsh et al. 1998). Variability in the N terminus of the ectodomain has been attributed to compensatory changes to mutations in the C terminus (Cane and Pringle 1995). Hence, positively selected sites 117 in HRSV A and 152 in HRSV B, which fall into the N terminus, may represent compensatory mutations. This is supported by the observation that site 117 was not found to be positively selected in the analysis of the partial HRSV A data set. Finally, it is interesting that the positively selected sites in HRSV B fell into two distinct classes, one with an associated d_N/d_S of 3.592 and the other where d_N/d_S was 10.726, revealing a far stronger selection pressure, although only site 223 belonged to the latter class at the 99% level (the possible functional relevance of this site is discussed below).

That positive selection could not be found in BRSV suggests that this virus induces a weaker host immune response than HRSV. However, it is possible that if the period of adaptive evolution is brief or the selection pressures very localized, then even codon-based methods like those employed here will fail to identify positive selection. Moreover, vaccines had been administered to some of the cattle from which viruses were isolated and this may have affected our analysis, although these strains do not appear to contain any distinct amino acids compared to those from non-vaccinated calves (Valarcher et al., personal communication).

To characterize further why natural selection might operate at the selected sites identified in HRSV, we determined if they were associated with parts of the G protein known to be functionally important. One obviously crucial group of amino acid residues are those that

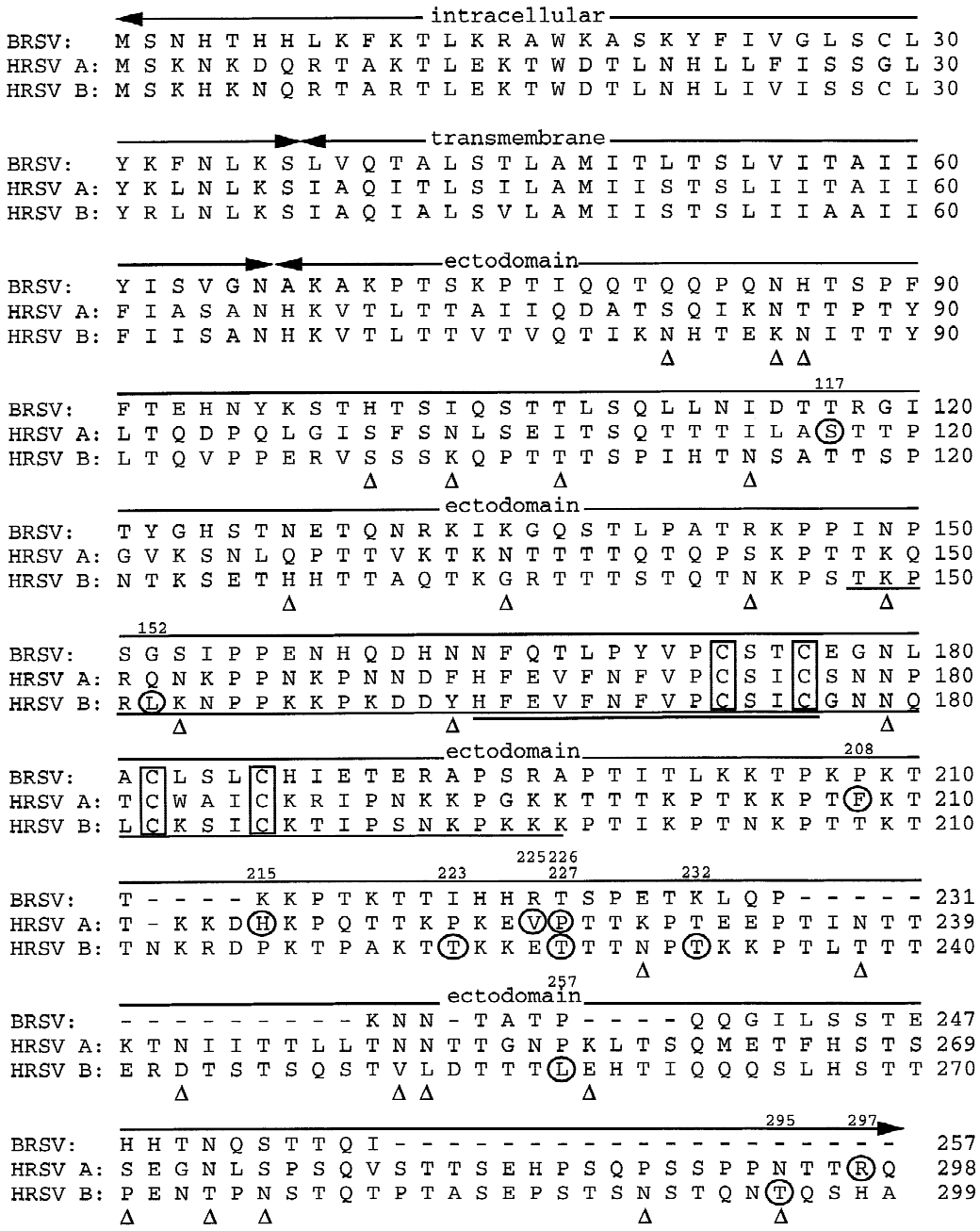

Fig. 1. Alignment of G protein amino acid sequences from strains 391-2 (BRSV), Long (HRSV A), and B1 (HRSV B). Notation above the sequences divides the protein into the intracellular domain (1–37), the transmembrane domain (38–66), and the ectodomain (67–298). The *underlined* region of the ectodomain delineates a conserved region (amino acids 148–197) and the *double-underlined* region identifies the hydrophobic receptor binding domain (amino acids 164–176). The conserved Cys residues in the ectodomain are boxed. The Δ symbol indicates both conserved and nonconserved N-glycosylation sites regardless of host or group. *Circled* amino acids represent those identified to be under positive selection above the 90% level in our analysis. It should be noted that although a positively selected site was identified at position 295 in HRSV B, the N-glycosylation marker at this position refers to a conserved N-glycosylation site at position 294 in HRSV A.

encode epitopes recognized by either the cellular or the humoral arms of the immune system. With respect to T-cell epitopes, although the G protein does not appear to induce a significant MHC class I restricted cytotoxic T-lymphocyte (CTL) response in humans (Cherrie et al. 1992), a CD4 T-cell epitope has been identified between amino acids 184–198 of HRSV A (Tebbey et al. 1998). This region was also found to be important by Sparer et al. (1998), who showed that amino acids 193–203 were responsible for inducing enhanced illness and lung eosinophilia and further predicted that two I-E^d epitopes lie between amino acids 185–193 and 189–197. Our selection analysis did not identify positively selected sites in any of these potential epitopes. It is therefore possible

that HRSV does not alter epitopes of the effector T cells currently identified in the G protein as a mechanism to evade immune responses.

We found a stronger association between positively selected sites and the location of epitopes for antibodies. Several epitopes have been detected in the G protein of HRSV A by analyzing the reaction of human convalescent sera with peptides based on the amino acid sequence of the ectodomain (Norrby et al. 1987; Cane 1997). Only positively selected sites 225 and 226 were located in one of these epitopes which was synthesized as a peptide fragment spanning amino acids 223–234 based on sequence data from the RSB89-6190 strain. Further studies have mapped antibody epitopes by sequencing escape mutants of the Long strain of HRSV A that were selected with anti-G monoclonal antibodies (MAbs) generated in mice (García-Barreno et al. 1990; Rueda et al. 1994; Walsh et al. 1998). One of these, epitope 63G, has an association with the position of positively selected sites determined here. Specifically, Cane and Pringle (1995) showed that HRSV A strain RSB89-1734 (mistakenly entered as RSB89-642 in GenBank) has lost epitope 63G integrity, whereas strain A2 maintained it. With reference to the Long strain, RSB89-1734 has amino acid substitutions at sites 208 (Phe to Pro), 214 (Asp to Gly), and 215 (His to Pro) and A2 has substitutions at sites 208 (Phe to Leu) and 215 (His to Pro). Both 215 and 208 were deemed to be under positive selection in our analysis. In this case it appears that site 208 plays the key role in epitope 63G determination such that a Phe-to-Pro substitution results in loss of epitope recognition, whereas a Phe-to-Leu substitution does not. Natural strains also exhibit amino acid substitutions of His to Leu, Ile, or Phe at site 215 in comparison to the Long strain and it is possible that these substitutions may likewise abolish antibody recognition of epitope 63G, although further epitope mapping is needed to confirm this.

Rueda et al. (1995) demonstrated that substitutions at the last three amino acids (296–298) in HRSV A influence the integrity of multiple overlapping strain-specific epitopes. Our analysis revealed that site 297 was under positive selection adding support to this finding. A similar overlapping epitope has been hypothesized for HRSV B strains, although they have not been identified to date (Martínez et al. 1999). Our analysis pinpointed selection on site 295, which lies just prior to the termination codon in some of the more recently circulating strains of HRSV B. We therefore suggest that site 295 may lie within an epitope and that substitutions here would likewise abolish the recognition of the G protein by strain-specific overlapping antibodies.

Modulation of glycosylation may also contribute to immune evasion by either abolishing G protein recognition by carbohydrate specific antibodies or by masking antigenic sites (Melero et al. 1997; Palomo 2000). None of the positively selected sites we identified corresponded to Asn positions with the potential for N-glycosylation. However, there was some support for positive selection at residues capable of being O-glycosylated. In HRSV A this was true of site 117, which is commonly Ser, and we predicted that this amino acid was O-glycosylated with a high potential (between 0.8857 and 0.9994; threshold of 0.5623 to 0.6575 depending on the strain analyzed). HRSV A strains WV23836, RSB89-642, RSB89-5857, RSB89-6190, RSB89-6256, RSB89-6614, MAD/6/93, and MAD/3/92 appear to have lost a Ser residue at this site through replacement with either Leu or Pro. Intriguingly, our phylogenetic analysis shows that this replacement has occurred at least four times, most notably in an entire clade of sequences comprising strains WV23836, RSB89-5857, RSB89-6256, RSB89-6614, and MAD/3/92, all of which were isolated in the late 1980s and early 1990s. It is therefore possible that these strains have abolished O-glycosylation at site 117 and thus lost immune recognition of a carbohydrate epitope. The positively selected site at position 152 in HRSV B also showed replacement of the amino acid Ser by Leu or Pro. Strikingly, only the three strains isolated prior to 1980 (18537, 8/60, and 9320) contained a Ser residue at this position, which we predict to have a fairly high potential of being O-glycosylated (0.7749; threshold = 0.5038), whereas all the more recently isolated strains contained a Leu or a Pro.

Evidence also exists for both the loss and the gain of O-glycosylation at positively selected sites containing Thr residues in HRSV B. This is particularly true of position 223, which was found to be subject to extremely strong positive selection ($\omega = 10.726$), although we could confidently predict only the Thr residue of strain WV4843 to be O-glycosylated at this site (potential = 0.7144, threshold = 0.5048). Positively selected site 232 in HRSV B showed loss of O-glycosylation in strains 18537 and WV10010 since the Thr amino acid at this site was replaced by Ala and Gln, respectively. The remaining HRSV B strains contained a Thr residue at this site and these were predicted to be O-glycosylated with a high potential (between 0.8848 and 0.8851; threshold of 0.4927 to 0.5614). Overall, these observations lend credence to the idea that HRSV modulates the extent of O-glycosylation of its G protein to evade the immune response.

Finally, previous work has suggested that A→G "hypermutation" events in the G gene are partly responsible for generating immune escape mutants of HRSV (García-Barreno et al. 1990; Martínez et al. 1997; Rueda et al. 1994). Only one of these escape mutants (19G/9A/2) (Martínez et al. 1997) corresponded to a positively selected site in our analysis, and this was at position 226. Here, an A→G substitution resulted in a Leu to Pro

change, which is also seen in natural isolates of HRSV A. Our analysis therefore concurs that limited A→G hypermutation events are partly responsible for creating selectively useful variation. Frameshifts, most likely caused by slippage of the viral polymerase during genome replication, have also been documented in RSV (García-Barreno et al. 1990; Rueda et al. 1995) and likewise may produce potentially useful genetic diversity. In our analysis, strain WV10010 of HRSV B contains a frameshift between amino acids 225 and amino acid 232, of which sites 227 and 232 were found to be selected, and this frameshift also abolishes a potential site of N-glycosylation.

Passage History and Positive Selection

One notable observation from our study was that the inclusion of viral strains which might have accumulated mutations following repeated passaging (hypothesised from their early isolation) can affect the analysis of selection pressures. However, these effects are clearly complex. If passaged strains do contain mutations which adapt them to cell culture, then these mutations might be different to those seen in natural viral isolates. Significantly, this was not the case in HRSV A, suggesting that the putative selected sites we identified in this group are *bona fide.* Furthermore, the stability of G gene sequences of HRSV A during tissue culture has been reported previously (Cane et al. 1994). The possible exceptions were site 117, as discussed earlier, and site 208, which contained a unique Phe amino acid in the Long strain (isolated in 1956). However, site 117 may be used to modulate O-glycosylation and site 208 does appear to play a role in epitope 63G determination, supporting our contention that the selection we document in HRSV A reflects *in vivo* evolution. With respect to HRSV B, models that accounted for positive selection could no longer unanimously reject those that did not when ancient strains were excluded. Although this could be an artifact of comparing a smaller number of sequences, thus reducing the analytical power, it is noteworthy that the ancient strains often possessed amino acid replacements at positively selected sites which were different from the variation seen in the rest of the data set. It is therefore difficult to determine if the six positively selected sites identified in HRSV B are naturally occurring positively selected sites. We conclude that studies of selection pressures in viral genomes should always consider the possible bias introduced by adaptation to cell culture.

Acknowledgments. We thank Dr. Ziheng Yang for his patience, two referees for useful suggestions, and Dr. Patricia A. Cane for providing partial HRSV A sequence data. This work was supported by the BBSRC and The Royal Society.

References

Baker JC, Ames TR, Werdin RE (1986) Seroepizootiologic study of bovine respiratory syncytial virus in a beef herd. Am J Vet Res 47:246–253

Cane PA (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 PA, Pringle CR (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 PA, Matthews DA, Pringle CR (1991) Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. J Gen Virol 72:2091–2096

Cane PA, Matthews DA, Pringle CR (1992) Analysis of relatedness of subgroup A respiratory syncytial viruses isolated worldwide. Virus Res 25:15–22

Cane PA, Matthews DA, Pringle CR (1994) Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. J Clin Micro 32:1–4

Cherrie AH, Anderson K, Wertz GW, Openshaw PJ (1992) Human cytotoxic T cells stimulated by antigen on dendritic cells recognize the N, SH, F, M, 22K, and 1b proteins of respiratory syncytial virus. J Virol 66:2102–2110

Collins PL, McIntosh FA (1995) Pneumovirinae. Raven Press, New York

Collins PL, McIntosh K, Chanock RM (1996) Respiratory syncytial virus. Lipponcott–Raven, Philadelphia

Elvander M, Vilcek S, Baule C, Uttenthal Å, Ballagi-Pordány A, Belák S (1998) Genetic and antigenic analysis of the G attachment protein of bovine respiratory syncytial virus strains. J Gen Virol 79:2939–2946

Furze JM, Roberts SR, Wertz GW, Taylor G (1997) Antigenically distinct G glycoproteins of BRSV strains share a high degree of genetic homogeneity. Virology 231:48–58

García O, Martin M, Dopazo J, Arbiza J, Frabasile S, Russi J, Hortal M, Perez-Brena P, Martinez I, García-Barreno B, Melero JA. (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

García-Barreno B, Portela A, Delgado T, Lopez J, Melero J (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

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S (1998) NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconjugate J 15:115–130

Johnson PR, Spriggs MK, Olmsted RA, Collins PL (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

Karron RA, Buonagurio DA, Georgiu AF, Whitehead SS, Adamus JE, Clements-Mann ML, Harris DO, Randolph VB, Udem SA, Murphy BR, Sidhu MS (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

Langedijk JP, Schaaper WM, Meloen RH, van Oirschot JT (1996) Proposed three-dimensional model for the attachment protein G of respiratory syncytial virus. J Gen Virol 77:1249–1257

Larsen LE, Uttenthal A, Arctander P, Tjornehoj K, Viuff B, Rontved C, Ronsholt L, Alexandersen S, Blixenkrone-Moller M (1998) Serological and genetic characterisation of bovine respiratory syncytial virus (BRSV) indicates that Danish isolates belong to the interme-

diate subgroup: no evidence of a selective effect on the variability of G protein nucleotide sequence by prior cell culture adaptation and passages in cell culture or calves. Vet Microbiol 62:265–279

Mallipeddi SK, Samal SK (1993) Sequence variability of the glycoprotein gene of bovine respiratory syncytial virus. J Gen Virol 74:2001–2004

Martínez I, Dopazo J, Melero JA (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

Martínez I, Valdes O, Delfraro A, Arbiza J, Russi J, Melero JA (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

Melero JA, García-Barreno B, Martínez I, Pringle CR, Cane PA (1997) Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. J Gen Virol 78:2411–2418

Mufson MA, Orvell C, Rafnar B, Norrby E (1985) Two distinct subtypes of human respiratory syncytial virus. J Gen Virol 66:2111–2124

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

Norrby E, Mufson MA, Alexander H, Houghten RA, Lerner RA (1987) Site-directed serology with synthetic peptides representing the large glycoprotein G of respiratory syncytial virus. Proc Natl Acad Sci USA 84:6572–6576

Palomo C (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

Peret TCT, Hall CB, Schnabel KC, Golub JA, Anderson LJ (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

Prozzi D, Walravens K, Langedijk JP, Daus F, Kramps JA, Letesson JJ (1997) Antigenic and molecular analyses of the variability of bovine respiratory syncytial virus G glycoprotein. J Gen Virol 78: 359–366

Rueda P, García-Barreno B, Melero JA (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

Rueda P, Palomo C, García-Barreno B, Melero JA (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

Schrijver RS, Langedijk JPM, Middel WGJ, Kramps JA, Rijsewijk FAM, van Oirschot JT (1998) A bovine respiratory syncytial virus strain with mutations in subgroup-specific antigenic domains of the G protein induces partial heterologous protection in cattle. Vet Microbiol 63:159–175

Sparer TE, Matthews S, Hussell T, Rae AJ, García-Barreno B, Melero JA, Openshaw PJ (1998) Eliminating a region of respiratory syncytial virus attachment protein allows induction of protective immunity without vaccine-enhanced lung eosinophilia. J Exp Med 187:1921–1926

Stine LC, Hoppe DK, Clayton CL, Kelling CL (1997) Sequence conservation in the attachment glycoprotein and antigenic diversity among bovine respiratory syncytial virus isolates. Vet Microbiol 54:201–221

Sullender WM (2000) Respiratory syncytial virus genetic and antigenic diversity. Clin Micro Rev 13:1–15

Sullender WM, Mufson MA, Anderson LJ, Wertz GW (1991) Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. J Virol 65:5425–5434

Sullender WM, Mufson MA, Prince GA, Anderson LJ, Wertz GW (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

Swofford DL (1998) Phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA

Tebbey PW, Hagen M, Hancock GE (1998) Atypical pulmonary eosinophilia is mediated by a specific amino acid sequence of the attachment (G) protein of respiratory syncytial virus. J Exp Med 188:1967–1972

Thompson JD, Higgins DG, Gibson TJ (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

Tolley KP, Marriott AC, Simpson A, Plows DJ, Matthews DA, Longhurst SJ, Evans JE, Johnson JL, Cane PA, Randolph VB, Easton AJ, Pringle CR (1996) Identification of mutations contributing to the reduced virulence of a modified strain of respiratory syncytial virus. Vaccine 14:1637–1646

Van der Poel W, Kramps J, Middel W, Van OJ, Brand A (1993) Dynamics of bovine respiratory syncytial virus infections: a longitudinal epidemiological study in dairy herds. Arch Virol 133:309–321

Walsh E, Falsey A, Sullender W (1998) Monoclonal antibody neutralization escape mutants of respiratory syncytial virus with unique alterations in the attachment (G) protein. J Gen Virol 79:479–487

Wertz GW, Collins PL, Huang Y, Gruber C, Levine S, Ball LA (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

Wertz GW, Krieger M, Ball LA (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

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15: 568–573

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449