Journal of General Virology (1998), **79**, 2221–2229. Printed in Great Britain

Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community

Teresa C. T. Peret,[1] Caroline B. Hall,[2] Kenneth C. Schnabel,[2] Joshua A. Golub[1] and Larry J. Anderson[1]

[1] Respiratory and Enteric Viruses Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA

[2] Departments of Pediatrics and Medicine, Infectious Diseases, University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Rochester, NY 14642, USA

Human respiratory syncytial virus (HRSV) is classified into two major groups, A and B, each of which contains multiple variants. To characterize the molecular epidemiology of HRSV strains over time, sequencing studies of a variable region of the attachment protein gene from a single community in the United States during 5 successive years were performed. Phylogenetic analysis revealed distinct clades (genotypes) that were further classified in subtypes based on $\geqslant$ 96% nucleotide similarity. Five genotypes and 22 subtypes among 123 group A HRSV isolates, and four distinct genotypes and six subtypes among 81 group B HRSV isolates were identified. One to two genotypes or subtypes accounted for $\geqslant$ 50% of isolates from a given year. A shift in the predominant genotype or subtype occurred each year such that no genotype or subtype predominated for more than 1 of the 5 study years. The consistency in the displacement of the predominant strain suggests that a shift, even within the same group, is advantageous to the virus. It was hypothesized that the 'novel' strain is better able to evade previously induced immunity in the population and consequently either circulates more efficiently or is more pathogenic. The yearly shift in HRSV strains may contribute to the ability of HRSV to consistently cause yearly outbreaks of HRSV disease. These results also suggest that isolates may need to be characterized as to both group and genotype to fully understand protective immunity after natural infection and efficacy studies of candidate vaccines.

Introduction

Human respiratory syncytial virus (HRSV) has long been recognized as a major cause of serious, life-threatening lower respiratory tract illness in infants and young children (Kim *et al.*, 1973; Institute of Medicine, 1985; Anderson *et al.*, 1990; Hall *et al.*, 1990), and has also been shown to be an important cause of lower respiratory tract illness in adults (Dowell *et al.*, 1996). On the basis of monoclonal antibody studies, HRSV has been classified into two major groups, A and B (Mufson *et al.*, 1985; Anderson *et al.*, 1985), and within these groups, antigenic variants or types of HRSV have also been identified (Hendry *et al.*, 1986; Åkerlind *et al.*, 1988; Mufson *et al.*, 1988; Anderson *et al.*, 1991; Storch *et al.*, 1991). Sequence studies of the nucleoprotein (N), phosphoprotein (P), small hydrophobic (SH) protein and attachment (G) protein genes have confirmed the division of HRSV into two groups and also identified numerous variants or lineages within each group (Johnson *et al.*, 1987*b*; Cane & Pringle, 1991, 1992, 1995*a*; Cristina *et al.*, 1991; Storch *et al.*, 1991; Sullender *et al.*, 1991; Cane *et al.*, 1992, 1994; Garcia *et al.*, 1994; Sanz *et al.*, 1994).

The HRSV G protein gene has been shown to be the most divergent, both between and within the major HRSV groups, with 67% similarity at the nucleotide and 53% similarity at the deduced amino acid levels between group A and B strains (Johnson *et al.*, 1987*b*). The G protein gene encodes a highly glycosylated protein that ranges from 289 to 298 amino acids (Wertz *et al.*, 1985; Johnson *et al.*, 1987*b*; Sullender *et al.*, 1991; Cane & Pringle, 1995*a*). Variability of the G protein gene primarily occurs in the extracellular domain, which contains two hypervariable regions separated by a 13 amino acid motif (residues 164–176), conserved among all RSV strains (Johnson *et al.*, 1987*b*; Cane & Pringle, 1991; Sullender *et al.*, 1991) and

Author for correspondence: Larry Anderson.
Fax +1 404 639 1307. e-mail lja2@cdc.gov

GenBank accession numbers for the reported sequence data: AF065250 and AF065251–AF065258.

Table 1. Seasonal distribution of HRSV isolates from Rochester, New York, used in the sequencing studies

Season	Year	Month	Patients tested for RSV (Rochester)*	Patients RSV-positive (Rochester)*	RSV isolate sequencing studies (CDC) (%)†
1990/91	1990	October	1	0	0 (0)
		November	29	14	8 (57)
		December	61	30	16 (53)
	1991	January	83	40	18 (45)
		February	44	15	7 (47)
		March	1	0	0 (0)
Total 1 ...			**219**	**99**	**49 (49)**
1991/92	1991	November	4	4	0 (0)
		December	22	22	0 (0)
	1992	January	112	71	24 (34)
		February	64	46	20 (43)
		March	45	25	9 (35)
		April	6	4	2 (50)
Total 2 ...			**253**	**172**	**55 (32)**
1992/93	1992	November	2	1	0 (0)
		December	27	10	5 (50)
	1993	January	50	23	9 (39)
		February	47	23	7 (30)
		March	42	15	6 (40)
		April	10	4	2 (50)
		May	6	1	0 (0)
Total 3 ...			**184**	**77**	**29 (38)**
1993/94	1993	November	4	1	1 (100)
		December	23	18	7 (39)
	1994	January	50	24	13 (54)
		February	49	23	13 (57)
		March	34	12	3 (25)
		April	12	0	0 (0)
Total 4 ...			**172**	**78**	**37 (47)**
1994/95	1994	December	14	14	3 (14)
	1995	January	29	28	8 (29)
		February	46	42	12 (29)
		March	29	27	11 (37)
		April	2	2	0 (0)
Total 5 ...			**120**	**113**	**34 (30)**
Total 1–5 ...			**948**	**539**	**204 (38)**

* Cell culture and rapid antigen detection.
† Percentage = RSV isolates sequenced at CDC/total RSV-positive in Rochester.

considered to be the putative site for virus attachment to the cell receptor (Johnson *et al.*, 1987*b*).

The genetic variability of HRSV has been best characterized for group A strains (Cristina *et al.*, 1991; Cane & Pringle, 1992, 1995*a*; Garcia *et al.*, 1994; Cane *et al.*, 1994; Sanz *et al.*, 1994). Studies have shown up to 20% diversity at the deduced amino acid level among group A strains (Cane *et al.*, 1991) and demonstrated that genetically similar strains seem to cluster by time rather than by geographical location (Garcia *et al.*, 1994; Sanz *et al.*, 1994; Cane & Pringle, 1995*a*). These studies suggest that immunological pressure may direct HRSV evolution (Garcia *et al.*, 1994; Cane & Pringle, 1995*a*). Although these studies have provided important insights into the epidemiology of HRSV strains, the circulation patterns of distinct strains in a single community and simultaneously in different communities have still not been fully characterized.

In this study, we evaluated the genetic diversity of both group A and B HRSV strains by sequencing a variable region

of the G protein gene of isolates collected during five consecutive epidemic seasons (1990/91–1994/95) in a single community (Rochester, New York, USA). These data provide a better understanding of the patterns of HRSV strain circulation and the possible importance of strain differences to the consistency of yearly community outbreaks of HRSV disease.

Methods

■ **Virus isolates.** The isolates were obtained from children under 3 years of age admitted with lower respiratory tract disease to the University of Rochester Medical Center Annual Inpatient Surveillance Program in Rochester, New York, USA. Isolates were chosen to represent the entire time span of each HRSV season. A total of 204 HRSV isolates, representing 38% of all HRSV-positive specimens identified during the study period, were sequenced (Table 1). The viruses were isolated in HEp-2 cells at the University of Rochester and propagated once in the same cell line at The Centers for Disease Control and Prevention for sequencing studies.

■ **Antigenic characterization.** Monoclonal antibody reactivity patterns based on a previously described capture enzyme immunoassay (Anderson *et al.*, 1991) were used to assign each isolate to group A or B.

■ **Sequencing data sources.** Published sequences of the SH, G and F (fusion) protein genes obtained from GenBank and EMBL as well as G gene sequences obtained in our laboratory were used in our studies.

■ **Oligonucleotide primers.** Primers complementary to SH, G, and F gene nucleotide sequences were used to amplify (SHA-SHB/F1; Table 2) and sequence the full-length G gene in selected isolates (SHA, GA1–GA4, F1 for A group; SHB, GB1–GB4, F1 for B group; Table 2). Primer sets GPA/F1-GSA/F1 and GPB/F1-GSB/F1 (Table 2) were used, respectively, to amplify and sequence the limited region of the G gene for group A and B strain studies.

■ **RT–PCR and nucleotide sequencing.** RNA was extracted using a guanidium isothiocyanate–phenol method (RNAzol B, Tel-Test) and viral RNA was amplified in an RT–PCR procedure. Up to 1 μg RNA was annealed with 20 mM F1 primer (Table 2) at 50 °C for 25 min and reverse-transcribed with avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) at 42 °C for 1 h. After cDNA synthesis, the PCR assay was carried out in a mix containing up to 5 μl template, 10 μl buffer (150 mM Tris–HCl, 500 mM KCl, 15 mM $MgCl_2$, DMSO 3%, glycerol 8%), 20 mM dNTPs, 40 mM of the forward and reverse primers, and 5 U *Taq* polymerase brought up to 100 μl with distilled water. Amplification conditions consisted of 2 min at 94 °C, followed by 30 cycles of PCR (94 °C for 1 min; 52 °C for 1 min; 72 °C for 2 min) and 7 min at 72 °C. The PCR products were purified with either a QIAquick gel extraction kit or a QIAquick PCR purification kit (Qiagen). Both strands were sequenced on ABI 373 or 377 Sequencer models, using a fluorescent dye terminator kit (Applied Biosystems). The nucleotide sequences were edited with SeqEd in the Wisconsin Package version 8.1 (GCG, Madison, Wisconsin, USA) for Unix, or Sequencher version 3.0 (Gene Codes Corporation) for Power Macintosh.

■ **Nucleotide sequencing approach.** Multiple alignments of the G gene nucleotide and deduced amino acid sequences were plotted to estimate the average similarities for A and B group sequences, using the 'Multiple Comparison Programs' in the GCG package. To determine if a limited region of the G gene could serve as a reliable proxy for the entire G gene variability, phylogenetic analysis of sequence alignments encompassing (i) the first variable region (nt 280–459, A and B groups), (ii) the second variable region (nt 649–918, A group; nt 652–921, B group) and (iii) both variable regions was performed and compared with an analysis conducted with full-length G gene sequences of group A and B viruses (data not shown). Nucleotide numbering was based on the prototype strains A2 and 18537 for A and B groups, respectively (Johnson *et al.*, 1987*b*). The second variable region, consisting of 270 nucleotides, was selected for the subsequent sequencing studies.

■ **Phylogenetic analysis.** Nucleotide sequences obtained from the second variable region of the G gene from the 204 HRSV isolates were compiled in multiple alignments, using the GCG program PILEUP. Unique representative sequences for both A and B group viruses were included in the phylogenetic analysis for computational efficiency. In this way, 123 A group and 81 B group sequences were abridged, respectively,

Table 2. List of primers used in PCR and nucleotide sequencing

Primer	Polarity	Position (nt)	Sequence (5′–3′)
SHA	+	264–284	TCGAGTCAACACATAGCATTC
SHB	+	278–298	CATAGTATTCTACCATTATGC
F1	−	3–22	CAACTCCATTGTTATTTGCC
GA1	+	258–281	CCAGATCAAGAACACAACC
GA2	−	258–281	GGTTGTGTTCTTGATCTGG
GA3	+	526–544	GTACCCTGCAGCATATGCA
GA4	−	526–544	TGCATATGCTGCAGGGTAC
GPA	+	511–530	GAAGTGTTCAACTTTGTACC
GSA	+	606–625	AACCACCACCAAGCCCACAA
GB1	+	226–245	CTAACAACGGTCACAGTTCA
GB2	−	226–245	TGAACTGTGACCGTTGTTAG
GB3	+	524–544	TTGTTCCCTGTAGTATATGTG
GB4	−	524–544	CACATATACTACAGGGAACAA
GPB	+	494–515	AAGATGATTACCATTTTGAAGT
GSB	+	603–623	AAAACCAACCATCAAACCCAC

in 46 and 21 sequences. Phylogenetic trees were constructed for group A and B alignments, using a maximum likelihood method (DNAML, PHYLIP version 3.5c package; Felsenstein, 1995), with transition-to-transversion ratio calculated with PUZZLE version 3.1 (Strimmer & von Haeseler, 1996). Confidence estimates were based on bootstrap re-sampling carried out with 100 replicates (Felsenstein, 1985). Trees were plotted using TREEVIEW (Page, 1996).

Results

Variation in HRSV: A and B genotypes and subtypes

Trees derived from the maximum likelihood reconstruction based on a limited region of the group A and B HRSV G genes are shown in Fig. 1. To facilitate analysis of the data, sequences were assigned to 'genotypes' defined by the major clades observed in the tree (Fig. 1). Phylogenetic analysis of the 46 group A unique sequences identified five clades, assigned as genotypes GA1 to GA5 (Fig. 1*a*). Bootstrap tests of the internal nodes at the interior branches in the tree that characterized the genotypes displayed values of 78–100%.

A phylogenetic analysis for 21 representative B group sequences showed four clades (GB1 to GB4; Fig. 1*b*), supported by 88–100% of bootstrap replicates. One isolate (140) which was weakly supported at bootstrapping analysis (60%), did not cluster in any of the four B group genotypes and, therefore, was not considered in the analyses of these four genotypes.

After reviewing the phylogenetic relationships among strains within the genotypes, we further divided closely related strains, based on nucleotide similarity $\geqslant$ 96%, into groups that we arbitrarily designated as 'subtypes'. We identified 22 distinct subtypes among the five group A genotypes composed of 123 HRSV sequences. Seventy-three sequences were placed in GA5; this genotype comprised half of the group A subtypes (A.5a to A.5k) and presented the greatest intra-genotype nucleotide variability (13%). GA4 exhibited the least intra-genotype nucleotide variability (1%), with only one subtype (A.4a), composed of five HRSV sequences.

Six subtypes were identified among the four group B genotypes, consisting of 81 HRSV sequences. GB1, represented by one subtype (B.1a), accounted for half of the B group HRSV sequences, while GB2 exhibited one subtype with eight RSV sequences. The intra-genotype nucleotide variability for the B genotypes was quite low, ranging from 1% (GB1 and GB2) to 3% (GB3 and GB4).

Variation in HRSV: amino acid sequences

The amino acid sequences deduced from the 270 nucleotide fragment of HRSV A and B G gene sequences were used to generate a consensus sequence in each genotype (Fig. 2). The predicted amino acid sequences in A genotypes consisted of 86 (GA1–GA4) or 87 residues (GA5). B group genotype sequences showed 83 (GB2–GB4) or 87 (GB1) residues. The difference in the length of the deduced amino acid sequences resulted from first-nucleotide-position mutations in the first stop codon in the carboxyl terminus of the G protein gene.

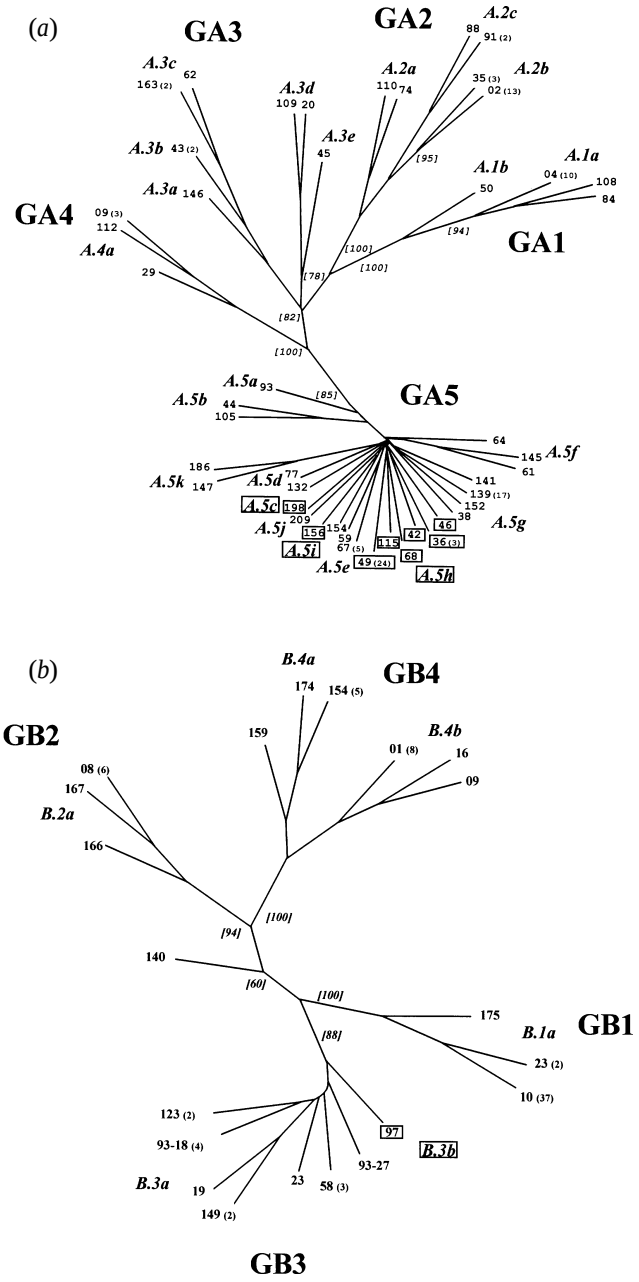

Fig. 1. Maximum likelihood unrooted trees (transition-to-transversion ratio of 5) showing the relationships for the HRSV A (*a*) and B (*b*) group strains. The italicized numbers in square brackets at the main interior branches show the bootstrap replicates supporting the analysis (n = 100). Numbers in parentheses at the terminal nodes correspond to number of identical sequences. Italic bold numbers (*a*, ***1a–5k***; *b*, ***1a–4b***) indicate subtypes. Squared boxes are used to associate virus sequences with the corresponding subtype when necessary. Genotypes are indicated by capital letters (GA1–GA5 and GB1–GB4). Branch lengths are not drawn to scale.

Group A strains showed the greatest variation between genotypes in the predicted amino acid sequences, with an average difference of 20% and a range from 10% between GA2 and GA3 to 28% between GA1 and GA4. The average

```
          1                                                                                                   5 0
Consensus K K D L K P Q T T K   P K E - L T T K P T   E K P T I N T T K T   N I R T T L L T - N   T T G N P E H T S Q
A1        . . . p . . . . . .   s . . a p . . . . .   . e . . . . . . . .   . . t . . p f . s .   . . r . . . l . . .
A2        . . . . . . . . . .   . . . v . . . . s .   . . . . . . . . . .   . . . . . . . . s .   . . . . . . . . . .
A3        . . . . . . . . . .   . . . v . . . . . .   . . . . . . . . . .   . . . . . . . . t .   . . . . . . h/y . . .
A4        . . . . . . . . . .   . . g t . . . . . .   . . s . . . a . . .   . . . . . m p i n .   . . . . . . n . . .
A5        . . . . . . . . . .   . . . a p . . . . .   . . . . . . i . . p   . . . . . . . . n s   . . . . l . . . . .

          5 1                                                           8 7
Consensus K E T L H S T S S E   G N P S P S Q V Y T   T S E Y P S Q P P S   P S N T T N
A1        m . . f . . . . . .   . . . . . . . f s .   . . . . . . . . s .   . p . . s r
A2        . . . . . . . t . .   . . . . . . . . . .   . . . . l . . s l .   . . . . . .
A3        . . . . . . . . p .   . . . . . . . . . .   . . . . . . . . . .   . . . . . d/n
A4        . . . . . . . . . .   . . . . . p . . . .   . . . . . . . . . .   . f . . . a
A5        e . . . . . . . . .   . . t . . . . . . .   . . . . l . . . . .   . . . i . . q


          1                                                                                                   5 0
Consensus K R D P K T P A K T   L K K E T T T N P T   K K P T L K T T E R   D T S T S Q S T V L   D T T T S K H T I Q
B1        . . a . . . . . . .   t . . . . . . . . .   . . . . . e . . . .   . . . . . . . . . .   . . . . l . . . . .
B2        . . . . . . . . . .   . . . . . . . . . .   . . . . . . . . . .   . . . . p . . . . .   . . . . p n . . . .
B3        . . . . . . l . . .   . . . . . . . . . .   . . . . p . . . . .   . . . . . . . . . .   . . . . . . . . . .
B4        . . . p/s . . s . . a   . r . . . . . d . .   . e . . . . . . . .   . . . . . r . . . .   . . . . . d . . v .

          5 1                                                           8 7
Consensus Q Q S L H S T T P E   N T P N S T Q T P T   A S E P S T S N S T   Q K A
B1        . . . . . . . . . .   . . . . . . . . . .   . . . . . . . . . .   . n t q s r d
B2        . . . . . . . . . .   . . . . . . . . . .   . . . . . . . . . .   . . .
B3        . . . . . . . . . .   . . . . . . . . . .   . . . . . . . . . .   . . l
B4        . . . . . . . . l .   . . . . . . . . . .   . . . . . . . . . .   . g .
```

Fig. 2. Alignment of predicted amino acid consensus sequences for A and B genotypes based on the 270 nucleotide fragment of the G gene. A consensus sequence generated from the genotype consensus amino acid sequences is shown for A and B groups, respectively. Only amino acids that differ from the group consensus sequence are represented. Amino acid positions with two different residues indicate that there was not a consensus for the genotype at that position and both amino acids were represented in equivalent proportion among the studied strains.

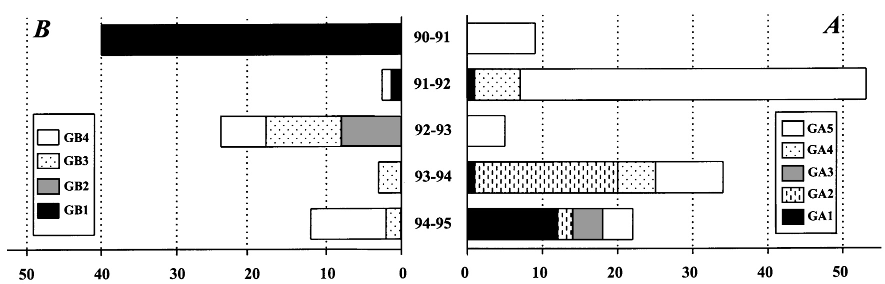

Fig. 3. Distribution of proposed HRSV genotypes (GA1–GA5 and GB1–GB4) over five HRSV seasons. Numbers in the abscissa represent absolute values.

difference in predicted amino acid difference between group B strains was 13 %, with a range from 7 % (GB2 and GB3) to 19 % (GB1 and GB4).

Certain amino acid substitutions were specific for the genotypes, i.e. every genotype could be distinguished from the others belonging to the same group based on particular amino acid substitutions at different positions along the deduced amino acid motif. For instance, GA1–GA5 could be differentiated by, respectively, Pro^{4}, Ser^{19}, Pro^{59}, Pro^{66}, and Thr^{63}, and GB1–GB4 could be distinguished by Ala^{3}, Pro^{35}, Pro^{25}, and Arg^{36} (Fig. 2).

Temporal circulation patterns of HRSV group A and B genotypes and subtypes

Several patterns in the temporal distribution of HRSV strains were noted when the strains were analysed in the context of HRSV group, genotype and subtype.

First, the relative ratio of group A and group B strains shifted each year (Fig. 3), i.e. group B strains predominated during the first and third years, group A strains predominated during the second and fourth years, and A and B group strains were present at nearly equal proportions during the last year.

Table 3. Subtype distribution of HRSV A and B genotypes (1990/95)

Year (*n*)	Subtype*	*n*†	%‡
1990/91 (49)	A.5h	9 (6)	18
	B.1a	**39 (37)**	**80**
1991/92 (55)	**A.5g**	**18 (15)**	**32**
	A.5h	**16 (15)**	**29**
	A.5f	3	5
	A.5i	3	5
	A.3c	3 (2)	5
	A.5k	2	4
	A.3b	2 (2)	4
	A.5b	1	2
	A.5c	1	2
	A.5d	1	2
	A.5j	1	2
	A.3a	1	2
	A.1b	1	2
	B.1a	1	2
	B.4a	1	2
1992/93 (29)	A.5e	4 (4)	14
	A.5d	1	3
	B.3a	**10 (4, 3, 2)**	**34**
	B.2a	**8 (6)**	**28**
	B.4a	6 (5)	21
1993/94 (37)	**A.2b**	**16 (13, 3)**	**43**
	A.4a	**5 (3)**	**14**
	A.2c	3 (2)	8
	A.5h	3 (3)	8
	A.5g	2 (2)	5
	A.5e	2	5
	A.5b	1	3
	A.5a	1	3
	A.1a	1	3
	B.3a	3	8
1994/95 (34)	**A.1a**	**12 (10)**	**35**
	A.5h	3 (3)	9
	A.2a	2	6
	A.3d	2	6
	A.3e	2 (2)	6
	A.5f	1	3
	B.4b	**10 (8)**	**29**
	B.3a	1	3
	B.3b	1	3

* Subtypes (defined as $\geqslant$ 96% nucleotide similarity) were denoted with a three-character code. The predominant subtypes for each RSV season are indicated in bold.
† *n* = total number of isolates, numbers in brackets indicate identical sequences (subtypes A.2b and B.3a showed two and three distinct groups of identical sequences, respectively).
‡ Percentage applies to the total number of isolates for that year.

Second, multiple genotypes and subtypes circulated each year. Two subtypes from two distinct genotypes circulated during the 1990/91 season, and five (1992/93) to 15 subtypes (1991/92) from four to six genotypes circulated during the remaining four RSV seasons (Fig. 3 and Table 3). Third, despite the multiplicity of subtypes circulating each year, one to two subtypes accounted for more than 50% of the total number of isolates for the respective year (predominant subtypes) (Table 3, bold letters).

The shift in predominant genotype and subtype each year was of particular interest. In addition to the change in the relative predominance of A and B group strains each year, a shift in genotype and subtype occurred each year such that no one genotype or subtype was among the predominant genotypes or subtypes in more than one of the study seasons. Several genotypes and subtypes, however, were present in multiple seasons. GA5 was identified in each season, GA1 in four of five seasons, and GB3 and GB4 were detected in three seasons (Fig. 3). Of the 28 group A and B subtypes, seven were present during two seasons, while subtype B.3a was present during three seasons and subtype A.5h during four seasons. The remaining 19 subtypes were present during only one season. Among the 10 predominant subtypes (Table 3), five were present in one additional season and one (B.3a) was present during two additional seasons.

Identical sequences within some subtypes were seen in multiple years (Fig. 1 and Table 3). One subtype, A.1a isolate (1993/94 season), had a sequence identical to 10 isolates in the following season. Six subtype A.5h isolates (1990/91 season) had a sequence identical to 15 isolates from the 1991/92 and three isolates from the 1993/94 seasons. Fifteen subtype A.5g isolates (1991/92 season) had a sequence identical to two isolates from the 1993/94 season. Among B group strains, 37 subtype B.1a isolates (1990/91 season) had the same sequence as one isolate in the following season.

Discussion

Extensive differences in HRSV strains have been demonstrated previously by different groups studying primarily two RSV membrane proteins, the G protein and the SH protein (reviewed in Cane & Pringle, 1995*b*). Cane *et al.* (1994) analysed HRSV strain variation in consecutive years in one community on the basis of SH gene sequencing for the A group and N gene restriction analysis for the B group. They observed a shift in the predominant genetic group, circulation of multiple 'lineages' of HRSV, and emergence of new lineages in the community during the study period. In addition, studies in several communities have shown yearly shifts in relative predominance of group A and B strains (Hall *et al.*, 1990; Waris, 1991). The data obtained in our study provide additional evidence that the occurrence of shifts in HRSV strains is a consistent and probably important feature of the epidemiology of HRSV infections.

Phylogenetic analysis of sequences obtained from the second variable region of the G gene allowed us to identify two levels of relationships between strains within the same group that we arbitrarily designated as genotypes and

subtypes. Genotypes probably identify differences between strains within group A or group B most likely to affect antigenic characteristics of these viruses, while subtypes should identify viruses which are antigenically similar.

To place our genotype designation in the context of previously reported studies, we examined the relationships of G gene sequences of the HRSV strains obtained in our study with published G gene sequences described by Cane *et al.* (1991), Sullender *et al.* (1991) and Garcia *et al.* (1994) (dendrograms not shown).

Phylogenetic analysis for A group strains showed that (i) strains 642 and 1734 (United Kingdom; Cane *et al.*, 1991) and Mon-9-91, Mon-9-92, Mon-5-90 and Mon-5-91 (Uruguay; Garcia *et al.*, 1994) clustered with GA1 strains; (ii) strains A2 (Australia) and Long (USA) were related to GA1 strains but were distinct from each other and from the other GA1 strains; (iii) strains Mon-3-88, Mon-1-89 (Uruguay) and Mad-3-89 (Spain) (Garcia *et al.*, 1994) clustered with GA2 strains; (iv) strains 5857, 6614 and 6256 (United Kingdom; Cane *et al.*, 1991), Mad-4-91, Mad-3-92 and Mad-5-92 (Spain; Garcia *et al.*, 1994) clustered with GA3 strains; and (v) strains represented by Mad-6-92, Mad-8-92 (Spain), Mon-1-90 and Mon-8-92 (Uruguay) (Garcia *et al.*, 1994) and 6190 (United Kingdom; Cane *et al.*, 1991) grouped with GA5 strains. None of the strains with published G gene sequences clustered with GA4 strains.

Since there are substantially fewer published B group G gene sequences, it was more difficult to define genotype relationships in the context of those sequences. Based on data from Sullender *et al.* (1991), we observed that (i) prototype strains 8-60 (Sweden) and 18537 (USA) formed a separate and distinct cluster, (ii) strains 9320 (USA), nm1355 (USA) and wv4843 (West Virginia, USA) were distinct from each other and did not cluster with any of our genotypes, and (iii) strains wv10010 and wv15291 (West Virginia, USA) grouped with GB1 strains.

This analysis highlights the fact that genotype and even subtype are preliminary designations and this, or other approaches, to categorize HRSV strains will probably be modified as additional strain sequence data becomes available.

Genotype and subtype designations have, however, proven to be very helpful in illustrating the multiplicity of strains circulating in Rochester, New York; the presence of predominant genotypes and subtypes during each HRSV season; and the shift in the relative proportion of group A and B strains and in the predominant genotype and subtype each year. The consistent shift in predominant genotypes and subtypes is emphasized by the fact that no genotype or subtype predominated for more than one HRSV season and suggests that strain differences either impact the efficiency of circulation or the likelihood that a strain will be isolated. We hypothesize that strain differences affect protective immunity such that a relative lack of immunity in the population to a 'novel' virus may allow it to transmit more efficiently or be more pathogenic.

A number of studies have demonstrated immunologically important differences between group A and group B viruses (Johnson *et al.*, 1987*a*; Hendry *et al.*, 1988), but very few studies have looked for immunological differences between strains in the same group, and no study has yet demonstrated such differences to be immunologically important. Two studies of antibody-mediated response induced by natural infection in humans do, however, show that variable regions of the G protein do induce antibodies specific to strains within group A. In these two studies, human antibodies reacted against expressed or synthetically produced peptides from variable regions of the G protein in a strain-specific fashion (Cane *et al.*, 1996; Cane, 1997). We have initiated studies to determine if these 'strain-specific' G protein antibodies influence which strains cause disease in infants during community RSV outbreaks. Since the G protein is highly glycosylated, it is possible that the immunological impact of amino acid differences between strains is amplified by changes that occur at glycosylation sites (Melero *et al.*, 1997). Mutations in potential glycosylation sites in the haemagglutinin protein have been shown to provide a selective advantage for influenza virus by either generating virus progeny which are more resistant to neutralization or 'covering' antigenic sites by the addition of a new glycosylation site (Schulze, 1997).

In addition to their potential importance to RSV immunity, strain differences have also helped us better understand RSV transmission in hospitals, in communities and globally. Previous studies have identified both community-specific and regional or global features of the epidemiology of HRSV strains. On the basis of monoclonal antibody reaction patterns, one study demonstrated that different communities had distinct patterns of circulating strains during the same year and found no regional or national epidemic strain(s), suggesting that the pattern of outbreak strains is dictated from within the community (Anderson *et al.*, 1991). Other investigators, using sequencing studies, have identified very similar strains present simultaneously in widely separated geographical regions (Cane & Pringle, 1995*a*; Garcia *et al.*, 1994), suggesting that outbreak strains may spread globally. The data from the present study suggest that both community factors and global spread impact circulation of strains in a community. First, the multiplicity of strains and the appearance of strains not previously seen in the community suggest that many strains probably come from outside the community. Second, the potential for G gene sequences to be stable over several years as shown in this study and other studies (Garcia *et al.*, 1994; Cane & Pringle, 1995*a*) requires that investigators use caution in determining source and links between outbreak strains based on G gene sequences. Thus, isolation of viruses with identical sequences from distant locations indicates a common source but the stability of G gene sequences means that spread from this common source may have occurred over years and not necessarily in an epidemic fashion like that for influenza A virus. We suspect that HRSV strains from all over the world

can be introduced into a community, but then local factors, such as population patterns of strain-specific immunity, determine which of these strains circulate efficiently and become 'outbreak strains'.

In summary, our data suggest that the multiplicity and shifting patterns of circulating HRSV strains may be important to the ability of HRSV to consistently cause yearly outbreaks of disease. Additionally, our data suggest that both group and genotype may need to be considered to fully understand protective immunity after natural infection and the results of vaccine efficacy studies.

The authors thank Sarah McKneally (CDC Biotechnology Core Facility, Computational Chemistry Laboratory) for the valuable help during the computational analysis and Brian Holloway and staff (CDC Biotechnology Core Facility, DNA Chemistry Section) for the oligonucleotide synthesis. T.C.T.P. was partially sponsored by CNPq, Brazil while obtaining her doctoral degree at UFESP-EPM, São Paulo, Brazil.

References

Åkerlind, B., Norrby, E., Örvell, C. & Mufson, M. A. (1988). Respiratory syncytial virus: heterogeneity of subgroup B strains. *Journal of General Virology* **69**, 2145–2154.

Anderson, L. J., Hierholzer, J. C., Tsou, C., Hendry, R. M., Fernie, B. F., Stone, Y. & McIntosh, K. (1985). Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *Journal of Infectious Diseases* **151**, 626–633.

Anderson, L. J., Parker, R. A. & Strikas, R. L. (1990). Association between respiratory syncytial virus outbreaks and lower respiratory tract deaths of infants and young children. *Journal of Infectious Diseases* **161**, 640–646.

Anderson, L. J., Hendry, R. M., Pierik, L. T., Tsou, C. & McIntosh, K. (1991). Multicenter study of strains of respiratory syncytial virus. *Journal of Infectious Diseases* **163**, 687–692.

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. *Journal of Medical Virology* **51**, 297–304.

Cane, P. A. & Pringle, C. R. (1991). Respiratory syncytial virus heterogeneity during an epidemic: analysis by limited nucleotide sequencing (SH gene) and restriction mapping (N gene). *Journal of General Virology* **72**, 349–357.

Cane, P. A. & Pringle, C. R. (1992). Molecular epidemiology of respiratory syncytial virus: rapid identification of subgroup A lineages. *Journal of Virological Methods* **40**, 297–306.

Cane, P. A. & Pringle, C. R. (1995*a*). Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino changes in the attachment protein. *Journal of Virology* **69**, 2918–2925.

Cane, P. A. & Pringle, C. R. (1995*b*). Molecular epidemiology of human respiratory syncytial virus. *Seminars in Virology* **6**, 371–378.

Cane, P. A., Matthews, D. A. & Pringle, C. R. (1991). Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. *Journal of General Virology* **72**, 2091–2096.

Cane, P. A., Matthews, D. A. & Pringle, C. R. (1992). Analysis of relatedness of subgroup A respiratory syncytial viruses isolated worldwide. *Virus Research* **25**, 15–22.

Cane, P. A., Matthews, D. A. & Pringle, C. R. (1994). Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. *Journal of Clinical Microbiology* **32**, 1–4.

Cane, P. A., Thomas, H. M., Simpson, A. F., Evans, J. E., Hart, C. A. & Pringle, C. R. (1996). Analysis of the human serological response to a variable region of the attachment (G) protein of respiratory syncytial virus during primary infection. *Journal of Medical Virology* **48**, 253–261.

Cristina, J., Moya, A., Arbiza, J., Russi, J., Hortal, M., Albo, C., Garcia-Barreno, B., Garcia, O., Melero, J. A. & Portela, A. (1991). Evolution of the G and P genes of human respiratory syncytial virus (subgroup A) studied by the RNase A mismatch cleavage method. *Virology* **184**, 210–218.

Dowell, S. F., Anderson, L. J., Gary, H. E., Erdman, D. D., Plouffe, J. F., File, T. M., Marston, B. J. & Breiman, R. F. (1996). Respiratory syncytial virus is an important cause of community-acquired lower respiratory infection among hospitalized adults. *Journal of Infectious Diseases* **174**, 456–462.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. (1995). PHYLIP: Phylogeny inference package, version 3.57c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA.

Garcia, O., Martin, M., Dopazo, J., Arbiza, J., Frabasile, S., Russi, J., Hortal, M., Perez-Brena, P., Martinez, I., Garcia-Barreno, B. & Melero, J. A. (1994). Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. *Journal of Virology* **68**, 5448–5459.

Hall, C. B., Walsh, E. E., Schnabel, K. C., Long, C. E., McConnochie, K. M., Hildreth, S. W. & Anderson, L. J. (1990). Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiological and clinical characteristics in hospitalized and ambulatory children. *Journal of Infectious Diseases* **162**, 1283–1290.

Hendry, R. M., Talis, A. L., Godfrey, E., Anderson, L. J., Fernie, B. F. & McIntosh, K. (1986). Concurrent circulation of antigenically distinct strains of respiratory syncytial virus during community outbreaks. *Journal of Infectious Diseases* **153**, 291–297.

Hendry, R. M., Burns, J. C., Walsh, E. E., Graham, B. S., Wright, P. F., Hemming, V. G., Rodriguez, W. J., Kim, H. W., Prince, G. A., McIntosh, K., Chanock, R. M. & Murphy, B. R. (1988). Strain-specific serum antibody responses in infants undergoing primary infection with respiratory syncytial virus. *Journal of Infectious Diseases* **157**, 640–647.

Institute of Medicine (1985). Diseases of importance in the United States. In *New Vaccine Development: Establishing Priorities*, vol. 1, pp. 397–409. Washington, DC: National Academy Press.

Johnson, P. R., Jr, Olmsted, R. A., Prince, G. A., Murphy, B. R., Alling, D. W., Walsh, E. E. & Collins, P. L. (1987*a*). Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity. *Journal of Virology* **61**, 3163–3166.

Johnson, P. R., Spriggs, M. K., Olmsted, R. A. & Collins, P. L. (1987*b*). The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proceedings of the National Academy of Sciences, USA* **84**, 5625–5629.

Kim, H. W., Arrobio, J. O., Brandt, C. D., Jeffries, B. C., Pyles, G., Reid, J. L., Chanock, R. M. & Parrott, R. H. (1973). Epidemiology of respiratory syncytial virus infection in Washington, DC. I. Importance of the virus in different respiratory tract disease syndromes and temporal distribution of infection. *American Journal of Epidemiology* **98**, 216–225.

Melero, J. A., García-Barreno, B., Martínez, I., Pringle, C. R. & Cane, P. A. (1997). Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. *Journal of General Virology* **78**, 2411–2418.

Mufson, M. A., Örvell, C., Rafnar, B. & Norrby, E. (1985). Two distinct subtypes of human respiratory syncytial virus. *Journal of General Virology* **66**, 2111–2124.

Mufson, M. A., Belshe, R. B., Örvell, C. & Norrby, E. (1988). Respiratory syncytial virus epidemics: variable dominance of subgroup A and B strains among children, 1981–1986. *Journal of Infectious Diseases* **157**, 143–148.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Sanz, M. C., Kew, O. M. & Anderson, L. J. (1994). Genetic heterogeneity of the attachment glycoprotein G among subgroup A respiratory syncytial viruses. *Virus Research* **33**, 203–217.

Schulze, I. T. (1997). Effects of glycosylation on the properties and functions of influenza virus hemagglutinin. *Journal of Infectious Diseases* **176** (Suppl. 1), s24–s28.

Storch, G. A., Anderson, L. J., Park, C. S., Tsou, C. & Dohner, D. E. (1991). Antigenic and genomic diversity within group A respiratory syncytial virus. *Journal of Infectious Diseases* **163**, 858–861.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13**, 964–969.

Sullender, W. M., Mufson, M. A., Anderson, L. J. & Wertz, G. W. (1991). Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. *Journal of Virology* **65**, 5425–5434.

Waris, M. (1991). Pattern of respiratory syncytial virus epidemics in Finland: two-year cycles with alternating prevalence of groups A and B. *Journal of Infectious Diseases* **163**, 464–469.

Wertz, G. W., Collins, P. L., Huang, Y., Gruber, C., Levine, S. & Ball, L. A. (1985). Nucleotide sequence of the G protein gene of respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proceedings of the National Academy of Sciences, USA* **82**, 4075–4079.

Received 6 January 1998; Accepted 29 April 1998