Epidemiol. Infect. (1999), **122**, 155–160. Printed in the United Kingdom

Molecular epidemiology of respiratory syncytial virus in The Gambia

P. A. CANE[1]*, M. WEBER[2]†, M. SANNEH[2], R. DACKOUR[2], C. R. PRINGLE[3] AND H. WHITTLE[2]

[1] *Division of Immunity and Infection, University of Birmingham Medical School, Birmingham B15 2TT, UK*
[2] *Medical Research Council Laboratories, Fajara, PO Box 273, Banjul, The Gambia*
[3] *Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK*

(*Accepted 24 August 1998*)

SUMMARY

Respiratory syncytial virus (RSV) infection in The Gambia occurs seasonally in association with the rainy season. This study examined the genetic variability of RSV isolates from four consecutive epidemics from 1993–6. Each epidemic was made up of a number of variants which were replaced in subsequent epidemics. Analysis of attachment (G) protein gene sequences showed that isolates were closely related to those observed in the rest of the world. However, many isolates from 1993 and 1994 were unlike other isolates observed in the developed world during this period and were more similar to isolates from 1984 in Europe. In addition, the most commonly observed genotype in the UK in the 1990s was not detected in The Gambia during this period.

INTRODUCTION

Respiratory syncytial virus (RSV) is well known as the major viral cause of lower respiratory tract infection in infants in developed countries with epidemics occurring annually during the winter months in temperate climates. Much less is known about RSV infection in developing countries. The significance of RSV in The Gambia over four outbreaks from 1993 to 1996 has recently been reported [1]. It was found that the majority of RSV cases occurred between August and November starting about 1–2 months after the onset of the rains. This result was similar to that described for other tropical countries with seasonal rainfall [2–5].

It has previously been shown that RSV isolates can be divided into two antigenic groups, A and B, on the basis of their reactions with monoclonal antibodies and that these groups are also distinct with respect to nucleotide sequence. It has also been demonstrated that the groups can be subdivided into a number of genotypes (reviewed in [6]), with the most variable gene between both the groups and genotypes being the attachment (G) glycoprotein gene. The group A genotypes have been designated SHL1–6, and the group B genotypes NP1 and NP3 [7]. The relative proportions of the genotypes in epidemics vary from year to year with steady replacement of the dominant genotype each year, suggesting that herd immunity may play a role in the abundance of particular genotypes [8]. The G protein appears to show accumulation of amino acid changes with time suggesting evolution under selective pressure [9]. Although RSV can be variable even within a single local epidemic, it has been found that very similar viruses circulate world-wide [7, 10–12]. Thus, RSV isolates analysed to date show temporal but little geographical clustering [9]. However, these analyses

* Author for correspondence:
† Present address: CHD/WHO, CH–1211 Geneva 27, Switzerland.

https://doi.org/10.1017/S0950268896001872 Published online by Cambridge University Press

have mainly been based on isolates from developed temperate countries: this report now examines the molecular epidemiology of RSV in a developing tropical country and compares the relatedness of RSV isolates from there with those from Europe.

MATERIALS AND METHODS

Patients

The Gambia occupies a narrow strip of land, about 25 miles wide by 300 miles long, in the lower and central Gambia River valley in West Africa. The climate is tropical with a rainy season that lasts from June to October. Surveillance for RSV disease in children < 2 years of age took place from 1993 to 1996 in the Royal Victoria Hospital in Banjul, the MRC hospital, and in the Sibanor Mission Hospital situated about 95 km from Banjul and serving a rural population. Diagnosis of RSV infection was by antigen detection by immunofluorescence on nasopharyngeal aspirates.

Virus isolation and culture

Samples positive by immunofluorescence were inoculated onto Hep-2 or HeLa cells and then further subcultured in either Hep-2 or MRC-5 cells.

Genetic analysis

A random selection of isolates was grouped and further genotyped as described previously [7]. Briefly, PCR products derived from parts of the nucleocapsid (N) and G genes were analysed by restriction digests, together with some nucleotide sequencing of the G gene [7, 8]. Nucleotides 857–1135 of the N gene were amplified and the PCR products digested with *Hin*dIII, *Pst*I, *Bgl*II, *Hae*III and *Rsa*I. This gives patterns for human RSV which have been designated previously NP1–NP6 with further patterns being found in animal pneumoviruses [10]. Nucleotides 1–584 of the G gene of group A isolates and 153–817 of group B isolates were also amplified and these products digested with *Alu*I, *Taq*I, *Mbo*I and *Mse*I. Nucleotides 297–514 of the G genes of representative isolates showing different N and/or G gene restriction patterns isolates were sequenced by automatic ABI sequencing of PCR products. In some instances where the initial sequencing indicated novel variants, the sequence of nucleotides 799–917 was also determined. Relatedness of sequences was determined using CLUSTAL [13].

RESULTS

Distribution of isolates during epidemics

Most cases of RSV infection observed from 1993 to 1996 were diagnosed in the rainy season from July to November with the peaks the epidemics occurring in August to October. The median age of the children was 6 months. Isolates were obtained throughout the area under study and there was little apparent geographic progression of the epidemics with time.

Variability of RSV isolates over four epidemics

RSV group

Eighty-four isolates over the 4 years were analysed with respect to group and genotype. In the case of the 1995 epidemic, only five isolates were examined due to loss of infectivity of many virus stocks. 27/30 (90%) of isolates examined from 1993 were group A, 8/29 (28%) from 1994, 3/5 (60%) from 1995 and 19/20 (95%) from 1996 (Table 1). Thus, from this study of only four epidemics, it appears that group A is the overall dominant group in most epidemics, as found in temperate climates. The occurrence of a predominantly group B epidemic in 1994 did not correlate temporally with group B epidemics in the UK which occurred in 1992–3 and 1995–6 ([8]; unpublished data).

Variants

Restriction endonuclease mapping of PCR products derived from parts of the N and G genes showed that the epidemics were not homogeneous as has also been described for epidemics in temperate climates. Many of the group A isolates showed restriction patterns distinct from those previously seen in isolates from developed countries so it was not always possible using just this technique to allocate isolates to particular genotypes [7], and so the term 'variants' is used in this report. For example, 14/30 of the isolates examined from 1993 showed an N gene restriction pattern not previously observed: an additional *Rsa*I site was present. This pattern has been designated NP10. These isolates also had a distinct G gene nucleotide sequence (see below). The relative incidence of each of the observed variants in each of the epidemics is shown in Table 1: the group A isolates

https://doi.org/10.1017/S0950268896001872 Published online by Cambridge University Press

Table 1. *Relative incidence of RSV groups and variants in epidemics 1993–6*

Group/ genotype	1993		1994		1995		1996	
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
B group	3	(10)	21	(72)	2	(40)	1	(5)
A group total	27	(90)	8	(28)	3	(60)	19	(95)
Variant G1	14	(47)	3	(10)	0		0	
Variant G2	6	(20)	0		0		0	
Variant G3	6	(20)	4	(14)	0		0	
Variant G4	0		1	(3)	0		9	(45)
Variant G5	0		0		3	(60)	9	(45)
Other*	1	(3)	0		0		1	(5)

* Two isolates gave restriction patterns different from all other isolates.

Table 2. *Restriction patterns of PCR products derived from Gambian group A RSV isolates*

Variant	N gene pattern*	G gene pattern†
G1	NP10	A^2BBE
G2	NP4	A^1ABE
G3	NP4	A^2ABE
G4	NP4	BABE
G5	NP2	BABE

* The N gene patterns (except NP10) are as described in ref. [7]. NP10 is similar to NP4 but with an additional *Rsa*I site.
† The G gene patterns are as described in ref. [7] (the first letter is the *Alu*I pattern, the second *Taq*I, the third *Mbo*I, and the fourth *Mse*I) with the exception that the *Alu*I A patterns have been subdivided based on mobility of a minor band.

have been arbitrarily designated variants Gambia (G)1–5 while the group B isolates have not been further subdivided. These variants were defined as showing differences in either their N gene or G gene restriction patterns. The restriction patterns shown by the variants are summarized in Table 2. Variants G4 and G5 showed the same G gene patterns and similar G gene nucleotide sequences (see below) but differed in their N gene pattern (NP4 and NP2 respectively). These designations are not the same as the genotypes described from temperate climates, although some of the viruses are very closely related to temperate viruses (see below). The variants were distributed throughout the country and during the epidemic seasons.

All the group B isolates showed the same N gene pattern, NP3. Nucleotide sequencing of part of the G gene from two group B isolates (data not shown) showed sequence distinct from those reported by Sullender and colleagues [14], but data are not available to compare these sequences with group B strains currently circulating in the rest of the world.

As has been observed in the UK [8], each epidemic was made up of different proportions of the observed variants, with the most abundant variants (G1–3) in the 1993 epidemic making a much reduced contribution to the 1994 epidemic, and then apparently disappearing, while the most frequent variants in the 1996 epidemic were not observed in 1993.

G gene nucleotide variability

A variable region of the G genes from 14 group A isolates was sequenced. The isolates were selected as representative of the different variants observed. As has been previously described for G gene sequence analysis [11, 14, 15], the majority of the nucleotide changes observed in this gene led to predicted amino acid changes in the G protein, including changes in areas that have been found to contain epitopes important in the human antibody response [16]. G gene nucleotide sequence data from representative examples of the variants are shown in Figure 1. The relatedness of these sequences to each other and to sequences from isolates from Europe, especially Birmingham, UK, is shown in Figure 2. All the Gambian isolates were found to cluster with isolates from Europe. The majority of isolates from the 1994, 1995 and 1996 epidemics were closely related to isolates circulating in the UK during the same period. For example, phylogenetic analysis of G gene nucleotide sequences of isolates of variants G4 and G5 showed that these clustered with a Birmingham isolate, Birm/896/94, which had been previously classified as belonging to genotype SHL6 which has been common

https://doi.org/10.1017/S0950268896001872 Published online by Cambridge University Press

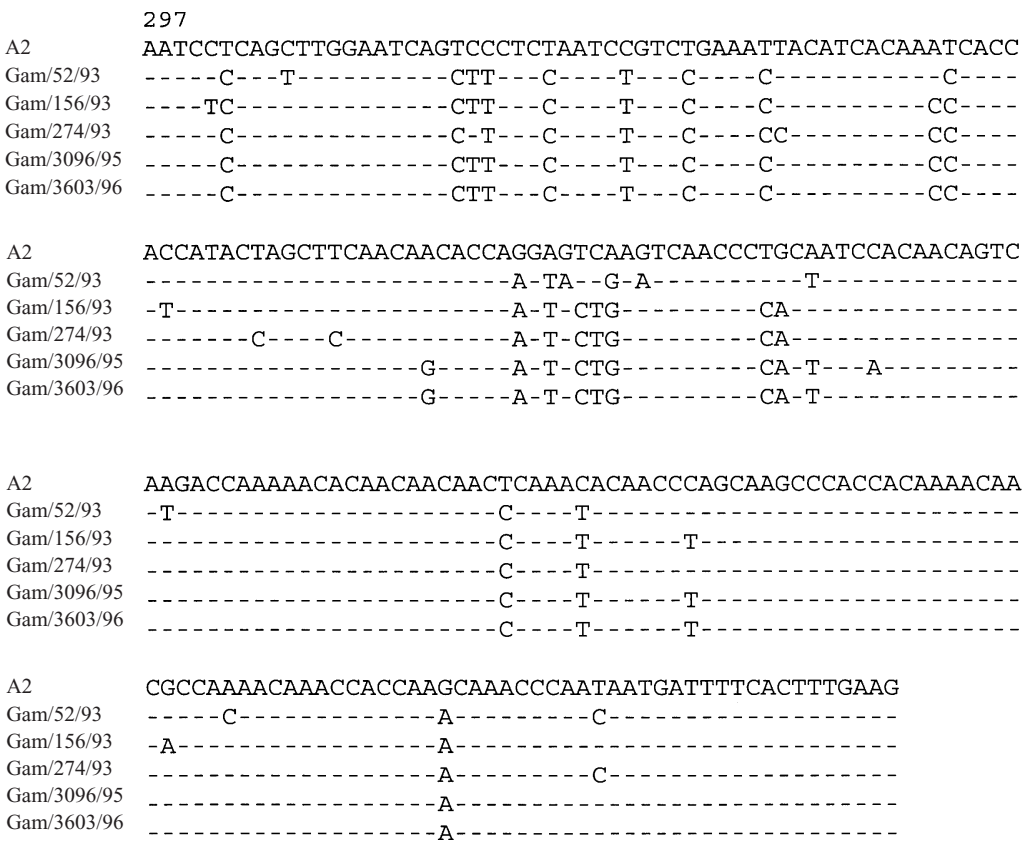

Fig. 1. Nucleotide sequence comparisons among G mRNAs (nucleotides 297–514) for Gambian RSV isolates from 1993 to 1996, aligned with sequence for strain A2 (17). G1 variants are represented by Gam/52/93. G2 variants by Gam/274/93, G3 variants by Gam/156/93, G4 variants by Gam/3096/95, and G5 variants by Gam/3603/96.

in Europe since 1991 ([8]; unpublished results). As shown in Figure 2 variants G4 and G5 had very similar G gene nucleotide sequences, although they had distinct N gene patterns. This discordance between N gene and G gene variability has not been observed previously.

The isolates from 1993 which showed a novel N gene restriction pattern (variant G1) were unlike any observed in Europe in the 1990s, and were initially thought to be unique to The Gambia. However, G gene sequencing of isolates Gam/52/93 and Gam/66/93 gave nucleotide sequence which clustered with unpublished sequence of a strain from Madrid isolated in 1984 (Mad/1/84) (J. Melero, personal communication) as illustrated in Figure 2. The amino acid sequence of the variable regions of the G gene of isolate Gam/52/93 is shown in Figure 3.

Isolates belonging to genotype SHL2 (isolates Birm/6190/89, Birm/8960/95, and Birm/9387/95 in Fig. 1) which has recently been very abundant throughout Europe and has also been frequently found in USA, Uruguay, Malaysia and Australia [6, 7], were not detected in The Gambia during 1993–6. Likewise, isolates of genotype SHL5 (isolates Birm/1734/89 and Birm/642/89 in Fig. 2), common in the developed world in the late 1980s, were not detected in The Gambia during the period examined.

DISCUSSION

RSV epidemics in The Gambia have been found to be seasonal and associated with the rainy season as has been described for other tropical countries with seasonal rainfall [2–4]. As in temperate climates, the epidemics were found to be made up of a number of variants of RSV and there was steady replacement year on year of the dominant variant.

Thus, the phenomenon observed in cities in temperate regions in both Uruguay [19] and UK [8] is also present in a less-developed tropical country. In addition, part of this study included isolates from patients living in rural areas and these also showed a mixture of variants even within the same village (data not shown). As with the previous studies, the samples described here were derived only from severely ill infants, so one cannot be certain that these entirely

https://doi.org/10.1017/S0950268896001872 Published online by Cambridge University Press

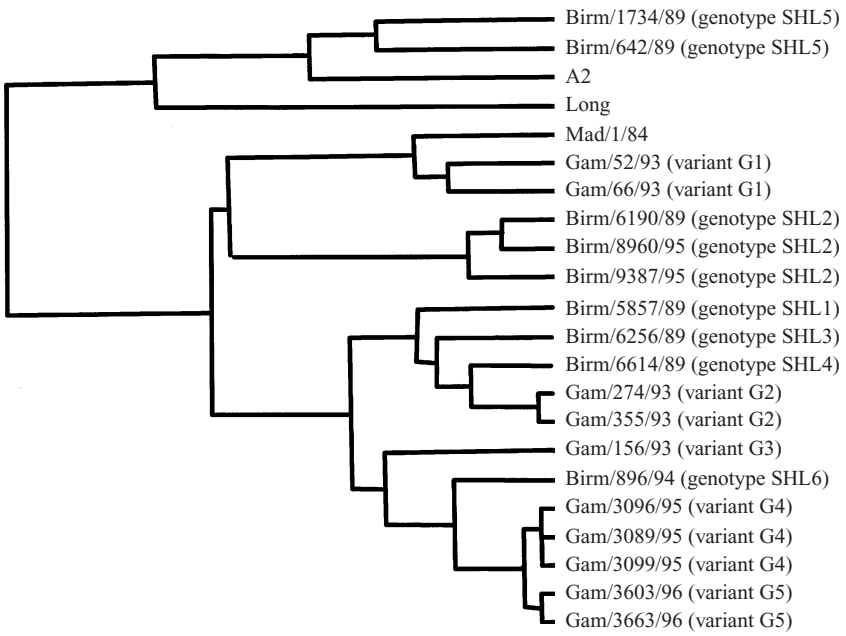

Fig. 2. Dendogram to show relatedness of group A RSV G genes (nucleotides 297–514), derived using CLUSTAL [13]. Letters at beginning of isolate designations indicate place of isolation (Gam = Gambia; Birm = Birmingham, UK; Mad = Madrid). Numbers at end of isolate designations indicate year of isolation. Sequence for strain A2 (Australia, 1961) was taken from Wertz and colleagues [17], for strain Long (USA, 1956) from Johnson and colleagues [18], for Mad/1/84 from personal communication from J. A. Melero, and for Birm/89 isolates from Cane and colleagues [15]. Other sequences are from this report or unpublished. Genotype designations for the Birmingham isolates and variant designations for the Gambian isolates are indicated in parentheses.

```
            89
A2          TYLTQNPQLGISPSNPSEITSQITTILASTTPGVKSTLQSTTVK
Gam/52/93   --------F---F--L--T---T---------SIE---L----M

            262
A2          METFHSTSSEGNPSPSQVSTTSEYPSQPSSPPNTPRQ
Gam/52/93   K--L--------------Y---------P--F-ITA-
```

Fig. 3. Comparison of deduced amino acid sequence of the variable regions (amino acids 89–132 and 262–end) of the G protein of Gambian isolate Gam/52/93 with that of strain A2 [17].

reflect the variants circulating within the community. However, it seems likely that the disappearance of particular variants is at least in part to increased levels of herd immunity and that this probably occurs with respect to RSV epidemics everywhere.

The variants present in The Gambian epidemics did not correlate with those present during the same years in the UK. However, many of the isolates were closely related to currently circulating European isolates (Fig. 2). The 1993 Gambian epidemic included isolates with G gene sequences that were very distinct from published sequences, but phylogenetic analysis showed that these isolates clustered with an isolate from Madrid from 1984. Presumably, this RSV variant was relatively common in Europe around 1984 but has not been detected since ([9, 11]; unpublished results), so the question arises as to where it has been transmitted in the period 1984–93.

Little is known about why RSV epidemics occur and whether the virus continues to circulate in the community at low levels between epidemics or whether it is necessarily introduced from outside a community in order to set off a new epidemic when social or climatic conditions are appropriate. All the analyses reported to date concerning the molecular epidemiology of RSV have concentrated on developed countries with good airline connections to the rest of the world. It is possible that the RSV isolates observed in The Gambia are effectively a mixture of viruses circulating in the developed world and entering the

https://doi.org/10.1017/S0950268896001872 Published online by Cambridge University Press

country via the coastal and urban districts and viruses that are derived from the more remote interior regions of this part of Africa.

ACKNOWLEDGEMENTS

Many thanks to Brian Greenwood and Kim Mulholland for help with organizing the study; Anna Simpson, Daniel Troup and Agnes Awemoyi for excellent technical assistance; Fatou Bah, Mariama Madi, Saikou Badgie, Foday Drammeh, and Saikou Mendy for collecting and transporting samples; and the physicians and the nurses of the study hospitals. This work was supported by the Wellcome Trust, the MRC, and the European Commission.

REFERENCES

1. Weber MW, Dackour R, Usen S, et al. The clinical spectrum of RSV disease in The Gambia. Pediatr Infect Dis J 1998; **17**: 224–32.
2. Hazlett DT, Bell TM, Tukei PM, et al. Viral etiology and epidemiology of acute respiratory infections in Nairobi, Kenya. Amer J Trop Med & Hyg 1988; **39**: 632–40.
3. Nwanko MU, Dym AM, Schuit KE, Offor E, Omene JA. Seasonal variation in respiratory syncytial virus infections in children in Benin-City, Nigeria. Trop Geog Med 1988; **40**: 309–13
4. Cherian T, Simoes EAF, Steinhoff MC, et al. Bronchiolitis in tropical South India. Amer J Child Dis 1990; **144**: 1026–30.
5. Weber MW, Mulholland EK, Greenwood BM. Respiratory syncytial virus infection in tropical and developing countries. Trop Med Int Health 1998; **3**: 268–80.
6. Cane PA, Pringle CR. Molecular epidemiology of human respiratory syncytial virus. Sem Virol 1995; **6**: 371–8.
7. Cane PA, Pringle CR. Molecular epidemiology of respiratory syncytial (RS) virus: rapid identification of subgroup A isolates. J Virol Meth 1992; **40**: 297–306.
8. Cane PA, Matthews DA, Pringle CR. Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. J Clin Microbiol 1994; **32**: 1–4.
9. Cane PA, Pringle CR. Evolution of subgroup A respiratory syncytial virus: evidence for accumulation of amino acid changes in the attachment protein. J Virol 1995; **69**: 2918–25.
10. Cane PA, Pringle CR. Respiratory syncytial virus heterogeneity during an epidemic: analysis by limited nucleotide sequencing (SH gene) and restriction mapping (N gene). J Gen Virol 1991; **72**: 439–57.
11. Garcia O, Martin M, Dopazo J, et al. Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. J Virol 1994; **68**: 5448–59.
12. Sanz MC, Kew OM, Anderson LJ. Genetic heterogeneity of the attachment glycoprotein G among group A respiratory syncytial viruses. Virus Res 1994; **33**: 203–17.
13. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 1988; **73**: 237–44.
14. Sullender WM, Mufson MA, Anderson LJ, Wertz GW. Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. J Virol 1991; **65**: 5425–34.
15. Cane PA, Matthews DA, Pringle CR. Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. J Gen Virol 1991; **72**: 2091–6.
16. Cane PA. 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 1997; **51**: 297–304.
17. Wertz, GW, Collins PL, Huang Y, Gruber C, Levine S, Ball LA. Nucleotide sequence of the G protein of human respiratory syncytial virus reveals an unusual type of viral membrane protein. Proc Natl Acad Sci USA 1985; **82**: 4075–9.
18. Johnson PR, Spriggs MK, Olmsted RA, Collins PL. 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 1987; **84**: 5625–9.
19. Cristina J, Moya A, Arbiza J, et al. Evolution of the G and P genes of human respiratory syncytial virus (subgroup A) studied by the Rnase A mismatch cleavage method. Virology 1990; **184**: 210–8.