Genetic variability among group A and B respiratory syncytial viruses in Mozambique: identification of a new cluster of group B isolates

Anna Roca,1, 3 Mari-Paz Loscertales,1, 3 Llorenç Quintó,1 Pilar Pérez-Breña,4 Neide Vaz,3 Pedro-Luis Alonso1, 3 and Juan-Carlos Saiz2

1, 2Epidemiologıa y Bioestadıstica1 and Hepatologıa-IMD2, Fundació Clinic, IDIBAPS, Hospital Clinic, Universidad de Barcelona, Villarol 170, 08036 Barcelona, Spain
3Centro de Investigacao en Saude da Manhiça, INS-Ministerio da Saude, Mozambique
4Servicio de Virologıa, Centro Nacional de Microbiologıa, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

Respiratory syncytial virus (RSV) is the major cause of acute lower respiratory tract infection in children and vulnerable adults, but little is known regarding RSV infection in Africa. In this report, a recent RSV outbreak in Mozambique was studied and results showed that 275 of 3192 (8.6%) nasopharyngeal aspirates tested were RSV-positive by ELISA. RSV presents two antigenic groups (A and B) with a high genetic and antigenic variability between and within them. Analysis by a new RFLP assay of RT–PCR amplified N protein gene products showed a higher prevalence of group B RSV than that of group A (85% versus 15%). However, genetic variability of the G protein gene was higher among group A RSV strains. The frequency and pattern of glycosylation sites were also quite different between both groups. In addition, two different phylogenetic clusters of Mozambican viruses were found within each group, but only sequences from cluster B-I were relatively distinct from previously described isolates. The implications of such differences in the antigenic and immunogenic characteristics of each group are discussed.

Introduction

Respiratory syncytial virus (RSV) is the major cause of acute lower respiratory tract infection (ALRI) in children and vulnerable adults, such as the elderly and immunocompromised individuals (reviewed by Collins et al., 1996). RSV is distributed worldwide and, in temperate climates, annual epidemics occur during the winter months (Collins et al., 1996). In tropical regions, RSV outbreaks are usually observed during the rainy seasons, although wide distribution throughout the year has also been described (Hazlett et al., 1988; Cane et al., 1999). However, little is known about the burden of the disease associated with RSV infection in developing countries (Weber et al., 1998; M.-P. Loscertales, A. Roca, J.-C. Saiz & P. L. Alonso, unpublished results), where it may follow a different epidemiological pattern (Heirholzer et al., 1994). In these geographical regions about one-third of all deaths in children under 5 years of age are thought to be due to ALRI (WHO, 1984; Forgie et al., 1992), a figure 30 to 70 times higher than that in developed countries.

RSV is a member of the genus Pneumovirus within the family Paramyxoviridae. It has a non-segmented, negative-stranded RNA genome encoding ten genes (Collins et al., 1996). Two viral proteins, the attachment glycoprotein G and the surface glycoprotein F, are the main antigens responsible for inducing a neutralizing immune response and resistance to infection (Collins et al., 1996). Two antigenic groups of RSV (groups A and B) have been described based on their reactivity with monoclonal antibodies (Anderson et al., 1985; Mufson et al., 1985; Hall et al., 1990). This classification was later confirmed by nucleotide sequence analysis (Cristina et al., 1990; Cane & Pringle, 1991). High genetic diversity of the G protein gene occurs between and within the two RSV groups (reviewed by Melero et al., 1997). For instance, the antigenic differences between RSV strains influence the immune response...
of children, which is mainly group-specific and poorly cross-reactive (Hendry et al., 1988; Sullender et al., 1990; Cane et al., 1996).

Different groups of RSV epidemics have been described: those in which only viruses from one group are detected, those in which one of the groups is predominant (more frequently group A), and those in which both groups co-circulate with a similar frequency. In any of these cases, multiple lineages of RSV co-circulate within a single outbreak with replacement of predominant genotypes during consecutive years (Hall et al., 1990; Cane et al., 1994). In addition, viruses isolated in geographically distant places and in different years may be genetically more closely related than viruses isolated in the same place on two consecutive days (reviewed by Melero et al., 1997).

While extensive analysis of RSV G protein gene variability has been performed with isolates from antigenic group A (Cane et al., 1991, 1999; García et al., 1994; Cane & Pringle, 1995, 1997; Martínez et al., 1997; Melero et al., 1997), little information is available about group B strains (Sullender et al., 1990, 1991; Cane et al., 1997; Coggins et al., 1998; Martínez et al., 1999). Only a single study about G protein gene variability of both RSV groups has been published to date (Coggins et al., 1998).

In the present study, we report the prevalence and the genetic variability of group A and B isolates obtained during an RSV outbreak in southern Mozambique, characterizing a new cluster of group B isolates and providing additional insights into the structure and evolution of RSV.

Methods

Study area and population. The clinical–epidemiological study was conducted at Manhiça town in the Maputo province of Mozambique. This rural coastal area has a characteristic subtropical climate with the main rainy season lasting from November to March. A total of 3,192 nasopharyngeal aspirates (NPA) was collected between July 1998 and June 1999 from children under 5 years of age who presented at the Manhiça District Hospital with an illness that met the age-dependent criteria proposed by WHO (1996). Informed consent was obtained from the parents or guardians of all the children enrolled in the study.

Clinical samples and diagnosis. NPA samples were collected in 2 ml of sterile buffered saline and kept at 4 °C until tested, within 24 h after collection, with a commercial ELISA kit (RSV detection kit, Abbott) that detects RSV of both groups A and B. RSV-positive samples and randomly selected negative samples were aliquoted and stored at −70 °C.

RSV RNA extraction and RT–PCR amplification of the N protein gene. Viral RNA was directly extracted from 500 ml of clinical samples using a commercial kit (SV Total RNA Isolation kit, Promega), dissolved in 100 µl of diethyl pyrocarbonate water and stored at −70 °C. To confirm diagnostic results and to carry out RSV-typing studies, one-fifth of the extracted RNA was used to amplify the N protein gene by nested RT–PCR. The RT–PCR mixture contained RSV RNA, 10 pmol sense and antisense outer oligonucleotides (see below), 200 µM dNTPs, 1.5 mM MgCl₂, 10 mM Tris–HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 10 U ribonuclease inhibitor (RNasIn, Promega), 2.5 U AMV reverse transcriptase (Promega) and 1.5 U Taq DNA polymerase (Gibco-BRL) in a total reaction volume of 50 µl and was incubated for 30 min at 42 °C. This was followed by one cycle of denaturation at 94 °C for 5 min and 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min and extension at 72 °C for 1 min, with a final extension at 72 °C for 8 min. An aliquot of the 274 bp RT–PCR product (1 µl) was reamplified in a 50 µl reaction mixture containing 10 pmol sense and antisense inner oligonucleotides (see below) and 1-25 U Taq DNA polymerase. Cycling parameters were as described above. Measures to prevent sample contamination were carefully observed and all PCRs were run with negative controls (Kwok & Higuchi, 1989). Amplified products of the expected size (176 bp) were visualized by electrophoresis on a 1.5% agarose gel stained with ethidium bromide. The following specific oligonucleotides were used: RSV-N(+), 5′ outer, 5′ CCTATGGT(G/T)-CAGGCAAGT, nt 760 to 778; RSV-N(−), 3′ outer, 5′ GGTGTACCTCTGTACCTC, nt 1033 to 1015; N1(+), 5′ inner, 5′ GCAGA-AATGGAACAAGTTG, nt 846 to 865; and N1(−), 3′ inner, 5′ TACTCTCCATTATGCCCTAG, nt 1021 to 1030. Nucleotide positions are according to Collins et al. (1985).

RSV-typing by RFLP analysis. Based on the N protein gene sequence obtained from the National Centre for Biotechnology Information, a new RFLP analysis, without the need for cell culture of the virus, was developed. Approximately 200 ng of the RT–PCR amplified N protein gene products were digested with 10 U of HinfI in a volume of 10 µl for 90 min at 37 °C. Digestion products (68, 62 and 40 bp for group A strains; 114 and 62 bp for group B strains) were analysed on a 2.5% agarose gel stained with ethidium bromide. For evaluation of the new RFLP approach, 30 virus isolates, already characterized in a Spanish reference laboratory (Centro Nacional de Microbiología, Spain) according to previously described methods (Cane & Pringle 1991), were used.

Amplification of the protein G gene of RSV. A total of 30 samples was selected from the entire surveillance period of the study (10 of group A and 20 of group B, randomly selected). The RSV G protein gene was directly amplified from NPA samples as described above, except that cycle parameters were as follows: 30 cycles of denaturation at 93 °C for 30 s, annealing at 45 °C for 30 s and extension at 72 °C for 50 s, with a final extension at 72 °C for 8 min. Amplified products of the expected size, 546 and 583 bp (for group A and B, respectively), were purified using a commercial kit (Concert Rapid PCR purification system, Life Technologies) and directly sequenced in a 310 DNA sequencer (Applied Biosystems) using the Rhodamine Terminator Cycle sequencing kit (Perkin Elmer). Relatedness of sequences was determined with the help of the software ClustalW2 (http://www.ebi.ac.uk/Tools/st/emboss/clustalw2/).

RSV sequencing. Amplified products of the N and G protein genes of the expected size were purified using a commercial kit (Concert Rapid PCR purification system, Life Technologies) and directly sequenced in a 310 DNA sequencer (Applied Biosystems) using the Rhodamine Terminator Cycle sequencing kit (Perkin Elmer). Inner oligonucleotides described above were used for sequencing reactions.

Data analysis. Sequence editing was performed using Sequence Navigator (Applied Biosystems). Relatedness of sequences was de-
RSV genetic variability
termined using CLUSTAL (Higgins & Sharp, 1988) and phylogenetic
analysis was carried out with the PHYLIP package (Felsenstein, 1990).
Sequences were arbitrarily assigned to a cluster when their genetic
distance to any other member of the same cluster was lower than 2%. Ks
and Ka values (where Ks is the frequency of synonymous substitutions
per synonymous site and Ka is the frequency of non-synonymous
substitutions per non-synonymous site) and $K_s/K_a$ ratios were calculated
with the MEGA program (Kumar et al., 1993) by comparison with
prototype strains of both RSV groups.
 Comparisons between sequences were programmed using Stata 6.0
software (Stata Corporation). Nucleotide and amino acid variation was
described in terms of mean, minimum and maximum.

**Results**

A total of 3192 NPA samples was collected, during the 1
year surveillance period (July 1998 to June 1999), from children
under 5 years of age who presented with signs and symptoms
compatible with ALRI (according to WHO criteria). ELISA
results showed that 275 (8.6%) of the samples were RSV-
positive. Virus infection was confirmed by specific RT–PCR
amplification of a 176 bp fragment of the N protein gene of
RSV in 273 of 275 ELISA-positive samples; thus proving the
high sensitivity of the amplification method.

A new and simple RFLP assay of the amplified N protein
gene products (Fig. 1) was developed for typing RSV, as
described in Methods. The assay was validated using a set of
30 previously well characterized samples of both RSV groups
and was further confirmed by sequencing randomly selected
samples of the studied Mozambican outbreak (data not shown).

Results of the RFLP typing showed a higher prevalence of
group B RSV strains than those of group A (85% versus 15%,
respectively). Further sequence analysis of the G protein gene
confirmed the results of the new RFLP assay.

To accurately define the extent of genetic variability within
and between groups, the nucleotide sequence of the C-terminal
end of the G protein gene was determined in randomly
selected samples (10 of group A and 20 of group B) collected
throughout the entire outbreak period. The sequences de-
termined correspond to nt 589 to 918 of the A2 prototype
strain (Wertz et al., 1985) and to nt 554 to 915 of the B8/60
prototype strain (Sullender et al., 1990).

The deduced amino acid (aa) sequence of the Mozambican
group A isolates indicated that all of them shared an expected
G protein gene length of 298 aa due to the presence of a UAG
stop codon (Table 1). The mean nucleotide variation detected
was 4.6% (ranging from 0% to 8.18%) within Mozambican

---

### Table 1. Usage of termination codons by Mozambican RSV isolates

<table>
<thead>
<tr>
<th>RSV isolates</th>
<th>292</th>
<th>293</th>
<th>294</th>
<th>295</th>
<th>296</th>
<th>297</th>
<th>298</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>CCA</td>
<td>CCC</td>
<td>AAC</td>
<td>ACA</td>
<td>CCA</td>
<td>GCC</td>
<td>CAG</td>
</tr>
<tr>
<td>Mon/1/89</td>
<td>CCA</td>
<td>UCC</td>
<td>AAC</td>
<td>ACA</td>
<td>ACA</td>
<td>AAC</td>
<td>UAG</td>
</tr>
<tr>
<td>Mozambican cluster A-I</td>
<td>CCA</td>
<td>UCC</td>
<td>AAC</td>
<td>ACA</td>
<td>ACA</td>
<td>AAC</td>
<td>CAG</td>
</tr>
<tr>
<td>Mozambican cluster A-II</td>
<td>CCA</td>
<td>UCC</td>
<td>AAC</td>
<td>UAA</td>
<td>ACA</td>
<td>AAC</td>
<td>CAG</td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B8/60</td>
<td>AAA</td>
<td>UAA</td>
<td>AAA</td>
<td>ACC</td>
<td>UAG</td>
<td>UCA</td>
<td>GCU</td>
</tr>
<tr>
<td>Mon/7/94</td>
<td>ACC</td>
<td>CAA</td>
<td>AAA</td>
<td>ACC</td>
<td>CAG</td>
<td>UGC</td>
<td>UAG</td>
</tr>
<tr>
<td>Vir/1/80</td>
<td>CCA</td>
<td>CAA</td>
<td>AAA</td>
<td>CCC</td>
<td>UAG</td>
<td>UCA</td>
<td>GCU</td>
</tr>
<tr>
<td>Mozambican cluster B-I</td>
<td>CCA</td>
<td>CAA</td>
<td>AAA</td>
<td>UAG</td>
<td>UCA</td>
<td>GCU</td>
<td>UAG</td>
</tr>
<tr>
<td>Mozambican cluster B-II</td>
<td>CCA</td>
<td>CAA</td>
<td>AAA</td>
<td>CUC</td>
<td>CAG</td>
<td>UCA</td>
<td>GCU</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** RFLP results of RT–PCR amplified N protein gene products after
Hinfl digestion. Lane 1, molecular mass ladder (bp); lanes 2 and 3,
undigested RSV group A and B amplified products of NPA samples from
Mozambique; lanes 4 to 7, Hinfl-digested isolates of group A yielding
fragments of 68, 62 and 46 bp; lanes 8 to 11, Hinfl-digested isolates of
group B yielding fragments of 114 and 62 bp.
**Fig. 2. Deduced amino acid sequences at the C-terminal end of the G protein gene from Mozambican group A (a) and B (b) isolates. Only amino acid changes between isolates are shown. Conserved (●) and non-conserved (○) putative N-glycosylation sites are indicated. Arrows indicate the ...KPX...TKK... motif that may be implicated in O-glycosylation of the native protein (Cane et al., 1991). Conserved motifs are underlined. Amino acid positions are described in Table 1.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Isolate</th>
<th>TTTT</th>
<th>E</th>
<th>I</th>
<th>PI</th>
<th>NS</th>
<th>L</th>
<th>E</th>
<th>TTTT</th>
<th>P</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>A-I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/9/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/24/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/12/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/33/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/40/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/167/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A-II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/27/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/201/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/169/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/170/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz26/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B-I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/47/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/202/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/205/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/211/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/198/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/204/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B-II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/11/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/14/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/25/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/28/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/29/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/41/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/44/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/48/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/267/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/43/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/46/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/50/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moz/177/99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

between Mozambican and non-Mozambican isolates. The mean proportions of nucleotide changes resulting in aa changes within the variable region analysed were 84.5% within Mozambique group A isolates and 74% within non-Mozambican group A isolates. Two clusters of viruses A-I and A-II (with six and four isolates, respectively) with similar aa sequences within each of them were detected among Mozambique group A strains (Fig. 2a). Sequences from these clusters differed by 15 to 19 aa over the sequenced length.

Phylogenetic analysis of group A sequences confirmed the presence of these two clusters (Fig. 3a). The genetic distance of sequences from Mozambican isolates within each cluster were similar to each other (from 0% to 1-92% within either cluster), but their mean genetic distances to some of the non-Mozambican isolates were also alike (for example, a mean of 2.02% between cluster I isolates and strain Mon/1/89, and a mean of 0.80% between cluster II viruses and strain Al-1-96) and lower than those found between both the Mozambican clusters (mean of 9.15%).

Contrary to what was observed among group A isolates, the deduced aa sequence of group B isolates from Mozambique indicated the presence of strains with G proteins of two different expected sizes (295 aa and 299 aa length). In all cases the stop codon was a UAG triplet (Table 1). The mean nucleotide variation observed within the group B isolates from Mozambique was 2.67% (ranging from 0% to 6.27%). The mean nucleotide variation observed between Mozambican and non-Mozambican isolates was 5.50% (ranging from 0.85% to 11.72%) and the mean proportions of nucleotide changes resulting in aa changes were 63.4% and 64%, respectively. Again, two clusters of viruses, B-I and B-II (with 6 and 13 isolates, respectively), with similar aa sequences were detected.
Fig. 3. Phylogenetic trees of group A and group B RSV isolates. Described worldwide sequences retrieved from GenBank (Wertz et al., 1985; Sullender et al., 1990, 1991; Cane et al., 1994; Coggins et al., 1998; Martínez et al., 1999) were included in the trees for comparison. Prototype RSV strains A2 (Wertz et al., 1985) and B8/60 (Sullender et al., 1990) were used as out-group sequences in the trees. The scale bar corresponds to 10% nucleotide sequence divergence. Numbers in italics represent bootstrap proportions in support of the adjacent node based on 100 resampling iterations; only bootstrap values greater than 80% are shown. The only Mozambican group B isolate not ascribed to any cluster is boxed.

among group B Mozambique strains (Fig. 2b). The remaining isolate Moz/26/99, although containing an identical length to that of sequences from cluster II, was not ascribed to any cluster of the Mozambican B group isolates (Fig. 2b). Sequences from these two clusters differed by 7 to 9 aa over the sequenced length. Isolate Moz/26/99 differed from sequences of clusters B-I and B-II by 9 to 10 and 5 to 7 aa, respectively.

Phylogenetic analysis confirmed these results (Fig. 3b). Genetic distances within group B sequences from each cluster varied from 0% to 0.62% and from 0% to 0.86% for cluster I and cluster II, respectively. Cluster II seems to be genetically related to previously reported sequences of very different origin (for example, a mean of 1.75% genetic distance with isolate Mad/1/93). Sequences from cluster I seem to be relatively separate from the rest of the sequences analysed (4.65% genetic distance to the more related Mon/7/94 sequence included in the tree), including Mozambique cluster II (6.44%, ranging from 6.20% to 6.98% between sequences from both clusters). However, our preliminary analysis did not show any particular epidemiological feature of these viruses. Genetic relatedness of Mozambican strain Moz/26/99 was higher with non-Mozambican isolates (0.78% genetic distance to strain Mad/11/92) than with Mozambique isolates (mean of 6.98% and 2.42% to cluster I and II sequences, respectively).

To further characterize the sequence heterogeneity observed among RSV isolates, a more extensive analysis of the type of nucleotide changes was performed. When compared to the A2 prototype strain of RSV, the mean values of synonymous changes for synonymous sites (Ks) and nonsynonymous changes for nonsynonymous sites (Ka) among Mozambican group A sequences were 0.048 and 0.053, respectively, with a Ks/Ka ratio of 0.90. Similarly, when compared to the B8/60 prototype strain, the Ks and Ka values among the Mozambican group B sequences were 0.133 and 0.078, respectively, with a Ks/Ka ratio of 1.7. The mean nucleotide variation between the Mozambique group A and B
isolates was 42.74% (ranging from 40.74% to 44.27%). This nucleotide diversity rises to 63.67% (ranging from 59.14% to 68.09%) aa difference when both groups of Mozambique RSV strains are compared.

Pair-wise comparisons of RSV sequences obtained from Mozambican isolates showed that transitions were more prevalent than transversions (70% versus 30% and 91.2% versus 8.8% among group A and B isolates, respectively). Moreover, AG+GA transitions were the most frequent transitions detected among group A viruses, while CU+UC transitions were the most common among group B isolates (Fig. 4a). Comparison of the sequences from each of the Mozambican clusters with previously described worldwide sequences of the same group showed that the proportion of CU+UC transitions was higher than AG+GA transitions only in sequences from cluster B-I (Fig. 4b). This result reinforces the singularity of this group of sequences.

Four putative N-glycosylation sites, two of them conserved, were observed among the sequenced group A viruses (Fig. 2a). A similar analysis of group B sequences showed only two putative N-glycosylation sites that were conserved in all sequenced strains (Fig. 2b).

Out of the 27 to 32 pair-wise threonine residues detected, seven were conserved among group A isolates from Mozambique (Fig. 2a). Among the 29 to 30 deduced threonine residues detected in group B sequences (Fig. 2b), only five were pair-wise (four of them conserved).

Up to five repetitions of the ...KPx...TTKX... motif, which may be associated with extensive O-glycosylation of the protein, have been previously described between aa 193 and 240 of group A viruses (Cane et al., 1991). Five repetitions of this motif, two conserved and three non-conserved, were found in Mozambican group A isolates (Fig. 2a). Such a motif was not found among the group B isolates (Fig. 2b).

**Discussion**

RSV is a major cause of lower respiratory tract infections in infants. However, little information regarding the prevalence and the genetic and antigenic characteristics of RSV epidemics in developing countries is available (Hazlett et al., 1988; Forgie et al., 1992; Heirholzer et al., 1994; Weber et al., 1998; Cane et al., 1999). RSV was detected in 8-6% of the NPA samples obtained during a recent Mozambican outbreak.

A new and simple RFLP assay, based on a single digestion of RT-PCR amplified products of the N protein gene with Hinfl, allowed a relatively quick and simple typing of RSV NPA samples without the need for tissue cultivation of the viruses. Although group A viruses are predominant in three out of four RSV outbreaks in other geographical regions (Hall et al., 1990), our RFLP data showed a higher prevalence of group B to group A viruses (85% versus 15%) collected during the Mozambican outbreak.

The G protein gene displays the highest capacity for differentiating between RSV strains (Melero et al., 1997). Extensive analysis of RSV G protein variability has been performed with isolates from antigenic group A (Cane & Pringle, 1991, 1995; Cane et al., 1991, 1999; Garcia et al., 1994; Martínez et al., 1998; Melero et al., 1997; Coggins et al., 1998), but little information is available about group B RSV strains, especially regarding the African continent (Sullender et al., 1990, 1991; Coggins et al., 1998; Martínez et al., 1999). These analyses have shown up to 9% and 20% variability in aa sequence between isolates of group B and A, respectively (Wertz et al., 1985; Sullender et al., 1990). Our data confirmed the lower variability, in terms of nt and aa changes, within group B isolates. This different pattern of variability may be related to the higher worldwide predominance of group A outbreaks (Hall et al., 1990; Coggins et al., 1998).

Viruses with G proteins of a given length share sequence changes in the termination codons and in other parts of the protein, suggesting that they originated by unique mutational events (Martínez et al., 1997). While three different lengths (292, 295 and 299 aa) of the G protein gene have been
described among B group viruses, only two (297 and 298 aa) have been found among group A isolates (García et al., 1994). In our study, all Mozambican group A sequences had G protein lengths of 298 aa, while two G protein gene lengths were observed among group B strains (6 isolates of 295 aa and 14 of 299 aa). Changes in stop codon usage have been associated with important antigenic variations found in RSV escape mutants that were selected with monoclonal antibodies recognizing strain-specific epitopes (Melero et al., 1993, 1997). Therefore, assessment of the length of the circulating strains may have important antigenic and immunogenic implications in the characterization of RSV outbreaks.

As expected from previous observations (García et al., 1994), phylogenetic analysis showed that RSV strains are grouped according to the length of their G protein gene (Fig. 3b). Two different lineages (or clusters) of Mozambican strains were detected within each of the groups. However, only sequences from Mozambican cluster B-I seem to be relatively genetically distant from previous isolates found in other regions. Since there is less information about group B sequences than about group A sequences, further analysis of group B isolates of different geographical origin will help to clarify the putative uniqueness of viruses from cluster B-I. In any case, our data confirm that strains of RSV circulating in a single outbreak may be genetically more closely related to strains isolated in very distant places in different years than to co-circulating strains of the same group isolated during the same outbreak in the same geographical region (Melero et al., 1997). Hence, these data reinforce the high capacity of RSV for worldwide spread (Melero et al., 1993; García et al., 1994).

Only sporadic aa changes were observed among sequences from the same cluster and most of them correspond to aa present in strains from the other cluster (Fig. 2). These data suggest that at certain positions, and despite the sequence heterogeneity of RSV, the acceptance of specific aa changes is restricted. This fact could have important implications in the design of broadly reactive synthetic vaccine candidates.

Analysis of the nucleotide changes observed among RSV strains, by comparison with archaic prototypes of each group, showed that the Ks/Ka ratio was almost twofold higher among group B strains than group A strains. This difference is reflected in the higher proportion of nucleotide changes resulting in aa variations found in group A isolates (81% versus 66% in group A and B viruses, respectively). All these data support the idea of a strong selective pressure operating in RSV. In fact, it has been proposed that these changes may favour virus escape from the host immune response and contribute to the capability of RSV to establish infections throughout life (Melero et al., 1993; García et al., 1994).

Therefore, the higher worldwide prevalence of group A viruses may have contributed to the differences observed in the Ks/Ka ratios.

Different patterns of putative N- and O-glycosylation sites, which may play an important role in the mechanisms involved in virus escape from host immune pressure, have been previously described in group A viruses (Cane et al., 1991). Our analysis of Mozambican group A RSV strains showed the presence of four N-glycosylation sites. A similar study in group B sequences has not been carried out before; however, our results indicate the presence of only two putative N-glycosylation sites. In addition, and despite the similar total number of threonines recorded in both groups of viruses, the distribution of glycosylation sites throughout the sequenced region is also quite different between both groups of RSV. Finally, while Mozambican group A viruses display five repetitions of the … KPX … TTKX … motif, this motif was not detected among group B strains. Therefore, and in agreement with the lower sequence heterogeneity of group B isolates, our data suggest that the potential antigenic variability mediated by changes in the putative N- and O-glycosylation patterns of group B isolates is different from that of group A isolates.

Whether these different patterns also have implications in the higher antigenic variability detected among group A isolates remains to be established.

Contrary to what was observed in a group B primary RSV infection, where cross-reactive antibodies were detected in only a minor proportion of children, antibodies recognizing epitopes of the G protein of group B viruses were raised by children with a group A primary RSV infection (Muñelenaer et al., 1991). Based on our data, and even though the three-dimensional structure of the G protein remains to be elucidated, it could be hypothesized that the different pattern and number of putative glycosylation sites found among both groups of viruses might contribute to such previous observations, since extensive glycosylation of group A strains may help to mask epitopes that are also presented in B strains.

Similarly, different features between both groups of RSV are observed when the frequency of transitions and transversions is analysed. As previously described (Martínez et al., 1997, 1999) and in our study, transitions were more frequently detected than transversions. Also, while transitions were twofold higher in group A strains, an extremely high frequency of transitions was observed among group B isolates. It has been described by Martínez et al. (1997, 1999) that, unlike the similar frequencies of transitions found in other RNA virus models (such as measles virus, foot-and-mouth disease virus and influenza virus), AG + GA transitions are three times more frequent than UC + CU transitions. Surprisingly, and in sharp contrast with these results, our analysis indicates that, in Mozambican group B transitions, UC + CU transitions are detected more frequently than AG + GA transitions (Fig. 4a). However, a more detailed study showed that this unusual feature was only observed in sequences from cluster B-I (Fig. 4b).

Different mechanisms for the A to G hypermutation events, and their possible implications on RSV generation of variants during natural evolution, have been previously discussed (Martínez et al., 1997). If, in fact, AG + GA hypermutation events play a role in RSV evolution, our data will suggest that
such mechanisms are not apparently involved in the generation of Mozambican cluster B-I viruses and thus reinforce again the unusual features of this cluster of sequences. In summary, our genetic analysis of a recent RSV outbreak in Mozambique confirms the previously described higher variability found among group A isolates, as well as the great differences shown between group A and B isolates when a detailed comparison of different genetic and antigenic features was performed. These differences may play a relevant role in the varied host immune response against the different RSV groups raised by infected children. In addition, although most of the Mozambique sequences analysed were genetically closely related to previously described worldwide viruses, a cluster of Mozambique group B-specific sequences, genetically distant from previously reported strains and with a different transition pattern, was detected. Our results therefore provide additional insight into the genetic and antigenic structure of both groups of RSV co-circulating in Africa and may be helpful for the implementation of further vaccination campaigns.

This study has been supported by grants from the Global Program for Vaccines, WHO and the Fondo de Investigaciones Sanitarias (FIS 99/0151) Spain. The CISM receives core funding from the Spanish Agency for International Cooperation (AECI). A. Roca is supported by Beca de Formacion e Investigacion (BEFI-9017/99-FIS). We are grateful to the laboratory and clinical and nursing staff of the Centro de Saude da Maniça and Centro de Investigacao em Saude da Maniça. We thank J. A. Melero and I. Martinez for their helpful discussions and critical reading of the manuscript.

References


RSV genetic variability


**WHO (1996).** Diagnosis of an acute lower respiratory tract infection. Vaccine and Research Development, general document 01.

Received 23 June 2000; Accepted 29 September 2000