Replacement and Positive Evolution of Subtype A and B Respiratory Syncytial Virus G-Protein Genotypes From 1997–2012 in South Africa

Marthi A. Pretorius,1,4 Stephanie van Niekerk,4 Stefano Tempia,2 Jocelyn Moyes,1 Cheryl Cohen,1 Shabir A. Madhi,1,3,5 and Marietjie Venter,1,4 for the SARI Surveillance Group

1Centre for Respiratory Disease and Meningitis, National Institute for Communicable Diseases (NICD), National Health Laboratory Service (NHLS), 2US Centers for Diseases Control and Prevention Attache to NICD-NHLS, and 3Medical Research Council: Respiratory and Meningeal Pathogens Research Unit, University of the Witwatersrand, Johannesburg, 4Zoonosis Research Unit, Department Medical Virology, University of Pretoria, Pretoria, and 5Department of Science and Technology/National Research Foundation: Vaccine-Preventable Diseases, Johannesburg, South Africa

Background. Of the respiratory syncytial virus (RSV) genotypes previously described in South Africa during 1997–2002, only GA2 and GA5 persisted until 2006, with BA having replaced all previous RSV-B genotypes. This poses the question whether RSV-A is more stable than RSV-B and whether positive selection drives evolution of genotypes.

Methods. RSV-positive specimens were randomly selected during 2009–2012, subtyped, sequenced, and compared to RSV recovered from specimens obtained during 1997–2001 and 2006–2009. Bayesian phylogenetic analysis was performed on the G-protein.

Results. Phylogenetic analysis indicated that RSV-A genotype GA2 dissolved to form SAA2 (unique to South Africa), NA1 and NA2 (identified in Japan), and ON1 (identified in Canada and having a 72-bp insertion) and that GA5 drifted from 1999–2012 to form 3 subgenotypes (GA5 I–III). RSV-B genotypes all had the 60-bp insertion typical of BA genotypes but clustered into subgenotypes BA8-10. Positive selection was identified in the G-protein of both subtypes, but RSV-A’s rate of evolution was slower than that of RSV-B, with the most recent common ancestors dating from 1945 and 1951, respectively. Seven new positively selected sites were identified in South African strains, 2 for RSV-A and 5 for RSV-B.

Conclusion. Positive selection drove both RSV-A and -B genotypes to evolve, resulting in replacement of all genotypes over the 15-year study period in South Africa.

Keywords. respiratory syncytial virus; positive selection; evolutionary rate; genotypes.
the BA genotype, which has a 60-bp insertion [13] and has subsequently been classified into subgenotypes BA1-10 [14, 15].

In previous studies conducted in South Africa during 1997–2001, we reported the circulation of GA2, GA5, GA7, GB3, and GA4, as well as novel genotypes such as SAA1, SAB1, SAB2, and SAB3 [7, 12, 16]. These novel genotypes have subsequently been identified worldwide [10, 17]. The BA genotype was first detected in South Africa during the investigation of a nosocomial outbreak, in 2006 [18], and we subsequently showed that BA had replaced all previously described RSV-B genotypes in South Africa [6]. Also, until 2009, GA2 and GA5 were the only subtype A genotypes still circulating in South Africa, suggesting that these RSV-A genotypes have been stable over the past 9 years. However, poor bootstrap and pairwise distance analysis suggested that drift may be occurring. The objective of this study was to compare the genotypic characteristics of RSV identified through a surveillance program in South Africa, conducted during 2009–2012 to investigate the prevalence of respiratory viruses among individuals hospitalized for severe acute respiratory infections (SARI), and to compare these to the genotypes of RSV identified in South Africa during 1997–2001 and 2006–2009, as well as to RSV identified elsewhere globally.

**MATERIALS AND METHODS**

**Study Design and Sample Collection**

**SARI Surveillance Program (2009–2012)**

Specimens were collected as part of a prospective hospital-based sentinel surveillance program conducted from February 2009 through May 2012, which aimed to determine respiratory viruses and risk factors associated with hospitalization for SARI in South African children of all age groups. The methods of this study have been previously described [19]. The participating SARI surveillance sites are Chris Hani Baragwanath Academic Hospital (CHBAH; Gauteng Province), Edendale Hospital (KwaZulu-Natal Province), Matikwana and Mapulaneng hospitals (Mpumalanga Province), and Tshepong and Klerksdorp hospitals (North-West Province). RSV-positive specimens from individuals of all age groups with a known human immunodeficiency virus infection status, and RSV from a percentage of positive specimens were selected at random retrospectively from the 4-year period for sequencing. The genotypic profile of RSV obtained from the SARI program during 2009–2012 was compared to that of RSV previously identified in South Africa.

**RSV Previously Identified in South Africa**

Samples were obtained from children aged <5 years from CHBAH (Soweto) during 1997–2000 [7] and from children <5 years from community clinics in Agincourt during 1998–2001 [12]. Additional samples were obtained from high-risk children from across the country during 2000–2001 [16] and primarily from hospitalized patients, mostly children, in Steve Biko and Kalafong academic hospitals (Pretoria, South Africa) during 2006–2009 [6].

**Detection of Respiratory Viruses**

All SARI specimens were tested using a real-time reverse-transcription polymerase chain reaction (rRT-PCR) protocol described by Pretorius et al [19] and were screened for parainfluenza types 1–3, RSV, influenza A and B viruses, adenovirus, enterovirus, human metapneumovirus, and rhinovirus.

**Full RSV G-protein Gene Amplification**

The complete or near-complete RSV G genes were amplified by nested PCR. Briefly, complementary DNA (cDNA) was synthesized with random primers using SuperScript reverse transcriptase (Invitrogen, Life Technologies, Carlsbad, CA, USA) according to manufacturer’s instructions. Full-length RSV G-protein genes were amplified using Platinum Pfx DNA polymerase (Invitrogen, Life Technologies) with primers G1-21 [20] and F164 [21]. In brief, how to set up the full g-protein amplification. 5 µL of cDNA was added to 7.5 µL of 10× reaction buffer, 10 mM each deoxynucleoside triphosphate (dNTP), 10 pmol each primer (G1-21 and F164), 50 mM MgSO4 1.25 U of Platinum Pfx DNA polymerase up to a 50-µl final volume. The following cycling conditions were used: 1 cycle at 94°C for 2 minutes; 40 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 2 minutes; and 1 cycle at 68°C for 10 minutes. This was followed by nested PCR with forward primers G32A(RSV-A), G32B(RSV-B), G598A(RSV-A), and G604B(RSV-B) [22] and reverse primers G665R [22] and F1 [5]. The nested PCR was conducted in a 50-µL reaction mixture volume. In brief, 1 µL of the first round product was added to 7.5 µL 10× reaction buffer, 10 mM each dNTP, 10 pmol each primer, 50 mM MgSO4 1.25 U of Platinum Pfx DNA polymerase, using the same cycling conditions as the first round. PCR products were analyzed on a 1% agarose gel against a 100-bp ladder as a molecular weight marker (DNA molecular marker XIV; Roche Diagnostics, Mannheim, Germany).

**Subtyping of RSV**

Subtyping-specific primers were used to determine the subtype as either RSV-A or RSV-B, using the amplifed full g-protein as described before on the full-length G-protein gene by nested PCR with forward primers G52B and G598A [22] and reverse primer F1 [5], or directly by sequencing using a combination of the primers listed above (Supplementary Table 1).

**Nucleotide Sequencing**

PCR products were purified using the ExoSAP-IT enzyme system (USB, Cleveland, OH). Cycle sequencing was performed with a BigDye terminator 3.1 cycle sequencing kit as recommended by the manufacturer (Applied Biosystems, Foster City, CA). Nucleotide sequencing was performed on both strands, and the editing
was performed with Sequencher, version 5 (Gene Codes, Ann Arbor, MI, USA).

**Phylogenetic Analysis**

A region spanning 270 nucleotides in RSV-A and 330 nucleotides in RSV-B, representing the second hypervariable region of the G-protein gene, was used for phylogenetic analysis [23]. The South African RSV sequences were compared against reference sequences of each genotype in the GenBank database (available at: [http://www.ncbi.nlm.nih.gov/GenBank/index.html](http://www.ncbi.nlm.nih.gov/GenBank/index.html)). All South African strains present in the database from our previous investigations in Pretoria (2006–2009) [6, 18], Soweto (1997–2001) [11], and Agincourt, Pretoria, Bloemfontein, Cape Town, and Durban (2000) [7, 12, 22] were also included. Nucleotide sequences of RSV-A and RSV-B viruses were aligned separately with ClustalX 1.81 [24] embedded in BioEdit Sequence Alignment Editor, version 7.0.4.1 [25]. Because most previously published sequences were based only on the second variable region, phylogenetic analysis was based on this region for all strains.

**Nucleotide and Amino Acid Sequence Analysis**

The pairwise distance (p-distance) was calculated between individual genotypes, as well as within each genotype, using Mega, version 5 [26]. Estimates of evolution were based on the ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks). These estimates were calculated with the modified Nei-Gojobori method in Mega, version 5 [27]; a ratio of Ka to Ks of > 1 indicates positive selection. Amino acid analysis for individual genotypes were calculated for the full G-protein ectodomain of representative strains of each genotype, using Genedoc, version 2.6.003 (available at: [http://www.nrbsc.org/gfx/genedoc/](http://www.nrbsc.org/gfx/genedoc/)).

**Evolutionary Rate and Most Recent Common Ancestor (MRCA)**

The evolutionary analyses over time were conducted using a coalescent Bayesian Markov chain Monte Carlo (MCMC) approach, as implemented in BEAST (Bayesian Evolutionary Analysis by Sampling Trees), version 1.4.6 [23], to estimate the rate of evolution. MCMC chains were run for a sufficient time to achieve convergence (as assessed using Tracer). Dates were introduced according to the year of isolation. Maximum clade credibility trees were generated using Tree Annotator, and all trees were plotted using FigTree, version 1.1.2 (available at: [http://tree.bio.ed.ac.uk/](http://tree.bio.ed.ac.uk/)). Using the root-to-tip regression analysis that displays the correlation between phylogenetic branch length and the time of sampling of the viral strains to calculate the MRCA for all of the sequences under analysis (Path-o-Gen version 1.3 available at: [http://tree.bio.ed.ac.uk/software/pathogen](http://tree.bio.ed.ac.uk/software/pathogen)); plot as performed with Path-o-Gen version 1.3 (available at: [http://tree.bio.ed.ac.uk/software/pathogen](http://tree.bio.ed.ac.uk/software/pathogen)); this displays the correlation between phylogenetic branch length and the time of sampling of the viral strains to calculate the MRCA for all of the sequences under analysis.

**Selective Pressure Analysis**

Selective pressure analysis was conducted by use of the Datamonkey Web server (available at: [http://www.datamonkey.org](http://www.datamonkey.org)), which estimated the rates of nonsynonymous and synonymous changes at each site in a sequence alignment, to identify sites under positive or negative selection [28–30].

**Statistical Analysis**

The frequency of RSV subtypes was calculated with the Fisher exact test, using Stata 10 (College Station, TX). P values of <05 were considered statistically significant.

**Nucleotide Sequence Accession Numbers**

Sequences of partial and full G-protein genes were submitted to GenBank (accession numbers KC476656–KC477097).

**Ethical Considerations**

The protocol was reviewed and approved by the University of the Witwatersrand Human Research Ethics Committee and the University of KwaZulu-Natal Human Biomedical Research Ethics Committee (protocols M081042 and BF157/08, respectively).

**RESULTS**

**Detection of Respiratory Viruses and Characteristics of RSV-Positive Patients Identified in the SARI Surveillance Program During 2009–2012**

During February 2009–May 2012, we enrolled 13 664 patients meeting the SARI case definition, of whom 13 432 (98%) were tested for respiratory viruses. RSV was detected in 2182 individuals (16%), including 1111 (51%) in whom RSV was the sole pathogen identified. The proportion testing positive for RSV varied by age group: 31% (1387/4542) among individuals aged 0–2 years, 20% (536/2631) among those aged 2–4 years, 7% (35/479) among those aged 5–14 years, 3% (17/525) among those aged 15–24 years, 4% (140/3295) among those aged 25–44 years, 4% (59/1568) among those aged 45–64 years, and 2% (7/373) among those aged >65 years (P < .001).

**Evolution of RSV Subtypes**

Among the 486 randomly selected RSV-positive specimens recovered from the SARI program between 2009 and 2012 that were further investigated, RSV-A was detected in 282 (58%), and RSV-B was detected in 204 (42%; P = .001). RSV-A predominated in 2 of the 4 seasons (Table 1). Data on subgroup switching and replacement between genotypes during 1997–2012 are indicated in Figure 1. These findings indicated that all genotypes initially identified in South Africa during 1997 had been replaced by 2012 for both subtypes A and B.
Phylogenetic Comparison of Strains Identified in 2006–2012 Relative to Earlier Periods

**Diversity**

Maximum likelihood (Figure 2) and Bayesian (Figure 3) phylogenetic comparison of 655 RSV strains identified in clinical specimens during 2006–2012 to strains identified in earlier studies (during 1997–2002) in South Africa, as well as to international reference sequences, indicated that 343 strains (52.4%) detected since 2006 clustered with strains that were formerly assigned to the GA2 genotype. However, none clustered with statistically significant bootstrap values to the older strains, while 67 (10.2%) clustered with GA5 (Figure 1). Several clusters seemed to have evolved from the GA2 genotype and grouped together with statistically significant bootstrap values to form new genotypes (Figure 1A). Some of these clusters were similar to genotypes NA1 (286 [43.7%]) and NA2 (20 [3.1%]), which were previously identified in Japan, and to genotype ON1 (3 [0.5%]), which was previously identified in Canada, with bootstrap support of 72, 82, and 98 respectively; 1 cluster was unique to South Africa and putatively named SAA2 (12 [1.8%]; bootstrap support of 83; Figure 2A). Although strong bootstrap values could be obtained for all strains clustering with GA5, 3 GA5 subclusters were identified since its emergence 19 years ago and were named GA5I–III (Figures 2A and 3A). All of the

---

**Table 1.** Frequencies of Respiratory Syncytial Virus Subtypes A and B in South Africa, 1997–2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>31 (62)</td>
<td>30 (42.9)</td>
<td>41 (74.5)</td>
<td>51 (87.9)</td>
<td>48 (63.2)</td>
<td>69 (81.2)</td>
<td>29 (90.6)</td>
<td>8 (7.8)</td>
<td>127 (84.1)</td>
<td>71 (42.8)</td>
<td>75 (82.4)</td>
</tr>
<tr>
<td>B</td>
<td>19 (38)</td>
<td>40 (57.1)</td>
<td>14 (25.5)</td>
<td>7 (12.1)</td>
<td>28 (36.8)</td>
<td>16 (18.8)</td>
<td>3 (9.4)</td>
<td>95 (92.2)</td>
<td>24 (15.9)</td>
<td>95 (57.2)</td>
<td>16 (17.6)</td>
</tr>
<tr>
<td>Total</td>
<td>50 (100)</td>
<td>70 (100)</td>
<td>55 (100)</td>
<td>58 (100)</td>
<td>76 (100)</td>
<td>85 (100)</td>
<td>32 (100)</td>
<td>103 (100)</td>
<td>151 (100)</td>
<td>166 (100)</td>
<td>91 (100)</td>
</tr>
</tbody>
</table>

a Samples were obtained primarily from children aged <5 years from Chris Hani Baragwanath Academic Hospital during 1997–2000 [15].

b Samples were obtained from children aged <5 years from community clinics in Agincourt during 2000–2001 [17], as well as from high-risk children from across the country [21].

c Samples were obtained from hospitalized children at Steve Biko and Kalafong academic hospitals (Pretoria, South Africa) during 2006–2009 [5, 10].

d Samples were obtained from the SARI surveillance program during 2009–2012 (in 4 provinces).

---

**Figure 1.** Frequencies of respiratory syncytial virus genotypes in South Africa during 1997–2012.
Figure 2. Maximum likelihood tree analysis. Midpoint-rooted maximum-likelihood trees for subtype A (A) and subtype B (B) constructed under the HKY codon position substitution model, using Mega, version 5 [51]. The image is drawn to scale, with the bars indicating 0.02 nucleotide substitutions. Estimates were based on bootstrap resampling performed with 1000 replicates. Only bootstrap values of ≥70 are shown. The virus designations refer to the molecular evolution of RSV in South Africa.
RSV-B isolates recovered since 2006 clustered within the BA genotype and had the characteristic 60-nucleotide duplication (Figure 2B); all previously identified RSV-B genotypes in South Africa were replaced by the BA genotype. Analysis showed that of 234 RSV-B strains selected since 2006, 132 (20.2%) clustered with BA9, 98 (14.9%) clustered with BA10, and 4 (0.6%) clustered with BA8 (Figure 1).

**Pairwise Distances**
SAA2 strains were distinct from strains belonging to genotypes GA2, NA1, and NA2, (average p-distances, 4.0%, 6.2%, and 6.1%, respectively). GA5 also displayed the largest intragenotypic range of p-distances (0%–12%). The intragenotypic p-distances for the different genotypes are shown on the diagonal of Table 2. The average intergenotypic p-distance between the different genotypes within RSV-A ranged from 4% to 18%, including a 13.2% difference between the 2 dominant genotypes, NA1 and GA5. Among the 378 RSV-A isolates, 5 termination codons were used: 20 (5.3%) used the UAG stop codon, 298 (78.8%) used UGA, and 2 (0.5%) used UAA at position 894, whereas 55 (14.6%) used stop codon UAG at position 895, and 3 (0.79%) used UGA in position 964.

For the RSV-B strains, the intragenotypic p-distances for the different genotypes detected are shown on the diagonal of Table 3. The average intergenotypic p-distance between the different genotypes within RSV-A ranged from 2% to 18%, with a 4.3% difference between the 2 dominant genotypes, BA9 and BA10. Among the 224 RSV-B isolates, 2 termination codons were used: 203 (90.6%) used UAA at position 931, and 21 (9.4%) used UAG at position 952. Figure 4 indicates the amino acid alignments of the full G-proteins of genotypes identified in South Africa during 1997–2012 relative to data for the reference strains.

**Analysis of Selective Pressure**

Full RSV G-protein bioinformatics calculations were based on data for the complete RSV G-proteins of strains sequenced from the SARI samples and for representative full RSV G-protein sequences for each of the genotypes previously identified in South Africa and elsewhere that were available on GenBank. The genetic changes observed because of nucleotide substitution were as follows: 196 substitutions in RSV-A strains, with 75 synonymous and 120 nonsynonymous; and 203 substitutions in RSV-B strains, with 94 synonymous and 109 nonsynonymous. The Nei-Gojobori method identified positive selection within both RSV-A and RSV-B. Genotypes SAA1, GA6, GA3, SAB1, GB4, and BA5 had Ka/Ks ratios of 1.304, 1.177, 1.051, 1.07, 1.26, and 1.42, respectively (a ratio of >1 indicates positive selection). Investigation of Ka/Ks ratios between genotypes that emerged from the GA2 genotype indicated intergenotypic positive selection between NA1, SAA2, and ONI. Intergenotypic positive selection was also identified within the BA genotype between subgenotypes BA1–BA5 and BA7 [28, 30, 31]. Calculations for positive selective sites identified 16 sites in the G-glycoprotein gene that had a posterior probability of >0.5, which indicates positive selection in both RSV-A and RSV-B [28–30]. For RSV-A, 2 positively selected sites were located in the first hypervariable region (amino acid positions 94 and 136), and 6 positively selected sites were located in the second hypervariable region (amino acid positions 233, 237, 250, 262, 274, and 290). Four amino acids (positions 237, 262, 274, and 290) had posterior probabilities for positive selection of >90%; 1 (position 233) had a posterior probability of >80%, 2 (positions 136 and 250) had a posterior probability of >70%, and 1 (position 94) had a posterior probability of >60%. For RSV-B, 1 positively selected site was located in the first hypervariable region (amino acid position 103), and 7 positively selected sites were located in the second hypervariable region (amino acid positions 209, 245, 248, 251, 268, 269, and 285). Four amino acids (positions 251, 268, 269, and 285) were identified to be under positive selection at >90%; 1 (position 248) had a posterior probability of >80%, 2 (positions 103 and 209) had a posterior probability of >70%, and 1 (position 245) had a posterior probability of >50%.

**Evolutionary Rates**
The evolutionary rate was estimated as 4.68 × 10⁻³ substitutions/site per year (95% highest posterior density [HPD] interval, 3.5499 × 10⁻³–5.8881 × 10⁻³) substitutions/site/year for RSV-A and 5.8907 × 10⁻³ substitutions/site per year (95% HPD interval, 14.2816 × 10⁻³–7.6298 × 10⁻³) for RSV-B. A root-to-tip transition plot estimation indicated the MRCA to date back to 1945 for RSV-A and to 1957 for RSV-B.

**DISCUSSION**
The implementation of SARI surveillance in 4 South African provinces with distinct geographic and climatic conditions provided an opportunity to analyze the molecular epidemiology of RSV at a national level, expanding on previous work undertaken in South Africa [7, 16, 21]. Phylogenetic analysis in this study confirmed previous observations that similar genotypes dominated in...
Figure 3. Bayesian analysis. A midpoint-rooted BEAST (Bayesian Evolutionary Analysis by Sampling Trees) phylogenetic tree of respiratory syncytial virus (RSV)–A (A) and RSV-B (B) genotypes identified worldwide since 1960, together with all of the South African genotypes identified during this study (2006–2012) and other South African sequences from 1997–2000. The lengths of the horizontal lines are proportionate to the genetic distance between viruses. The bar represents 7 substitutions per site. The virus designations refer to the place of isolation (SA, South Africa), isolate number, and year of isolation. South African strains were compared to strains from different continents. The sources of the reference sequences obtained from GenBank were as follows: NY, New York; AL, Alabama; MO, Missouri; TX, Texas; CH, Rochester; WI, Wisconsin; ON, Ontario, Canada; Que, Quebec, Canada; MON, Montevideo, Uruguay; MAD, Madrid, Spain; NG, Nagasaki, Japan; BE, Belgium; BA, Buenos Aires, Argentina; and Cam, Cambodia. The prototype strain for subtype A is strain A2 (A2AUS61, Australia).
Table 2. Average Percentage Nucleotide p-distance Between and Within Respiratory Syncytial Virus A South African Strains Used in Figure 2.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>GA1</th>
<th>GA2</th>
<th>SAA1</th>
<th>GA5</th>
<th>GA6</th>
<th>GA7</th>
<th>GA3</th>
<th>NA1</th>
<th>NA2</th>
<th>ON1</th>
<th>SAA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA1</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA2</td>
<td>12.6</td>
<td>2.7</td>
<td>(0-6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAA1</td>
<td>14.3</td>
<td>7.6</td>
<td>2.7</td>
<td>(0.4-7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA5</td>
<td>17.2</td>
<td>11.4</td>
<td>10.8</td>
<td>5.4</td>
<td>(0-12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA6</td>
<td>12.9</td>
<td>6.0</td>
<td>5.9</td>
<td>9.9</td>
<td>1.5</td>
<td>(0-2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA7</td>
<td>18.6</td>
<td>13.5</td>
<td>12.5</td>
<td>15.9</td>
<td>12.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA3</td>
<td>14.2</td>
<td>5.8</td>
<td>7.8</td>
<td>11.6</td>
<td>5.4</td>
<td>12.9</td>
<td>3.5</td>
<td>(0-6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA1</td>
<td>14.7</td>
<td>4.9</td>
<td>10.2</td>
<td>13.2</td>
<td>8.6</td>
<td>15.6</td>
<td>10.5</td>
<td>9.2</td>
<td>2.6</td>
<td>(0-7)</td>
<td></td>
</tr>
<tr>
<td>NA2</td>
<td>14.3</td>
<td>5.0</td>
<td>10.3</td>
<td>13.8</td>
<td>8.7</td>
<td>15.0</td>
<td>10.4</td>
<td>8.3</td>
<td>5.8</td>
<td>3.0</td>
<td>(0-10)</td>
</tr>
<tr>
<td>ON1</td>
<td>17.7</td>
<td>6.2</td>
<td>12.6</td>
<td>16.1</td>
<td>10.6</td>
<td>18.1</td>
<td>13.1</td>
<td>10.4</td>
<td>4.6</td>
<td>7.6</td>
<td>0.2</td>
</tr>
<tr>
<td>SAA2</td>
<td>13.3</td>
<td>4.0</td>
<td>9.3</td>
<td>12.5</td>
<td>7.8</td>
<td>15.1</td>
<td>9.0</td>
<td>7.4</td>
<td>6.2</td>
<td>6.1</td>
<td>7.5</td>
</tr>
</tbody>
</table>

p-distances calculated between the individual genotypes are shown in bold below the diagonal in each column of the table, and the range of p-distances within each genotype are in parentheses. Abbreviation: n/c: could not calculate the p-distance range within the genotype.

Table 3. Average Percentage Nucleotide p-distance Between and Within Respiratory Syncytial Virus B South African Strains Used in Figure 2.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>GB1</th>
<th>GB2</th>
<th>GB3</th>
<th>GB4</th>
<th>SAB1</th>
<th>SAB2</th>
<th>SAB3</th>
<th>BA1</th>
<th>BA2</th>
<th>BA3</th>
<th>BA4</th>
<th>BA5</th>
<th>BA6</th>
<th>BA7</th>
<th>BA8</th>
<th>BA9</th>
<th>BA10</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB1</td>
<td>2.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB2</td>
<td>8.10</td>
<td>n/c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB3</td>
<td>11.47</td>
<td>6.30</td>
<td>n/c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB4</td>
<td>15.06</td>
<td>8.42</td>
<td>11.47</td>
<td>3.16</td>
<td>(0.3-5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAB1</td>
<td>10.72</td>
<td>5.79</td>
<td>7.75</td>
<td>11.50</td>
<td>5.27</td>
<td>(n/c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAB2</td>
<td>14.47</td>
<td>9.02</td>
<td>7.94</td>
<td>11.77</td>
<td>5.33</td>
<td>(n/c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAB3</td>
<td>11.80</td>
<td>5.90</td>
<td>3.96</td>
<td>10.13</td>
<td>7.80</td>
<td>6.66</td>
<td>0.53</td>
<td>(n/c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA1</td>
<td>10.86</td>
<td>5.14</td>
<td>3.03</td>
<td>10.50</td>
<td>7.12</td>
<td>6.78</td>
<td>2.78</td>
<td>0.66</td>
<td>(0-1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA2</td>
<td>11.51</td>
<td>4.74</td>
<td>4.73</td>
<td>11.42</td>
<td>7.78</td>
<td>8.50</td>
<td>4.50</td>
<td>2.54</td>
<td>0.59</td>
<td>(0.6-1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA3</td>
<td>12.26</td>
<td>6.04</td>
<td>4.98</td>
<td>10.68</td>
<td>9.07</td>
<td>8.74</td>
<td>4.75</td>
<td>2.75</td>
<td>3.74</td>
<td>0.44</td>
<td>(n/c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA4</td>
<td>12.80</td>
<td>5.98</td>
<td>4.91</td>
<td>12.45</td>
<td>8.45</td>
<td>8.71</td>
<td>4.71</td>
<td>2.72</td>
<td>3.20</td>
<td>3.04</td>
<td>1.85</td>
<td>(0.3-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA5</td>
<td>11.82</td>
<td>5.07</td>
<td>4.13</td>
<td>11.50</td>
<td>8.10</td>
<td>7.88</td>
<td>3.89</td>
<td>2.09</td>
<td>3.11</td>
<td>2.41</td>
<td>2.38</td>
<td>0.84</td>
<td>(0-2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA6</td>
<td>10.47</td>
<td>4.75</td>
<td>3.68</td>
<td>11.16</td>
<td>7.78</td>
<td>7.44</td>
<td>3.44</td>
<td>1.65</td>
<td>2.65</td>
<td>2.86</td>
<td>2.80</td>
<td>2.20</td>
<td>n/c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA7</td>
<td>11.86</td>
<td>6.15</td>
<td>5.09</td>
<td>12.57</td>
<td>9.19</td>
<td>8.86</td>
<td>4.85</td>
<td>3.20</td>
<td>3.60</td>
<td>3.52</td>
<td>2.72</td>
<td>2.86</td>
<td>3.31</td>
<td>1.50</td>
<td>(0.3-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA8</td>
<td>11.75</td>
<td>5.92</td>
<td>5.33</td>
<td>12.51</td>
<td>9.02</td>
<td>8.98</td>
<td>5.09</td>
<td>3.03</td>
<td>3.15</td>
<td>3.35</td>
<td>2.69</td>
<td>2.70</td>
<td>3.14</td>
<td>2.33</td>
<td>1.81</td>
<td>(0.3-5)</td>
<td></td>
</tr>
<tr>
<td>BA9</td>
<td>11.45</td>
<td>5.85</td>
<td>4.80</td>
<td>12.15</td>
<td>8.46</td>
<td>8.58</td>
<td>4.58</td>
<td>3.34</td>
<td>3.76</td>
<td>3.83</td>
<td>3.03</td>
<td>3.16</td>
<td>3.35</td>
<td>2.54</td>
<td>2.65</td>
<td>2.06</td>
<td>(0-7)</td>
</tr>
<tr>
<td>BA10</td>
<td>12.39</td>
<td>6.67</td>
<td>5.61</td>
<td>13.05</td>
<td>9.72</td>
<td>9.32</td>
<td>5.37</td>
<td>3.47</td>
<td>3.71</td>
<td>3.94</td>
<td>3.26</td>
<td>3.26</td>
<td>3.42</td>
<td>2.70</td>
<td>2.84</td>
<td>3.07</td>
<td>2.23</td>
</tr>
</tbody>
</table>

p-distances calculated between the individual genotypes are shown in bold below the diagonal in each column of the table, and the range of p-distances within each genotype are in parentheses. Abbreviation: n/c: could not calculate the p-distance range within the genotype.
different locations in South Africa and globally during 1 season [7–10, 12]. We also identified strains that were identical to a new genotype (ON1, identified in Canada in 2011 with a 72-bp insertion) not previously seen in South Africa, confirming the global spread of RSV strains [32]. The SARI surveillance data confirmed that although the RSV season peaked between February and May each year, ongoing low-level circulation outside of the epidemic season was detected and may contribute to the persistence of

Figure 4. The amino acid alignment of respiratory syncytial virus (RSV)-A (A) and RSV-B (B) indicates the genotypes identified in South Africa from 1997–2012 relative to those of the reference strains.
strains between seasons. Although drift within genotypes was also visible over the seasons, strains from later seasons were further from the root of each genotype across the tree. Figure 4 illustrates genotype distribution over 15 years and indicates that different genotypes cocirculate within a season, with certain genotypes dominating and then declining before being replaced with a different genotype; all genotypes originally described for both subtype A and B were replaced by 2012 [7, 10, 33]. This suggests reestablishment of annual epidemics and reinfection throughout life, indicating that vaccines targeting the G-protein would need to be evaluated for cross-protection against recent RSV strains.

The GA2 genotype was first detected in 1998 and was continuously detected up to 2009 [6]. In 2007, new clusters emerged closely related to GA2 but with insignificant bootstrap values and p-distances of >4%. One of these clusters was unique to South Africa, and we named it SAA2, to follow on the previous genotype identified here in 1997, SAA1. The other clusters resembled genotypes identified in Japan before identification of NA1, the dominant genotype from 2010–2012. The ON1 genotype identified in 2011 was the first instance in which a duplication insertion was described for RSV-A [32]. It will be interesting to follow this genotype over the next seasons to see whether it will have the same impact on RSV-A evolution as the BA genotype had for RSV-B, and whether the 72-nucleotide duplication will provide an evolutionary advantage to the virus. Drift also occurred in the GA5 genotype, with p-distance analysis suggesting that 3 subclusters, GA5I–III, could be identified.

Investigation of RSV-B evolution over the 15 years showed that the BA genotype not only had replaced all the previously identified RSV-B genotypes in South Africa [6], but is also undergoing positive selection and evolving into new genotypes. The BA genotype has become dominant since it was first described in 1999 across most of the world [6, 7, 11, 18, 34–36]. In the early investigation of the natural history of the BA genotype, it was divided into 6 subgenotypes, BA1–BA6 [20]. Further genetic drift since its emergence 13 years ago resulted in 4 new subgenotypes, BA7–BA10 [15]. The South African RSV-B strains identified since 2006 belong to 2 main genotypes, BA9 and BA10, and none of the South Africa–specific RSV-A and RSV-B genotypes identified in earlier studies (SAA1, SAA2, and SAB1–3) were identified to be circulating during 2009–2012 [6]. This may be due to community-wide immunological pressure induced by previous exposure.

We identified 7 novel positively selected sites in RSV-A (2 sites) and RSV-B (5 sites) located mainly in the second hyper-variable region for the G-protein gene, confirming the value of this region in phylogenetic analysis [8, 9, 37]. This suggests that evolution of both RSV-A and RSV-B is driven by the selection of new variants because of positive selective pressure at certain codon positions. Unlike other respiratory viruses, such as influenza virus, it would be impossible to make predictions about future evolutionary changes to the G-protein, since the mechanism that drives changes in the neutralization site is not completely understood; however, immune pressure seems to be a likely explanation because positive selection has been identified, although random changes may also occur, probably because of mistakes introduced by the polymerase, and retained mutations could have an evolutionary advantage for the virus. Use of alternative stop codons and tolerance to large insertion mutations contribute to the higher evolutionary rate of RSV-B [8–10]. Emergence of a subgroup A strain with a 72-bp insertion could also provide evolutionary advantage to subgroup A.

Our study is subject to a number of limitations. We have described data from several different studies, and differences in study designs and populations (in terms of age and severity of disease) made it difficult to assess the potential association of different genotypes with disease severity among the different studies. There is also a 4-year gap (2002–2005) in the RSV molecular data, preventing a report on the molecular changes in RSV during this period.

We have been investigating RSV genetic and antigenic variation in patients over the past 15 years. Data from these studies reflect findings from other parts of the world but also identified differences unique to our population and may provide some insight about the molecular evolution of RSV in the rest of Africa. We identified several genotypes for the first time in South Africa that have been identified across the world, suggesting that the emergence of new RSV strains in distant areas could contribute to RSV strains circulating globally and to the recurrence of annual epidemics.

**Supplementary Data**

Supplementary materials are available at The Journal of Infectious Diseases online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

**Notes**

**Acknowledgments.** We thank the following Units and individuals at the National Institute for Communicable Diseases a division of the National Health Laboratory Service: Shabir Madhi (Executive Director NICD), Amelia Buys, Cardia Fourie, Xolisa Stuurman, Orianka Helftsersee, Jocelyn Moyes, Florette K Treurnicht, Sibongile Walaza (Centre for Respiratory Diseases and Meningitis), Andrew Black (Chris Han BARagwanath Hospital), Michelle Groome, David Moore, Suzan Nzenze (Department of Science and Technology/National Research Foundation: Vaccine Preventable Diseases Unit), Meera Chhagan, Halima Dawood, Sumayya Haifejee, Fathima Naby (Edendale Hospital), Erna du Plessis, Omphile Mekgoe, Ebrahim Variava (Klerksdorp/Tshepong Hospital Complex), Kathleen Kahn, Stephen Tollman, Rhian Twine (MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt)), Adam Cohen (US Centers for Disease Control and Prevention), and the whole SARI surveillance team.

**Financial support.** This work was supported by the Centers for Disease Control and Prevention (cooperative agreement SU51/IP000155).

**Potential conflicts of interest.** All authors: No reported conflicts.
All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


