Replacement of previously circulating Respiratory Syncytial Virus (RSV) subtype B strains with the BA genotype in South Africa

Stephanie van Niekerk[1] and Marietjie Venter[1,2,*]

Author affiliations:

[1] Department Medical Virology, University of Pretoria, Pretoria, Gauteng, South Africa;

[2] Respiratory Virus Unit, National Institute for Communicable Disease, Johannesburg, Gauteng, South Africa

*Corresponding author:

Prof Marietjie Venter (BSc, BSc(Hons), MSc, PhD)

Associate Professor/Head of Respiratory Virus Unit

National Institute for Communicable Diseases

Private Bag X 4

Modderfonteinroad

Sandringham 2131

South Africa

Tel: +27 12 3192660

Fax: +27 12 325 5550

E-mail: marietjiev@nicd.ac.za

Institute where the work was done: Department of Medical Virology, University of Pretoria, Pretoria, South Africa

S. van Niekerk (BSc, BSc (Hons): Department of Medical Virology, University of Pretoria, Pretoria, South Africa.

M. Venter (BSc, BSc (Hons), MSc, PhD): Department of Medical Virology, University of Pretoria, National Institute for Communicable Disease, National Health Laboratory Services, Sandringham, South Africa.

Word count abstract: 250

Word count text: 3,216

Accession numbers: HQ711628-HQ711839 (Partial G protein sequences)

JF704219-JF704234 (Full G protein sequences)

Abstract

Respiratory Syncytial Virus (RSV) is a major cause of bronchiolitis and pneumonia in infants, immunocompromised and the elderly in both developed and developing countries. Re-infections are common and G protein variability is one mechanism to overcome herd immunity. This is illustrated by the appearance of the BA genotype with a 60 nucleotide duplication dominating the subtype B genotypes in epidemics worldwide. To investigate the evolution of subtype B in South Africa (SA) since 2002 the genetic variability of the G

protein was analyzed in all recent strains isolated over four years (2006-2009) in SA hospitals. Bayesian analysis revealed a replacement of all subtype B genotypes previously identified in SA with the BA genotype since 2006, while subtype A genotypes identified in previous years are still circulating. Compared to BA strains from other countries, the evolutionary rate of the SA BA genotype was shown to be 2.305 x 10^{-3} nucleotide substitutions/site/year and drift was evident. The most recent common ancestor (MRCA) of the SA BA viruses was determined to date back to 1996. All SA BA isolates clustered with the BA-IV sub genotype and the appearance of new sub-genotypes within this branch may occur if drift continues. Sequencing of the complete G protein of selected SA strains revealed an additional 6 nucleotide deletion. Acquisition of the 60 nucleotide duplication appeared to have improved the fitness of this virus and more recent subtype B strains may need to be included in experimental vaccines to evaluate their efficacy in the current setting of evolved circulating strains.

Key words: Respiratory Syncytial Virus (RSV), Subtype, Genotype, Evolution, South Africa (SA)

Introduction

Respiratory Syncytial Virus (RSV), member of the genus *Pneumovirus*, in the *Paramyxoviridae* family, is a major cause of severe paediatric respiratory tract disease in infants, the immunocompromised and the elderly (3, 10, 41). In moderate climates, RSV epidemics occur yearly in the winter months, whereas outbreaks are associated with the rainy season in humid climates (5). Although the mortality rate for RSV infections has decreased significantly over the past 20 years, approximately 500 deaths still occur annually in the U.S., of which 80% occur in children <1 year (29). Globally, the World Health Organization estimates that RSV causes 64 million infections and 160 000 deaths annually (44). A few studies have characterized the disease burden of RSV in developing countries. A study in Kilifi, Kenya, estimated that 85,000 infant cases of severe lower respiratory tract infection (LRTI) were due to RSV per year (23). Re-infection is known to occur throughout life. Children are infected in the presence of maternal antibodies and natural infection affords only partial protection (31, 43).

RSV has a single-stranded negative sense RNA genome containing 10 genes encoding 11 proteins (6). Two antigenic subtypes (A and B) exist with little cross protection (2). Major antigenic differences between subtypes are a feature of the attachment protein G, a type II transmembrane glycoprotein with a conserved central region with four cysteines

postulated to be a receptor binding site. Variability is concentrated in two hypervariable regions of the ectodomain (15). Several G protein genotypes within the two subtypes have been identified including GB1-GB4 (24). SAA1, SAB1, SAB2 and SAB3 were identified in SA (39) and subsequently in various other geographic locations (26, 30). Venter *et al*. (38) also showed that identical RSV genotypes were identified in different regions in South Africa during one season. A new BA genotype has been identified in Buenos Aires in 1999 characterized by a 60 nucleotide duplication starting after residue 791 of the G protein (36). Subsequently strains with this duplication have been found in clinical specimens from distantly related places in the world (16, 21, 27, 28, 46, 47) including Kenya in East Africa (28). This BA genotype was first detected in SA during investigation of a nosocomial outbreak in Pretoria in 2006, and motivated us to re-evaluate the current RSV molecular epidemiology in SA (40).

It is unclear why some children experience severe disease and others develop milder disease. It may be due to host factors, maternal immunity or to differences in the virus itself. Genotypes show complex circulation patterns likely facilitated by herd immunity to certain genotypes which might influence disease severity (24, 25).

Despite the importance of RSV as a respiratory pathogen, there is currently no vaccine or effective therapy available. Experimental vaccines have been based on prototype strains identified shortly after its discovery in the 1960's(7, 8, 20). Protective vaccines may however have to take recent strains into account based on drift demonstrated relative to these strains. To determine the evolution of the subtype B genotype in SA hospitals since the emergence of the BA genotype, the molecular epidemiology of RSV epidemics in children in SA hospitals was investigated. Correlation between disease severity and infecting strains was also investigated.

Materials and methods

Study population

The study population consisted of all children < 1 year of age who were diagnosed between February 2006 and May 2009 with RSV by immunofluorescence with the Respiratory Panel 1 IF Assay (Chemicon, Hampshire, UK) or with the Directigen^{TM} RSV rapid test (Becton Dickinson Microbiology Systems, Franklin Lakes, NJ), at the department Medical Virology University of Pretoria National Health Laboratory Service (NHLS), Tshwane academic division, and confirmed to have RSV by RT-PCR at our laboratory. The NHLS serves three secondary and tertiary hospitals which serves the public sector in the Pretoria region and acts as referral hospitals for the Northern Parts of South Africa:

Kalafong Secondary Hospital, Steve Biko Academic Hospital and the 1-Military Hospital. For the disease severity associations, patients were divided into two classes: Mild disease was classified as out-patients not requiring hospitalization and severe disease were patients requiring hospitalization with severe LRTI or requiring intensive care treatment. This study was reviewed, approved and monitored by the human ethics committee, University of Pretoria (25/2006).

Laboratory analysis

RSV positive specimens were confirmed by RT-PCR and screened for co-infections by multiplex PCR (17). Specimens were labelled according to the following system: SA (isolate number)(Out-patient/Hospitalized)(year of isolation).

RNA Extraction

RNA was extracted directly from nasopharyngeal aspirates with QIAamp viral RNA mini kit according to the manufacturer's recommendations (Qiagen, Valencia, CA).

Subtype Identification: Multiplex RT-PCR and nested PCR

The specimens were subtyped with a multiplex nested RT-PCR method as described before (38). Full length G protein RT-PCR were amplified with primers: G1-21 (37) and F164 (32). In brief, 10µl of RNA was added to 10µl 10x reaction buffer, 10mM of each dNTP, 20pmol of each primer (G1-21, F164), 5mM DTT solution, 10 U RNase inhibitor, and 1µl of the Titan^{TM} enzyme mix; the volume made up the 50µl with distilled water. The following cycling conditions were used: 50ºC for 30min, 94ºC for 2min, (94ºC for 10sec, 53ºC for 30sec, 68ºC for 1min) X35 cycles, 68ºC for 7min. This was followed by nested PCR for weak specimens with primers: G32B (32) and F1 (25). The nested PCR was conducted in a 100-µl reaction. Two µl of the RT-PCR product was mixed with 10mM of each dNTP, 20pmol of each primer (G32B, F1), 2.6 U Expand high-fidelity PCR enzyme mix, 10x Expand HF buffer with 15mM $MgCl_2$ according to the following conditions: 94ºC for 2min, (94ºC for 30sec, 53ºC for 30sec, 72ºC for 1min) X35 cycles, 72ºC for 7min. PCR products were analyzed on a 1.5% agarose gel, against a 100-bp ladder as molecular weight marker (DNA molecular marker XIV, Roche Diagnostics, Mannheim Germany).

Nucleotide sequencing

The Wizard SV Gel and PCR Clean-up system were used for PCR product purification (Promega, Madison, WI). Cycle sequencing was performed with the BigDye Terminator 3.1

Cycle Sequencing kit as recommended by the manufacturer (Applied biosystems, Foster City, CA). Primers were as described by Venter et al. (38). Nucleotide sequencing was carried out on both strands and the editing was performed with Sequencher™ Version 4.6, (Gene Codes Corporatoion, Ann Arbor, MI).

Phylogenetic Analysis

A region spanning 270 nucleotides (330 nucleotides for the BA strains), representing the second hypervariable region of the G protein gene was used for phylogenetic analysis as described before (25). This region corresponds to nucleotides 649-918 of prototype strain A2 for subtype A and 652-921 of prototype strain 18537 for subtype B (15). SA sequences were compared against reference sequences of each genotype, available in the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/index.html). Nucleotide sequences of subtype A and B viruses were aligned separately with Clustal X 1.81, using the multiple alignment option (35). Midpoint-rooted maximum likelihood trees were constructed under the HKY codon position substitution model using PhyML (12). Confidence estimates were based on bootstrap re-sampling carried out with 1000 replicates.

Bayesian evolutionary analysis was carried out with BEAST (Bayesian Evolutionary Analysis by Sampling Trees) version 1.4.8 (9) using the Bayesian Markov chain Monte Carlo method (MCMC) to generate maximum clade credibility trees with the Tree Annotator software and showing the highest posterior probability values for each node of the tree for the BA genotype identified in SA from 2006 to 2009 together with 63 unique BA sequences from the rest of the world since 1999. According to Trento *et al*. (37) the BA genotype could be further divided into six different branches (sub genotypes) (I-VI). These guidelines were used to assign SA BA strains to specific sub genotypes. All the trees were plotted using Figtree version 1.1.2 (http://O-tree.bio.ed.ac.uk.innopac.up.ac.za/software/figtree).

Nucleotide and amino acid sequence analysis

Pairwise distances (P-distance) were calculated between individual genotypes as well as within each genotype using Mega version 4 (34). Positive selection was calculated by Nei/Gojobori method indicated by Ka/Ks >1, where Ka is the nonsynonymous substitution rate and Ks the synonymous substitution rate (22). BioEdit Sequence Alignment Editor version 7.0.4.1 was used for amino acid analysis (13).

Evolutionary rate and date of most recent common ancestor (MRCA)

The evolutionary rate was calculated using a Bayesian MCMC method available in the BEAST package, assuming an uncorrelated relaxed lognormal molecular clock that assumes independent rates on different branches (9). MCMC chains were run for sufficient time to achieve convergence (as assessed using TRACER programme). Statistical uncertainty in parameter estimates is given by the 95% highest probability density (HPD) values. The root-to-tip regression plot displays the correlation between phylogenetic branch length and the time of sampling of the viral strains. This was performed with Path-O-Gen version 1.1.2 (http://tree.bio.ed.ac.uk/software/pathogen). The crossing point was taken as the date of the MRCA for all the sequences under analysis.

Statistical analysis

Association between genetic variability and disease severity was calculated with Fisher exact using STATA 10 (Washington, USA). P values <0.05 were considered statistically significant.

Results

Frequency of RSV subtypes and genotypes over four consecutive years in South Africa

Out of 621 specimens that tested positive for RSV at the NHLS Tshwane laboratory over the four years (2006-2009), 245 specimens met the criteria for RT-PCR and 209 were confirmed as RSV positive. Of the 209 patients, 81 (39%) were HIV seropositive (54 were DNA PCR negative and 27 DNA PCR positive for HIV); 26 (12%) were HIV negative and 102 (49%) had unknown HIV status. One hundred and eighty-five (89%) patients had single RSV infections, while 24 (11%) had co-infections with other viruses as tested with a respiratory multiplex PCR (17).

As shown in Table 1A, subtype A was the more prevalent subtype in 2006, 2007 and 2008 whereas subtype B was more prevalent in 2009. Several dual infections with both subtypes were also seen over two of the four years. Two genotypes within subtype A were detected: GA2 and GA5 (Figure 1A_/Table 1B). Genotype GA5 predominated the subtype A genotypes in 2006 with 88%, whereas GA2 predominated in 2007 and 2008 with 82% and 83% respectively. All our subtype B isolates clustered within the BA genotype and had the characteristic 60 nucleotide duplication (Figure 1B_/Table 1B). All previously identified

subtype B genotypes in SA were replaced by the BA genotype for all four years investigated.

Characterization of the South African subtype A strains

Subtype A strains identified in previous epidemics are still circulating in SA although the previously identified SAA1 genotype was not detected. The intragenotypic P-distances were 0.040 and 0.031 for GA2 and GA5 respectively, and the average intergenotypic P-distance between the different genotypes within subtype A fell in the range of 5 to 16% with 11% difference between the two dominant genotypes, GA2 and GA5. The percentage of non-synonymous changes among SA subtype A isolates within the C-terminal of the G protein was 74%. Genotype specific amino acid substitutions could also be identified. The Nei/Gojobori method identified positive selection within genotype GA2 with Ka/Ks as 1.08. Using BEAST, the evolutionary rate of SA subtype A strains was estimated as 3.382×10^{-3} (95% HPD: 1.911 to 4.954×10^{-3}) substitutions/site/year. A root-to-tip transgression plot estimation of SA subtype A strains indicated the MRCA to date back to 1980. Subtype A isolates utilised four termination codons. In total 18% used the UAG stop codon, 36% used UGA and 2% used UAA, at position 909, while 44% used stop codon UAG at position 912.

Characterization of the South African subtype B strains

All subtype B genotypes previously identified in SA have been replaced by the BA genotype since 2006. The overall nucleotide- and amino acid identity levels were shown to be 97.7% and 96.6% respectively for all the SA BA strains. It was shown that the same sequence in different patients was stable for three successive years. Using the reference strains in Figure 1B, differences of 4 to 15% were observed between the subtype B genotypes. The BA strains currently circulating in SA differ by 8%, 11% and 5% from SAB1, SAB2 and SAB3 that were previously identified in SA.

Except for the 60 nucleotide duplication, only substitutions, and no deletions, insertions, or frame shift mutations were observed in the C-terminal of the G protein. The percentage of non-synonymous substitutions among SA subtype B isolates was 52% within the C-terminal of the G protein. The full G protein was sequenced for selected BA strains for each of the four years including five sequences for 2006, three sequences for 2007, two for 2008 and six for 2009. Compared to other complete G protein sequences on GenBank, all sixteen SA BA strains sequenced here had a 6 nucleotide (2 amino acid) deletion at position 490 resulting in deletion of a Proline and Lysine. SA BA isolates used the UAA stop codon at nucleotide position 946 (79%) resulting in a protein length of 315 amino

acids, while 21% of the isolates used the UAG stop codon at position 967, resulting in a length of 322 amino acids.

Evolutionary relationship between South African BA strains and BA strains isolated in other countries

Figure 2 illustrates Bayesian analysis of BA strains isolated over the world since 1999 together with all the BA strains isolated during this study period (2006-2009) in SA. It was shown that all SA BA strains isolated during the four years clustered within group four (BA-VI). Differences ranging from 0.3% to 6.4% were seen between SA BA-IV strains. All the SA BA strains were unique with respect to the other strains within this group from other countries. SA strains were most closely related to strains from Brazil, Argentina and India. Strains from Belgium, Canada and Kenya within group IV clustered separately from the SA strains. Differences from 1.7% to 4.7% were seen between the different BA branches. Positive selection was identified in BA-II, BA-III and BA-VI with Ka/Ks 1.25, 1 and 2 respectively. Positive selection could not be identified within BA-IV.

Genetic drift is visible within the BA genotype since the emergence eleven years ago, as can be seen by the dates of isolation and the location on the tree. The older strains are closer to the root than the more recently isolated strains. Drift is visible within the SA BA-IV strains. When the BA viruses were first isolated in 1999, the 60 nucleotide duplication was an exact replicate of the preceding 60 nucleotides (37). However, numerous changes were observed between these two regions in the SA BA strains (Figure 3). Using BEAST, the average nucleotide substitution rate across all sites in the alignment was calculated as 2.305×10^{-3} (95% HPD: 1.112 to 3.665 x $10^{-}3$) substitutions/site/year for the SA BA strains. With the use of the programme Path-O-Gen, the MRCA of the SA BA genotype was determined to date back to 1996.

Discussion

Our study showed that the BA genotype has replaced all previously identified B genotypes in SA and illustrates the ability of RSV to evolve and overcome herd immunity. Such variability complicates human RSV vaccine development. Strong evidence was shown for a common ancestry of all BA viruses with this 60 nucleotide duplication (37). Since its discovery, strains with this duplication have been found in clinical specimens from distantly related places (16, 21, 27, 28, 46, 47).

In the first three years studied in SA, subtype A predominated, whereas subtype B predominated in the fourth year. Two genotypes (GA2 and GA5) dominated the subtype A

specimens that have previously been isolated in SA (39). The SAA1 genotype was not detected again, suggesting that GA2 and GA5 are the current stable genotypes. All the subtype B specimens clustered with the BA genotype and had the characteristic 60 nucleotide duplication replacing all B genotypes previously identified in SA. Referring to Venter *et al*. (38), identical genotypes were previously found in different regions in South Africa during one season, and thus strains identified in one region likely represent the strains circulating in South Africa. The fact that only BA subtype B strains were identified over all four years in major public sector referral hospitals in the capital of the country is also strong evidence that the previously identified genotypes are not circulating any more. Investigations of strains from the rest of Africa will determine if this is true for the whole continent. The BA genotype has also been detected in Kenya in 2006 but co-circulated with the SAB1 genotype in that epidemic(28). When compared to other genotypes, SA BA strains were more closely related to the SAB3 genotype previously detected in SA. It is not certain if new genotypes in SA occurred as a result of spontaneous mutations or of importation. Apart from the 60 nucleotide duplication, no deletions, insertions or frame shift mutations were detected in the C-terminal region of the G protein although the full G protein sequence identified a 6 nucleotide (2 amino acid) deletion at position 490 resulting in a missing Proline and Lysine. This deletion has also been detected in strains from Kenya and Buenos Aires (1).The dominance of this genotype indicates that the duplication might give this genotype an evolutionary advantage. Sequencing of the full gene suggests further changes in the G gene may occur. Full genome sequence for the new BA strains would elucidate changes in other proteins.

To determine the evolutionary rate of the SA BA genotype as well as the MRCA of these strains, Bayesian evolutionary analysis was conducted on all SA BA isolates as well as BA isolates from all over the world since it was first discovered in 1999. Firstly it was shown that all the strains isolated in SA clustered within branch IV. According to Trento et al. *(37),* this is the most heterogeneous branch regarding the date and the place of isolation. Thus, it is possible that this 60 nucleotide duplication changed the antigenic structure of these BA genotypes, giving them an evolutionary advantage to re-infect individuals previously exposed to subtype B.

When the BA viruses were first isolated in 1999, the 60 nucleotide duplication was an exact replicate of the preceding 60bp (37). This region accumulated further nucleotide substitutions over time and certain nucleotide positions were shown to be under positive selection (4, 19, 45-47). These changes might enhance the fitness of these viruses which would explain why they are now dominating worldwide and replacing all other B

genotypes. This region is believed to be involved in changing/replacing the antigenic epitopes which contribute to the evasion of the immune response of the population. Certain amino acids in this region have been shown to be under positive selection, as well as the change 952CAA-UAA (Q313-STOP) that was present in the most recent lineages of the BA-IV branch (4, 46). Most of the SA BA strains (79%), all from BA-IV, had this stop codon change. However, it was at position 946 because of the 6 nucleotide deletion. Changes in stop codon usage are thought to be associated with antigenic variation that allows immune evasion, especially in the subtype B isolates that are less variable than the subtype A isolates (18).

Positive selection was shown to occur within some of the different BA branches. The dates when the strains were isolated and their placement on the tree indicate how much this genotype changed over time, especially the SA strains (BA-IV). The strains first isolated in 1999 lie closer to the root of the tree than the strains isolated in 2009. The appearance of new sub genotypes within the BA-IV branch may occur as drift continues with four different clusters already existing within BA-IV (Figure 2). Substitution rates have been estimated for only a limited number of viruses. In general, values lie close to 1×10^{-3} substitutions/site/year, although considerable variation exists. This variation is partly due to their error prone replication which causes mutations to accumulate quickly (14). The rate of evolution for SA BA strains across all sites in the alignment was calculated as 2.305×10^{-3} substitutions/site/year, which correlates with findings by Trento *et al*. (37). According to them, the origin of the MRCA was dated between 1998 and 1999, shortly before it was discovered in 1999. Our data indicate that the MRCA of the SA BA sequences date back to 1996.

This study did have some limitations and were not able to fully identify possible associations between disease severity and genotype. Although information on hospitalization was available, the number of out-patients was small and no significant associations between subtype/genotype and disease severity were found. Nevertheless most other studies only evaluated this during one epidemic season whereas the present study investigated this over four consecutive seasons. The accumulation of herd immunity may determine the disease association of different strains in a community from season to season. Also, information on HIV infection and immune status for most study participants was lacking which limited our ability to address the role of HIV in diseases association with the different RSV genotypes.

To conclude, G protein variability may play a significant role in RSV pathogenesis by allowing immune evasion. Certain substitutions or alterations may enhance the fitness of

viruses as is evident with the BA strains that replaced all other B genotypes previously identified in SA. The G protein's ability to accommodate such substantial changes and facilitate immune evasion may complicate vaccine development. It remains to be seen if the BA genotype will remain dominant or if dominance will eventually fade because of herd immunity. Our study suggests that subtype B strains may need to be updated with recent strains in experimental vaccines to re-evaluate their efficacy.

Acknowledgements

This study was reviewed, approved and monitored by the human ethics committee, University of Pretoria (25/2006) and funded by the Poliomyelites research foundation. A special thanks to Perry Smith and Kariuki Njenga of the CDC for critical review of this paper.

References

1. **Agoti, C. N., J. L. Mbisa, A. Bett, G. F. Medley, D. J. Nokes, and P. A. Cane.** Intrapatient variation of the respiratory syncytial virus attachment protein gene. J Virol **84:**10425-8.
2. **Anderson, L. J., J. C. Hierholzer, C. Tsou, R. M. Hendry, B. F. Fernie, Y. Stone, and K. McIntosh.** 1985. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. J Infect Dis **151:**626-33.
3. **Boeck, K. D.** 1996. Respiratory syncytial virus bronchiolitis: clinical aspects and epidemiology. Monaldi Arch Chest Dis **51:**210-3.
4. **Botosso, V. F., P. M. Zanotto, M. Ueda, E. Arruda, A. E. Gilio, S. E. Vieira, K. E. Stewien, T. C. Peret, L. F. Jamal, M. I. Pardini, J. R. Pinho, E. Massad, O. A. Sant'anna, E. C. Holmes, and E. L. Durigon.** 2009. Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. PLoS Pathog **5:**e1000254.
5. **Cane, P. A.** 2001. Molecular epidemiology of respiratory syncytial virus. Reviews in Medical Virology **11:**103-116.
6. **Collins, P. L., and J. E. Jr Crowe.** 2007. Respiratory syncytial virus and metapneumovirus, p. 1601-1646. *In* D. M. Knipe and P. M. Howley (ed.), Fields Virology, 5th ed, vol. 2. Lippincott Williams and Wilkens, Philidelphia.
7. **Collins, P. L., and B. R. Murphy.** 2002. Respiratory syncytial virus: reverse genetics and vaccine strategies. Virology **296:**204-11.
8. **Crowe, J. E., Jr.** 2001. Respiratory syncytial virus vaccine development. Vaccine **20 Suppl 1:**S32-7.
9. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **7:**214.
10. **Falsey, A. R., P. A. Hennessey, M. A. Formica, C. Cox, and E. E. Walsh.** 2005. Respiratory syncytial virus infection in elderly and high-risk adults. N Engl J Med **352:**1749-59.
11. **Garcia, O., M. Martin, J. Dopazo, J. Arbiza, S. Frabasile, J. Russi, M. Hortal, P. Perez-Brena, I. Martinez, B. Garcia-Barreno, and et al.** 1994. Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. J Virol **68:**5448-59.
12. **Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel.** 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **33:**W557-9.
13. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95-98.

14. **Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes.** 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol **54:**156-65.
15. **Johnson, P. R., M. K. Spriggs, R. A. Olmsted, and P. L. Collins.** 1987. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. Proc Natl Acad Sci U S A **84:**5625-9.
16. **Kuroiwa, Y., K. Nagai, L. Okita, I. Yui, T. Kase, T. Nakayama, and H. Tsutsumi.** 2005. A phylogenetic study of human respiratory syncytial viruses group A and B strains isolated in two cities in Japan from 1980-2002. J Med Virol **76:**241-7.
17. **Lassauniére, R., T. Kresfelder, and M. Venter.** 2010. A novel multiplex real-time RT-PCR assay with FRET hybridization probes for the detection and quantification of 13 traditional and newly identified respiratory viruses. Journal of Virological Methods.
18. **Martinez, I., O. Valdes, A. Delfraro, J. Arbiza, J. Russi, and J. A. Melero.** 1999. Evolutionary pattern of the G glycoprotein of human respiratory syncytial viruses from antigenic group B: the use of alternative termination codons and lineage diversification. J Gen Virol **80:**125-30.
19. **Melero, J. A., B. Garcia-Barreno, I. Martinez, C. R. Pringle, and P. A. Cane.** 1997. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. J Gen Virol **78 (Pt 10):**2411-8.
20. **Murata, Y.** 2009. Respiratory syncytial virus vaccine development. Clin Lab Med **29:**725-39.
21. **Nagai, K., H. Kamasaki, Y. Kuroiwa, L. Okita, and H. Tsutsumi.** 2004. Nosocomial outbreak of respiratory syncytial virus subgroup B variants with the 60 nucleotides-duplicated G protein gene. J Med Virol **74:**161-5.
22. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:**418-26.
23. **Nokes, D. J., E. A. Okiro, M. Ngama, R. Ochola, L. J. White, P. D. Scott, M. English, P. A. Cane, and G. F. Medley.** 2008. Respiratory syncytial virus infection and disease in infants and young children observed from birth in Kilifi District, Kenya. Clin Infect Dis **46:**50-7.
24. **Peret, T. C., C. B. Hall, G. W. Hammond, P. A. Piedra, G. A. Storch, W. M. Sullender, C. Tsou, and L. J. Anderson.** 2000. Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. J Infect Dis **181:**1891-6.
25. **Peret, T. C., C. B. Hall, K. C. Schnabel, J. A. Golub, and L. J. Anderson.** 1998. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. J Gen Virol **79:**2221-9.
26. **Riccetto, A. G., L. H. Silva, F. R. Spilki, A. M. Morcillo, C. W. Arns, and E. C. Baracat.** 2009. Genotypes and clinical data of respiratory syncytial virus and metapneumovirus in brazilian infants: a new perspective. Braz J Infect Dis **13:**35-9.
27. **Sato, M., R. Saito, T. Sakai, Y. Sano, M. Nishikawa, A. Sasaki, Y. Shobugawa, F. Gejyo, and H. Suzuki.** 2005. Molecular epidemiology of respiratory syncytial virus infections among children with acute respiratory symptoms in a community over three seasons. J Clin Microbiol **43:**36-40.
28. **Scott, P., R. Ochola, M. Ngama, E. Okiro, D. Nokes, G. Medley, and P. Cane.** 2004. Molecular epidemiology of respiratory syncytial virus in Kilifi district, Kenya. J Clin Microbiol **74:**344-354.
29. **Shay, D. K., R. C. Holman, G. E. Roosevelt, M. J. Clarke, and L. J. Anderson.** 2001. Bronchiolitis-associated mortality and estimates of respiratory syncytial virus-associated deaths among US children, 1979-1997. J Infect Dis **183:**16-22.
30. **Shobugawa, Y., R. Saito, Y. Sano, H. Zaraket, Y. Suzuki, A. Kumaki, I. Dapat, T. Oguma, M. Yamaguchi, and H. Suzuki.** 2009. Emerging genotypes of human respiratory syncytial virus subgroup A among patients in Japan. J Clin Microbiol **47:**2475-82.
31. **Sullender, W. M.** 2000. Respiratory syncytial virus genetic and antigenic diversity. Clin Microbiol Rev **13:**1-15, table of contents.
32. **Sullender, W. M., L. Sun, and L. J. Anderson.** 1993. Analysis of respiratory syncytial virus genetic variability with amplified cDNAs. J Clin Microbiol **31:**1224-31.

33. **Sullender, W. M., and G. W. Wertz.** 1991. Synthetic oligonucleotide probes differentiate respiratory syncytial virus subgroups in a nucleic acid hybridization assay. J Clin Microbiol **29:**1255-7.
34. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:**1596-9.
35. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-82.
36. **Trento, A., M. Galiano, C. Videla, G. Carballal, B. Garcia-Barreno, J. A. Melero, and C. Palomo.** 2003. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. J Gen Virol **84:**3115-20.
37. **Trento, A., A. Viegas, M. Galiano, C. Videla, G. Carvallal, A. Mistchenko, and J. Melero.** 2006. Natural History of Human Respiratory Syncytial Virus Inferred from Phylogenetic Analysis of the Attachment (G) Glycoprotein with a 60-Nucleotide Duplication. Journal of Virology **80:**975-984.
38. **Venter, M., M. Collinson, and B. D. Schoub.** 2002. Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: Comparison of viruses and genotypes responsible for different disease manifestations. J Med Virol **68:**452-61.
39. **Venter, M., S. A. Madhi, C. T. Tiemessen, and B. D. Schoub.** 2001. Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new subgroup A and B genotypes. J Gen Virol **82:**2117-24.
40. **Visser, A., S. Delport, and M. Venter.** 2008. Molecular epidemiological analysis of a nosocomial outbreak of respiratory syncytial virus associated pneumonia in a kangaroo mother care unit in South Africa. J Med Virol **80:**724-732.
41. **Welliver, R. C.** 2003. Review of epidemiology and clinical risk factors for severe respiratory syncytial virus (RSV) infection. J Pediatr **143:**S112-7.
42. **Wertz, G. W., P. L. Collins, Y. Huang, C. Gruber, S. Levine, and L. A. Ball.** 1985. Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. Proc Natl Acad Sci U S A **82:**4075-9.
43. **Wertz, G. W., and R. M. Moudy.** 2004. Antigenic and genetic variation in human respiratory syncytial virus. Pediatr Infect Dis J **23:**S19-24.
44. **WHO Department of Vaccines and biologicals(V&B), V. a. a. m. V.** 2001. Report of A workshop on the epidemiology of Respiratory syncytial virus (RSV) in developing countries. V&B/VAM/SCEFR/00.01. WHO.
45. **Woelk, C. H., and E. C. Holmes.** 2001. Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). J Mol Evol **52:**182-92.
46. **Zlateva, K. T., P. Lemey, E. Moes, A. M. Vandamme, and M. Van Ranst.** 2005. Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein. J Virol **79:**9157-67.
47. **Zlateva, K. T., P. Lemey, A. M. Vandamme, and M. Van Ranst.** 2004. Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup a: positively selected sites in the attachment g glycoprotein. J Virol **78:**4675-83.
48. **Zlateva, K. T., L. Vijgen, N. Dekeersmaeker, C. Naranjo, and M. Van Ranst.** 2007. Subgroup prevalence and genotype circulation patterns of human respiratory syncytial virus in Belgium during ten successive epidemic seasons. J Clin Microbiol **45:**3022-30.

List of figures and tables:

Figure 1 Midpoint-rooted maximum likelihood trees constructed under the HKY codon position substitution model using PhyML for **A)** subtype A and **B)** subtype B (12). It was drawn to scale with the bars indicating 0.03 nucleotide substitutions. Estimates were based on bootstrap re-sampling carried out with 1000 replicates. Only bootstrap values >70 are shown. Names of the viruses refer to the place/number/disease severity/year of isolation. The number of identical sequences is indicated as x n. The genotypes assigned are indicated at the right by brackets. The reference sequences obtained from GenBank were as follows: from the USA (NY, New York; Al, Alabama; MO, Missouri; Tx, Texas; CN, Canada; CH, Rochester, New York) (24, 25). WV, West Virginia (33); Montevideo, Uruguay (MON) and Madrid, Spain (MAD) (11, 18). Prototype strains for subtype A: strain A2 (Australia) (42); subtype B: Swed8-60 (Sweden) (33); USA (18537) (15).

Figure 2 A midpoint-rooted BEAST phylogenetic tree of BA genotypes isolated over the world since 1999 together with all the BA genotypes isolated during this study (2006-2009) (9). The length of the horizontal lines is proportional to the genetic distance between viruses. The bar represents 2 substitutions per site. The numbers on the branches represent the highest posterior probability value for each node of the tree. Names of viruses refer to the place/number/disease severity/ year of isolation. SA strains were compared to strains from Buenos Aires (BA), Japan, Kenya (Ken), Quebec (QUE), Jundiai Brazil (JU), Sau Paulo Brazil (SP), Ribeirao Preto Brazil (RP), India (PN) and Belgium (BE) (28, 37, 48). The number of identical sequences is indicated as x n. Branches are indicated at the right by brackets. The scale at the bottom indicates years since first isolation.

Figure 3 Amino acid alignment of the duplicated region within the BA genotype. The alignments are shown relative to the BE12445 strain isolated in Belgium in 1999.

Table 1 Frequency of Respiratory Syncytial virus **A)** subtype A and B **B)** genotypes over four consecutive years in South Africa

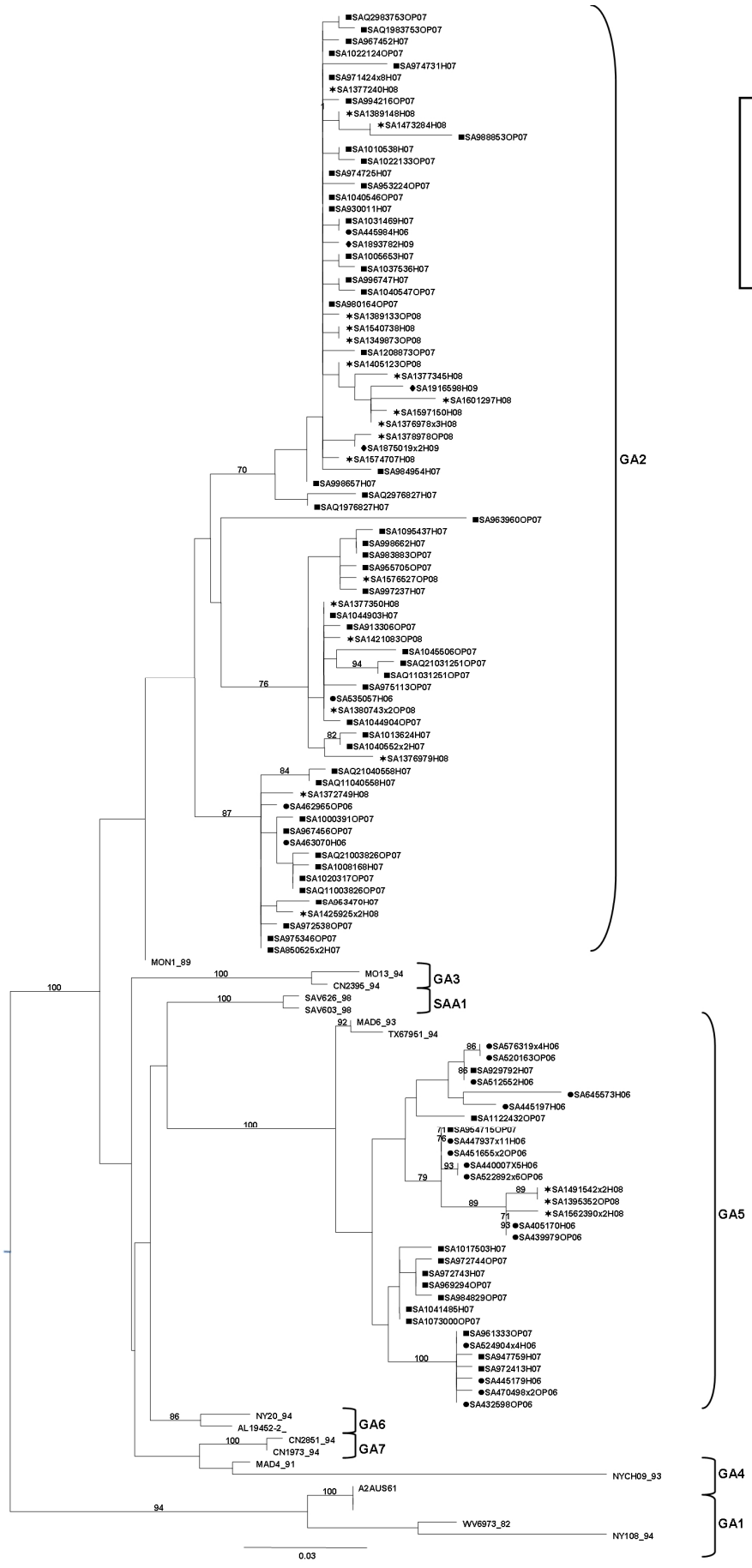


Figure 1A

Van Niekerk & Venter, 2010

Figure 1B

Van Niekerk & Venter, 2010

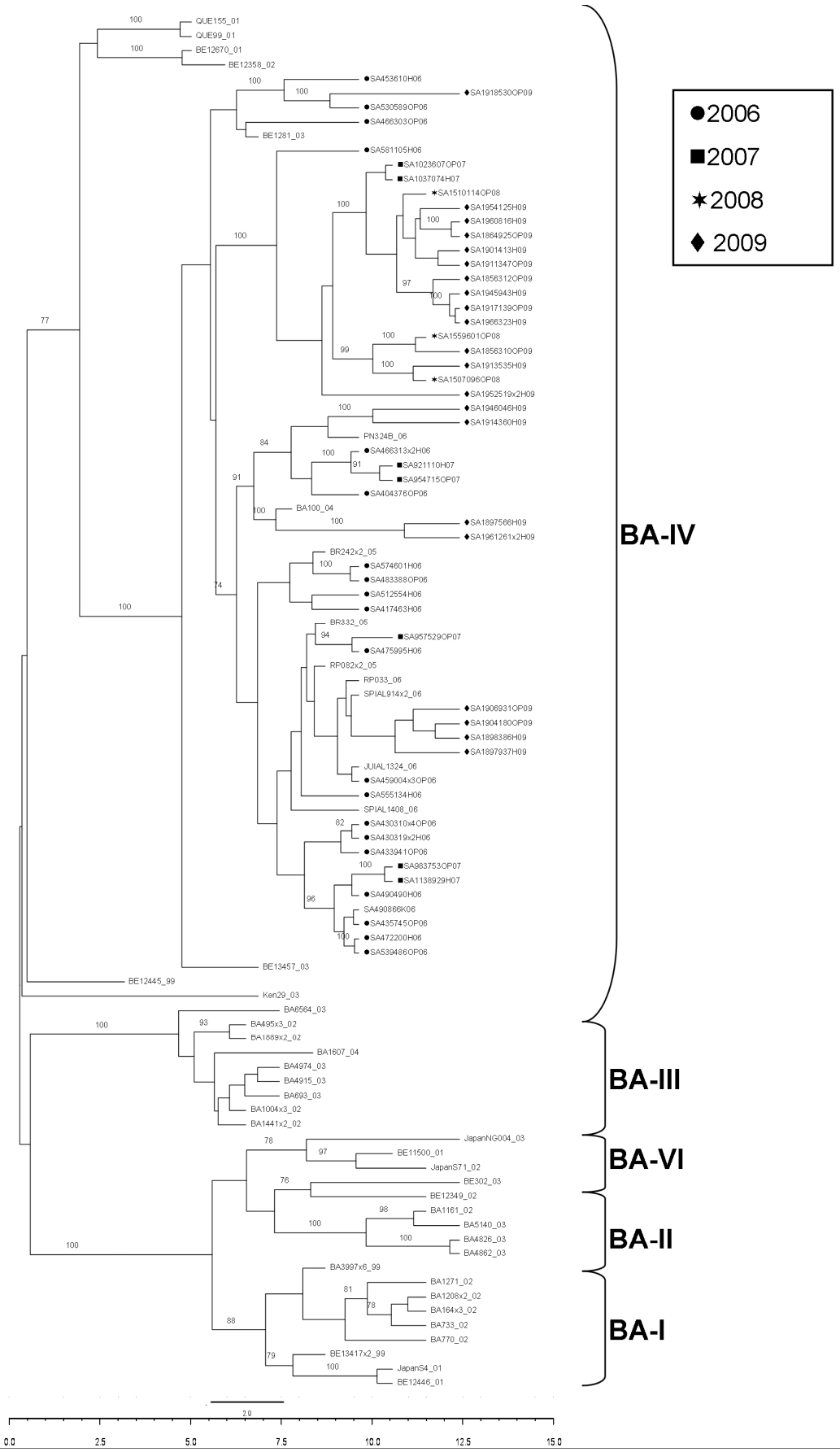
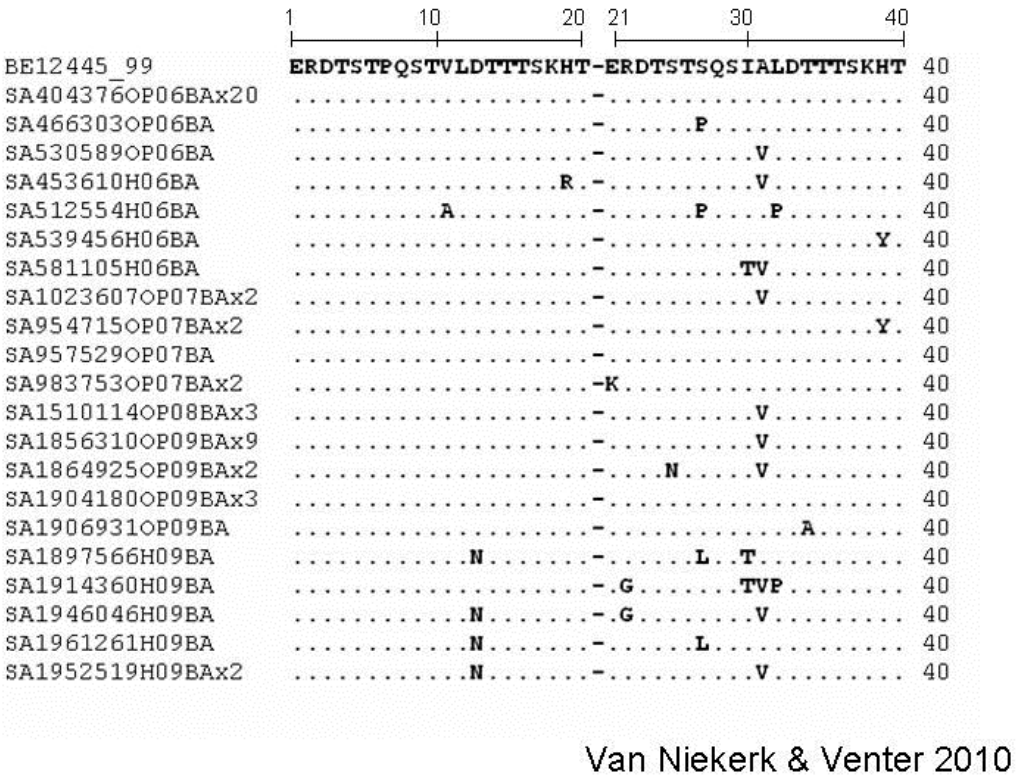


Figure 2

Van Niekerk & Venter, 2010



Van Niekerk & Venter 2010

Figure 3

A

	2006 * (76 SPECIMENS)	2007 * (74 SPECIMENS)	2008 * (32 SPECIMENS)	2009 * (25 SPECIMENS)
Subtype A	45 (59%)	67 (90%)	29 (91%)	4 (16%)
Subtype B	27 (36%)	5 (7%)	3 (9%)	21 (84%)
Dual infections	4 (5%)	2 (3%)	0	0

***Total specimens screened per year**

B

	2006	2007	2008	2009
Subtype A	49 *	69 *	29 *	4 *
GA2	4 (8%)	56 (81%)	24 (83%)	4 (100%)
GA5	43 (88%)	13 (19%)	5 (17%)	0
Untypable	2 (4%)	0	0	0
Subtype B	31 *	7 *	3 *	23 *
BA	30 (97%)	7 (100%)	3 (100%)	23 (100%)
Untypable	1 (3%)	0	0	0

*** Total strains sequenced. Due to dual-infections in some patients these totals are higher than the amount of specimens tested.**

Table 1A and B van Niekerk & Venter, 2010