

Correlation of cytokine profiles, disease severity and RSV
genotypes in an archive of nasal aspirate samples

Table of contents

1. Introduction.....	7
1.1 Can different RSV strains cause clinical disease of varying severity?	10
1.2 Methods of virus genotyping	11
1.3 The role of variation in the N and P proteins in viral replication and cytokine activation ...	11
1.4 The role of genotype variation in mediating virus escape from Cytotoxic T Lymphocyte (CTL) responses.	12
1.5 The effect of G gene variation on host immune responses	13
1.6 Cytokine production profiles as predictors of clinical illness.....	13
1.7 Samples used in this study.....	14
2. Materials and Methods.....	16
2.1 Virus Genotyping by Nucleoprotein gene Restriction Fragment Length Polymorphism (RFLP) patterns.	16
2.1.1 RNA Extraction	16
2.1.2 Complimentary DNA (cDNA) Synthesis	17
2.1.3 Primary N gene PCR from cDNA.....	17
2.1.4 Secondary <i>N Gene PCR</i>	18
2.1.5 Agarose Gel Electrophoresis	19
2.1.6 Purification of PCR products.	20
2.1.7 Restriction Fragment Length Polymorphism (RFLP) analysis of purified N gene products through restriction enzyme digestion of purified PCR products.	20
2.2.1 Virus Genotyping of the attachment protein (G) gene.	21
2.2.2 G gene PCR using a cDNA template.....	22
2.2.3 Restriction Fragment Length Polymorphism (RFLP) analysis of purified G gene	

products through restriction enzyme digestion of purified PCR products.	23
2.3.1 Phylogenetic analysis of sequenced products from purified PCR products.....	23
2.4.1 Cytokine concentration measurement using the Luminex assay.....	24
3. Results	27
3.1 Effect of viral co-infection on the duration of illness.	27
3.2 Effect variation in the N gene characterized by RFLP.....	28
3.2.2 Effect of RFLP based N gene variation on viral load.....	29
3.2.3 Effect of RFLP based N gene variation on Duration of clinical illness	30
3.3 Effects of phylogenetic cluster based N-P (nucleoprotein-phosphoprotein) gene variation on viral load and duration of illness	31

Table of figures

Figure 1 (a) Proportion of RSV infected children with co-infections from other respiratory pathogens.

(b) The effect of viral co-infection on the duration of respiratory symptoms..... 24

Figure 2 (a) NP genotype distribution patterns among males and females

(b) Age distribution of different NP genotypes..... 25

Figure 3 L gene copy numbers among the various NP genotypes. The y axis represents the L gene copy number, while the X axis represents various NP genotypes. U represents the viral isolates that could not be typed by RFLP. The dotted line represents the detection limit of this assay.

.....26

Figure 4 The duration of illness in children infected by different N-P genotypes.

(a) Mean duration of illness was compared for different NP genotypes regardless of age.

(b) Comparison of duration of illness was restricted to infants under 8 months..... 27

Figure 5 Phylogenetic tree generated based on analysis of sequence data by the computer program

DNAml. The numbers represent the anonymised identifiers of individual children samples.

The tree diagram was generated by the computer program Treeview. Closely related clusters

were grouped according data shown in table 1..... 28

Figure 6. Comparison of (a)duration of illness and (b) viral load among different phylogenetic clusters

derived from the N and P intergenic region. Viral load was derived from the copy number of the polymerase (L) gene.29

Figure 7 Phylogenetic tree generated from G gene sequences representing the gene segment between nucleotide 1 and 584. The tree was generated using the maximum likelihood model by the DNAmI software. Clusters of closely related sequences were selected and are shown in table 2.30

Figure 8. Comparison of the effect of different phylogenetic clusters on the duration of respiratory illness31

Figure 9 Cytokine/chemokine production profiles among different NP RFLP genotypes..... 32

Figure 10. Cytokine/chemokine production profiles among different G gene phylogenetic clusters.34

Figure 11 The ratio of IFN- α /IL6 cytokine concentrations among different G gene phylogenetic clusters.

Acknowledgments

I owe a debt of gratitude to all members of the department of Immunology, Imperial College London who were always at hand with invaluable help and advice. I also thank members of the Wellcome Trust Centre for Respiratory Infections, Imperial College London who constantly offered much valued help, encouragement and the benefit of their expertise during the conduct of the experiments.

My sincerest thanks also go to Prof. Hans Bisgaard and all members of the COPSAC project in Copenhagen for kindly providing the samples on which this work is based. Many thanks also for their very kind help and advice and also for hosting me in their facilities during my visit to Copenhagen.

I would especially like to thank my supervisor Prof. Peter Openshaw who always offered a hand of friendship and encouragement at every stage of the project work. His insights and mentorship have greatly influenced my understanding of the immunology of RSV. I would also like to thank my other supervisors Prof. Sebastian Johnston, Drs. Pat Cane and James Nokes, whose unfailing guidance and support have made a career in science an exciting and fulfilling undertaking.

Finally, my sincerest thanks to the Wellcome Trust who supported this work.

Abstract

Genotype variation is a well-characterised feature of Respiratory Syncytial Virus (RSV) epidemiology. Successive RSV epidemics are predominated by individual genotypes and these are replaced in subsequent epidemics. It is not clear whether some of these genotypes are responsible for more severe illness. In animal studies, there is evidence that some strains have the capacity to induce more severe illness than others. Data implicating particular human genotypes with severe disease has been conflicting. In the present study we examined an archive of RSV positive nasal samples to determine the genotype of the infecting virus based on variation in the nucleoprotein (N) gene. We then correlated the genotyped to both disease outcome and cytokine/chemokine production profiles.

RESULTS: We did not find any positive associations between virus genotypes characterised by variation in the N gene and disease severity. Further, no associations were found between these genotypes and viral load as well as cytokine/chemokine production. We propose that in future, studies should use more robust methodologies such as sequence analysis to examine variation within viral genes.

1. Introduction

Respiratory syncytial virus (RSV) was first isolated over half a century ago and has been identified as the major cause of bronchiolitis in infants, common colds in adults and a major cause of winter deaths in the elderly [1, 2]. Despite forty years of concerted research, a safe and effective vaccine has not been found for any vulnerable group [3]. The disease exerts a disproportionate illness burden at age extremities, with infants and the elderly forming the bulk of clinical RSV disease patients [4-6]. In infants the risk for severe RSV disease peaks at after 6 months of age, coinciding with the decline of maternal antibodies[7].

RSV pathology is thought to represent excessive host inflammatory responses, but is driven by viral ingress into cells, the transcription of viral genes, the expression of their protein products and the budding of new virus. However, the level of viral replication at a given point in time after infection correlates poorly or not at all with severe illness. Possibly, this is because the levels of virus are highly dependent on the time after infection, and in studies of naturally infected infants, the time after infection cannot be controlled. The poor correlation between viral load and disease is also evident in some other infections; for example, infants infected with human immunodeficiency virus type 1 (HIV1) exhibit prolonged viral shedding (over 199 days) without substantial disease[8]. Monoclonal antibody treatment in children with severe RSV has been shown to drastically reduce viral titre without a concomitant decrease in clinical symptoms [9]. Another line of evidence in support of the immunopathology model is a model of human RSV disease, in which exacerbated disease was observed in a group of infants immunized with a formalin inactivated vaccine, resulting in 80% hospitalization of vaccinees compared to the relatively mild symptoms exhibited by the control group[10]. In this situation, an increased cellular exudate was associated with severe disease and even death.

RSV causes disease of varying clinical severity in infants. Known risk factors for severe disease include premature birth (infants born at 29 to 32 weeks gestational age) [11], congenital heart disease [12], old age [6], immunosuppression [13], bronchopulmonary dysplasia[14], congenital heart disease[12] and an immunocompromised immune status[15]. In addition to these risk factors, viral load has been independently associated with severe RSV disease by some investigators[16, 17] although these findings have been disputed by others[18]. It has also been demonstrated that RSV disease is exacerbated in cases where there is co-infection with other viruses. For example, the risk of severe RSV disease is increased five fold in the presence of rhinovirus coinfection [19].

Variation in the host's immune genes has also been found to determine susceptibility to severe disease. For example, single nucleotide polymorphisms (SNPs) in the VDR, JUN, IFNA5, and NOS2 innate immune genes have been shown to correlate with susceptibility to severe RSV disease [20]. Polymorphisms in TLR4, IL-10, IL-8 IL-13, IL-4 and surfactant proteins A and D have been shown to be predisposing factors to severe disease [21-25].

Environmental factors may also determine the course of clinical RSV disease. In the Yukon-Kuskokwim (YK) delta in Alaska, patients are at a five fold increased risk for hospitalization due to RSV infection as compared to the general US population [3]. This differential susceptibility has been attributed household crowding as well as the extended epidemic periods experienced in this area. Exposure to passive tobacco smoke is another important environmental risk factor for the development of severe RSV disease [14].

Differential T helper cell response profiles have also been suggested by some authors to determine the clinical course of illness. A Th2 skewed profile has been associated with severe disease, while a

predominant Th1 response has been associated with a relatively less severe course of illness [26]. These findings have however been disputed by other studies which have failed to demonstrate differences in clinical outcome based on T helper cell cytokine response profiles [27].

One possible determinant of severe clinical illness is infection with different strains of the RSV. It is known that progressive changes in genome during selection of temperature sensitive mutants in associated with attenuation of disease inducing capacity in human challenge. In mice, different human isolates of RSV (e.g. A2, Long, 8/60 and Line 19) have very different abilities to cause disease. While a biologically plausible argument can be made that different strains of the virus, varying both in their antigenicity and virulence, lead to different manifestations of clinical disease, clear and definitive evidence implicating specific viral strains, genes or gene products as the determinants of severe disease during natural infection of children has not been found [28].

1.1 Can different RSV strains cause clinical disease of varying severity?

RSV can be classified into two distinct antigenic groups, A and B. This distinction is based upon reactivity with and ferret sera and monoclonal mouse antibodies[29]. Some monoclonal antibodies are able to bind certain epitopes in one cluster of viruses but fail to recognise an alternative, distinct cluster of viruses. Each of these major groups are made up of a heterogeneous mix of viruses known as 'genotypes'. These genotypes exhibit either nucleotide or amino acid sequence variation within one or more of the viral genes. They are thought to arise through a process of positive selection, a mechanism by which the virus out-evolves the host's protective responses both at the cellular and humoral levels [30].

Although many studies have been conducted to try and establish a link between certain clinical RSV

phenotypes and distinct viral genotypes, many of these studies have failed to establish such a link. Most of these studies have been hospital based surveillance studies, which recruit infants with a broad spectrum of clinical illness [28]. Such studies isolate virus from infants and assign these viruses to genotype designations that are based on minor variations in one or two of the virus's genes. These genotypes are then correlated with varying disease phenotypes. However common genotype assignment strategies fail to take into account the functional roles of the viral gene regions that are being genotyped and thus genetic regions of functional importance in terms of viral replication or interaction with the host's immune system could possibly be spread across several genotypes.

1.2 Methods of virus genotyping

The most common RSV genotyping strategy focuses on variability within a number of viral genes[31, 32]. The gene in which variation has been most extensively characterised is the nucleoprotein gene (N gene). The N gene region that is used in genotyping experiments is moderately conserved and bears five restriction enzyme sites. Mutations that affect these sites have been used to characterise gross variation within the gene based on fragment length polymorphism (RFLP) patterns [32, 33]. Another common genotyping strategy focuses on variability within the viral attachment protein gene (G gene)[34]. This gene is the most variable gene in wild type clinical isolates[30]. Three regions with distinct sequence characteristics have been identified in this gene; a conserved, highly glycosylated central region, sandwiched between two variable, mucin-like domains[35].

1.3 The role of variation in the N and P proteins in viral replication and cytokine activation

Like all paramyxoviruses, the functional unit for viral replication and transcription for RSV is the ribonucleoprotein (RNP). RNPs are a complex consisting of a helical an RNA-N protein nucleocapsid

complex (NC) and the viral P, L and M2-1 proteins. The viral polymerase (the L protein) is only active when the viral RNA is bound to the NC in the association with P and M2-1 proteins[36]. Recent studies have shown that when certain residues on the P protein undergo mutation, the binding efficiency of this protein to the RNP complex is reduced and thereby leading to a reduction in the efficiency of viral replication [37]. It is therefore biologically plausible that wild type variants incorporating such mutations as a result of immune pressure could equally result in lower viral loads *in vivo* possibly modulating disease outcome.

RSV infection has also been shown to lead to persistent activation of the nuclear transcriptional factor kappa B (NF-kB) in a P protein dependant manner [38]. Certain residues in the P protein sequester phosphate groups and resist dephosphorylation by protein phosphatase 2A. This leads to a constant dephosphorylation state of the inhibitory subunit of Nf-KB, IkBb, thereby causing Nf-KB to remain in a constantly activated state. The downstream consequences of this induction include upregulation of cytokine transcription. Variations within the P protein could therefore alter its phosphate sequestration capacity leading to differential patterns of Nf-KB activation and downstream cytokine production profiles.

1.4 The role of genotype variation in mediating virus escape from Cytotoxic T Lymphocyte (CTL) responses.

Previous studies have mapped CTL epitopes on a number of viral proteins. Human specific RSV epitopes on the N protein restricted by HLA-A and B alleles have been described [39]. CTL epitopes have also been mapped for the F proteins of group A RSV. Evidence of immune selection through the occurrence of non synonymous mutations in successive field strains within the F protein CTL epitopes, have led to the suggestion that genotype variation may be a mechanism by which the host's immune

responses [40]. Similar studies in the mouse influenza model have generated convincing data showing that the virus escapes robust and protective CTL responses by incorporating non synonymous mutations within CTL epitopes on the N protein of the influenza virus, thus mediating immune escape [41]. However studies to demonstrate this effect within the N protein of circulating RSV strains have failed to show that virus can escape specific CTL responses mounted against epitopes on the N protein by altering the amino acid sequence of this protein[42].

1.5 The effect of G gene variation on host immune responses

The G gene codes for the G protein, the viral attachment protein. It is heavily glycosylated and is the target of protective neutralising antibody responses in humans[43]. The nucleotide and amino acid sequence of the G gene is remarkably variable in clinical isolates. Much of this variability is focused on the carboxyterminal third and to a lesser extent on the N terminal portion of the gene. A central, thirty-nine nucleotide motif (translated into a 13 amino acid stretch, ie residues 164 to 176) is conserved in virtually all known RSV isolates[44]. The central region has been found to bear immunodominant epitopes that are restricted by HLA-DP4 [45]. Overlapping peptides within this region have been shown to elicit strikingly different cytokine production profiles in T cells obtained from healthy adults. T cells challenged with G165-179 produced high levels of IL-5 and IL-13 and only minor quantities of IFN-g. However G162-175 was associated with a predominant release of IFN-g[45].

1.6 Cytokine production profiles as predictors of clinical illness.

In some diseases such as leprosy, the T helper 1 / T helper 2 (Th1/Th2) cytokine skew profile can reliably predict the course of clinical illness[46]. Attempts to find whether certain cytokine production profiles can reliably predict the course of clinical RSV infection have resulted in conflicting data from different groups. Bont *et al* have reported that severely ill, ventilated infants in general have

significantly lower lymphoproliferative responses and a lower production of IFN-gamma and IL-4. The levels of these cytokines are in many cases completely undetectable in these patients. IL-8 levels in the plasma of ventilated patients were also much higher than in non ventilated patients who were less severely ill [47].

1.7 Samples used in this study.

Bronchial aspirate samples used in this study were obtained from the Copenhagen Study on Asthma in Children (COPSAC). This is a prospective surveillance study investigating the causes of asthma in childhood and its subsequent progression. Infants from parents with a history of asthma or atopy are recruited at birth and followed up prospectively at regular, pre-determined intervals and upon manifestations of acute respiratory signs. During the follow-up visits, a nasal aspirate sample is taken by use of a catheter that is inserted into the patient's trachea, and a aspirate taken. This aspirate is then placed in a volume of a transport medium, containing normal saline.

The cohort used in this study gave rise to a total of 739 nasal samples, 165 of which contained RSV by PCR testing. The cohort was followed up for a period of eight years, covering 1998 and 2005. In some cases more than one RSV positive sample was collected from one child, representing a repeat infection.

A 'relative clinical illness' score was used to scale the degree clinical signs following infection. This score was based on the degree of difficulty in breathing and wheeze at the time of infection. The final score in each case was based on the total number of days that the patient was observed to have difficulty in breathing. Laboratory diagnoses of RSV as well as other respiratory pathogens were done by use of established Polymerase Chain Reaction (PCR) techniques targeting conserved regions within these pathogens. In addition to RSV, rhinovirus, mycoplasma, chlamydia, parainfluenza, adenovirus,

influenza, corona virus, human metapneumovirus and boca virus were detected using PCR.

In this project, we identified different RSV genotypes and correlated them to different clinical outcomes.

2. Materials and Methods

2.1 Virus Genotyping by Nucleoprotein gene Restriction Fragment Length Polymorphism (RFLP) patterns.

Virus genotyping based on variation in the N gene was carried by using distinguishing RFLP patterns within a region of the gene with five restriction enzyme recognition sites. This method is adopted from that established by [32].

2.1.1 RNA Extraction

RNA was extracted from nasal aspirate samples obtained from children with respiratory illness using the QIAamp viral RNA mini kit. All extraction procedures were carried out according to the manufacture's instructions. Briefly, 140ul of nasal aspirate samples were mixed with 560ul of a lysis buffer (AVL) containing carrier RNA. This was then mixed by pulse vortexing for 15 seconds, followed by a 10 minute incubation at room temperature. This was then followed by a brief centrifugation to remove drops from the inside of the microcentrifuge tube containing the mixture. 560UI of 99% ethanol was then added and mixed by pulse vortexing for 15 seconds, followed by a brief centrifugation. This mixture was then aliquoted into a QIAamp mini column and centrifuged at 8000 revolutions per minute (rpm) for 1 minute. 500UI of the wash buffer AW1 was then added into the spin column, which was centrifuged at 800rpm for 1 minute. 500UI of a second buffer (AW2) was added into the spin column, after wich the column was centrifuged at 14000 rpm for 3 minutes. 60UI of the elution buffer AVE were then used to elute RNA bound on the QIAamp spin column into a clean microcentrifuge tube. The purified RNA was then immediately stored at -80C.

2.1.2 Complimentary DNA (cDNA) Synthesis

cDNA was synthesised from RNA extracted from NPA samples using the Qiagen Omniscript Reverse Transcription kit. All of the reagents (10X buffer, 5mM dNTPs, 250ng/ul random primers and nuclease free water) were thawed and vortexed to mix, prior to use. 0.5ml tubes were labelled with the sample names and dates. The reagents were added to one 0.5ml tube according to the following table (Table 1), resulting in a mix for the number of samples that were to be used (e.g. if working with 5 samples and a negative control, then the volumes will be multiplied by 6). Reagent mixing and preparation were done in the PCR clean room.

<i>Reagent</i>	<i>Volume per sample</i>	<i>Final concentration</i>
10X RT buffer	4.0 ul	1X
5mM dNTPs*	4.0ul	
250ng/ul random primers	1.0ul	
Omniscript RT enzyme	1.0ul	
Water (provided)	10.0ul	

The contents of the tubes were mixed by vortexing and then 20ul aliquoted into each of the sample tubes. The caps on the tubes were then closed and taken to the main lab where 20ul of RNA for each sample was added to the corresponding tube of cDNA. The tubes were then capped and incubated at 37C for 1 hour and then either used for PCR directly or frozen until needed at -20C.

2.1.3 Primary N gene PCR from cDNA

The N gene PCR was carried out using the Qiagen Taq mastermix kit. Reagents were taken to the PCR

clean room and mixed as shown in table 2. This reagent mix was then taken to the main lab where cDNA was added.

<i>Reagent</i>	<i>Volume per sample</i>
Qiagen PCR Mastermix	50ul
25mM MgCl ₂	8ul
5uM first round primers	1ul
Water	21ul
cDNA	20ul

The primers used (forward primer - 5'GTCTTACAGCCGTGATTAGG', reverse primer - 5'GGGCTTTCTTTGGTACTTC'.) amplified the intergenic segment between nucleotides 532 of the N gene 164 of the P gene. The resultant PCR product, was about 808 base pairs in length and was used as a template for the amplification of a secondary PCR product. The primers were reconstituted to a working concentration of 5uM. The PCR thermocycling reaction was then carried out under the following conditions; 94C for 2 minutes, 94C for 1 minute, 50C for 1 minute, 72C for 1minute, The second to fourth step were repeated for 35 cycles and finally, 72C for 5 minutes followed by cooling at 4C. Once the reaction was complete, the PCR products were used directly for the secondary PCR amplification.

2.1.4 Secondary N Gene PCR

The secondary N gene PCR was carried out using the Qiagen Taq PCR mastermix kit. Primers (foward primer (N1) 5'GGAACAAGTTGTTGAGGTTTATGAATATGC3' and reverse primer (N2) 5'

CTTCTGCTGTCAAGTCTAGTACACTGTAGT3') were used to amplify part of the carboxy terminal part of the N gene. The region amplified in the secondary PCR was on the N gene between nucleotides 858 and 1135.

25ul of the mastermix was added to 23ul water and 1ul N1+N2 primer mix per sample to a tube and mix in the PCR clean room. 49ul of this PCR mix was aliquoted to separate sample tubes. These reaction mixtures were then taken to the main lab and 1ul of the first round PCR product was added. The tubes were vortexed to mix and put on a PCR machine which was then run with the following reaction conditions; 95C for 2 minutes, 95C for 45 seconds, 54C for 45 seconds, 72C for 1minute, steps 2-4 were repeated for 30 cycles and finally 72C for 5 minutes (final extension) followed by cooling at 4C. Once the reaction was completed, the samples were either kept at -20C for storage or used directly for agarose gel electrophoresis.

2.1.5 Agarose Gel Electrophoresis

N gene PCR products were subjected to agarose gel electrophoresis to determine whether amplification had taken place. First a loading dye consisting of 3g glycerol, 7ml of sterile distilled water and a drop of bromophenol blue was prepared. 3ul of the loading dye was then added 10ul of the PCR product and mixed by briefly vortexing. A 2% agarose gel was prepared by adding 2 grams of agarose to 200ul of 1X Tris Borate EDTA (TBE) buffer in a conical flask. This was then heated to high power in a microwave to dissolve the agarose. Once dissolved, the mixture was placed under a stream of cold water to cool it down. Once cooled, 3ul of 10mg/ml ethidium bromide was added and the flask swirled around for a few seconds to mix. A gel tank that had been previously prepared was placed on a flat surface and the gel poured in. A gel comb was inserted into place to create sample loading wells. This was left for about 30 minutes to allow the gel to set. Once set, 8ul of the prepared samples were

aliquoted into the well. In addition to the sample a 1kb DNA ladder (consisting of the ladder, purified, sterile distilled water and the loading dye) was added to the first well on the gel. Once the ladder and all the samples had been loaded, the gel tank was connected to a power pack and run at 100mA for 1 hour. The gel was then visualised using a ultraviolet (UV) light reader.

2.1.6 Purification of PCR products.

PCR products that were clearly visible on the agarose gel were then subjected to purification using the QIAquick PCR purification kit. All the procedures were carried out according to the manufacturer's protocol. Briefly, 100ul of the PCR product was added to 500ul of the purification buffer (PB). The sample was then applied into a QIAquick spin column and centrifuged at 13000rpm for 1 minute. The bound products in the spin column were washed using 750ul of the wash buffer PE. This was followed by a 1 minute centrifugation at 13000rpm. 50ul of the elution buffer EB was finally used to elute the purified PCR product by pipetting it into the spin column followed by a 1 minute centrifugation at 13000rpm.

2.1.7 Restriction Fragment Length Polymorphism (RFLP) analysis of purified N gene products through restriction enzyme digestion of purified PCR products.

To determine unique RFLP patterns that would be used to assign individual clinical samples to NP (N Pattern) genotype groups, restriction enzymes were used to digest purified N gene products. The enzymes used were HindIII, PstI, BglII and RsaI. The restriction digestion reaction mix for each sample was set up according to the table below:

<i>Reagents</i>	<i>Volume per sample</i>				
	<i>HindIII</i>	<i>PstI</i>	<i>BglIII</i>	<i>HaeIII</i>	<i>RsaI</i>
10X digestion buffer	2ul	2ul	2ul	2ul	2ul
Restriction enzyme	0.5ul	0.5ul	0.5ul	0.5ul	0.5ul
Sterile Distilled Water	10.5ul	10.5ul	10.5ul	10.5ul	10.5ul
Cleaned PCR product	7ul	7ul	7ul	7ul	7ul

Once prepared, the reaction tubes were incubated in a 37C water bath for 1 hour. After this incubation period, the samples were subjected to agarose gel electrophoresis using a 2% gel and later visualised on a UV reader. The resultant patterns were interpreted using the following table and assigned into the appropriate genotypes accordingly.

NP Group	Enzymes (cut=+; uncut=-)					Group A or B RSV
	<i>HindIII</i>	<i>PstI</i>	<i>BglIII</i>	<i>HaeIII</i>	<i>RsaI</i>	
NP1	-	-	-	-	+	B
NP2	-	-	-	+	+	A
NP3	-	-	+	-	+	B
NP4	-	-	+	+	+	A
NP5	+	-	-	+	+	A
NP6	-	-	+	-	+	B
NP7	-	+	-	-	*ND	
NP8	+	+	-	-	-	

*ND - Not Determined

2.2.1 Virus Genotyping of the attachment protein (G) gene.

The virus was also genotyped using variation within the attachment protein gene. Two methods were used for this purpose; RFLP of purified PCR products representing part of the gene as well as phylogenetic clustering of sequenced gene products followed by phylogenetic cluster analysis using the DNAmI software.

2.2.2 G gene PCR using a cDNA template.

cDNA was used to generate amplicons of the G genes. Two sets of primers were used to target two different regions of the RSV G gene. In the first set of reactions the forward primer (AG1), 5' GGA TCC CGG GGC AAA TGC AAA CAT GTC C 3' and the reverse primer (AG4) 5' GGT ATT CTT TTG CAG ATA GC 3' were used to amplify G gene nucleotides 1 to 584 of RSV group A isolates. In the second set of reactions, the forward primer (BG1) 5' CCT GCA GGC AAT GAT AAT CTC AAC CTC 3' and the reverse primer (BG3) 5' GAA TTC TCG AGT GGA GGG ATT GCT GTT GG 3' were used to amplify G gene from nucleotide 154 to the end of the G gene of group B RSV isolates.

Reaction mixtures for this PCR reaction were set up in the PCR clean room according to the following table:

<i>Reagent</i>	<i>Volume per sample</i>
<i>Taq Mastermix</i>	25ul
Primers (AG1+AG4) or (BG1+BG3)	2ul
Sterile Distilled Water	18ul

45ul of this mix was this mix was then taken to the main lab where 5ul of cDNA was then added. The samples were then loaded onto the PCR machine and thermocycling done according to the following parameters:

1. 95C for 2 minutes
2. 95C for 45 seconds
3. 54C for 45 seconds
4. 72C for 1minute
5. Repeat 2-4 for 30 cycles
6. 72C for 5 minutes (final extension) followed by cooling.

Once the reaction was complete, 8ul of each sample was run on a 1% agarose gel, to determine the presence of bands. The positive samples were then subjected to purification using the QIAamp PCR purification kit.

2.2.3 Restriction Fragment Length Polymorphism (RFLP) analysis of purified G gene products through restriction enzyme digestion of purified PCR products.

Digestion of G gene PCR products was done using four restriction enzymes. These are AluI, TaqI, MboI and MseI. The resultant patterns were then analysed using a method established by Cane and Pringle [32] and subsequently assigned into G gene RFLP based genotype groups, referred to as SHL genotypes.

The reaction mixtures for the digestion reactions were set up according to the following table:

<i>Reagents</i>	<i>Volume to add per sample</i>			
	<i>AluI</i>	<i>TaqI</i>	<i>MboI</i>	<i>MseI</i>
10X buffer	2ul	2ul	2ul	2ul
Restriction enzyme	0.5ul	0.5ul	0.5ul	0.5ul
Sterile Distilled Water	10.5ul	10.5ul	10.5ul	10.5ul
Cleaned PCR product	7ul	7ul	7ul	7ul

Once the reaction had been set up, the reaction tubes were incubated on a 37C water bath for 1 hour (except for TaqI) which was incubated at 60C for 1 hour. The reaction products were then run on a 2% gel. The resultant patterns were then analysed based on the method outlined by Cane and Pringle [32].

2.3.1 Phylogenetic analysis of sequenced products from purified PCR products

Purified G gene and N-P transgenic PCR products were sent to a commercial sequencing service for sequencing. A primer walk sequencing reaction was done using the forward primers used to amplify

both gene G and N-P gene segments. The results of the sequencing reactions were edited using Bioedit software to correct basecaller and ambiguous nucleotide assignment errors. The sequences were then aligned using ClustalX software.

The sequence alignments were then subjected to phylogenetic analysis by maximum likelihood using DNAmI software. Tree diagrams were drawn from DNAmI output files using Treeview software. The trees were unrooted. Using the diagrammatic tree output, closely related clusters were chosen and sequences from clinical samples that clustered in a closely related group were assigned to a phylogenetic genotype.

Using this classification methodology, four G gene and four N-P transgenic gene phylogenetic genotype classifications were generated.

2.4.1 Cytokine concentration measurement using the Luminex assay.

Cytokine levels in nasal aspirate samples were measured using the Luminex assay. A 9 plex assay kit was used to measure the level of cytokines in nasal aspirate samples. This kit was designed to measure the levels of IFN- α , IP10, RANTES, MDC, IL13, IL5, IL6, TNF and IL8. A standard sample containing known concentrations of these mediators was supplied by the kit manufacturer as a lyophilized sample. This standard was reconstituted using 250 μ l deionized water into a glass vial to give a 10,000 pg/ml. The vial was inverted a few times and left to sit for 5 minutes. The standard was then transferred to a microcentrifuge and vortexed. After vortexing, a five times times serial dilution was performed as follows; five microcentrifuge tubes were labeled 2000, 400, 80, 16, and 3.2 pg/ml. 200 μ l of Assay Buffer was added to each of the five tubes. 50 μ l of the 10,000 pg/ml standard was then added to the 2000pg/ml tube and mixed well. 50 μ l of the 2000 pg/ml standard was transferred to the 400pg/ml tube

and mix well. 50 μ l of the 400 pg/ml standard was transferred to the 80pg/ml tube and mixed well. 50 μ l of the this was added to the 16pg/ml, tube and mixed well. 50 μ l of the 16 pg/ml standard was then added to the 3.2pg/ml tube. The 0 standard (Background) was the Assay Buffer.

To ensure that assay accuracy, quality control samples supplied by the manufacturer were prepared. Two quality control samples were reconstituted with 250 μ l deionized water. The vials were allowed to set for about 5 minutes then inverted several times times to mix and finally vortexed. The controls were then transferred to microcentrifuge tubes, to be used later on in the immunoassay.

Antibody immobilized beads with antibodies specific for the mediators of interest were then prepared. The beads were sonocated for 30 seconds and then vortexed for 1 minute. 0.15 ml was aliquoted from antibody bead tube to a mixing bottle into which 2.85 ml assay buffer was added, resulting in a final volume of 3 ml. This was then vortexed and stored at -4C.

To carry out the Luminex immunoassay, A filter plate was blocked by pipetting 200 μ l of assay buffer into each well of the microtiter plate. The plate was sealed and placed on a plate shaker for 10 minutes at room temperature to mix. The assay buffer was removed by use of a vacuum pump and any excess wash buffer removed from the bottom of the plate by blotting on an absorbent paper towels. 25 μ l of assay buffer was then added to the 0 standard (Background). 25 μ l of assay buffer was then added to the sample wells followed by addition of 25 μ l of each standard or control into the appropriate wells. 25 μ l of sample was then added into the appropriate wells followed by 25 μ l of mixed beads to each well. The plate was then sealed with aluminium foil, and incubated with agitation on a plate shaker overnight at 2C. After overnight incubation, the plates were allowed to warm to room temperature before continuing with the assay. The fluid on the plates was gently removed by vacuum and washed 3 times with 200 μ l/well of wash buffer (supplied by the manufacturer). The buffer was removed by

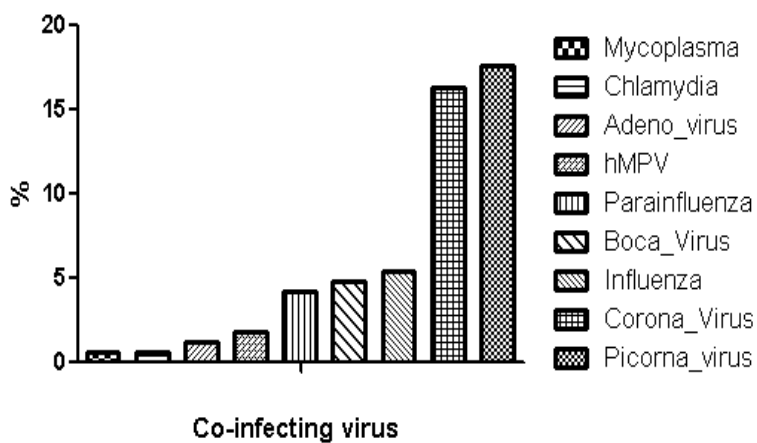
vacuum filtration between each wash. 50 μ l of detection antibody was added into each well. The plate was then sealed with aluminium foil, and incubated with agitation on a plate shaker for 30 minutes at room temperature. 50 μ l of Streptavidin-Phycoerythrin (SAPE) was added to each well containing the 50 μ l of detection antibody. The plate was again sealed with aluminium foil, and incubated with agitation on a plate shaker for 30 minutes at room temperature. The fluid in the plates was then removed by vacuum and the plate washed 3 times with 200 μ l/well wash buffer, removing wash buffer by vacuum filtration between each wash. 150 μ l of sheath fluid was added to all wells. The plates were then sealed with aluminium foil and beads resuspended on a plate shaker for 5 minutes. The plate was then read on a Luminex instrument.

3. Results

3.1 Effect of viral co-infection on the duration of illness.

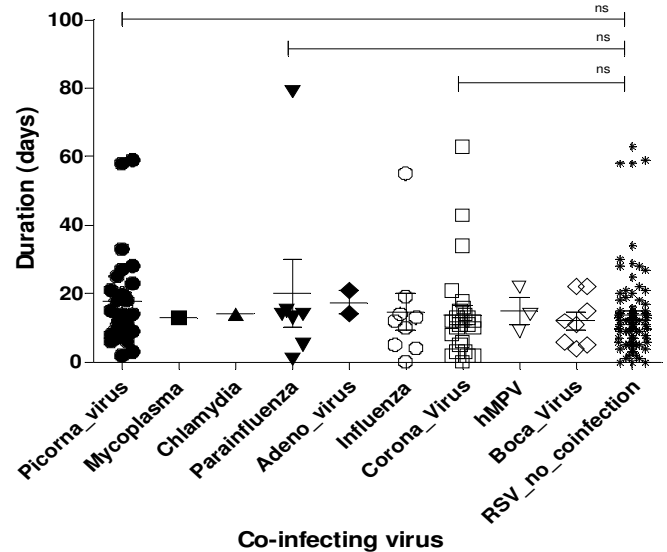
In order to determine whether viral co-infection could result in increased disease severity and potentially confound the possible role of genotype-mediated severe disease, a comparison of viral co-infection and disease severity was carried out. The results are shown in figure 1.

RSV positive cases with viral co-infections



(a)

Effect of coinfection on duration of illness



(b)

Figure 1 (a) Proportion of RSV infected children with co-infections from other respiratory pathogens. (b) The effect of viral co-infection on the duration of respiratory symptoms.

As shown in figure 1a the most prevalent viral co-infection in the RSV positive cohort was picorna virus. 73 children out of 165 RSV positive cases had viral co-infections, representing 44.2%. 28(16.97%) had picornavirus, 1(0.61%) had mycoplasma, 1(0.61%) had chlamydia, 7(4.24%) had parainfluenza, 2(1.21%) had adenovirus, 10(6.06%) had influenza, 26(15.76%) had corona virus, 3(1.82%) had (human metapneumovirus)hMPV, while 8(4.85%) had bocavirus. A one way analysis of variance (ANOVA) was carried out comparing the mean duration of illness between the group of RSV infected children who had no co-infections and the various groups of children who had one or more co-

infections in addition to RSV. The Kruskal-Wallis non parametric test was used at a significance level (α) of 0.05. The mean duration of illness in the group of RSV infected children without co-infections did not vary significantly from any of the groups with viral co-infections.

3.2 Effect of variation in the N gene characterized by RFLP

3.2.1 General cohort characteristics

Based on variation in the N gene, RFLP patterns of PCR products were examined. The resulting patterns were assigned into genotype groups, based on a classification established by Cane and Pringle [32]. The sex and age distribution of the various genotypes were analysed as shown in figures 2a and 2b.

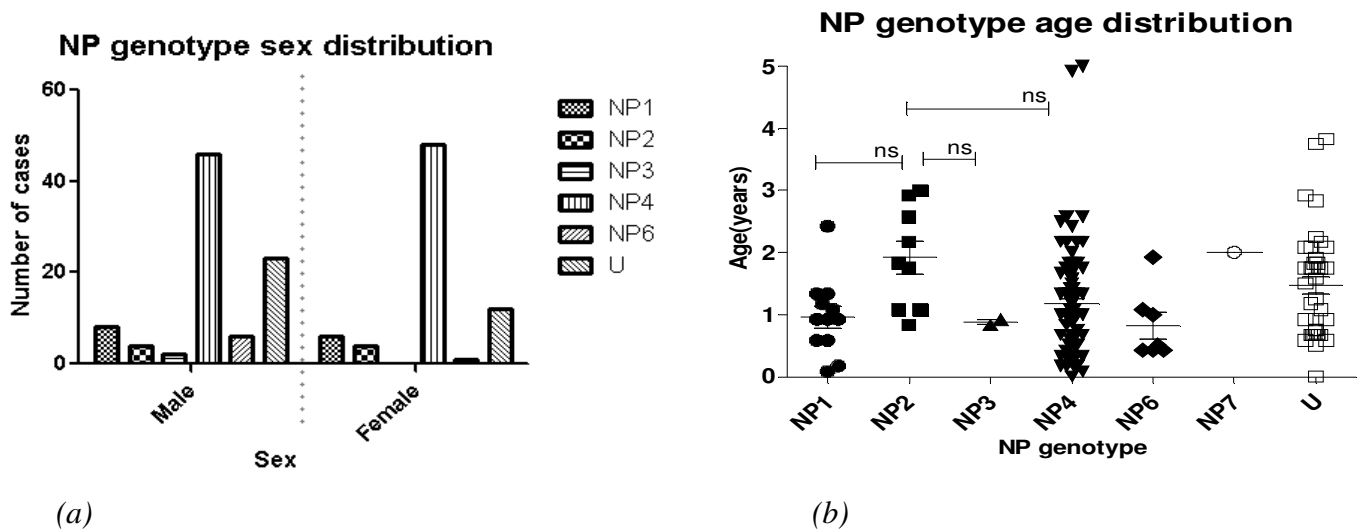


Figure 2 (a) NP genotype distribution patterns among males and females (b) Age distribution of different NP genotypes. (U refers to the isolates that could not be typed due to the absence of an N gene PCR product)

Similar genotype infection patterns were seen both in males and females (fig 2a). There were also no significant differences in the mean age of infection by the different genotypes (fig 2b).

3.2.2 Effect of RFLP based N gene variation on viral load

To determine whether variation in the N gene had any effect on viral load, a quantitative real time TaqMan PCR was carried out. This PCR targeted conserved regions of the L (viral polymerase protein) gene. An L gene standard of known concentration was used to determine the L gene copy number in the clinical samples.

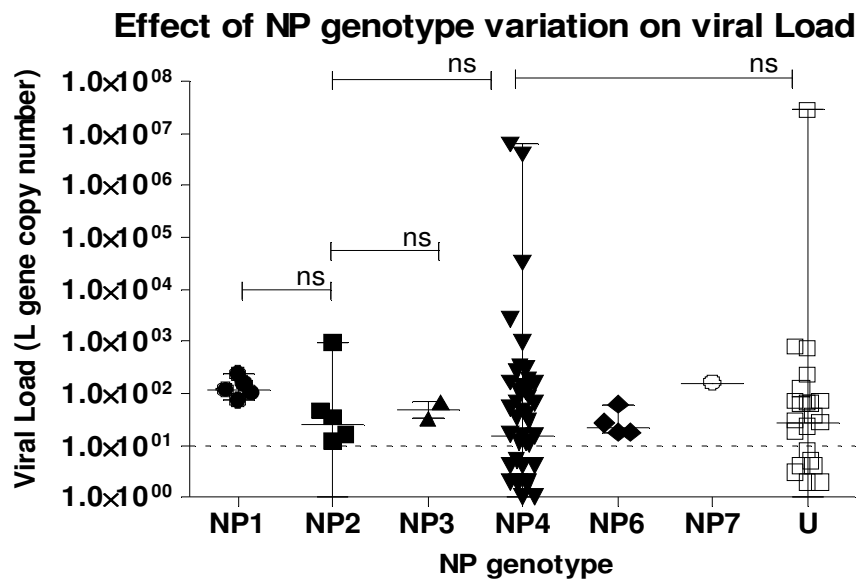


Figure 3 L gene copy numbers among the various NP genotypes. The y axis represents the L gene copy number, while the X axis represents various NP genotypes. U represents the viral isolates that could not be typed by RFLP. The dotted line represents the detection limit of this assay.

Mean L gene copy numbers from different NP genotypes were compared using one way ANOVA ($\alpha=0.05$) (Fig 3). There were no statistically significant differences in viral load among the NP genotypes.

3.2.3 Effect of RFLP based N gene variation on Duration of clinical illness

To determine whether genetic variations within the N gene resulted in varying manifestations of clinical disease, we compared NP genotype patterns to duration of illness.

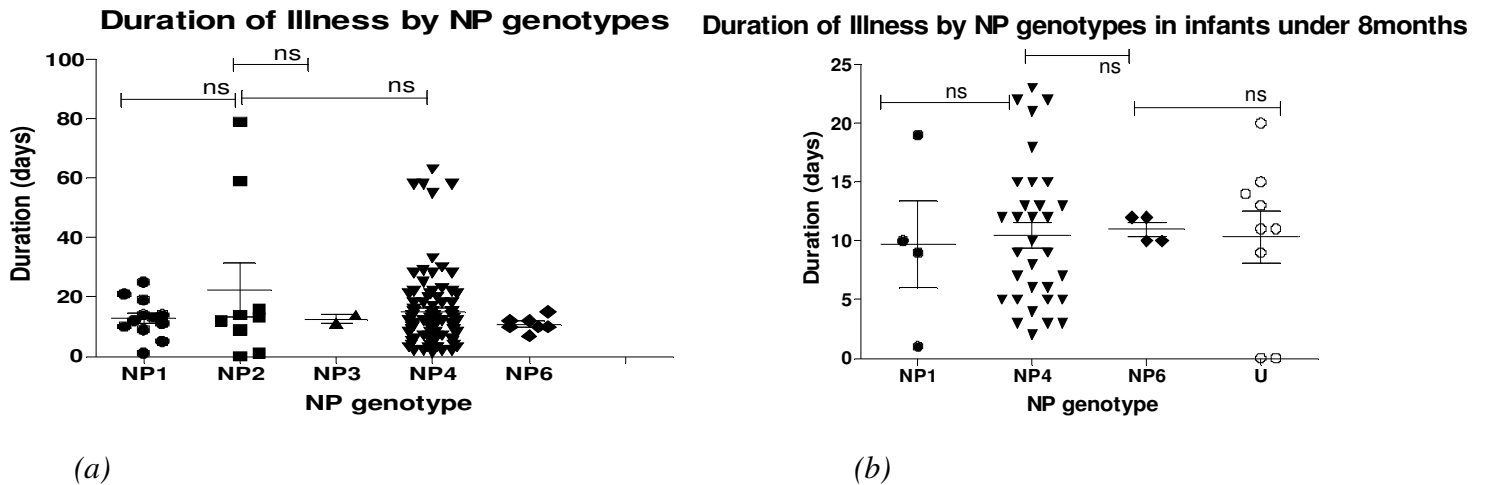


Figure 4 The duration of illness in children infected by different N-P genotypes. (a) Mean duration of illness was compared for different NP genotypes regardless of age. (b) Comparison of duration of illness was restricted to infants under 8 months.

The clinical measure used to determine the severity of illness was the duration (in days) that the patient had difficulty in breathing (duration of illness). The mean duration of illness was grouped according to NP genotype and compared using one way ANOVA ($\alpha=0.05$).

The data were analysed using two approaches. In one group mean duration of illness in all infants infected with different NP genotypes was compared regardless of age. In the other group, this analysis was restricted to infants aged 8 months or less. No significant differences in mean duration of illness among the various genotypes were seen in either group (Fig. 4).

3.3 Effects of phylogenetic cluster based N-P (nucleoprotein-phosphoprotein) gene variation on viral load and duration of illness

In addition to RFLP, N-P PCR amplicons were subjected to sequencing reactions and the resultant sequences clustered using maximum likelihood phylogeny. The sequences used in this analysis encompassed the carboxyterminal end of the N gene and the amino terminal residues of the P gene. Samples whose sequences were closely related were clustered together in a phylogenetic tree. The phylogenetic tree derived from this analysis is shown in Fig 5, while the deduced, closely related clusters that were selected for subsequent analysis are shown in table 1.

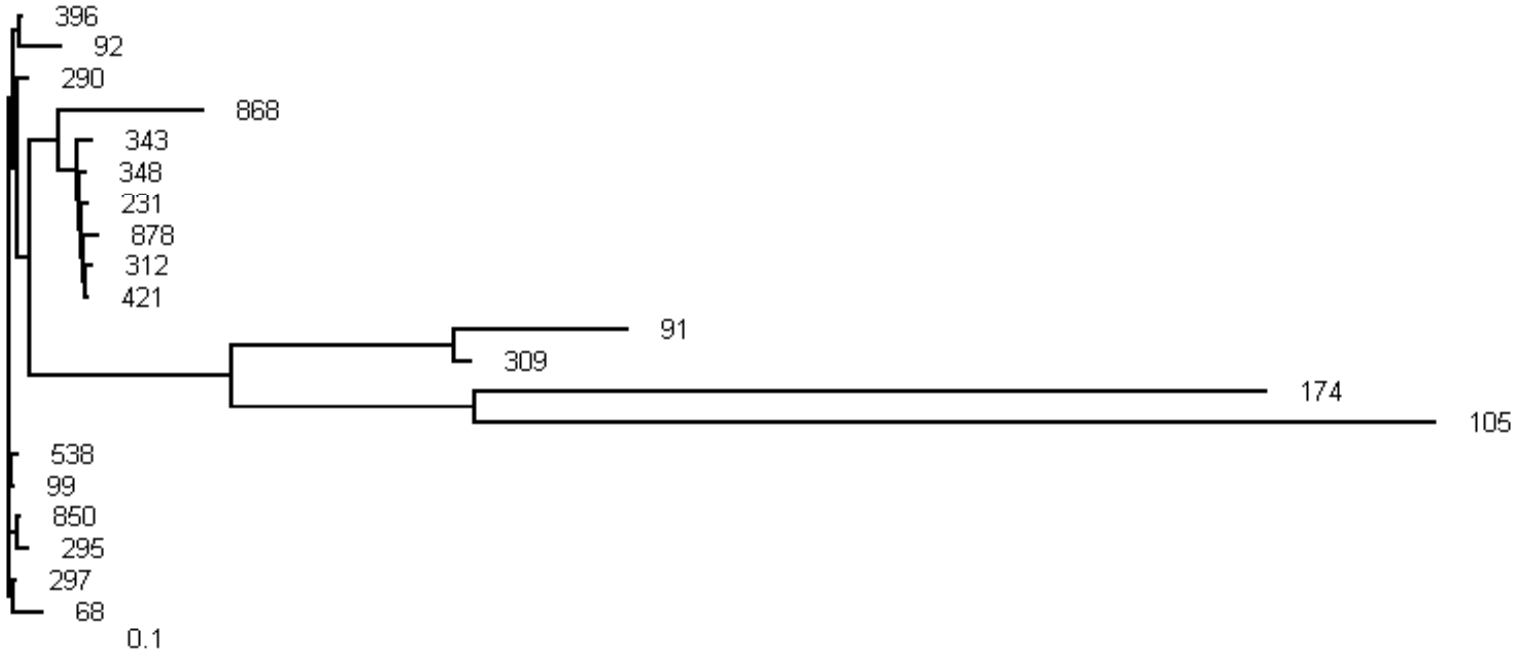


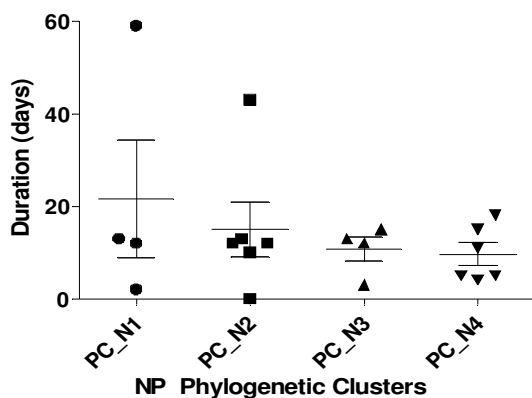
Figure 5 Phylogenetic tree generated based on analysis of sequence data by the computer program DNAmI. The numbers represent the anonymised identifiers of individual children samples. The tree diagram was generated by the computer program Treeview. Closely related clusters were grouped according data shown in table 1.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
396	848	91	53
92	343	309	99
290	348	174	850
	231	105	295
	878		297
	312		68
	421		

Table 1. Selected N-P Phylogenetic Clusters based on maximum likelihood phylogenetic analysis of N-P gene sequences.

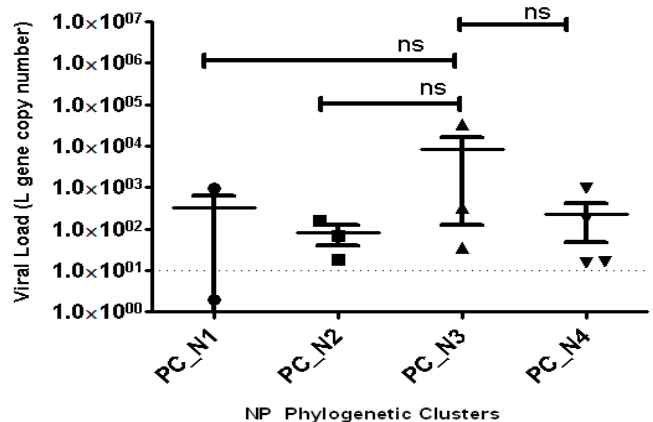
To determine whether variation in the N-P intergenic segment analysed by phylogenetic analysis had any effect on clinical outcome, the duration of illness was grouped according to the phylogenetic cluster and compared by one way ANOVA ($\alpha=0.05$). Similarly, to assess the effect of N-P genotype variation as on viral load, L gene copy numbers of the infecting viruses were grouped according to the N-P phylogenetic genotype and compared using one way ANOVA. The results of these analyses are shown in Fig 6a and b.

NP Phylogenetic Clusters vs Duration of illness



(a)

NP Phylogenetic genotypes vs Viral Load



(b)

Figure 6. Comparison of (a) duration of illness and (b) viral load among different phylogenetic clusters derived from the N and P intergenic region. Viral load was derived from the copy number of the polymerase (L) gene.

3.4 Clinical effects of G gene variation

3.4.1 G gene Phylogenetic cluster Analysis.

Purified products representing nucleotide 1 to 584 of group A RSV isolates were sequenced. The sequences were then aligned using clustalX software and subjected to phylogenetic cluster analysis using the DNAmI program. The output from this program was used in the TreeView software to generate the phylogenetic tree shown in Fig 7.

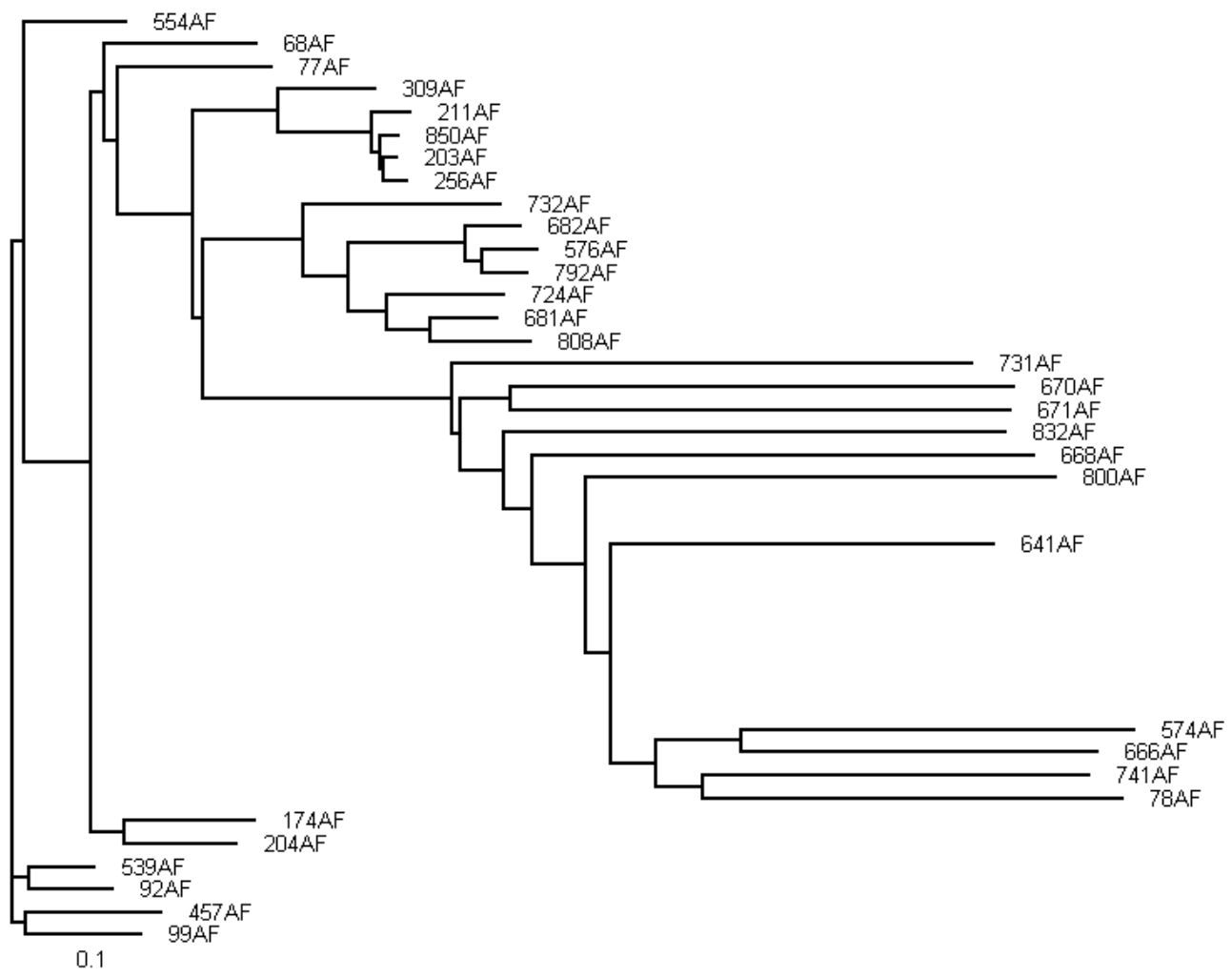


Figure 7 Phylogenetic tree generated from G gene sequences representing the gene segment between nucleotide 1 and 584. The tree was generated using the maximum likelihood model by the DNAmI software. Clusters of closely related sequences were selected and are shown in table 2.

Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8
309	68	554	732	731	641	174	539
211	77		682	670	574	204	92
850			576	671	666		457
203			792	832	741		99
256			724	668	78		
			681	800			
			808				

Table 2. *G* gene Phylogenetic Clusters based on maximum likelihood pylogenetic analysis of nucleotide 1-584 of *G* gene sequences.

Using these results, six clusters of closely related samples were selected and defined as *G* gene phylogenetic clusters. These clusters were then used to determine whether any particular group of closely related viruses was associated with increased duration of illness. This is shown in Fig. 8.

G Phylogenetic Clusters Vs Duration of Illness

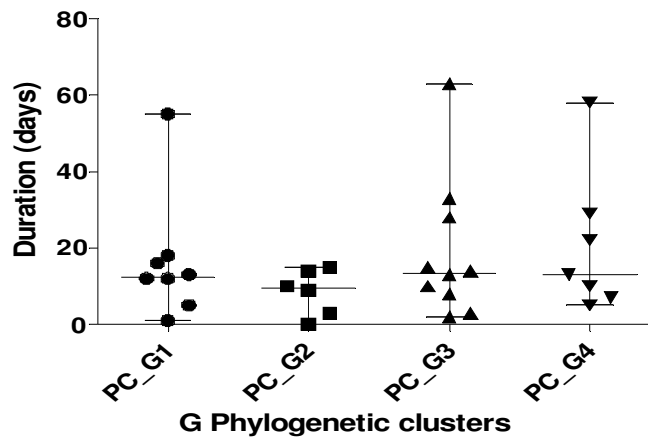


Figure 8. Comparison of the effect of different phylogenetic clusters on the duration of respiratory illness

A one way ANOVA was used to test for a statistically significant difference in the mean duration of illness from each group at a significance level of 0.05. No significant associations were found.

3.5 Measurement of cytokine levels in nasal samples using the luminex assay

The levels of different immune mediators within nasal aspirate samples obtained from children with an antigen confirmed diagnosis of RSV was measured using the Luminex assay. The assay detection range was from 2000pg/ml to 3.2pg/ml. Figure 9 is a diagrammatic representation of the general production patterns of the different cytokine measured classified according to RFLP based variation of the N gene.

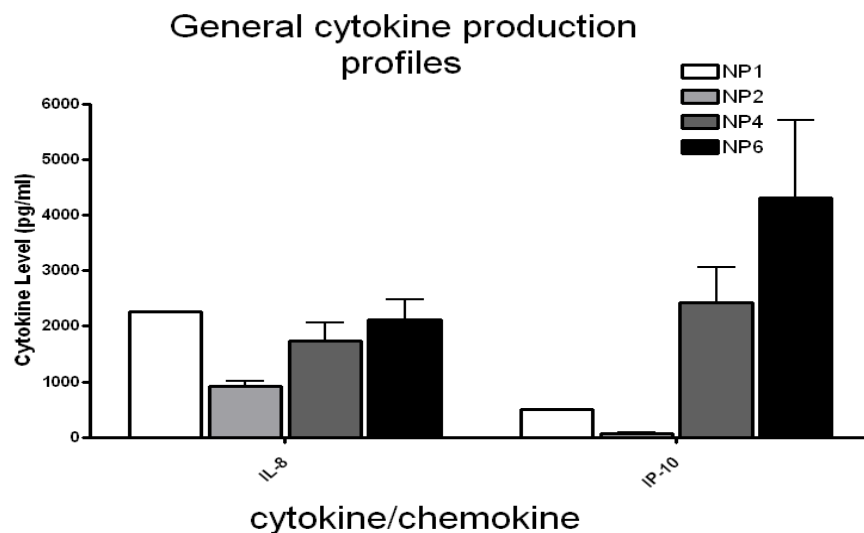
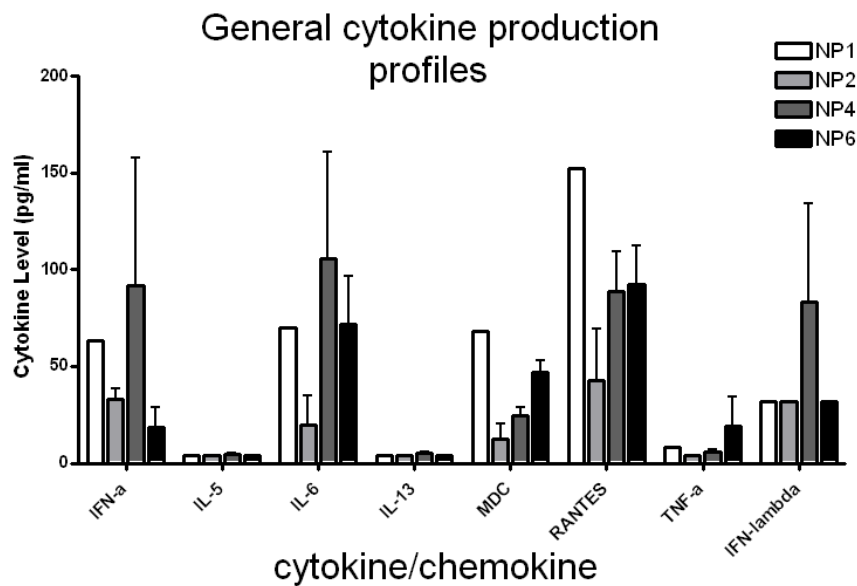


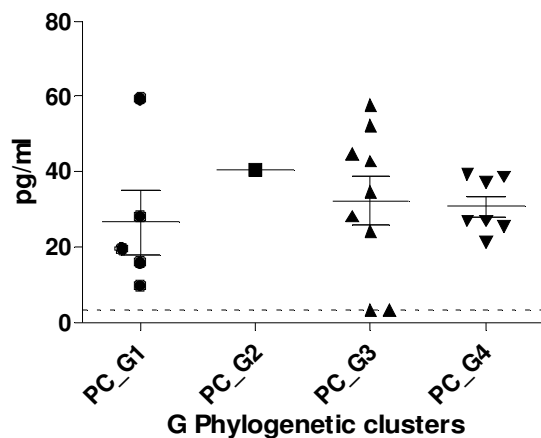
Figure 9 Cytokine/chemokine production profiles among different NP RFLP genotypes

The NP4 genotype was generally associated with elevated production of IFN- α , IL-6 and IFN- λ . NP6 genotype was associated with increased production of the chemokines MDC and RANTES. While the NP1 genotype appears to be associated with elevated levels of MDC and RANTES only one sample from this genotype was available for analysis making it difficult to make meaningful statistical inference from the data. In general IL-8 and IP10 were produced at much higher levels than other cytokines assayed. In contrast IL-5 and IL-13 were produced at extremely low concentrations among all genotypes tested.

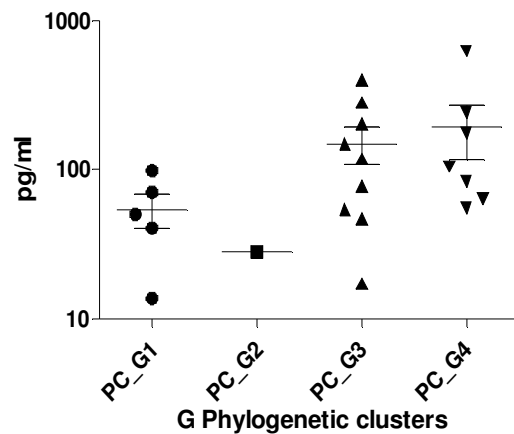
3.6 Cytokine production by G gene phylogenetic Clusters

To determine whether any group of closely related viruses was associated with a significantly different profile of cytokine/chemokine production, the cytokine/chemokine concentration in the different G gene phylogenetic clusters were analysed. The levels of IFN- γ , IL6, IL-8, IP-10, MDC, RANTES, IFN- λ and TNF- α were analysed. The results are shown in Fig 10.

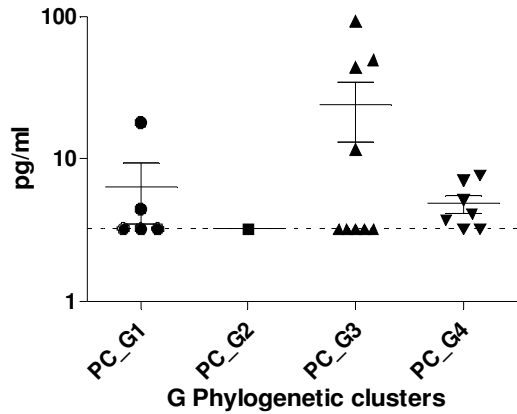
MDC Production by G phylogenetic Clusters



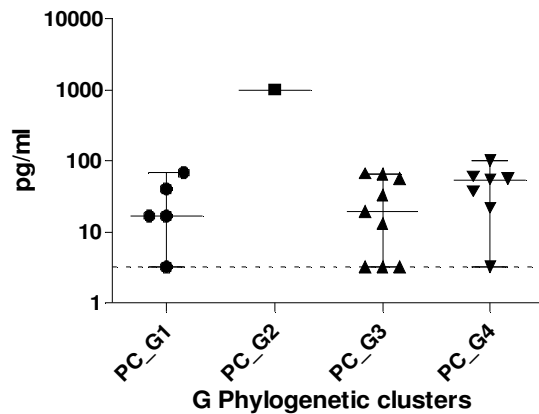
RANTES Production by G phylogenetic Clusters



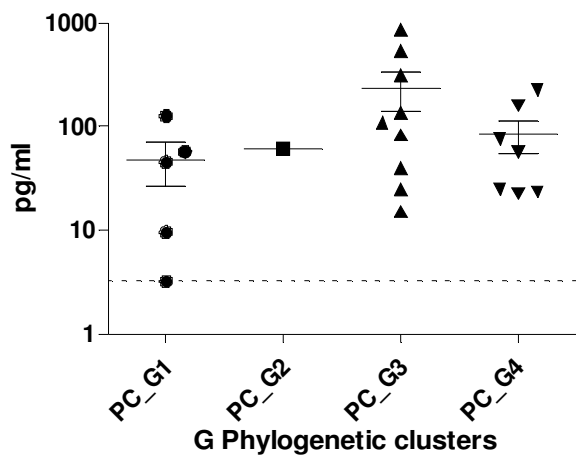
TNF- α Production by G phylogenetic Clusters



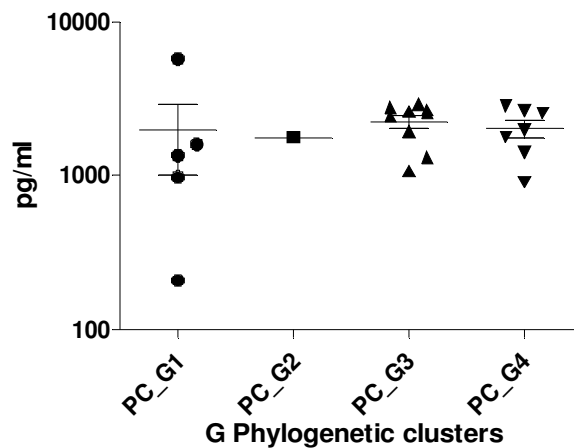
IFN- γ Production by G phylogenetic Clusters



IL-6 Production by G phylogenetic Clusters



IL-8 Production by G phylogenetic Clusters



IP-10 Production by G phylogenetic Clusters

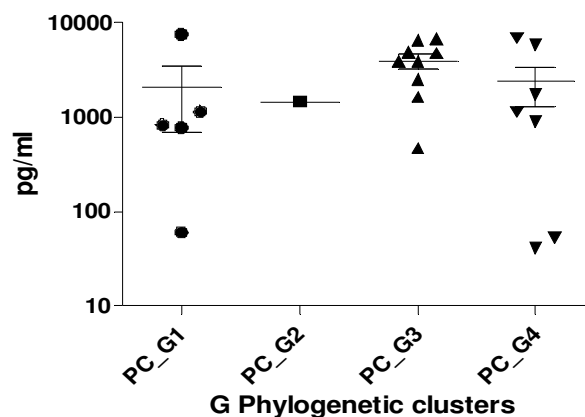


Figure 10. Cytokine/chemokine production profiles among different G gene phylogenetic clusters.. The dotted line represents the lowest measurable concentration (3.2pg/ml). No data is shown for IFN- λ since all genotypes produced cytokines below the detection limit of the assay. The y axis is logged.

To determine whether any G gene phylogenetic clusters were associated with a significant increase in cytokine/chemokine production a one way ANOVA ($\alpha=0.05$) was carried out using the Kruskal-Wallis Test non parametric test. Dunn's post test was used to control for the familywise error rate that arises from multiple comparisons.

No significant associations were found between G gene phylogenetic clusters and elevated cytokine /chemokine responses.

3.7 Th1/Th2 Cytokine skew Pattern among G phylogenetic clusters

Using IFN- α as a Th1 cytokine marker and IL-6 as a Th2 cytokine marker, a ratio of these responses was calculated for each respective G gene phylogenetic cluster. The results were compared using a one way ANOVA ($\alpha= 0.05$). The results of this analysis are shown in Fig 11 below. This analysis excluded the second G phylogenetic cluster, since it only had one observation.

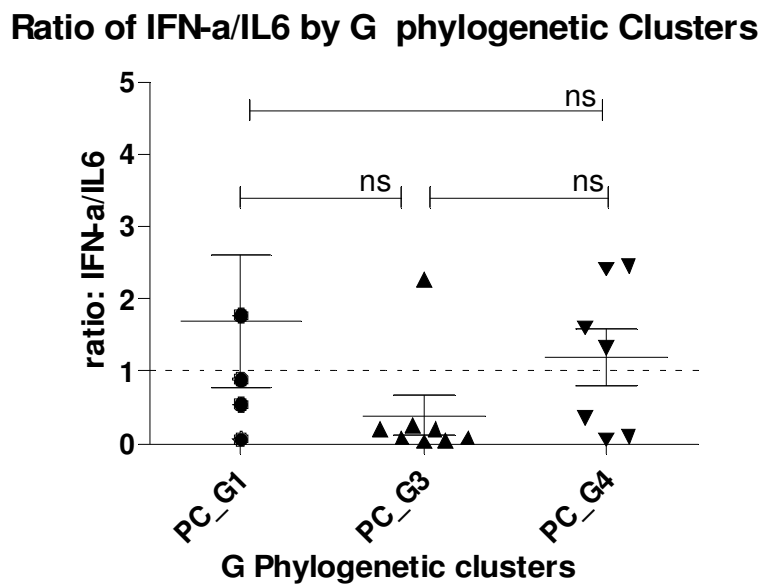


Figure 11 The ratio of IFN- α /IL6 cytokine concentrations among different G gene phylogenetic clusters. The dotted line is set to 1.

There was no significant difference in the IFN- α /IL-6 cytokine ratio among the various G gene phylogenetic clusters investigated.

4. Discussion.

4.1 Defining effect confounders of severe disease

The severity of RSV disease must depend on a multitude of factors: host genetics, the past and present environment, the previous history of infection, the age of the host and the exact nature of the infecting pathogen can all be assumed to influence disease. In this project an attempt was made to associate variation in viral genes with disease severity. To do this, it was necessary to determine whether other possible determinants of severity might have played a role in the final clinical outcome of infection. To this end, other respiratory co-infections at the time of RSV challenge were investigated.

Viral co-infection has been proposed to increase the risk of severe disease significantly. For example rhinovirus infections are associated with a five-fold increase in RSV disease [19]. Antigen confirmed diagnoses of nine other pathogens that cause respiratory illness in children had been carried out on each of the COPSAC samples by use of a polymerase chain reaction targeting conserved gene regions in these viruses. However, our analysis showed that there was no significant risk of prolonged respiratory symptoms associated with an RSV infection in the presence of a co-infecting respiratory pathogen.

This result conflicts with previous reports that suggest increased severity of illness is associated with multiple infections. A possible reason for this is fact that the definition of clinical severity varies from study to study. It is unfortunate that in our study, clinical severity could only be defined by the number of days that the child was deemed to have suffered difficulty in breathing as assessed by its primary care giver. This interpretation is subjective and prone to wide inter-observer variance. Another possible explanation might have been the role of other confounding contributors of severe disease that were not investigated in this project. Among these are host genetic variation that are known to be strongly associated with severity of clinical disease. If the relative contribution of host genetic polymorphisms

are a more salient determinant of severe disease than individual virus genotypes, then it is likely that the effect of genotype variation in contributing to severe disease might have been obscured. It would be better in future studies to make allowance for host gene variation, but this would require a very large and expensive study.

Since our results did not show any effect of co-infection, we decided not to stratify our analysis based on co-infection with other respiratory pathogens. Based on N gene RFLP genotypes we then sought to determine whether there were any variations in genotype prevalence based on age and sex. No genotype was associated with an increased infection rate based on either age or sex. This means that neither age nor sex are independent risk factors for genotype selection and consequently subsequent analyses made no attempt to stratify responses based on these phenotypes.

4.2 Effect of genotype variation on viral load.

We then attempted to explore the possible role of variation in the nucleoprotein gene as well as the phosphoprotein gene in determining the viral load. The RSV nucleoprotein as well as the phosphoprotein form part of the virus's ribonucleoprotein complex (RNP), the functional subunit of viral replication. During the process of viral replication, viral RNA is anchored onto the nucleoprotein, which is associated with the phosphoprotein. It is only when the RNA is within this complex, that the viral reverse transcriptase can initiate its polymerizing activity [36].

It has previously been shown that mutations within a region of the phosphoprotein gene lead to dramatic reductions in the viral copy load [37]. Since viral load has been independently associated with severe disease by some authors [16], we sought to determine whether variation in the N and P genes was associated with increased or decreased viral load in children infected by RSV.

Viral load in the clinical samples was measured using a Taqman real time PCR assay which targets conserved regions of the polymerase (L) gene. There were no significant associations between N gene viral genotypes and viral load. To further characterise the viral load in the context of variation within genes that code for RNP protein constituents, PCR products that span the carboxyterminal end of the N gene and the amino terminal end of the P gene were sequenced. This region spans nucleotides 532 of the N gene 164 of the P gene (the entire length of the N gene is 1173 nucleotides while that of the P gene is 723 nucleotides). The sequenced products were aligned and subjected to phylogenetic analysis by maximum likelihood. Closely related phylogenetic clusters of the clinical isolates were then used to compare variations in viral load. No significant association was found between the phylogenetic clusters of closely related sequences and viral load.

While this result was unexpected, it is important to note that the gene region examined in these analyses encompassed only part of both the N and P genes. If this region does not encompass functionally relevant amino acid sequences responsible for viral replication, then the effect of variation there might be minimal. This notion is supported by the work of Asenjo et al who showed that variation in the P gene region encompassing nucleotides 294 and 474 were associated with variations in viral load, suggesting that the functionally important nucleotide residues lie within this region.

4.3 Effect of genotype variation on duration of illness.

Using both RFLP and phylogenetic clustering as markers of variation in the N, P and G genes, we sought to find whether there was an association between variation within these genes and duration of clinical illness in children. NP genotypes that reflected variation within the N gene were used to compare the duration of illness among children regardless of their ages. No statistically significant association linked individual genotypes to increased duration of disease. We then restricted this

analysis to infants less than eight months of age because these infants have been shown to be much more prone to severe illness than older children. However, we could find no significant association between different N gene genotypes and duration of illness. We then clustered the samples based on their phylogenetic relatedness and used these clusters to attempt to find a correlation between any of them and prolonged duration of illness. However no statistically significant associations were found between phylogenetic clusters encompassing the carboxyterminal part of the N gene and the amino terminal part of the P gene. We then moved on to assess whether variation within part of the G gene was associated with increased duration of clinical symptoms. Clusters of phylogenetically related sequences of the G gene, encompassing the region between nucleotide 1 and 584 were used in the analysis. No statistically significant association was found between the G gene phylogenetic clusters and increased duration of disease.

These results indicate that within the gene regions examined, viral variation did not correlate to prolonged illness. The most probable explanation of this observation is the possibility that no relevant immune epitopes were located within the areas examined. An alternative explanation is the possibility that these regions represented viral sequences that were not under functional constraint and could thus be varied without imposing a severe fitness cost to the virus. In other words, though the different viruses differed subtly in the regions examined, the functional consequences of these specific variations to the virus's infection cycle and pathogenicity were undetectable in our analysis.

4.4 Effect of genotype variation on host cytokine production profiles.

We used cytokine production as a general indicator of the nature of the immune response mounted by the host in response to genetically variable viruses. Using the Luminex cytokine measurement assay, we measured the levels of IFN- α , IL-8, RANTES, IP10, IL-6, IFN- λ , IL-13 TNF- α and MDC in nasal

samples derived from children with an antigen confirmed RSV diagnosis. Most children mounted robust cytokine responses with the exception of IL-5, IL-13 and TNF- α . When these responses were classified according to G gene phylogenetic clusters, no association was found between these different genotypes and an elevated cytokine response. Previous work focusing on certain epitopes found in this region, found that they activated T cells and skewed cytokine production profiles in either a Th1 or Th2 direction *in vitro* [45]. When that nucleotide region (nucleotide 162 to 175 of the G gene) was examined in the samples sequenced within this cohort, we found that it was very highly conserved, in agreement with previously published data[44]. This high degree of conservation could mean that similar clonal subsets of T cells were activated by these viruses leading to similar patterns of cytokine responses.

We also examined the effect of genotype variation based on G gene variation on the ability to skew the immune response in a Th1 and Th2 manner. Various investigators have implicated an imbalance in the T helper response profiles in severe RSV disease [26]. We selected IFN- α as a marker for type 1 responses while IL-6 was selected as a marker of type 2 responses. A ratio of the concentration of IFN- α and IL-6 was calculated for each sample. These ratios were then grouped according to the G gene phylogenetic cluster into which the infecting viruses fell. A one way ANOVA ($\alpha=0.05$) was used to test whether there was a statistically significant difference between the Th1/Th2 ratio among the different groups. No statistically significant difference was found, suggesting that the different G gene phylogenetic genotypes did not skew the immune response of the host in either a Th1 or Th2 direction.

This result could again be attributed to the high degree of similarity within epitopes of the G protein that have been shown to affect the outcome of T helper cell profiles. Another possibility is that epitopes that may play an even greater role in this skewage were either located downstream of the G

gene region that was genotyped or may be located on other RSV genes that were not examined in this project.

Overall, it appears from our results, that genotype variation does not influence disease severity, host cytokine immune responses or viral load. However it is important to note that this project only focused on a small part of the RSV genome, whose role in influencing disease or host immunity outcomes may not be as significant as that played by other RSV proteins. To definitively address the role of genotype variation in influencing these outcomes it is important to look holistically at the entire RSV genome as this would provide a more complete assessment on the relative roles played by different genes in the disease process.

References

1. Murata, Y., *Respiratory syncytial virus infection in adults*. Curr Opin Pulm Med, 2008. **14**(3): p. 235-40.
2. Bosis, S., et al., *Role of respiratory pathogens in infants hospitalized for a first episode of wheezing and their impact on recurrences*. Clin Microbiol Infect, 2008. **14**(7): p. 677-84.
3. Power, U.F., *Respiratory syncytial virus (RSV) vaccines--two steps back for one leap forward*. J Clin Virol, 2008. **41**(1): p. 38-44.
4. Elliot, A.J. and D.M. Fleming, *Influenza and respiratory syncytial virus in the elderly*. Expert Rev Vaccines, 2008. **7**(2): p. 249-58.
5. Falsey, A.R., *Respiratory syncytial virus infection in adults*. Semin Respir Crit Care Med, 2007. **28**(2): p. 171-81.
6. Falsey, A.R. and E.E. Walsh, *Respiratory syncytial virus infection in elderly adults*. Drugs Aging, 2005. **22**(7): p. 577-87.
7. Eick, A., et al., *The role of neutralizing antibodies in protection of American Indian infants against respiratory syncytial virus disease*. Pediatr Infect Dis J, 2008. **27**(3): p. 207-12.
8. King, J.C., Jr., et al., *Respiratory syncytial virus illnesses in human immunodeficiency virus- and noninfected children*. Pediatr Infect Dis J, 1993. **12**(9): p. 733-9.
9. Malley, R., et al., *Reduction of respiratory syncytial virus (RSV) in tracheal aspirates in intubated infants by use of humanized monoclonal antibody to RSV F protein*. J Infect Dis, 1998. **178**(6): p. 1555-61.
10. Kim, H.W., et al., *Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine*. Am J Epidemiol, 1969. **89**(4): p. 422-34.
11. Cunningham, C.K., J.A. McMillan, and S.J. Gross, *Rehospitalization for respiratory illness in infants of less than 32 weeks' gestation*. Pediatrics, 1991. **88**(3): p. 527-32.

12. MacDonald, N.E., et al., *Respiratory syncytial viral infection in infants with congenital heart disease*. N Engl J Med, 1982. **307**(7): p. 397-400.
13. Miller, R.B. and B.M. Chavers, *Respiratory syncytial virus infections in pediatric renal transplant recipients*. Pediatr Nephrol, 1996. **10**(2): p. 213-5.
14. Groothuis, J.R., K.M. Gutierrez, and B.A. Lauer, *Respiratory syncytial virus infection in children with bronchopulmonary dysplasia*. Pediatrics, 1988. **82**(2): p. 199-203.
15. Whimbey, E. and S. Ghosh, *Respiratory syncytial virus infections in immunocompromised adults*. Curr Clin Top Infect Dis, 2000. **20**: p. 232-55.
16. DeVincenzo, J.P., C.M. El Saleeby, and A.J. Bush, *Respiratory syncytial virus load predicts disease severity in previously healthy infants*. J Infect Dis, 2005. **191**(11): p. 1861-8.
17. DeVincenzo, J.P. and S.C. Buckingham, *Relationship between respiratory syncytial virus load and illness severity in children*. J Infect Dis, 2002. **186**(9): p. 1376-7; author reply 1377.
18. Wright, P.F., et al., *Illness severity, viral shedding, and antibody responses in infants hospitalized with bronchiolitis caused by respiratory syncytial virus*. J Infect Dis, 2002. **185**(8): p. 1011-8.
19. Papadopoulos, N.G., et al., *Association of rhinovirus infection with increased disease severity in acute bronchiolitis*. Am J Respir Crit Care Med, 2002. **165**(9): p. 1285-9.
20. Janssen, R., et al., *Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes*. J Infect Dis, 2007. **196**(6): p. 826-34.
21. Tulic, M.K., et al., *TLR4 polymorphisms mediate impaired responses to respiratory syncytial virus and lipopolysaccharide*. J Immunol, 2007. **179**(1): p. 132-40.
22. Puthothu, B., et al., *Association between severe respiratory syncytial virus infection and IL13/IL4 haplotypes*. J Infect Dis, 2006. **193**(3): p. 438-41.
23. Choi, E.H., et al., *A common haplotype of interleukin-4 gene IL4 is associated with severe respiratory syncytial virus disease in Korean children*. J Infect Dis, 2002. **186**(9): p. 1207-11.

24. Wilson, J., et al., *Genetic variation at the IL10 gene locus is associated with severity of respiratory syncytial virus bronchiolitis*. J Infect Dis, 2005. **191**(10): p. 1705-9.
25. Hull, J., et al., *Unusual haplotypic structure of IL8, a susceptibility locus for a common respiratory virus*. Am J Hum Genet, 2001. **69**(2): p. 413-9.
26. Legg, J.P., et al., *Type 1 and type 2 cytokine imbalance in acute respiratory syncytial virus bronchiolitis*. Am J Respir Crit Care Med, 2003. **168**(6): p. 633-9.
27. Chen, Z.M., et al., *Association of cytokine responses with disease severity in infants with respiratory syncytial virus infection*. Acta Paediatr, 2002. **91**(9): p. 914-22.
28. Smyth, R.L., et al., *Respiratory syncytial virus bronchiolitis: disease severity, interleukin-8, and virus genotype*. Pediatr Pulmonol, 2002. **33**(5): p. 339-46.
29. Anderson, L.J., et al., *Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies*. J Infect Dis, 1985. **151**(4): p. 626-33.
30. Cane, P.A. and C.R. Pringle, *Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein*. J Virol, 1995. **69**(5): p. 2918-25.
31. Cane, P.A. and C.R. Pringle, *Respiratory syncytial virus heterogeneity during an epidemic: analysis by limited nucleotide sequencing (SH gene) and restriction mapping (N gene)*. J Gen Virol, 1991. **72** (Pt 2): p. 349-57.
32. Cane, P.A. and C.R. Pringle, *Molecular epidemiology of respiratory syncytial virus: rapid identification of subgroup A lineages*. J Virol Methods, 1992. **40**(3): p. 297-306.
33. Scott, P.D., et al., *Molecular epidemiology of respiratory syncytial virus in Kilifi district, Kenya*. J Med Virol, 2004. **74**(2): p. 344-54.
34. Ostlund, M.R., et al., *Molecular epidemiology and genetic variability of respiratory syncytial virus (RSV) in Stockholm, 2002-2003*. J Med Virol, 2008. **80**(1): p. 159-67.
35. Melero, J.A., et al., *Antigenic structure, evolution and immunobiology of human respiratory*

- syncytial virus attachment (G) protein*. J Gen Virol, 1997. **78** (Pt 10): p. 2411-8.
36. Crowe, C.P.L.a.J.E.J., *Respiratory syncytial virus and metapneumovirus*, in *Field's Virology*, D.M. Knipe, Editor. 2007, Lippincott Williams & Wilkins: Philadelphia, PA. p. 1601-1646.
37. Asenjo, A., et al., *Residues in human respiratory syncytial virus P protein that are essential for its activity on RNA viral synthesis*. Virus Res, 2008. **132**(1-2): p. 160-73.
38. Bitko, V. and S. Barik, *Persistent activation of RelA by respiratory syncytial virus involves protein kinase C, underphosphorylated IkappaBbeta, and sequestration of protein phosphatase 2A by the viral phosphoprotein*. J Virol, 1998. **72**(7): p. 5610-8.
39. Goulder, P.J., et al., *Characterization of a novel respiratory syncytial virus-specific human cytotoxic T-lymphocyte epitope*. J Virol, 2000. **74**(16): p. 7694-7.
40. Agenbach, E., C.T. Tiemessen, and M. Venter, *Amino acid variation within the fusion protein of respiratory syncytial virus subtype A and B strains during annual epidemics in South Africa*. Virus Genes, 2005. **30**(2): p. 267-78.
41. Price, G.E., et al., *Viral escape by selection of cytotoxic T cell-resistant variants in influenza A virus pneumonia*. J Exp Med, 2000. **191**(11): p. 1853-67.
42. Venter, M., et al., *Respiratory syncytial virus nucleoprotein-specific cytotoxic T-cell epitopes in a South African population of diverse HLA types are conserved in circulating field strains*. J Virol, 2003. **77**(13): p. 7319-29.
43. Shinoff, J.J., et al., *Young Infants Can Develop Protective Levels of Neutralizing Antibody after Infection with Respiratory Syncytial Virus*. J Infect Dis, 2008.
44. Johnson, P.R., et al., *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, 1987. **84**(16): p. 5625-9.
45. de Graaff, P.M., et al., *HLA-DP4 presents an immunodominant peptide from the RSV G protein to CD4 T cells*. Virology, 2004. **326**(2): p. 220-30.

46. Modlin, R.L., *Th1-Th2 paradigm: insights from leprosy*. J Invest Dermatol, 1994. **102**(6): p. 828-32.
47. Bont, L., et al., *Peripheral blood cytokine responses and disease severity in respiratory syncytial virus bronchiolitis*. Eur Respir J, 1999. **14**(1): p. 144-9.