RSV-directed therapeutic strategies and evolutionary dynamics

Lydia Tan

Cover design: Lydia Tan “Back to the roots, back to basics”

Print: Gildeprint drukkerijen, Enschede, The Netherlands

ISBN: 978-90-393-6079-8

RSV-directed therapeutic strategies and evolutionary dynamics

RSV-gerichte therapeutische strategieën en evolutionaire dynamiek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het
besluit van het college voor promoties in het openbaar te verdedigen op
6 februari 2014 des middags te 2.30 uur

door

Lydia Tan

geboren op 25 december 1981 te Oosterhout

Promotor: Prof.dr. E.J.H.J. Wiertz

Co-promotor: Dr. F.E.J. Coenjaerts

Printing of this thesis was financially supported by Merus B.V. and the Infection & Immunity Center Utrecht

"Like an infant so naive, looking into the world with utter amazement,

so should we be amazed by nature's life and keep in mind that all is plausible.

Even, if it does not make sense according to the knowledge we have so far.

Let all aspects be reviewed and all experiences be a guidance.

This is how we learn, adapt and how we survive"

Lydia Tan

Voor pap en mam

Commissie: Prof.dr. P.J.M. Rottier
Prof.dr. L. Koenderman
Prof.dr. J.A.G. van Strijp
Prof.dr. L. Meijaard

Paranimfen: Odile van Kan
Malou Tan

CONTENTS

CHAPTER ONE

General Introduction

Partially published as

The Role of Toll-like Receptors in Regulating the Immune Response against Respiratory Syncytial Virus

Lydia Tan*, Peter Klein Klouwenberg*, Wendy Werkman, Grada M van Bleek & Frank E.J. Coenjaerts;

* Authors contributed equally to the paper

Critical Reviews in Immunology; 2009,29(6):531-550

1. Respiratory Syncytial Virus: An infectious agent causing respiratory disease

1.1. Discovery and disease

In 1956, a cytopathogenic agent involved in acute respiratory illness causing runny nose, coryza and malaise within chimpanzees was recovered and named Chimpanzee Coryza Agent (CCA) [1]. Not long after the discovery of CCA in these vertebrates, this cytopathogenic agent was also isolated from infants with croup. The agent was then renamed "Respiratory Syncytial Virus" due to its ability to form syncytia (giant multinucleated cells) of infected epithelial cells (Figure 1) [2].

RSV is primarily described as the common viral cause of severe bronchiolitis and pneumonia in infants and is responsible for the majority of hospitalizations in early childhood [3]. By the age of two, most children have been infected or even re-infected by this virus [4]. The majority of young children infected with RSV have mild symptoms like rhinitis and otitis media. However, when infection transfers to the lower respiratory tract, eventually causing life-threatening obstructive pulmonary disease, supportive care is required as in mechanical ventilation. The mortality associated with primary RSV infection in otherwise healthy children is estimated to be 0.005% to 0.020%, but in hospitalized RSV-infected children up to 3% succumbs due to RSV-associated disease [5]. The latter group especially involves

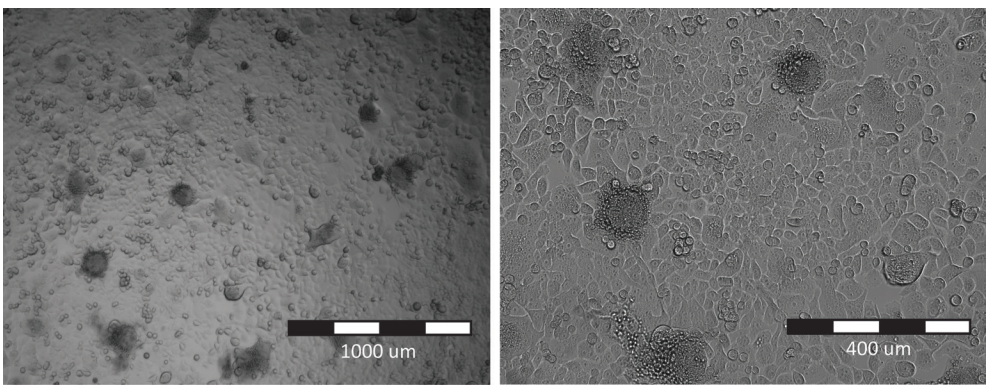


Figure 1. RSV Syncytia formation on Hep2 cells. Two days post-infection with multiplicity of infection (moi) 10. Images were made on an EVOS microscope.

infants with pre-existing high risk factors like prematurity, congenital heart disease, bronchopulmonary dysplasia and congenital or acquired immunodeficiency [6-8]. RSV is also and more often related to severe respiratory disease in immuno-compromised adults and elderly with cardiopulmonary diseases [9-11]. High attack rates are reported from nursing homes, often with fatal outcome. In the USA, around 10,000 death cases are annually reported for persons above 64 years of age [11].

1.2. Viral structure and replication cycle

RSV is an enveloped non-segmented negative sense single-stranded RNA (ssRNA) virus classified into the genus pneumovirus of the *Paramyxoviridae* family. Its 15.2kb RNA genome consists of 11 genes that encode Non-Structural proteins (NS1, NS2), a nucleocapsid protein (N), a nucleocapsid phosphoprotein (P), a matrixprotein (M), a small hydrophobic protein (SH), an attachment glycoprotein (G), a fusion glycoprotein (F), transcription regulation proteins (M2-1, M2-2), and a large polymerase protein (L) (Figure 2).

The viral envelope, giving the virus a spherical and sometimes pleiomorphic appearance, harbors the M, SH, G and F structural proteins (Figure 3A). It surrounds and protects the ribonucleocapsid (RNP) complex, serving as a carrier for viral genome transfer from host-to-host cell [12]. The G protein enables close contact of the virus to the host cell by binding carbohydrates. Viral entry occurs when the virus envelope and target cell membrane fuse, thereby transmitting the RNP complex into the cytosol. The F protein mediates this process and fusion activity requires cleavage of the F0 form by furin-like proteases (Figure 3B). These proteases remove a 27-amino acid peptide fragment (p27) from the fusion protein and separate it into the disulfide-liked F1 and F2 protein domains [13]. The host specificity is determined by its F2 domain [14].

Directly after entry, transcription is initiated and new viral proteins are produced that either support genome replication or form the structural basis of new virions. Gradually, during infection, transcription is switched towards genome replication, which is mediated by low levels of the M2-2 protein. This process is tightly regulated since overexpression of this protein in M2-2 transiently transfected cells showed impairment of RSV replication [15]. Especially the first two codons at the N terminus

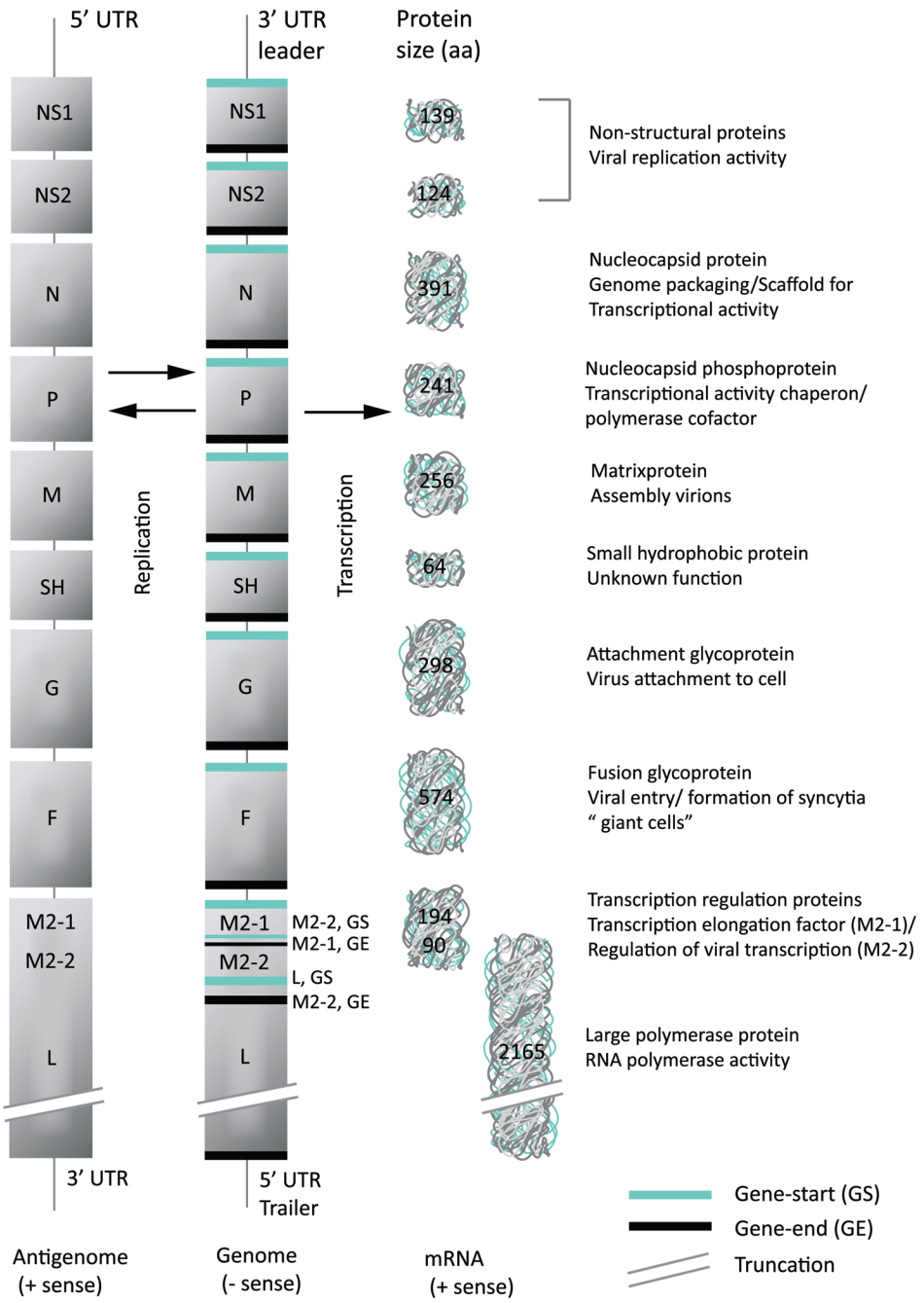
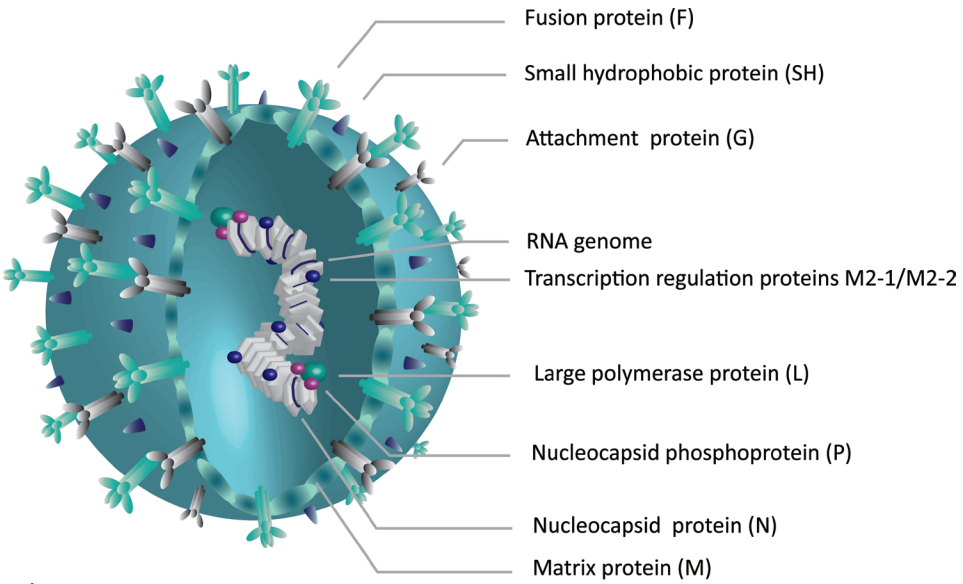
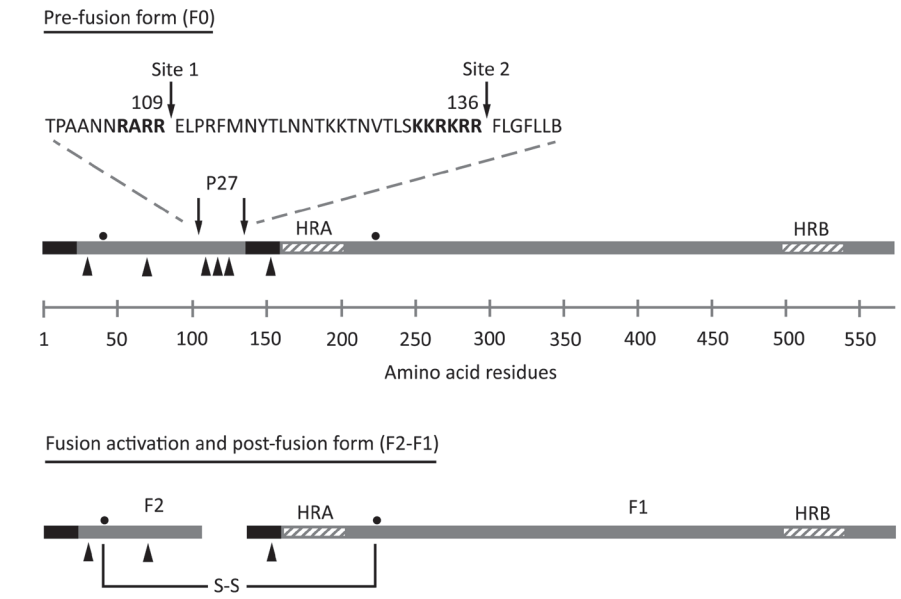


Figure 2. RSV genome structure and overview of the replication and transcription process.



A



B

Figure 3. Schematic representation of the RSV virion (A) and the two structural forms of the fusion protein (B). The fusion protein is cleaved at site 1 and 2 by furin-like proteases for fusion activation. Indicated are the potential N-glycosylation sites (black triangle), cysteine residues (black dots), hydrophobic domains (black boxes) and Heptad Repeat sequence A (HRA; grey striped box) and B (HRB). The F1 and F2 domains are linked by a disulfide bridge.

and the last four amino acids at the C terminus of M2-2 are essential for its activity. Genome replication results in single-stranded RNA (-sense) replicates that are obtained via double stranded RNA intermediates with an antigenome (+ sense) (Figure 2) [12].

The RSV genome is embedded in a multimeric N protein complex, which serves as scaffold for RNA synthesis and also protects the viral RNA from RNAse activity. Proteins M2-1, M2-2, P and L are incorporated in the RNA-Nucleocapsid (RNA-N) complex, which facilitates the rapid transcription and replication of the genome upon host cell entry [16]. The M2-1 transcriptional processivity and anti-termination factor regulates RSV transcription and it requires tetramerization at the His33 to Gly62 residue domain in order to accomplish this [17]. Protein P and RNA bind to the M2-1 region covering amino acid residues 59 to 178 in a competitive manner.

Progeny virus formation is mediated by the M2-1 protein and involves the association of RNP complexes with the M protein, in which the N-terminal 110 amino acids play an important role [18]. The M protein is the key factor in viral filament maturation and absence of this protein results in accumulation of RNP complexes in inclusion bodies as well as impaired transport to sites of budding [19]. RSV F binds the M protein in order to establish the incorporation of other viral proteins into virions [20]. The phenylalanine residue at the C terminal end of the RSV F cytoplasmic tail is critical for the integration of M, N and P proteins [21]. These RSV assembly mechanisms occur in an ESCRT (endosomal sorting complexes required for transport)-independent manner and contribute to budding of the virus.

The function of the type II transmembrane SH protein is not well known. It might play a role in inhibition of apoptosis, which could be due to interference with the TNF-α pathway [22]. SH is also suggested to function as a viroporin by oligomerization of SH proteins via their transmembrane domains. The pore structure has ion channel activity, which also controls the apoptosis of cells [23,24].

NS1 and NS2 proteins are not part of RSV virions, but are only present in infected cells where they are most abundantly expressed due to their promoter-proximal location [25,26]. These two non-structural proteins cooperatively counteract the antiviral state of host cells by impeding the activation and nuclear translocation of IFN-regulatory factor IRF-3, and by mediating proteosome degradation of signal

transducer and activator of transcription 2 (STAT2) with elongin E3 ligase, thereby interfering with the type I IFN signaling cascade [27-29].

1.3. Antigenic variants and epidemiology

Antibody cross-reactivity analyses have led to the identification of two major antigenic RSV subgroups (A and B), which have been further classified into genotypes based on G gene genetic divergence [30,31]. Both subgroups can circulate concurrently during annual epidemics, although the predominant genotypes may be replaced in time over successive epidemic seasons [32,33]. Primary infection with an RSV strain from one of the two subgroups is frequently followed by a secondary infection with an RSV strain from the other subgroup [34]. RSV re-infections occur frequently throughout life [35], which is the consequence of limited cross-immunity induced by different RSV genotypes. This evasion of pre-existing host immune responses is most likely facilitated by the high antigenic variability of the G protein, both within and between the two antigenic subgroups [36,37]. The community-specific behavioral and cultural patterns and also variation in herd immunity influence the heterogeneity in genotypic distribution patterns among different human populations [38]. A correlation between severity of disease and infection with a specific antigenic subgroup is still a matter of debate, as previously described studies had conflicting results [39-41]. However, the degree of viral load and the balance in host immunity are associated with the outcome of disease [42,43].

Annually, a peak in RSV infection incidence is observed during the winter in temperate countries and during the rainy season in tropical countries. Meteorological factors such as increased humidity (45-65%) and optimal temperature (2-6ºC for temperate climates and 24-30ºC for tropical countries) have been described to be favorable for RSV transmission [44].

2. Host defense mechanisms against RSV

2.1. Role for RSV-mediated immune responses in RSV pathogenesis

RSV infection initiates in the epithelial cells of the upper respiratory tract, but can transfer to the lower respiratory tract. The typical RSV syncytia formation, the viral RNA replication and cytotoxicity are pathogen-related factors that contribute to the disease manifestation. However, the triggered host immune response is broadly regarded as the determining factor in the RSV pathogenesis and severity of disease. This was particularly observed when immune stimulation by a formalin-inactivated RSV vaccine triggered enhanced disease in vaccinees [45,46]. In addition, RSV disease is especially prevalent during infancy when host immunity is least mature or in adults that have immunologic deficiencies, which implies that RSV is an opportunistic pathogen that triggers respiratory disease in weakened persons. The fact that RSV disease severity is not only a direct result of the invading pathogen, but even more determined by immunologic factors resulted in numerous studies of the RSV-associated immune response that partly explain the complex course of disease.

2.2. Innate immunity – Pattern recognition [47]

RSV transmission from person to person occurs via inhalation of contaminated aerosols and contact with contaminated surfaces. Initially, RSV infects the upper respiratory tract, but can spread to the lower respiratory tract via cell-cell contact. Epithelial cells from the small airways and alveoli, forming the first line barrier of the host, and underlying antigen-presenting cells (APC) like dendritic cells (DC) and alveolar macrophages are the first to be infected during natural RSV exposure. Upon pathogen invasion, these cells increase the expression of pattern recognition proteins (PRPs). This is a semi-specific manner of the innate immune system to restrict an infection and activate the adaptive immune system, which then specifically targets the offending pathogens for elimination and builds up immunologic memory for a rapid immune response following re-infection. The innate immune system recognises microorganisms by a limited number of PRPs [48,49] enabling it to recognise and respond to these pathogens without any previous exposure to the pathogens. This is especially important for RSV recognition and protection against RSV infection, as maternal antibodies wane quickly after birth

[50] and young infants still have an immature immune system. The retinoic acid-inducible gene I (RIG-I)-like helicases (RLHs), nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs), and Toll-like receptors (TLRs) encompass the three PRP classes most important in initial virus detection and modulation of the innate immune response [51].

The early antiviral response and induction of TLR3 expression in RSV-infected airway epithelial cells are mediated by the RIG-I signalling pathway, which recognizes the 5' triphosphate structure on both ssRNA and dsRNA [52,53]. However, the NS2 protein of RSV antagonizes antiviral interferon ß (IFN-ß) transcription by interacting with the caspase recruitment domain (CARD) of RIG-I soon after infection [54]. Relative gene expression levels of RIG-I are significantly higher in RSV-infected infants compared to virus-negative infants [55]. The ssRNA RSV genome can also be recognized by NOD2 receptors and the antigen-receptor complex triggers IFN-ß production and IRF-3 activation [56]. IFN-ß expression induced via NOD2 signalling results in increased expression of RIG-I and TLR3 virus receptors [57].

In contrast to the RLHs and NLRs, TLRs include many variants that play different roles during RSV infection and have been studied in more detail. To date, 10 different TLRs (TLR1-10) have been identified in humans [58,59]. TLRs are type I transmembrane receptor proteins composed of an extracellular domain containing multiple leucine-rich repeats, a transmembrane region, and a cytoplasmic tail containing the conserved TIR domain [60]. The extracellular domain forms the ligand interaction site. The C-terminus, located on the internal site of the membrane, contains a structure for signal transduction, which is similar to the IL1 receptor, and is called the Toll/IL1 receptor (TIR) domain [58,59]. Not all TLRs are expressed on the plasma membrane. TLR3, TLR7, TLR8, and TLR9 are, at least in part, intracellular proteins located in endosomes [59]. Most tissues express TLRs, but the expression pattern is dependent on cell type and activation state. The greatest variety and level of expression of TLRs is found on macrophages, DCs, neutrophils, and eosinophils [61,62]. TLRs are also constitutively expressed on epithelial cells lining the skin, as well as respiratory, intestinal, and genitourinary tracts, and finally on endothelial cells [61]. Accompanied by recruited adaptor proteins, stimulation of TLRs results in activation of intracellular signalling pathways leading to the production of specific sets of pro-inflammatory molecules [58,63] and thereby an adequate activation of

cells of the innate immune system, such as macrophages, natural killer (NK) cells, and neutrophils. The induction of cytokines, chemokines, and interferons (IFNs) in the lung after binding viral components leads to an antiviral state and stimulates the adaptive immune response [27,64,65].

The role of the TLR family members in the immunopathogenesis during RSV infection has been widely documented, but is still under debate. Clear associations between RSV infection and the up-regulation of TLR4 [66] or TLR3 [67], as well as murine experiments showing impaired immune responses in the absence of these TLRs, have demonstrated the involvement of TLRs [67-69]. In addition, two polymorphisms in the TLR4 gene [70-73] and one polymorphism in the CD14 co-receptor gene [74] have been shown to be related to RSV survival or immunologic potentiation during infection, respectively. Despite these positive correlations, other reports rejected a role for TLR4 in RSV disease or suggested that TLR4 genetic variations may protect against enhanced RSV disease [22,74-76]. In addition, TLR2, 3, 6, 7/8, and 9 only affect the severity of RSV disease in mice; TLR2 and 6 appear to play a role in regulating tumour necrosis factor-alpha (TNF-α) production, which can promote viral clearance and enhance TLR2 expression via an autocrine effect. The RSV small hydrophobic protein can block TNF-α production, thereby evading this component of the immune response [22]. Activation of TLR3 during RSV infection may have a down-regulating effect on immunity, but long-term impaired TLR3 activity could result in mucus accumulation, causing severe airway congestion [67]. Intracellular TLR7/8 and TLR9 stimulate IFN production, which can be strongly inhibited by the RSV non-structural proteins, NS1 and NS2 [27]. Overall, these studies create an image of TLRs that do contribute to the pathogenesis in RSV disease, but the outcome of TLR activation strongly depends on the virus population size and replication rate of individual infections. Both blockage of TLR signalling pathways by viral non-structural proteins and polymorphisms in TLR genes, leading to altered or deficient TLR expression, affect the balance of the RSV-induced immune response and thus determine the severity of disease.

2.3. Adaptive immunity – Cellular response

Generally, the adaptive immune response mediates viral clearance either by preventing virus entry into host cells via neutralizing antibodies or by direct

killing of virus-infected cells via T cells. The latter mechanism requires antigen presentation by ubiquitous professional antigen presenting cells (APCs) such as alveolar macrophages and dendritic cells (DCs). These APCs capture antigens and present them as antigenic peptides to effector $CD8^+$ and regulatory $CD4^+$ T cells via major histocompatibility complex (MHC) class I and MHC class II, respectively [77,78]. All nucleated cells express MHC class I molecules, which present peptides from cytosolic antigens that have been cleaved by the proteasome [79,80]. This is particularly important for the presentation of viral antigens in infected cells. Via the transporter associated with antigen processing (TAP), antigenic peptides are transferred into the endoplasmic reticulum and loaded on MHC class I molecules [81]. On the other hand, MHC class II expression is restricted to APCs and these molecules only present internalized antigens that are part of the endosomal pathway. After antigen uptake, the class II-invariant chain peptide (CLIP) located in the binding groove of MHC Class II dissociates and is replaced by the antigenic peptide, a process that is catalyzed by HLA-DM [82,83]. Peptide-loaded complexes involving both MHC classes are subsequently translocated to the cell surface for recognition by effector cells.

$CD8^+$ T cells target and directly kill virus-infected cells, while $CD4^+$ T cells produce cytokines that endorse $CD8^+$ T cell and B cell activation [84]. Alveolar macrophages and DCs are therefore fundamental in the coordination of adaptive immunity by means of initiating and modulating T cell responses against pathogens such as RSV. Lung DCs are located directly underneath the epithelial cell layer and sample the airways using cell extrusions that cross the epithelial barrier via tight junctions [85]. Both epithelial and dendritic cells can be infected by RSV, which triggers an IFN-mediated anti-viral state and recruitment of inflammatory cells [86]. RSV is able to infect and replicate within DCs and renders these cells to be ineffective in stimulating specific T cell proliferation and IFN-g secretion [87,88]. Although a considerable increase in mature DCs in the lungs was seen upon RSV infection in mice [89], RSV most likely evades the T cell mediated adaptive immunity by impairing functional immunological synapse assembly between DCs and T cells [90]. The unbalanced immunity leads to enhanced disease and each polymorphism or deficiency in the host defense system contributes to the pathogenesis [43].

A shift towards the T helper 2 (Th2)-type immune response involving $CD4^+$ T cells is observed upon RSV infection. Though significant numbers of $CD8^+$ T cells are found

in the human peripheral blood after primary RSV infection, low frequency of RSV-specific $CD8^+$ T cells are detected in the memory pool, which might be the reason for the re-infections throughout life [91,92]. Mouse model studies revealed a dominant role for $CD8^+$ T cells in viral load reduction. However, adoptive transfer of these cells in RSV-primed mice resulted in enhanced lung pathology [93]. The regulatory T cell (Treg) $CD4^+$ subpopulation has a suppressive function on the adaptive immune response. They prevent autoimmune diseases, limit chronic inflammatory disease and maintain peripheral tolerance in order to prevent immune hyperstimulation [94]. Reduced illness, but prolonged viral replication is observed when both T cell classes are depleted prior to RSV infection. These findings emphasize the importance of a balance in $CD8^+$ and $CD4^+$ T cell activation for optimal RSV viral clearance. Furthermore, they stress the role of unbalanced immune responses in enhanced disease.

2.4. Adaptive immunity – Humoral response

In general, the humoral immune response encompasses various B cell-derived antibody isotypes that can battle pathogens at different time points during infection, at distinct sites of infection and with distinctive affinity. Initial infection drives T-cell independent B cell activation outside the germinal centers of the secondary lymphoid tissue and stimulates the production of low affinity antibodies. T-cell dependent B cell activation in the germinal centers of B cell follicles follows later during infection and commences clonal expansion, antibody class switching (from IgM or IgD to IgA, IgG, and IgE), and somatic hypermutation. These processes result in the production of high affinity antibodies [95-97].

Mucosal and serum antibodies are involved in the protection against RSV infection in the upper respiratory tract (IgA) and lower respiratory tract (IgG), respectively. These antibodies either prevent viral spread by direct blockage of the virus-host interaction (neutralizing antibodies) or by opsonizing virus particles (eventually non-neutralizing) for recognition by APCs. Lowering the viral load with neutralizing antibodies helps to reduce pathogenic effects induced by viral infection. Additionally, both neutralizing and non-neutralizing antibodies, in complex with the antigen, contribute not only to the antigen presentation process, but also determine the degree of $CD4^+$ and $CD8^+$ T cell activation [98-100].

Adverse effects of humoral responses have been observed with increasing levels of RSV-specific IgE that were associated with the development of allergy-like conditions such as asthma [101,102]. Furthermore, antibody-virus immune complexes might also enhance infection by promoting close proximity of the opsonized virus particle and the host cell, a process called intrinsic antibody dependent enhancement (iADE) [103].

Full-term newborns are protected against RSV infection in the first two months of life by maternally derived IgG neutralizing antibodies that have been transferred via the placenta. However, maternal antibodies rapidly decline after birth and an increase in incidence of RSV-associated bronchiolitis is observed after 4-6 months [50,104,105]. Only high serum titers of antibodies derived from the mother or acquired via natural infection are protective against RSV. Maternal serum IgG includes higher levels of anti-RSV F than anti-RSV G, and pre-existing significant levels of IgG F antibody inversely correlate with the generation of mucosal IgA F antibody during primary infection in young infants [106,107]. This suggests that a decline in maternal IgG favors the priming of humoral immune responses, although this transition makes infants temporarily more vulnerable to severe RSV infection. The lack of RSV specific antibodies in immunocompromised adults results in a similar sensitivity to RSV infection. For elderly people it was shown that they have comparable baseline antibody levels as to middle-aged adults, but they elicit a significantly stronger antibody response upon RSV re-infection [108]. Disease severity in this group is therefore not caused by a defect in the humoral response. The exact underlying mechanism responsible for increased sensitivity to severe RSV infection is still not known.

RSV re-infections occur frequently throughout life due to the fact that naturally acquired adaptive immunity against RSV is both inefficient as a result of partial cross-immunity (immunity against different RSV variants) and transient as acquired antibody levels decline over time. Although otherwise healthy young children and adults experience only mild disease symptoms, patients with risk factors such as immaturity of the host immune response or immunodeficiencies, require intervention tools to counteract RSV infections and reduce disease severity. Immunologic studies have led to the development of such therapeutic and prevention strategies against RSV infections.

3. RSV intervention strategies

3.1. Vaccine

Over the last decades, many attempts have been made to develop an efficient RSV vaccine with minimal immunopathogenic impact. Formalin-inactivated RSV (FI-RSV) presented the first candidate vaccine that was taken into a clinical trial in the 1960s. Unfortunately, this attenuated virus not only failed to induce good protective neutralizing antibodies, but also induced enhanced disease upon subsequent natural infection [45,109]. Hospital admission was required for 80% of the infants that received FI-RSV, whereas only 5% of the parainfluenza control vaccine subjects needed hospitalization [46]. A robust T cell proliferation with eosinophil and neutrophil influx into the lung tissue was observed during post mortem examination of two fatal cases, which demonstrates that FI-RSV causes significant immune pathology. The FI-RSV vaccine trial failure has set off a major drawback in vaccine development and led to different follow-up studies with animal models in order to elucidate specific underlying mechanisms responsible for FI-RSV mediated enhanced disease [110-113]. From these studies, it appeared that FI-RSV is more prone to induce Th2-type immune responses, causes hyper-reactivity of the airways and induces inadequate levels of neutralizing antibodies [113,114]. The latter finding might be the consequence of formalin-induced RSV epitope disruption, which results in poor binding of antibodies to the natural antigen [115,116]. Furthermore, the carbonyl groups on formalin-treated vaccine antigens were also found to be responsible for the boost in Th2 responses. In addition to the epitope disruption, a lack of TLR stimulation by FI-RSV resulted in deficient B cell maturation and led to the production of low affinity antibodies only. These non-neutralizing antibodies caused deposition of immune complexes in the lungs upon RSV challenge and thereby enhanced respiratory disease [117].

So far, further research in vaccine design with alternative antigen components was unsuccessful as attenuated live vaccines were over-attenuated (i.e. not evoking an immune response) and subunit vaccines triggered immune pathology [118]. Many bottlenecks in RSV vaccination strategies hamper the search for an optimal vaccine. The neutralizing competence of maternal antibodies prevents efficient priming of the neonate immune system [106,107]. Moreover, young infants, who still lack somatic

hypermutation in early life, are not capable of producing high affinity IgG and IgA antibodies against RSV. On the other hand, pre-existing immune responses against RSV in older children might interfere with the development of the anti-viral $CD8^+$ T cell response. These restrictions might suggest that prevention of infection via prophylactic treatment with high levels of neutralizing antibodies is more effective and less of a burden for patients.

3.2. Therapeutics

Both RSV G and F proteins are responsible for the induction of RSV-specific humoral immune responses and only high antiviral titers against these proteins are effective in dampening the RSV infection [107]. Development of therapeutic antibodies is therefore focused on these two antigenic targets. For both proteins, numerous RSV epitopes and antigenic sites for binding of (non-)neutralizing antibodies have been identified by competitive binding assays and viral plaque reduction assays (Figure 4) [119-121]. Especially for the highly conserved F protein, epitope clusters like site A (residues 255-275) and C (residues 422-438) have been extensively studied [122-124], which resulted in the selection of the neutralizing antibody Palivizumab (Synagis).

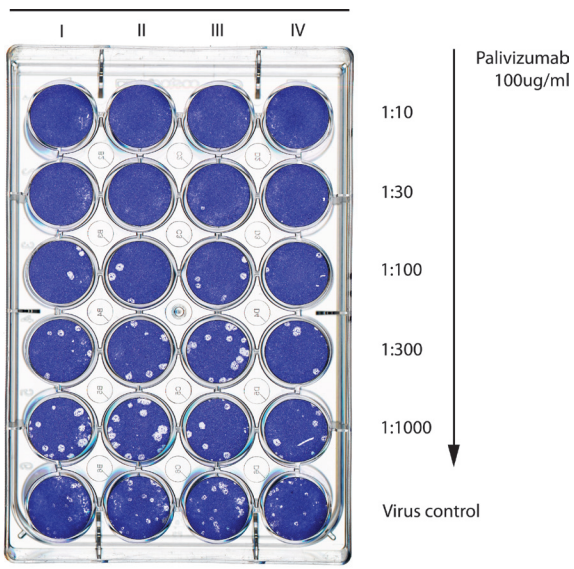


Figure 4. Plaque reduction assay showing four replicates of Palivizumab dilution series applied to the RSV Long strain. The neutralization titer is the antibody concentration that causes a 50% reduction in virus plaques (white spots)

The highly genetically variable G protein induces many native antibodies, but most of them provide only poor cross-immunity. However, antibodies targeting the central conserved domain of the RSV G extracellular domain with CX3C-associated activity have shown to be efficient in reducing viral titers in mice [125,126]. In contrast with antibodies against RSV F, anti-RSV G antibodies are effective in post-infection treatment, have higher efficacy in reducing lower respiratory tract infections and are more efficient in reducing pro-inflammatory cytokine production and pulmonary cell inflammation [125,126]. These properties can have important implications for the design of new therapeutic antibodies.

Palivizumab, a humanized monoclonal antibody that neutralizes RSV infection by targeting site A of the fusion protein, is the only therapeutic antibody available in the clinic. It requires prophylactic administration in order to be effective [127]. Post-infection RSV treatment is currently performed with the only approved antiviral drug, Ribavirin. However, its efficacy and impact on reducing the length of stay in the hospital is not clear [128]. The lack of post-infection effectiveness by Palivizumab, the ineffective RSV neutralization within the upper respiratory tract and the existence of Palivizumab escape mutants have been the reason for proceeding research into new therapeutic intervention strategies [129,130]. The ultra-potent Motavizumab, which originated from antibody affinity maturation of Palivizumab, had a 44-fold enhanced neutralization activity *in vitro* [131]. Nevertheless, *in vivo* experiments in cotton rats only showed a two-fold increase, which was caused by poor serum pharmakinetics and lack of lung bio-availability due to unexpected extensive tissue binding.

To date, there are neither successors of Palivizumab, nor effective RSV vaccines available. Therefore, research continues in order to develop the optimal RSV intervention strategy. Immunologic studies in relation to RSV disease only partly explain the underlying mechanisms that cause RSV-triggered immune-enhanced disease and provide some information on the fact that re-infections frequently occur during life. However, information regarding the impact of immunologic responses on RSV evolution, RSV transmission dynamic patterns and on disease outcome is incomplete. Increased knowledge on the genetic diversity of circulating patient strains (i.e. the RSV genetic blueprint) – including detailed information on the specific position and condenseness of mutational sites, the impact of evolutionary

changes on neutralizing epitopes and the degree of sequence similarity among RSV subgroups and genotypes – would provide us with valuable information on viral genes and proteins that play a key role in host immune adaptation and/or preservation of RSV viability. These data will give us better understanding of the complex interplay between RSV and the host immune response and will contribute to improved selection of therapeutic targets for intervention strategies.

4. Phylodynamic inference: A computational tool for studying viral evolution and epidemiologic patterns

Phylodynamics refers to the processes by which viral evolution and epidemiology are shaped through time [132]. Studying these two topics concurrently can elucidate numerous questions regarding the effect of viral molecular changes and natural selection of new strain variants with advantageous mutations, for example resulting in immune escape of host defense mechanisms [133]. This information can be highly valuable for the determination of antigenic targets for high-affinity therapeutic intervention strategies and vaccine development. In addition, extensive research of temporal (rapidness of viral evolution) and spatial (population size) dynamics in relation to virus genetic diversity might also support prediction of emerging epidemics.

Phylodynamic inference is conducted on whole genome sequences or gene sequence partitions from viral strains and is used to reconstruct the strain evolutionary dynamics in single host individuals that experienced multiple re-infections, in populations with a specific geographic location or even on global scale. Factors such as the control measures of viral spread (e.g. therapeutic intervention, vaccines or herd-immunity), local persistence of infection, size and frequency of epidemics, and molecular adaptation of viruses towards host immune responses all have an impact on these dynamics and the epidemiological pattern. Strain divergence, resulting from molecular changes, can be observed in a maximum clade credibility tree (Figure 5). From this phylogenetic tree representation, emergence of genotype strain variants can be determined via the occurrence of new branches that strongly deviate from the existing clades; an analysis called genotyping. Furthermore, the degree of evolution from a certain virus within a population can be estimated from

the association between tree branch root-to-tip distance and the exact sample date of each strain. Effects of population dynamics on genetic diversity through time are graphically depicted by a Bayesian Skyride shown in a Bayesian Skyline plot [134]. A peak in relative genetic diversity in such a plot may reflect a particular outbreak event [135] and vast peaks through time can be associated with seasonal epidemics as previously observed for influenza (Figure 6, modified from [133] Holmes *et al.* 2009) [136]. Epidemic expansions such as previously described for norovirus GII.4 can also be deduced from these analyses [135].

The combination of phylodynamic inference and viral sequence variability analysis support the search for genetically conserved genes and domains that are suitable therapeutic targets. Additionally, the finding of highly variable regions and hyper-mutational sites via these analyses might suggest a role in immune escape for certain genes and their translation products.

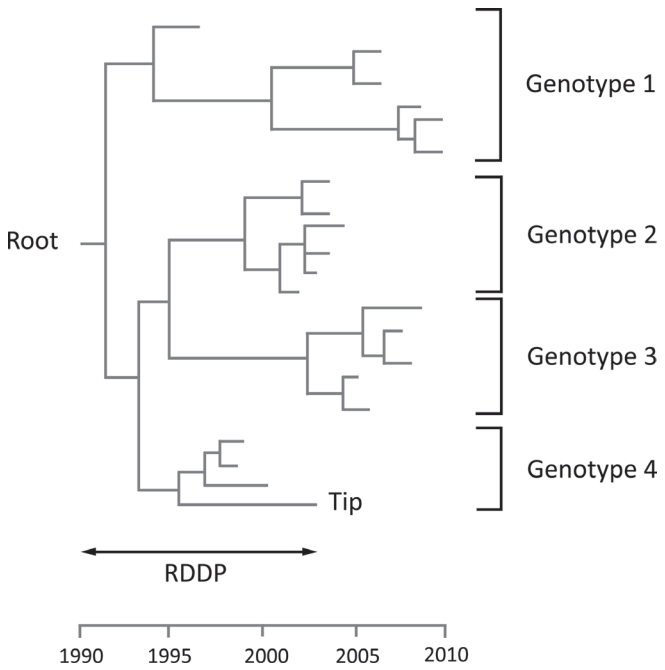


Figure 5. Example of a Maximum Clade Credibility tree (MCC tree). Each branch tip represents one virus strain isolated on a specific date. The root is the common ancestor of all strains that are included in the MCC tree and the root-to-tip divergence (RTTD) indicates the genetic divergence between one strain and the common ancestor of all strains. Branches that strongly deviate from each other are classified by different genotypes.

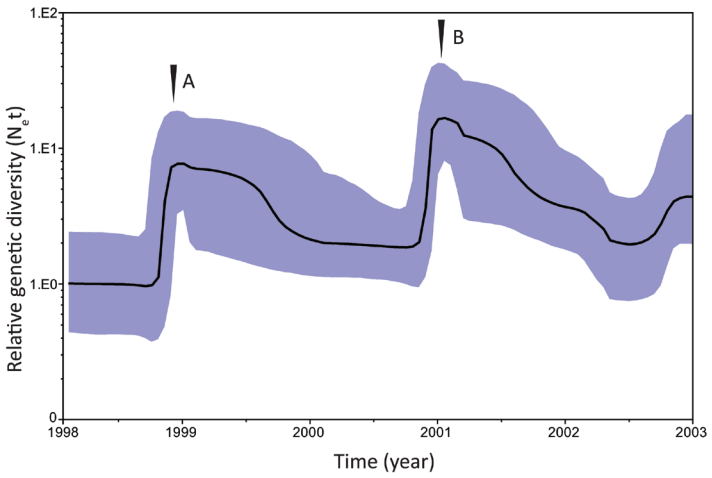


Figure 6. Bayesian Skyline plot showing the effect of population dynamics on genetic diversity through time. The Bayesian Skyride depicts two outbreak events for influenza. Modified from Holmes *et al*. 2009

5. Outline of this thesis

This thesis provides essential information regarding whole genome RSV evolutionary dynamics and combines sequence variability information with the screening of functional and effective antibody therapeutics against two important RSV antigenic targets.

Via phylodynamic inference, a relatively novel computational tool, we estimated the rapidity of RSV molecular evolution based on RSV whole genome and gene partition sequences. In addition, whole genome variability within RSV subgroup A clinical strains has been studied by plotting gene and protein substitutions in order to determine mutational hotspots that could influence protein functionality and even the course of infection. Evolutionary analysis show that RSV-A has a low degree of evolution and that directional selection is primarily found in the highly variable regions of the RSV attachment protein G. The results of this study have been described in **chapter 2**.

In **chapter 3** we further extend this work by comparing subgroup A data with data from RSV subgroup B clinical strains. This study shows a similar degree of evolution for both subgroups, but the sequence variability in RSV subgroup B is considerably lower compared to subgroup A.

A selection of RSV-A and -B clinical strains from different clades was tested in neutralization and binding assays to determine the impact of genetic diversity on antibody recognition and neutralization capacity. These analyses have been described in **chapter 4** for B cell derived antibodies directed against conformational epitopes of the RSV F protein. An outline of the neutralization capacity in both *in vitro* and *in vivo* analyses with these antibodies has been given. The impact of combinatorial antibody therapy on genetic divergent RSV strains has been described in **chapter 5**. In this chapter, bi-specific antibodies and antibody mixtures directed against both RSV F and G were compared. A summarizing discussion is presented in **chapter 6.**

References

1. Blount RE, Jr., Morris JA, Savage RE (1956) *Recovery of cytopathogenic agent from chimpanzees with coryza*. Proc Soc Exp Biol Med 92: 544-549.

2. Chanock R, Roizman B, Myers R (1957) *Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (CCA). I. Isolation, properties and characterization*. Am J Hyg 66: 281-290.

3. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, et al. (2010) *Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis*. Lancet 375: 1545-1555.

4. Glezen WP, Taber LH, Frank AL, Kasel JA (1986) *Risk of primary infection and reinfection with respiratory syncytial virus*. Am J Dis Child 140: 543-546.

5. Ogra PL (2004) *Respiratory syncytial virus: the virus, the disease and the immune response*. Paediatr Respir Rev 5 Suppl A: S119-126.

6. Thorburn K (2009) *Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection*. Arch Dis Child 94: 99-103.

7. Meissner HC (2003) *Selected populations at increased risk from respiratory syncytial virus infection*. Pediatr Infect Dis J 22: S40-44; discussion S44-45.

8. MacDonald NE, Hall CB, Suffin SC, Alexson C, Harris PJ, et al. (1982) *Respiratory syncytial viral infection in infants with congenital heart disease*. N Engl J Med 307: 397-400.

9. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE (2005) *Respiratory syncytial virus infection in elderly and high-risk adults*. N Engl J Med 352: 1749-1759.

10. Englund JA, Anderson LJ, Rhame FS (1991) *Nosocomial transmission of respiratory syncytial virus in immunocompromised adults*. J Clin Microbiol 29: 115-119.

11. Falsey AR, Walsh EE (2000) *Respiratory syncytial virus infection in adults*. Clin Microbiol Rev 13: 371-384.

12. Cowton VM, McGivern DR, Fearns R (2006) *Unravelling the complexities of respiratory syncytial virus RNA synthesis*. J Gen Virol 87: 1805-1821.

13. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, et al. (2001) *Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion*. Proc Natl Acad Sci U S A 98: 9859-9864.

14. Schlender J, Zimmer G, Herrler G, Conzelmann KK (2003) *Respiratory syncytial virus (RSV) fusion protein subunit F2, not attachment protein G, determines the specificity of RSV infection*. J Virol 77: 4609-4616.

15. Cheng X, Park H, Zhou H, Jin H (2005) *Overexpression of the M2-2 protein of respiratory syncytial virus inhibits viral replication*. J Virol 79: 13943-13952.

16. Grosfeld H, Hill MG, Collins PL (1995) *RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA*. J Virol 69: 5677-5686.

17. Tran TL, Castagne N, Dubosclard V, Noinville S, Koch E, et al. (2009) *The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner*. J Virol 83: 6363-6374.

18. Li D, Jans DA, Bardin PG, Meanger J, Mills J, et al. (2008) *Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein*. J Virol 82: 8863-8870.

19. Mitra R, Baviskar P, Duncan-Decocq RR, Patel D, Oomens AG (2012) *The Human Respiratory Syncytial Virus Matrix Protein is Required for Maturation of Viral Filaments*. J Virol.

20. Batonick M, Wertz GW (2011) *Requirements for Human Respiratory Syncytial Virus Glycoproteins in Assembly and Egress from Infected Cells*. Adv Virol 2011.

21. Shaikh FY, Cox RG, Lifland AW, Hotard AL, Williams JV, et al. (2012) *A critical phenylalanine residue in the respiratory syncytial virus fusion protein cytoplasmic tail mediates assembly of internal viral proteins into viral filaments and particles*. MBio 3.

22. Fuentes S, Tran KC, Luthra P, Teng MN, He B (2007) *Function of the respiratory syncytial virus small hydrophobic protein*. J Virol 81: 8361-8366.

23. Gan SW, Tan E, Lin X, Yu D, Wang J, et al. (2012) *The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels*. J Biol Chem 287: 24671-24689.

24. Carter SD, Dent KC, Atkins E, Foster TL, Verow M, et al. (2010) *Direct visualization of the small hydrophobic protein of human respiratory syncytial virus reveals the structural basis for membrane permeability*. FEBS Lett 584: 2786-2790.

25. Collins PL, Wertz GW (1985) *Nucleotide sequences of the 1B and 1C nonstructural protein mRNAs of human respiratory syncytial virus*. Virology 143: 442-451.

26. Evans JE, Cane PA, Pringle CR (1996) *Expression and characterisation of the NS1 and NS2 proteins of respiratory syncytial virus*. Virus Res 43: 155-161.

27. Spann KM, Tran KC, Collins PL (2005) *Effects of nonstructural proteins NS1 and NS2 of human respiratory syncytial virus on interferon regulatory factor 3, NF-kappaB, and proinflammatory cytokines*. J Virol 79: 5353-5362.

28. Elliott J, Lynch OT, Suessmuth Y, Qian P, Boyd CR, et al. (2007) *Respiratory syncytial virus NS1 protein degrades STAT2 by using the Elongin-Cullin E3 ligase*. J Virol 81: 3428-3436.

29. Lo MS, Brazas RM, Holtzman MJ (2005) *Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness*. J Virol 79: 9315-9319.

30. Anderson LJ, Hierholzer JC, Tsou C, Hendry RM, Fernie BF, et al. (1985) *Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies*. J Infect Dis 151: 626-633.

31. Mufson MA, Orvell C, Rafnar B, Norrby E (1985) *Two distinct subtypes of human respiratory syncytial virus*. J Gen Virol 66 (Pt 10): 2111-2124.

32. Cane PA, Matthews DA, Pringle CR (1994) *Analysis of respiratory syncytial virus strain variation in successive epidemics in one city*. J Clin Microbiol 32: 1-4.

33. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ (1998) *Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community*. J Gen Virol 79 (Pt 9): 2221-2229.

34. Mufson MA, Belshe RB, Orvell C, Norrby E (1987) *Subgroup characteristics of respiratory syncytial virus strains recovered from children with two consecutive infections*. J Clin Microbiol 25: 1535-1539.

35. Hall CB, Walsh EE, Long CE, Schnabel KC (1991) *Immunity to and frequency of reinfection with respiratory syncytial virus*. J Infect Dis 163: 693-698.

36. Johnson PR, Spriggs MK, Olmsted RA, Collins PL (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: 5625-5629.

37. Sullender WM (2000) *Respiratory syncytial virus genetic and antigenic diversity*. Clin Microbiol Rev 13: 1-15, table of contents.

38. Peret TC, Hall CB, Hammond GW, Piedra PA, Storch GA, et al. (2000) *Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America*. J Infect Dis 181: 1891-1896.

39. Walsh EE, McConnochie KM, Long CE, Hall CB (1997) *Severity of respiratory syncytial virus infection is related to virus strain*. J Infect Dis 175: 814-820.

40. Hornsleth A, Klug B, Nir M, Johansen J, Hansen KS, et al. (1998) *Severity of respiratory syncytial virus disease related to type and genotype of virus and to cytokine values in nasopharyngeal secretions*. Pediatr Infect Dis J 17: 1114-1121.

41. Devincenzo JP (2004) *Natural infection of infants with respiratory syncytial virus subgroups A and B: a study of frequency, disease severity, and viral load*. Pediatr Res 56: 914-917.

42. DeVincenzo JP, El Saleeby CM, Bush AJ (2005) *Respiratory syncytial virus load predicts disease severity in previously healthy infants*. J Infect Dis 191: 1861-1868.

43. Houben ML, Coenjaerts FE, Rossen JW, Belderbos ME, Hofland RW, et al. (2010) *Disease severity and viral load are correlated in infants with primary respiratory syncytial virus infection in the community*. J Med Virol 82: 1266-1271.

44. Meerhoff TJ, Paget JW, Kimpen JL, Schellevis F (2009) *Variation of respiratory syncytial virus and the relation with meteorological factors in different winter seasons*. Pediatr Infect Dis J 28: 860-866.

45. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE (1969) *An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine*. Am J Epidemiol 89: 405-421.

46. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, et al. (1969) *Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine*. Am J Epidemiol 89: 422-434.

47. Klein Klouwenberg P, Tan L, Werkman W, van Bleek GM, Coenjaerts F (2009) *The role of Toll-like receptors in regulating the immune response against respiratory syncytial virus*. Crit Rev Immunol 29: 531-550.

48. Meylan E, Tschopp J, Karin M (2006) *Intracellular pattern recognition receptors in the host response*. Nature 442: 39-44.

49. Koyama S, Ishii KJ, Coban C, Akira S (2008) *Innate immune response to viral infection*. Cytokine 43: 336-341.

50. Brandenburg AH, Groen J, van Steensel-Moll HA, Claas EC, Rothbarth PH, et al. (1997) *Respiratory syncytial virus specific serum antibodies in infants under six months of age: limited serological response upon infection*. J Med Virol 52: 97-104.

51. Takeuchi O, Akira S (2007) *Recognition of viruses by innate immunity*. Immunol Rev 220: 214-224.

52. Liu P, Jamaluddin M, Li K, Garofalo RP, Casola A, et al. (2007) *Retinoic acid-inducible gene I mediates early antiviral response and Toll-like receptor 3 expression in respiratory syncytial virus-infected airway epithelial cells*. J Virol 81: 1401-1411.

53. Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, et al. (2006) *5'-Triphosphate RNA is the ligand for RIG-I*. Science 314: 994-997.

54. Ling Z, Tran KC, Teng MN (2009) *Human respiratory syncytial virus nonstructural protein NS2 antagonizes the activation of beta interferon transcription by interacting with RIG-I*. J Virol 83: 3734-3742.

55. Scagnolari C, Midulla F, Pierangeli A, Moretti C, Bonci E, et al. (2009) *Gene expression of nucleic acid-sensing pattern recognition receptors in children hospitalized for respiratory syncytial virus-associated acute bronchiolitis*. Clin Vaccine Immunol 16: 816-823.

56. Sabbah A, Chang TH, Harnack R, Frohlich V, Tominaga K, et al. (2009) *Activation of innate immune antiviral responses by Nod2*. Nat Immunol 10: 1073-1080.

57. Vissers M, Remijn T, Oosting M, de Jong DJ, Diavatopoulos DA, et al. (2012) *Respiratory syncytial virus infection augments NOD2 signaling in an IFN-beta-dependent manner in human primary cells*. Eur J Immunol 42: 2727-2735.

58. Kaisho T, Akira S (2006) *Toll-like receptor function and signaling*. J Allergy Clin Immunol 117: 979-987; quiz 988.

59. Finberg RW, Wang JP, Kurt-Jones EA (2007) *Toll like receptors and viruses*. Rev Med Virol 17: 35-43.

60. Vercammen E, Staal J, Beyaert R (2008) *Sensing of viral infection and activation of innate immunity by toll-like receptor 3*. Clin Microbiol Rev 21: 13-25.

61. Sandor F, Buc M (2005) *Toll-like receptors. II. Distribution and pathways involved in TLR signalling*. Folia Biol (Praha) 51: 188-197.

62. Liu G, Zhao Y (2007) *Toll-like receptors and immune regulation: their direct and indirect modulation on regulatory CD4+ CD25+ T cells*. Immunology 122: 149-156.

63. Banerjee A, Gerondakis S (2007) *Coordinating TLR-activated signaling pathways in cells of the immune system*. Immunol Cell Biol 85: 420-424.

64. Garofalo R, Sabry M, Jamaluddin M, Yu RK, Casola A, et al. (1996) *Transcriptional activation of the interleukin-8 gene by respiratory syncytial virus infection in alveolar epithelial cells: nuclear translocation of the RelA transcription factor as a mechanism producing airway mucosal inflammation*. J Virol 70: 8773-8781.

65. Jamaluddin M, Wang S, Garofalo RP, Elliott T, Casola A, et al. (2001) *IFN-beta mediates coordinate expression of antigen-processing genes in RSV-infected pulmonary epithelial cells*. Am J Physiol Lung Cell Mol Physiol 280: L248-257.

66. Gagro A, Tominac M, Krsulovic-Hresic V, Bace A, Matic M, et al. (2004) *Increased Toll-like receptor 4 expression in infants with respiratory syncytial virus bronchiolitis*. Clin Exp Immunol 135: 267-272.

67. Rudd BD, Smit JJ, Flavell RA, Alexopoulou L, Schaller MA, et al. (2006) *Deletion of TLR3 alters the pulmonary immune environment and mucus production during respiratory syncytial virus infection*. J Immunol 176: 1937-1942.

68. Kurt-Jones EA, Popova L, Kwinn L, Haynes LM, Jones LP, et al. (2000) *Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus*. Nat Immunol 1: 398-401.

69. Haynes LM, Moore DD, Kurt-Jones EA, Finberg RW, Anderson LJ, et al. (2001) *Involvement of toll-like receptor 4 in innate immunity to respiratory syncytial virus*. J Virol 75: 10730-10737.

70. Awomoyi AA, Rallabhandi P, Pollin TI, Lorenz E, Sztein MB, et al. (2007) *Association of TLR4 polymorphisms with symptomatic respiratory syncytial virus infection in high-risk infants and young children*. J Immunol 179: 3171-3177.

71. Puthothu B, Forster J, Heinzmann A, Krueger M (2006) *TLR-4 and CD14 polymorphisms in respiratory syncytial virus associated disease*. Dis Markers 22: 303-308.

72. Tal G, Mandelberg A, Dalal I, Cesar K, Somekh E, et al. (2004) *Association between common Toll-like receptor 4 mutations and severe respiratory syncytial virus disease*. J Infect Dis 189: 2057-2063.

73. Mandelberg A, Tal G, Naugolny L, Cesar K, Oron A, et al. (2006) *Lipopolysaccharide hyporesponsiveness as a risk factor for intensive care unit hospitalization in infants with respiratory syncitial virus bronchiolitis*. Clin Exp Immunol 144: 48-52.

74. Inoue Y, Shimojo N, Suzuki Y, Campos Alberto EJ, Yamaide A, et al. (2007) *CD14 -550 C/T, which is related to the serum level of soluble CD14, is associated with the development of respiratory syncytial virus bronchiolitis in the Japanese population*. J Infect Dis 195: 1618-1624.

75. Janssen R, Bont L, Siezen CL, Hodemaekers HM, Ermers MJ, et al. (2007) *Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes*. J Infect Dis 196: 826-834.

76. Paulus SC, Hirschfeld AF, Victor RE, Brunstein J, Thomas E, et al. (2007) *Common human Toll-like receptor 4 polymorphisms--role in susceptibility to respiratory syncytial virus infection and functional immunological relevance*. Clin Immunol 123: 252-257.

77. Itano AA, Jenkins MK (2003) *Antigen presentation to naive CD4 T cells in the lymph node*. Nat Immunol 4: 733-739.

78. Banchereau J, Briere F, Caux C, Davoust J, Lebecque S, et al. (2000) *Immunobiology of dendritic cells*. Annu Rev Immunol 18: 767-811.

79. Kloetzel PM (2004) *Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPPII*. Nat Immunol 5: 661-669.

80. Rock KL, York IA, Goldberg AL (2004) *Post-proteasomal antigen processing for major histocompatibility complex class I presentation*. Nat Immunol 5: 670-677.

81. Elliott T, Williams A (2005) *The optimization of peptide cargo bound to MHC class I molecules by the peptide-loading complex*. Immunol Rev 207: 89-99.

82. van Niel G, Wubbolts R, Stoorvogel W (2008) *Endosomal sorting of MHC class II determines antigen presentation by dendritic cells*. Curr Opin Cell Biol 20: 437-444.

83. Watts C (2004) *The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules*. Nat Immunol 5: 685-692.

84. Doherty PC, Topham DJ, Tripp RA, Cardin RD, Brooks JW, et al. (1997) *Effector CD4+ and CD8+ T-cell mechanisms in the control of respiratory virus infections*. Immunol Rev 159: 105-117.

85. Sung SS, Fu SM, Rose CE, Jr., Gaskin F, Ju ST, et al. (2006) *A major lung CD103 (alphaE)-beta7 integrin-positive epithelial dendritic cell population expressing Langerin and tight junction proteins*. J Immunol 176: 2161-2172.

86. Beaty SR, Rose CE, Jr., Sung SS (2007) *Diverse and potent chemokine production by lung CD11bhigh dendritic cells in homeostasis and in allergic lung inflammation*. J Immunol 178: 1882-1895.

87. de Graaff PM, de Jong EC, van Capel TM, van Dijk ME, Roholl PJ, et al. (2005) *Respiratory syncytial virus infection of monocyte-derived dendritic cells decreases their capacity to activate CD4 T cells*. J Immunol 175: 5904-5911.

88. Guerrero-Plata A, Casola A, Suarez G, Yu X, Spetch L, et al. (2006) *Differential response of dendritic cells to human metapneumovirus and respiratory syncytial virus*. Am J Respir Cell Mol Biol 34: 320-329.

89. Beyer M, Bartz H, Horner K, Doths S, Koerner-Rettberg C, et al. (2004) *Sustained increases in numbers of pulmonary dendritic cells after respiratory syncytial virus infection*. J Allergy Clin Immunol 113: 127-133.

90. Gonzalez PA, Prado CE, Leiva ED, Carreno LJ, Bueno SM, et al. (2008) *Respiratory syncytial virus impairs T cell activation by preventing synapse assembly with dendritic cells*. Proc Natl Acad Sci U S A 105: 14999-15004.

91. Heidema J, Lukens MV, van Maren WW, van Dijk ME, Otten HG, et al. (2007) *CD8+ T cell responses in bronchoalveolar lavage fluid and peripheral blood mononuclear cells of infants with severe primary respiratory syncytial virus infections*. J Immunol 179: 8410-8417.

92. Isaacs D, Bangham CR, McMichael AJ (1987) *Cell-mediated cytotoxic response to respiratory syncytial virus in infants with bronchiolitis*. Lancet 2: 769-771.

93. Cannon MJ, Openshaw PJ, Askonas BA (1988) *Cytotoxic T cells clear virus but augment lung pathology in mice infected with respiratory syncytial virus*. J Exp Med 168: 1163-1168.

94. Vignali DA, Collison LW, Workman CJ (2008) *How regulatory T cells work*. Nat Rev Immunol 8: 523-532.

95. MacLennan IC (1994) *Germinal centers*. Annu Rev Immunol 12: 117-139.

96. MacLennan IC, Toellner KM, Cunningham AF, Serre K, Sze DM, et al. (2003) *Extrafollicular antibody responses*. Immunol Rev 194: 8-18.

97. Rajewsky K (1996) *Clonal selection and learning in the antibody system*. Nature 381: 751-758.

98. Hamano Y, Arase H, Saisho H, Saito T (2000) *Immune complex and Fc receptor-mediated augmentation of antigen presentation for in vivo Th cell responses*. J Immunol 164: 6113-6119.

99. de Jong JM, Schuurhuis DH, Ioan-Facsinay A, Welling MM, Camps MG, et al. (2006) *Dendritic cells, but not macrophages or B cells, activate major histocompatibility complex class II-restricted CD4+ T cells upon immune-complex uptake in vivo*. Immunology 119: 499-506.

100. Kruijsen D, Bakkers MJ, van Uden NO, Viveen MC, van der Sluis TC, et al. (2010) *Serum antibodies critically affect virus-specific CD4+/CD8+ T cell balance during respiratory syncytial virus infections*. J Immunol 185: 6489-6498.

101. Becker Y (2006) *Respiratory syncytial virus (RSV) evades the human adaptive immune system by skewing the Th1/Th2 cytokine balance toward increased levels of Th2 cytokines and IgE, markers of allergy--a review*. Virus Genes 33: 235-252.

102. Han J, Takeda K, Gelfand EW (2011) *The Role of RSV Infection in Asthma Initiation and Progression: Findings in a Mouse Model*. Pulm Med 2011: 748038.

103. Ubol S, Halstead SB (2010) *How innate immune mechanisms contribute to antibody-enhanced viral infections*. Clin Vaccine Immunol 17: 1829-1835.

104. Englund JA (1994) *Passive protection against respiratory syncytial virus disease in infants: the role of maternal antibody*. Pediatr Infect Dis J 13: 449-453.

105. Piedra PA, Jewell AM, Cron SG, Atmar RL, Glezen WP (2003) *Correlates of immunity to respiratory syncytial virus (RSV) associated-hospitalization: establishment of minimum protective threshold levels of serum neutralizing antibodies*. Vaccine 21: 3479-3482.

106. Yamazaki H, Tsutsumi H, Matsuda K, Nagai K, Ogra PL, et al. (1994) *Effect of maternal antibody on IgA antibody response in nasopharyngeal secretion in infants and children during primary respiratory syncytial virus infection*. J Gen Virol 75 (Pt 8): 2115-2119.

107. Shinoff JJ, O'Brien KL, Thumar B, Shaw JB, Reid R, et al. (2008) *Young infants can develop protective levels of neutralizing antibody after infection with respiratory syncytial virus*. J Infect Dis 198: 1007-1015.

108. Walsh EE, Falsey AR (2004) *Age related differences in humoral immune response to respiratory syncytial virus infection in adults*. J Med Virol 73: 295-299.

109. Fulginiti VA, Eller JJ, Sieber OF, Joyner JW, Minamitani M, et al. (1969) *Respiratory virus immunization. I. A field trial of two inactivated respiratory virus vaccines; an aqueous trivalent parainfluenza virus vaccine and an alum-precipitated respiratory syncytial virus vaccine*. Am J Epidemiol 89: 435-448.

110. De Swart RL, Kuiken T, Timmerman HH, van Amerongen G, Van Den Hoogen BG, et al. (2002) *Immunization of macaques with formalin-inactivated respiratory syncytial virus (RSV) induces interleukin-13-associated hypersensitivity to subsequent RSV infection*. J Virol 76: 11561-11569.

111. Prince GA, Jenson AB, Hemming VG, Murphy BR, Walsh EE, et al. (1986) *Enhancement of respiratory syncytial virus pulmonary pathology in cotton rats by prior intramuscular inoculation of formalin-inactiva ted virus*. J Virol 57: 721-728.

112. Piedra PA, Faden HS, Camussi G, Wong DT, Ogra PL (1989) *Mechanism of lung injury in cotton rats immunized with formalin-inactivated respiratory syncytial virus*. Vaccine 7: 34-38.

113. Waris ME, Tsou C, Erdman DD, Zaki SR, Anderson LJ (1996) *Respiratory synctial virus infection in BALB/c mice previously immunized with formalin-inactivated virus induces enhanced pulmonary inflammatory response with a predominant Th2-like cytokine pattern*. J Virol 70: 2852-2860.

114. Murphy BR, Walsh EE (1988) *Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity*. J Clin Microbiol 26: 1595-1597.

115. Delgado MF, Coviello S, Monsalvo AC, Melendi GA, Hernandez JZ, et al. (2009) *Lack of antibody affinity maturation due to poor Toll-like receptor stimulation leads to enhanced respiratory syncytial virus disease*. Nat Med 15: 34-41.

116. Moghaddam A, Olszewska W, Wang B, Tregoning JS, Helson R, et al. (2006) *A potential molecular mechanism for hypersensitivity caused by formalin-inactivated vaccines*. Nat Med 12: 905-907.

117. Polack FP, Teng MN, Collins PL, Prince GA, Exner M, et al. (2002) *A role for immune complexes in enhanced respiratory syncytial virus disease*. J Exp Med 196: 859-865.

118. Graham BS (2011) *Biological challenges and technological opportunities for respiratory syncytial virus vaccine development*. Immunol Rev 239: 149-166.

119. Langedijk JP, Meloen RH, van Oirschot JT (1998) *Identification of a conserved neutralization site in the first heptad repeat of the fusion protein of respiratory syncytial virus*. Arch Virol 143: 313-320.

120. Arbiza J, Taylor G, Lopez JA, Furze J, Wyld S, et al. (1992) *Characterization of two antigenic sites recognized by neutralizing monoclonal antibodies directed against the fusion glycoprotein of human respiratory syncytial virus*. J Gen Virol 73 (Pt 9): 2225-2234.

121. Anderson LJ, Hierholzer JC, Stone YO, Tsou C, Fernie BF (1986) *Identification of epitopes on respiratory syncytial virus proteins by competitive binding immunoassay*. J Clin Microbiol 23: 475-480.

122. Beeler JA, van Wyke Coelingh K (1989) *Neutralization epitopes of the F glycoprotein of respiratory syncytial virus: effect of mutation upon fusion function*. J Virol 63: 2941-2950.

123. Swanson KA, Settembre EC, Shaw CA, Dey AK, Rappuoli R, et al. (2011) *Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers*. Proc Natl Acad Sci U S A 108: 9619-9624.

124. McLellan JS, Yang Y, Graham BS, Kwong PD (2011) *Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes*. J Virol 85: 7788-7796.

125. Haynes LM, Caidi H, Radu GU, Miao C, Harcourt JL, et al. (2009) *Therapeutic monoclonal antibody treatment targeting respiratory syncytial virus (RSV) G protein mediates viral clearance and reduces the pathogenesis of RSV infection in BALB/c mice*. J Infect Dis 200: 439-447.

126. Collarini EJ, Lee FE, Foord O, Park M, Sperinde G, et al. (2009) *Potent high-affinity antibodies for treatment and prophylaxis of respiratory syncytial virus derived from B cells of infected patients*. J Immunol 183: 6338-6345.

127. Group TI-RS (1998) *Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants.* . Pediatrics 102: 531-537.

128. Law BJ, Wang EE, MacDonald N, McDonald J, Dobson S, et al. (1997) *Does ribavirin impact on the hospital course of children with respiratory syncytial virus (RSV) infection? An analysis using the pediatric investigators collaborative network on infections in Canada (PICNIC) RSV database*. Pediatrics 99: E7.

129. Zhu Q, Patel NK, McAuliffe JM, Zhu W, Wachter L, et al. (2012) *Natural polymorphisms and resistance-associated mutations in the fusion protein of respiratory syncytial virus (RSV): effects on RSV susceptibility to palivizumab*. J Infect Dis 205: 635-638.

130. Zhao X, Chen FP, Megaw AG, Sullender WM (2004) *Variable resistance to palivizumab in cotton rats by respiratory syncytial virus mutants*. J Infect Dis 190: 1941-1946.

131. Wu H, Pfarr DS, Johnson S, Brewah YA, Woods RM, et al. (2007) *Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract*. J Mol Biol 368: 652-665.

132. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) *Unifying the epidemiological and evolutionary dynamics of pathogens*. Science 303: 327-332.

133. Holmes EC, Grenfell BT (2009) *Discovering the phylodynamics of RNA viruses*. PLoS Comput Biol 5: e1000505.

134. Minin VN, Bloomquist EW, Suchard MA (2008) *Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics*. Mol Biol Evol 25: 1459-1471.

135. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, et al. (2010) *Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants*. PLoS Pathog 6: e1000884.

136. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, et al. (2008) *The genomic and epidemiological dynamics of human influenza A virus*. Nature 453: 615-619.

CHAPTER TWO

Genetic Variability among Complete Human Respiratory Syncytial Virus Subgroup A Genomes: Bridging Molecular Evolutionary Dynamics and Epidemiology

Lydia Tan[1], Philippe Lemey[2], Lieselot Houspie[2], Marco C.Viveen[1], Nicolaas J.G. Jansen[3], Anton M. van Loon[1], Emmanuel Wiertz[1], Grada M. van Bleek[3], Darren P. Martin[4], Frank E. Coenjaerts[1,#]

[1] Department of Medical Microbiology, University Medical Center Utrecht, The Netherlands

[2] Rega Institute for Medical Research, Department of Microbiology and Immunology, KU Leuven, Belgium

[3] Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht

[4] Institute of Infectious Diseases and Molecular Medicine, Computational Biology Department, Faculty Of Health Sciences, University of Cape Town, Observatory 7925, South Africa

PLoSONE; 2012, 7(12): 1-15

Abstract

Human respiratory syncytial virus (RSV) is an important cause of severe lower respiratory tract infections in infants and the elderly. In the vast majority of cases, however, RSV infections run mild and symptoms resemble those of a common cold. The immunological, clinical, and epidemiological profile of severe RSV infections suggests a disease caused by a virus with typical seasonal transmission behavior, lacking clear-cut virulence factors, but instead causing disease by modifying the host's immune response in a way that stimulates pathogenesis. Yet, the interplay between RSV-evoked immune responses and epidemic behavior, and how this affects the genomic evolutionary dynamics of the virus, remains poorly understood. Here, we present a comprehensive collection of 33 novel RSV subgroup A genomes from strains sampled over the last decade, and provide the first measurement of RSV-A genomic diversity through time in a phylodynamic framework. In addition, we map amino acid substitutions per protein to determine mutational hotspots in specific domains. Using Bayesian genealogical inference, we estimated the genomic evolutionary rate to be $6.47x10^{-4}$ (credible interval: $5.56x10^{-4}$, $7.38x10^{-4}$) substitutions/site/year, considerably slower than previous estimates based on G gene sequences only. The G gene is however marked by elevated substitution rates compared to other RSV genes, which can be attributed to relaxed selective constraints. In line with this, site-specific selection analyses identify the G gene as the major target of diversifying selection. Importantly, statistical analysis demonstrates that the immune driven positive selection does not leave a measurable imprint on the genome phylogeny, implying that RSV lineage replacement mainly follows nonselective epidemiological processes. The roughly 50 years of RSV-A genomic evolution are characterized by a constant population size through time and general co-circulation of lineages over many epidemic seasons – a conclusion that might be taken into account when developing future therapeutic and preventive strategies.

Introduction

Human respiratory syncytial virus (RSV) is the single most important cause of severe lower respiratory tract infections (LRTI) in infants and young children. As a consequence, RSV infections are the most frequent cause of hospitalization of infants and young children in industrialized countries. For example, RSV is responsible for at least 100,000 infant hospitalizations for pneumonia or bronchiolitis every year in the USA alone [1]. The RSV disease spectrum ranges from mild symptoms such as rhinitis and otitis media, to severe illness such as bronchiolitis or pneumonia, which require supportive care such as mechanical ventilation [2]. By the age of two, most children have been infected with RSV, with about half having experienced two or more infections [3]. Elderly people, patients suffering from cardiopulmonary diseases and immuno-compromised individuals also are at risk for severe RSV disease [4-6]. RSV attack rates in nursing homes in the USA are approximately 5-10% per year with a 2-8% case fatality rate, amounting to 10,000 deaths per year among persons above 64 years of age [4]. Among elderly persons followed for 3 consecutive winters, RSV infection accounted for 10.6% of hospitalizations for pneumonia, 11.4% of hospitalizations for obstructive pulmonary disease, 5.4% for congestive heart failure, and 7.2% for asthma [7].

RSV is classified in the genus *Pneumovirus* belonging to the *Paramyxoviridae* family and has an enveloped, non-segmented, single-stranded, negative sense RNA genome of approximately 15,000 nucleotides. The virus has 11 genes coding for non-structural proteins 1 and 2 (NS1 and -2), the nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), small hydrophobic protein (SH), attachment glycoprotein (G), fusion glycoprotein (F), transcription regulatory proteins M2-1 and M2-2 and, finally, a large polymerase (L).

In Europe and North America, RSV disease occurs as well-defined seasonal epidemic outbreaks during the winter and spring months. On the other hand, studies in tropical countries have often reported an increase in RSV during the rainy season, but this has not been a consistent finding [8]. The exact factors responsible for the typical RSV transmission dynamics are not well understood. Multiple genotypes can be present in a single population while new genotypes may replace older predominating genotypes over successive epidemic seasons. Among different

populations, variation in herd immunity has been suggested to play a role in the observed genotypic distribution patterns [9-11]. Community-specific cultural and behavioral patterns might also affect the acquisition and spread of RSV infection.

Antibody cross-reactivity patterns have led to the identification of two antigenic subgroups (A and B) for RSV and these subgroups have been further classified into genotypes based on genetic divergence within the G gene [12-14]. Subgroup A generally dominates during epidemics and subgroup B is mostly involved in re-infection, although both can co-circulate [15-17]. Re-infections with RSV occur frequently throughout life and it has been suggested that the evasion of pre-existing host immune responses is particularly facilitated by antigenic variability of the G protein both between and within the two major antigenic subgroups [18,19].

Development of vaccines to prevent RSV infection has been complicated by several factors, including the risk of potentiating naturally occurring disease. Early attempts at vaccinating children in the 1960s with a formalin-inactivated RSV vaccine showed that vaccinated children suffered from more severe disease on subsequent exposure to the virus as compared to the control-vaccinated group [20-23]. The enhanced severity of disease has been reproduced in animal models and is thought to result from inadequate levels of serum-neutralizing antibodies, antibody-mediated CD4(+) memory T cell response enhancement, and infection mediated CD8(+) memory T cell suppression [24-27]. Live, attenuated RSV vaccines have apparently not exacerbated pulmonary disease upon natural infection, but have also failed to provide any appreciable degree of protection [28-30]. On the basis of these and other observations, it has been hypothesized that the development of disease in RSV infected individuals is principally caused by immune responses induced by the virus [31,32]

From the clinical, immunological and epidemiological profile of RSV a picture emerges of a severe disease that is caused by a virus with typical seasonal transmission behavior, no clear-cut virulence factors, and the capacity to modify host immune responses in a way that affects pathogenesis (e.g. through interfering with Toll-like receptor-signaling [33]). However, despite its potentially important impact on global patterns of RSV genetic diversity, the interplay between RSV induced immune responses and the epidemiological and evolutionary dynamics of

RSV remain poorly understood.

Phylodynamic analyses of viral genetic data are increasingly being used to illuminate how evolutionary and ecological processes can jointly drive fluctuations in the genetic diversity of viral populations [34]. Such analyses hold the potential to reveal viral population dynamics based on phylogenetic patterns, to determine the likely short- and long-term evolutionary impacts of mutations that enable viruses to escape host immune responses, and to describe how these factors might collectively influence the epidemic behavior of viruses. Although the molecular epidemiology and evolutionary dynamics of RSV have been actively studied in the past, much of our current knowledge in this area has been determined through the analysis of single genes. Little is known on how the evolutionary patterns of these individual genes – and in particular the most actively studied G gene – reflect the genomic dynamics of RSV. Complete genome sequences are only starting to be generated for isolates of this species [35,36], but such efforts may become more common with the widespread use of next generation sequencing technologies. As recently pointed out for viruses in general [37], generating RSV genomes is only the first step on the path to understanding the processes that shape the epidemiology and evolution of this virus. Here, we present analyses of the most comprehensive collection of RSV-A complete genomes to date, including thirty three strains sampled over the last decade, and provide the first measurements made within a phylodynamic framework of RSV-A genomic diversity through time. In addition to detailed analyses of genome-wide variability we present estimates of key evolutionary parameters such as the genomic rate of evolution. In particular we demonstrate that relative to the remainder of the RSV genome, the G gene has a larger number of sites evolving under diversifying selection and is, probably as a consequence of this, the fastest evolving RSV-A gene. Finally, we test the degrees to which host immune responses have impacted both patterns of selection detectable within RSV genomes and global-scale RSV population dynamics.

Materials and Methods

Ethics statement

This study was performed according to the guidelines of the institutional ethics committees (METC for Dutch samples; CME for Belgian samples) and in concordance with Dutch privacy legislation. All anonymized clinical strains originated from the routine diagnostic process prior to the current study. The institutional review board (IRB) confirmed (protocol 12/320) that viral strains are not regarded as patient-owned material and consequently the use of these strains is not restricted in the applicable Dutch law ("Law Medical Scientific Research with People", WMO; art. 1b).

Viral isolates

Nasopharyngeal swabs and tracheal aspirates were collected between 2001 and 2011 from children ranging in age from 2 weeks to 5 years, who were hospitalized with respiratory distress symptoms on the pediatric wards of the Wilhelmina Children's Hospital in Utrecht, The Netherlands or the Gasthuisberg University Hospital of Leuven, Belgium (Table S1). Collected samples were cultured immediately and checked for the formation of syncytia in HEp2 cell cultures, a typical RSV infection phenotype. Those exhibiting RSV specific cytopathological effects (CPE) and testing positive for RSV by real-time PCR, were harvested and stored at -80°C.

Viral RNA extraction, cDNA synthesis and real-time Taqman PCR

RSV strains were routinely cultured on HEp2 cells for 2 passages at 33ºC. Virus was harvested from cultures that exhibited CPE in 80% of cells and viral genomic RNA was isolated using a MagnaPure LC total nucleic acid kit (Roche Diagnostics, Mannheim, Germany). The isolated viral RNA was reverse transcribed using a MultiScribe reverse transcriptase kit and random hexamers (Applied Biosystems, Foster City, CA), according to the manufacturer's guidelines. Primers and probes designed on the basis of highly conserved genomic regions of the N gene for both RSV subgroup A (RSV-A) and B (RSV-B) were used for subtyping of the RSV patient strains. The following primers and probes were used:

RSA-1: 5'-AGATCAACTTCTGTCATCCAGCAA-3'

RSA-2: 5'-TTCTGCACATCATAATTAGGAGTATCAAT-3'

RSB-1: 5'-AAGATGCAAATCATAAATTCACAGGA-3'

RSB-2: 5'-TGATATCCAGCATCTTTAAGTATCTTTATAGTG-3'

RSA probe: 5'-CACCATCCAACGGAGCACAGGAGAT-3'

RSB probe: 5'-TTCCCTTCCTAACCTGGACATAGCATATAACATACCT-3'

Murine encephalomyocarditis virus was used as an internal control. Samples were assayed in a 25 µl reaction mixture containing 10 µl of cDNA, TaqMan universal PCR master mix (Applied Biosystems, ABI), primers (900 nM RSV-A primers and 300 nM RSV-B primers), and fluorogenic probes (58.3 nM RSV-A probe and 66.7 nM RSV-B probe) labeled with the 5' reporter dye 6-carboxy-fluorescein (FAM) and the 3' quencher dye 6-carboxy-tetramethyl-rhodamine (TAMRA). Amplification and detection were performed with an ABI 7900 HT system for 2 min at 50°C, 10 min at 95°C, and 45 cycles of 15 sec at 95°C and 1 min at 60°C. Samples were monitored for the presence of possible inhibitors of the amplification reaction by the indicated internal control, signals of which had to range within a clear-cut interval.

Genomic RSV cDNA synthesis and sequencing

For sequencing purposes, human RSV PCR fragments were obtained by fractional amplification of MagNAPure LC genomic RNA isolates using the Superscript III one-step RT-PCR System with Platinum Taq High Fidelity kit (Invitrogen) and a 9800 Fast thermal cycler (ABI) according to the manufacturers protocol. PCR products were applied to a 1% agarose gel and purified from the gel by Qiaquick spin columns (Qiagen). Isolated fragments were used for whole genome sequencing.

The majority of the listed RSV strains (Table S1) was sequenced according to the whole genome sequence protocol recently described by Kumaria *et al* [35] using the conventional Sanger technique. Fragments ranging between 650 and 1,400 nucleotides were sequenced on an ABI 3730 48-capillary DNA analyzer using Big-Dye Terminator 3.1 (ABI). Supplementary Table S2 provides an overview of the primers used. Indicated strains (Table S1) were (partly) sequenced via commercial 454-sequencing (Keygene, Wageningen, The Netherlands). The resulting sequence

information was assembled into RSV whole genome sequences through alignment with the reference RSV A2 strain (M74568) using Seqman software (DNASTAR lasergene 8).

Protein substitution analysis

Gene sequences were extracted from the whole genome sequences and the consensus of each gene was derived from the set of patient and reference strains using Seqman software. Gene sequences were translated into protein coding sequences using Seaview4 and subsequently aligned via the EMBL-EBI ClustalW2-Multiple Sequence Alignment tool. For each strain, sequence variability scores were calculated per gene relative to the consensus sequence for all clinical and reference strains. Protein substitutions per site were mapped using the program Plot0.997. For the RSV fusion protein, substitutions were marked in the 3RKI_PDG crystal structure with MacPyMOL (available at http://www.pymol.org/pymol). The NetNGlyc 1.0 server [38] was used to predict the gain and loss of N-glycosylation sites.

Data set compilation and recombination analysis

To study the complete genome evolutionary dynamics of RSV, we combined our newly obtained complete genomes with the genomes from the recent studies of Kumaria *et al.* (2011) and Rebuffo-Scheer *et al.* (2011). The latter (JF920046-JF920070), were recovered from nasopharyngeal and nasal swabs collected from patients in the Milwaukee metro area. A total of 72 RSV-A complete genomes were aligned using Mafft [39] and manually edited in Se-Al (available at http://tree.bio.ed.ac.uk/software/seal/). Recombination was analyzed using the RDP, GENECONV, RECSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ recombination detection methods implemented in RDP3 [40,41]. Only potential recombination events detected by two or more of these methods, together with phylogenetic evidence of recombination, were considered robust evidence of likely recombination. Based on the considerable recombination signal detected in the complete genomes obtained by Kumaria *et al*. (2011), further analyses were performed without these 14 sequences (on an alignment of 58 sequences). Since six recombination events evident within five of the remaining sequences may potentially undermine the validity of any phylogeny based analysis (the evolutionary history of recombinant sequences can in many cases not be

adequately described by a single phylogenetic tree), we took the necessary steps to appropriately account for these recombination events. We noted that the five recombinant genomes all had a clear majority of their sequences inherited from a single major parent and a minority of sequences inherited from one or more minor parents (Fig. S1). We pursued two strategies to account for these recombination events. In the first strategy we simply discarded the portion of recombinant sequences that had been inherited from their minor parents (corresponding to the B and C regions in Figure S1) and replaced these "deleted" genomic sites in the five aligned recombinant sequences with gap characters (i.e. we treated these sites as though they are unobserved in the probabilistic inferences). In the second strategy rather than simply deleting the genomic sites apparently derived from the minor parent, we separated the genomic regions of the recombinants derived from their major and minor parents and treated these separated regions as though they were separately sampled sequences. Importantly, a single phylogenetic history could be safely assumed for the genomic sequence datasets yielded by both strategies. For the selection pressure analyses of the complete genome, we stripped the non-coding regions from the alignments and extracted a concatenated coding gene alignment. To scrutinize the evolutionary dynamics of the G gene, we also aligned the G gene sequences of our complete genomes to a data set previously analyzed by Zlateva *et al*. (2004; Table S3), resulting in an alignment of 185 sequences encompassing amino acid position 91 to 297 of the G protein.

Bayesian estimation of RSV evolutionary history

To evaluate whether our complete genome data set showed a clear signal of nucleotide divergence throughout the sampling time interval, we explored linear regressions of root-to-tip divergence as a function of sampling time using Path-O-Gen (available at http://tree.bio.ed.ac.uk/software/pathogen/) [42]. For this purpose, we employed maximum likelihood trees reconstructed using PhyML [43] based on a general time-reversible (GTR) substitution model and a discretized gamma distribution to model rate variation among sites. The exact sampling data was known for all genomes except for six sequences obtained from Rebuffo-Scheer *et al.* (2011), which were set at the midpoint of the reported sampling year for this exploratory analysis.

We reconstructed the RSV evolutionary and demographic history using Bayesian Markov chain Monte Carlo (MCMC) in BEAST v1.7 [44]. BEAST infers a posterior distribution of time-measured genealogies based on a full probabilistic model including a molecular clock model, a nucleotide substitution model and a coalescent model as prior distribution for the tree. We used a general time-reversible (GTR) substitution model and a discretized gamma distribution to model rate variation among sites. To examine substitution rate variability across the genome, we also applied a partition model that allowed for separate substitution rates for the different genes and the single non-coding genome region. The fit of strict and relaxed clock models, assuming homogenous and heterogeneous substitution rates across the branches of the phylogeny respectively [45], was assessed based on a range of statistics including marginal likelihoods estimated using the harmonic mean, Akaike's information criterion through MCMC (AICM), path sampling (PS) estimators and stepping-stone sampling (SS) estimators [46,47]. For the six sequences for which the exact sampling date was not available, we estimated the tip ages constraining them within a time interval of one year (using a uniform prior, [48]). We used the Bayesian skyline plot model as a flexible demographic prior allowing us to reconstruct how the effective population size changed through time [49]. MCMC analyses were run until convergence could be safely assumed, as evaluated using Tracer (available at http://tree.bio.ed.ac.uk/software/tracer/). Marginal posterior distributions for the evolutionary rate and divergence times, such as the times to most recent common ancestors (TMRCA) of various groups of sequences, were summarized using means and 95% highest posterior density intervals (HPDs). We represented the posterior tree distribution using a maximum clade credibility (MCC) tree annotated with divergence time and evolutionary rate summaries and provided a visualization of this tree using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) [45].

To test whether the phylogenetic tree shape in the posterior distribution deviated from neutral expectations, we performed a genealogical test based on posterior predictive simulation [50]. Briefly, we summarize the tree shapes obtained by our Bayesian coalescent analysis using the genealogical Fu and Li statistic (D_F). This statistic compares the length of terminal branches to the total length of the coalescent genealogy and returns negative values for long terminal branch lengths,

which in turn indicates an excess of slightly deleterious mutations on these branches and hence a deviation from neutrality. The posterior predictive simulation randomly simulates trees with the same number of tips, the same tip ages, and under the same neutral coalescent model as the one applied to the real data. We employed the posterior predictive simulation extension that allows simulation of trees under the Bayesian skyline plot model [51]. Finally, we derive *P*-values for this genealogical neutrality test based on the frequency at which the D_F of the inferred trees was more extreme than the D_F of the simulated trees.

Selective pressure analyses

For the concatenated coding genes, we identified sites under diversifying selection using a fixed-effects likelihood (FEL) approach [52] and a recently developed renaissance counting (RC) procedure [53]. The FEL method fits codon models to each site independently and performs a likelihood ratio test to evaluate whether a model that assumes equal non-synonymous and synonymous rates (d_N=d_S) can be rejected in favor of a model that allows for different d_N and d_S rates. The RC approach combines stochastic mapping under nucleotide substitution models and empirical Bayes to estimate site-specific d_N/d_S ratios, and quantifies their uncertainty while accommodating phylogenetic uncertainty. Sites yielding 95% d_N/d_S credible intervals (CI) that do not include 1 reject neutrality and were considered to be evolving under positive selection. Following the advice of Kosakovsky Pond et al. (2005), we considered a consensus approach to identify sites evolving under positive diversifying selection. We only listed sites for which the FEL approach produces *P*-values < 0.1 and for which the lower 95% CI for the RC estimate is larger than 1.

To detect diversifying positive selection in the G gene data set, we also used a random-effects likelihood (REL)[54] approach in addition to FEL and RC. This method fits a codon model to the entire alignment but allows the d_N/d_S ratio to vary among sites. Positively selected sites were then identified using an empirical Bayes approach. We first tested the fit of three codon models of different complexity: a model that allows for different d_N rates among sites but keeps the d_S rate constant among sites ('Non-synonymous'), a model that allows both d_N and d_S to vary among sites ('Dual'), and, finally, a variant of the Dual model that also allows d_N to vary

among branches in the phylogeny ('Lineage Dual'). We used a general discrete distribution with three categories of sites to model d_N and d_S variation among sites in these substitution models. We applied the empirical Bayes method to identify positively selected sites to the best fitting codon model according to the AICM and used a log Bayes Factor larger than three as a cut-off. We only listed sites that met the cut-offs for at least two methods applied to the G gene data set and, in the results section below, listed in bold those that met the criteria for all three methods (FEL *P*-value < 0.1, RC 95% lower CI value > 1, REL ln BF > 3).

Finally, we also investigated both pervasive an episodic diversifying selection using a recently developed mixed-effects model of evolution (MEME)[55]. MEME relaxes the assumption that the strength and direction of natural selection is constant across all lineages by allowing the distribution of d_N/d_S to vary from site to site and also from branch to a branch at a site. As an extension of FEL, MEME fits a codon model with two categories of lineage-specific d_N rates to each site, one that is $\leq d_S$ (referred to as β^-) and one that is unrestricted (β^+), and tests this against a model where the unrestricted d_N is also $\leq d_S$ using a likelihood ratio test. We report positively selected sites with a P-value ≤ 0.05, which controls the rate of false positives fairly well[55], and include the estimate of β^+ and the proportion of branches estimated to be part of this d_N rate class (p^+).

Nucleotide sequence accession numbers

The nucleotide sequences from the Dutch and Belgian RSV isolates were submitted to GenBank under the accession numbers JQ901447-JQ901458, JX015479-JX015499. Prototype reference strains A2 (M74568.1), Long (AY911262.1), Line19 (FJ614813.1) and RSS-2 (NC_001803.1) were included in protein analysis studies.

Results

Human RSV-A protein substitution mapping

Protein sequences of thirty-three Dutch-Belgian clinical strains (Table S1) and four reference lab-strains were aligned and compared to the RSV consensus. Changes in amino acid residues from all strains were mapped for each RSV protein to

determine the protein substitution density for all proteins (Fig. 1). The percentage of sequence variability indicates the degree of variation from the consensus for each strain (Fig.1 and Table S4). G and M2-2 are the most variable proteins with 10-18% and 9-20% variability, respectively. The relatively high protein sequence variation observed for RSV G suggests more relaxed selective constraints in this protein and could even indicate, as has been previously identified [56], instances of molecular adaptation. The ectodomain mucin-like regions of RSV G are two highly variable regions that are separated by a central conserved region containing the heparin-binding domain (HBD) (Fig.2). The majority of non-conserved substitutions are located in these two regions. The C-terminal part of the RSV G immunogenic domain [57] has the highest substitution density, but non-conserved substitutions are detected throughout this domain. The RSV G immunogenic domain (a.a. 159-186), previously shown to be homologous to the TNF receptor fourth subdomain

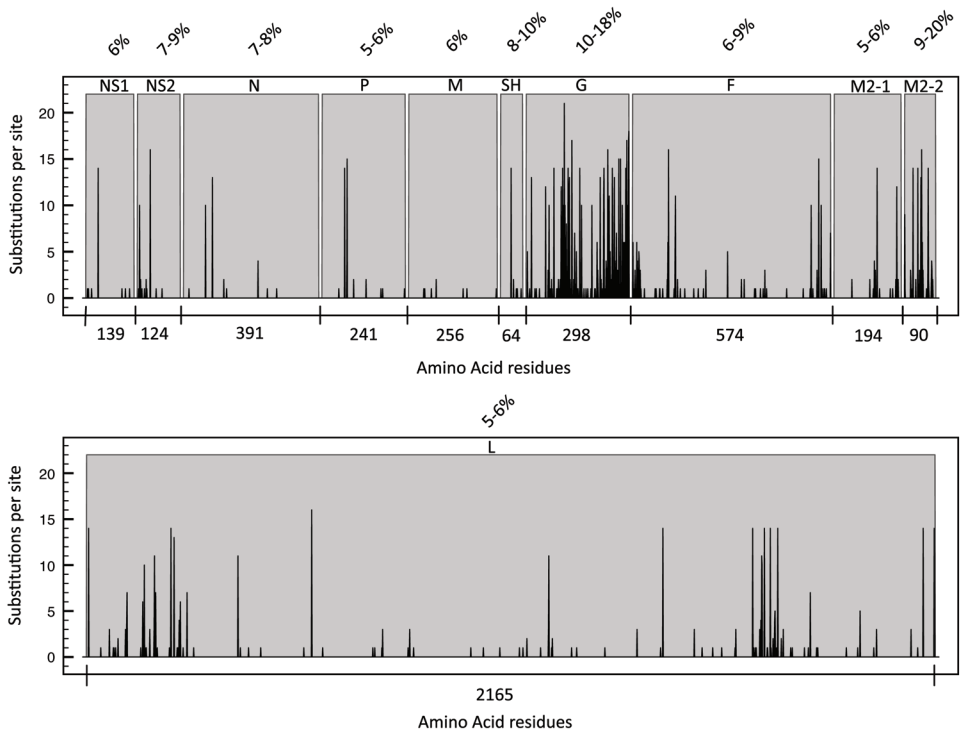


Figure 1. Sequence variability in RSV proteins. The number of substitutions per site (black bars) and the sequence variability (%) in each RSV protein calculated per strain relative to the concensus.

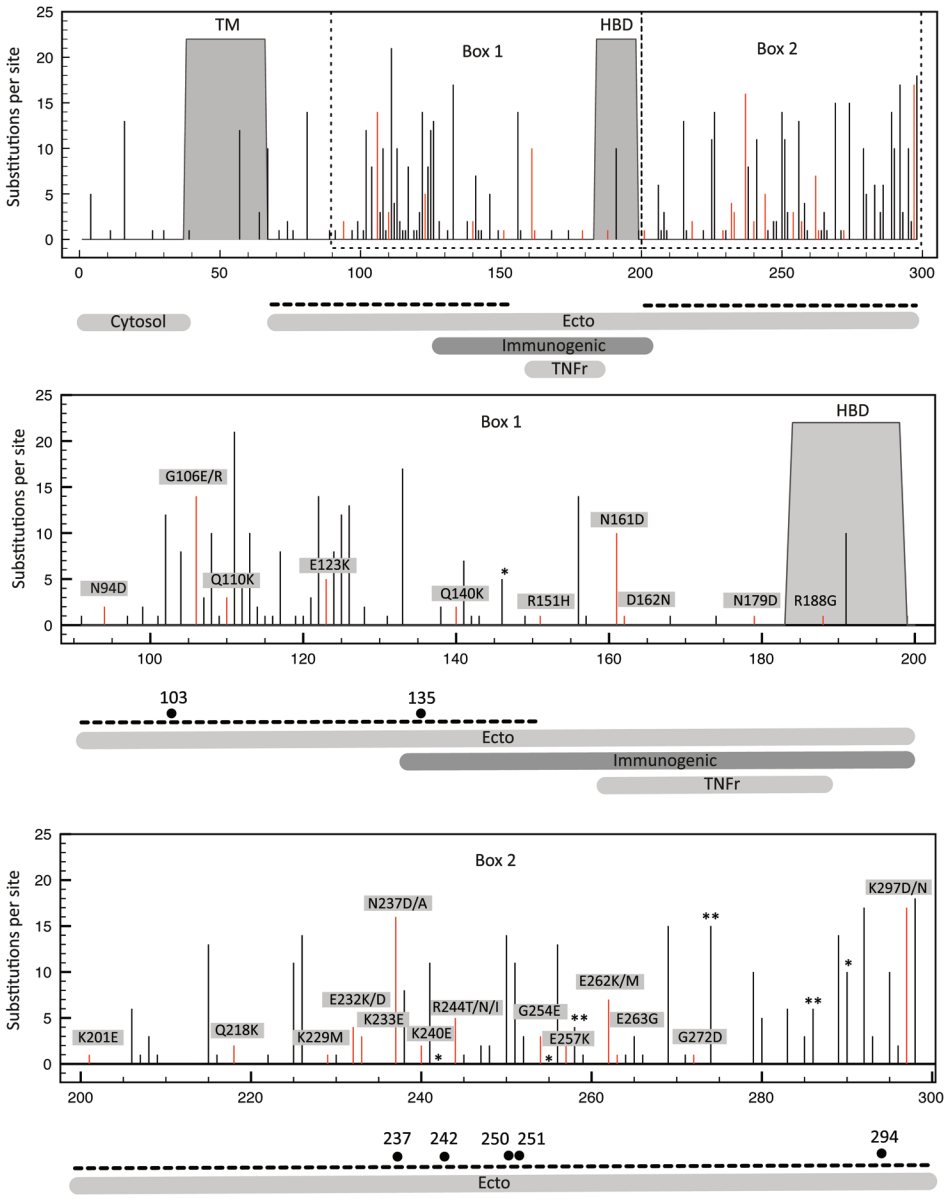


Figure 2. Amino acid substitutions in specific domains of the RSV G protein. Indicated are the transmembrane domain (TM), heparin binding domain (HBD), N-terminal cytosolic domain, C-terminal ectodomain, the immunogenic internal domain and the region showing partial homology with the fourth subdomain of the 55kDa TNFr. Dashed lines represent the mucin-like regions. N-glycosylation sites are indicated by black dots, conserved substitutions by black bars and non-conserved substitutions by red bars. Positively selected sites are marked by asteriks.

[58], shows only a few substitutions. However, the N161D substitution, as detected in both the RSV G immunogenic domain and the TNFr homologous domain, was present in ten out of thirty-seven strains. Interestingly, this residue is located in the protective epitope 152-163 of the central conserved domain of RSV G [59]. For several RSV strains, asparagine substitutions in the G protein result in the gain or loss of predicted N-glycosylation sites either by direct changes in the Asn-Xaa-Ser/Thr N-glycosylation sequons or by alterations in the surrounding sequence (Table S5).

The RSV F signal peptide that is involved in F protein translocation to the ER [60] is a major substitution hotspot (Fig.3A) whereas other parts of RSV F are rather conserved. Most substitutions in the mature F protein are located at the surface of this fusion protein (Fig.3B), which, therefore, might affect either antibody binding or interactions with cellular receptors. However, previously recognized antigenic regions [61] appear to be highly conserved. Non-conserved substitutions in the fusion protein are only found in the prototype strains RSV-Long (K80N, S213R) and RSV-line19 (E66K, S213R, T357K), as well as in the clinical strain 05-000417 (S213R). The E66K and K80N replacements are located in the F2 domain, which is involved in direct interaction with host cells and, accordingly, is a determinant of host species specificity. N-glycosylation sites are predicted for amino acid residues 27, 70, 116, 120 and 126. All of these are situated within the F2 domain and, in contrast to the N-glycosylation sites predicted for RSV G, are highly conserved with the exception of reference strain RSV-A2 and clinical strain 01-002279 that did not show N-glycosylation potential at position 120. The site at position 500 which was earlier predicted to have glycosylation potential [62] was not identified as such here (glycosylation potential <0.5).

Potentially artifactual recombination signals within some RSV genomes

An initial screen for recombination within a dataset containing the 39 RSV sequences available from the study of Kumaria *et al.* (2011) and Rebuffo-Scheer *et al.* (2011), together with the 33 newly determined sequences described here, revealed evidence of approximately 64 unique recombination events within the evolutionary histories of these sequences. Collectively 55 of these recombination events were clustered within the 14 RSV genome sequences deposited by Kumaria *et al*. (2011).

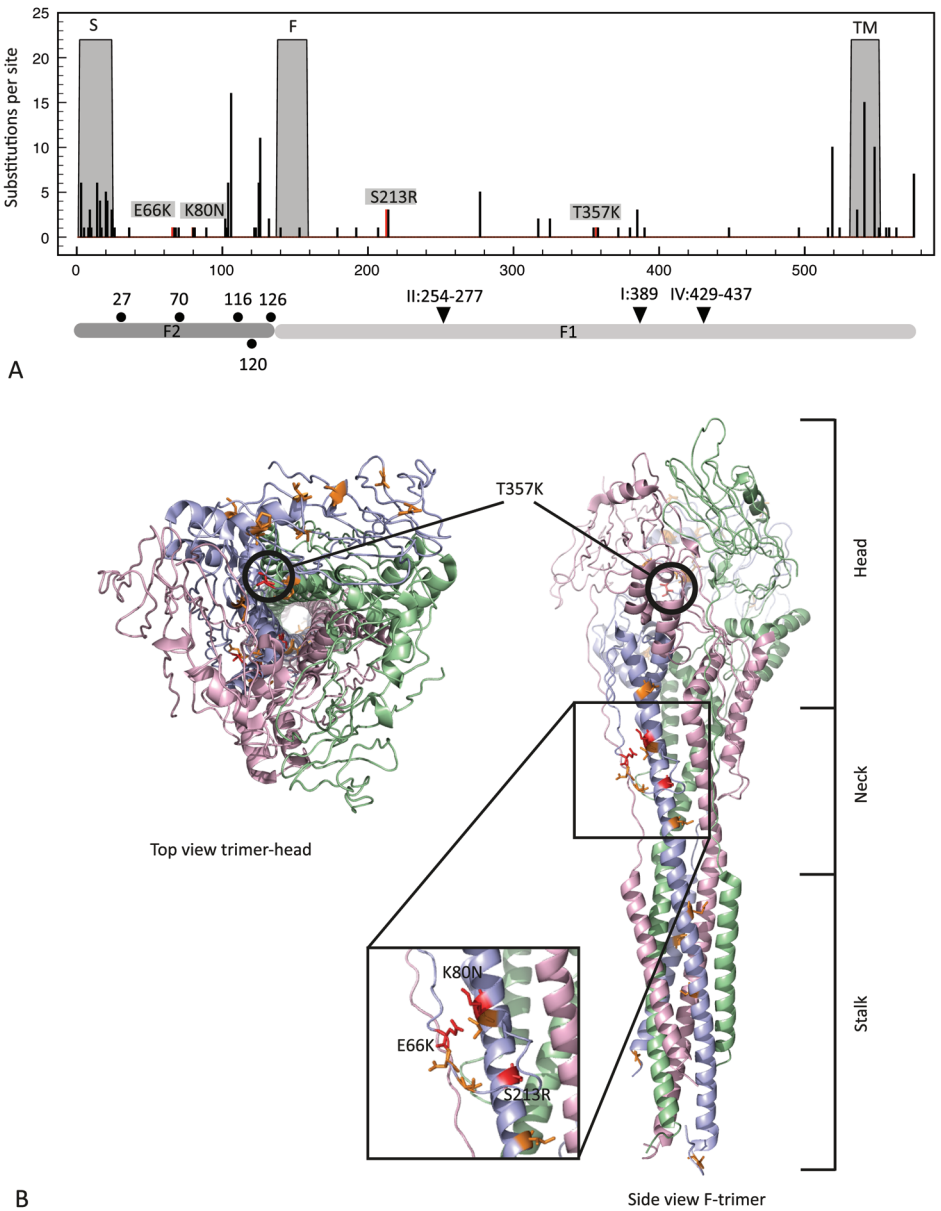


Figure 3. Amino acid substitutions in specific domains of the RSV F protein. A) Indicated are the signal peptide (S), fusion peptide (F). transmembrane domain (TM), F1 and F2 domains, antigenic sites I, II, IV (black triangles), and N-glycosylation sites (black dots). B) Substitutions marked in the 3D structure of the trimeric fusion protein. Amino acid class changes are indicated in red, orange indicating similar class.

The pattern of recombination within these sequences was extremely complex with, for example isolate RSV-12 (accession number GU591769) containing at least 10 recombination breakpoints and genome regions apparently derived from at least five different parental viruses.

The fact that signals of recombination were mostly concentrated within a set of genomes all determined within a single research laboratory suggested that these genomes might all be recombinant sequence assembly or amplification artifacts. Using the permutation-based recombination breakpoint-clustering test employed in [63], we detected an extremely significant association ($p<0.0001$) between breakpoint positions and the locations of the primer binding sites used to sequence these genomes. This very strongly suggests that the recombination signals detectable within these sequences are sequence assembly artifacts and we therefore excluded these sequences from subsequent recombination analyses.

Further recombination analyses indicated that five of the Milwaukee isolates, A/WI/629-4302/98, A/WI/629-4285/98, A/WI/629-3734/98, A/WI/629-4239/98 and A/WI/629-4111/98 [36] were detectably recombinant (collectively containing evidence of six recombination events; Figure S1). As the majority of these recombinant genomes (>70% of their sequences) were apparently derived from a single major parental virus, they were included in subsequent analyses with, in a "removed" dataset, the fragments derived from the minor parental viruses removed or, in a "distributed" dataset, the fragments derived from the different parents being separated into two different sequences.

Seven out of eleven Dutch-Belgian strains that were sequenced via the 454-sequencing method also showed some degree of recombination. After re-analyzing these (Table S1) strains by conventional Sanger sequencing, recombination events were no longer present confirming that short-read sequencing techniques and their associated contig-assembly methods can, in many cases produce artifactual recombinant viral genomes: A fact that should encourage all future studies of RSV whole genome sequences to explicitly account for recombination.

A Bayesian reconstruction of RSV-A complete genome evolutionary history

An exploratory linear regression analysis of root-tip-divergence as a function

of sampling time indicated that there is a clear temporal signal for nucleotide divergence in the RSV-A complete genomes sampled over the last thirteen years (Fig. S2), justifying the application of a dated tip molecular clock model in the Bayesian evolutionary reconstruction. We compared a strict and relaxed clock model using different approaches to assess model fit. Although this did not provide consistent results (Table S6), path sampling and stepping-stone sampling, which have recently been shown to provide the most accurate results [47], both rejected the assumption of a constant substitution rate across phylogenetic tree branches and hence we used this model in further analyses. We represent the posterior tree distribution for the 'distributed' RSV-A complete genome data set using a maximum clade credibility (MCC) tree in Fig. 4. The corresponding parts of each mosaic genome identified using RDP3 do not cluster together in this tree (green and black arrows/stars in the figure), which confirms that these genome regions are likely derived from genetically divergent viruses. The smaller genomic stretches extracted from these mosaic genomes are generally very closely related to other Milwaukee genomes, with four of them nearly identical to one another, highlighting the nonrandom nature of these mosaic patterns. The MCC tree for the 'removed' RSV-A complete genome data set had the same topology as for the 'distributed' RSV-A data set (data not shown).

The complete genome phylogeny does not indicate geographical or temporal clustering patterns because the Dutch-Belgian and Milwaukee strains are phylogenetically interspersed and strains from different epidemic seasons are present in the same clusters. For further classification of our new strains, we compared the G gene sequences extracted from the complete genomes to previously described sequences that have been assigned to specific subgroup A genotype lineages GA1, GA2, GA4, GA5 or GA7 [36,64] (Fig. S3). The Dutch-Belgian strains predominantly belong to the GA2 and GA5 lineages, and in this study, six Milwaukee strains (A/WI/629-4071/98, A/WI/629-3248/98, A/WI/629-4302/98, A/WI/629-4255/98, A/WI/629-3868/98 and A/WI/629-4266/98), which were previously assigned to the GA7 lineage, were classified as genotype GA2 in our analysis.

The RSV-A complete genome evolutionary history dates back to 1964 (95% credibility interval between 1956 and 1973). The inclusion of older strains sampled in 1954 and 1961 in the G gene pushes the MRCA further back in time (1943 [1923,1954],

Table 1). The Bayesian skyline model indicated that RSV populations over the past 70 years have been characterized by a constant degree of relative genetic diversity through time (for both the distributed and the removed data sets, Fig. S4). The genomic evolutionary rate was estimated to be 6.47 x 10^{-4} substitutions per site per year (95% credibility intervals ranging from 5.56 x 10^{-4} to 7.38 x 10^{-4}), which is

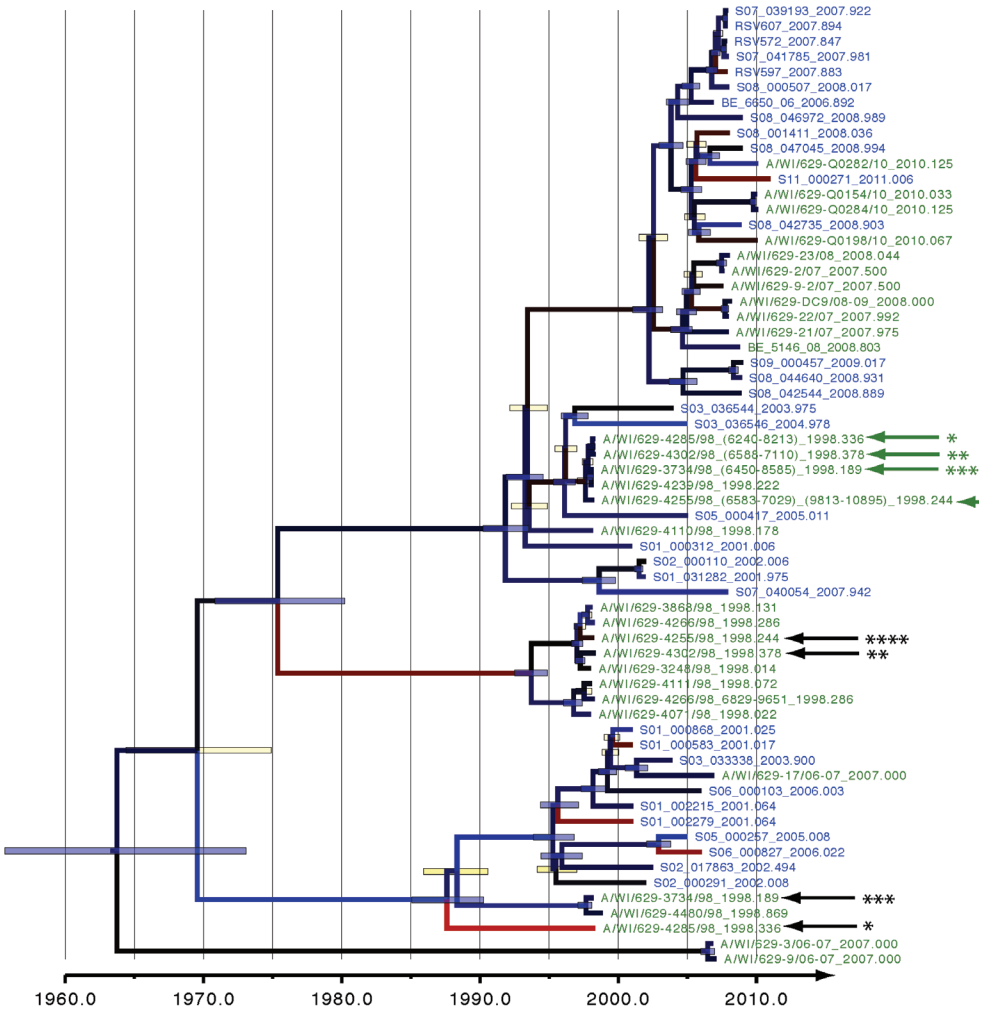


Figure 4. Phylogeny based on whole genome sequences. Distribution of Dutch-Belgian strains (blue) and Milwaukee strains (green). A color gradient (blue = slow, black = average, red = fast) reflects the variation in evolutionary rates among branches; node bars depict the credibility intervals for nodes showing a posterior probability support >95% (blue bar) or <95% (yellow bar). Recombinant partitions are indicated with arrows and asterisks.

considerably slower than rates previously estimated using only the G gene (1.83 x 10^{-3};[64]). Therefore, we investigated evolutionary rate variability across the genome by partitioning the various genes and the non-coding regions in our Bayesian estimation approach (Fig. 5). This revealed elevated substitution rates in the non-coding regions and the G gene. The latter also has the highest sequence variability score (5-9% variability; Fig S5). The relatively high rate we have estimated for the G gene partition is in good agreement with previous estimates of substitution rates in this gene [64] and indicates relaxed selective pressures operating on the G gene relative to other RSV genes. Specifically, other genes were considerably more conserved than the G gene and showed substitution rates that were 3- to 4-fold lower than those of the G gene.

Site-specific patterns of selective pressure

To identify positively selected sites in the RSV-A genome, we used a fixed-effects likelihood (FEL) approach and a recently developed renaissance counting (RC) method to evaluate nonsynonymous/synonymous substitution rate ratios (d_N/d_S) at each codon position. Table 2 lists sites that are apparently evolving under positive selection given the consensus cut-off values we propose for both methods (see Methods). The sites marked in bold also meet more stringent cut-offs for one of

Table 1. Evolutionary rate and TMRCA estimates.

			95% HPD	
		mean	**Lower**	**Upper**
Mean rate (subst./site/yr)	Distributed	6.47E-4	5.56E-4	7.38E-4
	Removed	6.44E-4	5.52E-4	7.46E-4
	G gene	2.22E-3	1.93E-3	2.56E-3
TMRCA (year)	Distributed	1964	1957	1972
	Removed	1964	1956	1973
	G gene	1942	1929	1953

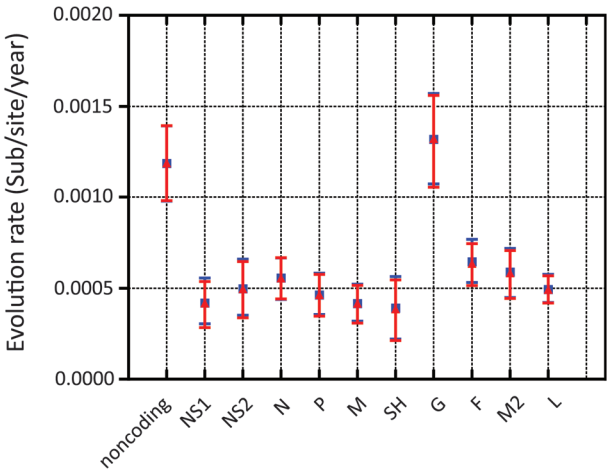


Figure 5. Evolutionary rate estimates for the different genes and for the non-coding partitions. Distributed (blue square) and removed (red triangle) data sets.

the two methods. As expected from the evolutionary rate variability across the genome, the majority of positively selected sites are found in the G gene, which can be interpreted as evidence for the action of diversifying selection at these sites. Single positively selected sites were detected in the L gene (site 104), the N gene (site 216) and the M2 gene (site 117; with support for the latter two sites being lower than that for the L gene site).

To further scrutinize the instances of diversifying selection in the G gene, we analyzed the larger G gene data sets with the same methods and an additional random-effects likelihood (REL) approach. For the latter, we first tested, using an AIC-based approach, whether the applied codon substitution model needed to account for nonsynonymous and synonymous rate variability across sites and/or branches (Table 3). We found strong evidence for both nonsynonymous and synonymous rate variability across sites, but no support for a model that also accommodates a separate nonsynonymous rate for each branch. Sites that were found positively selected according to the cut-offs for at least two methods are listed in Table 2; sites that meet the criteria for all three methods are listed in bold. There is only partial agreement between the sites identified in the G gene data (underlined in Table 2) and those found to be positively selected based on the G genes sampled from

Table 2. Site-specific positive selection pressure.

Data set	Positive selected site	Gene	RC	FEL		REL	
			dN/dS	dN-dS	P value	E[dN/dS]	Log(BF)
Full genome	**286**	G	2.62 (2.06,3.95)	9.49	<0.01	N/A	N/A
	244	G	2.54 (2.04,3.72)	8.12	0.1	N/A	N/A
	274	G	2.51 (2.05,3.69)	6.43	0.02	N/A	N/A
	250	G	2.11 (1.71,3.09)	5.17	0.01	N/A	N/A
	104	L	1.73 (1.37,2.65)	3.55	0.08	N/A	N/A
	101	G	1.72 (1.29,2.59)	3.67	0.04	N/A	N/A
	258	G	1.72 (1.38,2.51)	3.86	0.04	N/A	N/A
	216	N	1.32 (1.05,1.99)	4.15	0.1	N/A	N/A
	117	M2	1.30 (1.00,1.99)	4.31	0.1	N/A	N/A
	237	G	1.29 (1.05,1.89)	4	0.04	N/A	N/A
G gene	**274**	G	4.08 (2.63,6.39)	3.11	0.01	2.73	4.3
	290	G	3.19 (2.04,4.95)	1.81	0.03	2.71	4.14
	244	G	2.81 (1.83,4.34)	1.99	0.11	2.48	3.26
	258	G	2.52 (1.65,3.94)	1.58	0.03	2.68	3.99
	146	G	2.19 (1.33,3.38)	1.31	0.04	2.47	3.18
	286	G	2.00 (1.32,3.16)	1.67	0.05	2.7	4.13
	285	G	1.92 (1.25,2.98)	1.05	0.08	1.65	1.37
	242	G	1.91 (1.24,2.99)	1.92	0.02	2.55	3.45
	255	G	1.90 (1.25,2.92)	1.72	<0.01	2.55	3.41
	132	G	1.16 (0.74,1.82)	2.12	0.03	2.71	4.15

Sites marked in bold are positive selected sites that meet the criteria of all applied methods. Underlined sites are positive selected sites that agree with the G gene and full genome analysis. RC = Renaissance Counting, FEL = Fixed-Effect Likelihood, REL = Random-Effects Likelihood, dN = non-synonymous substitution, dS = synonymous substitution and log(BF) = Bayes Factor logarithm. REL could not be acquired for the whole genome data set (N/A).

Table 3. Model-fit of random-effect likelihood (REL) models.

Model	Log L	AIC	cv(dS)
Nonsynonymous	-6376.14	13506.28	0.00
Dual	-6345.39	13452.78	7.29
Lineage Dual	-6181.06	13856.12	13.18

Log L = logarithm of the maximum lilelihood value for the model
AIC = value of the Akaike Information Criterion model selection index
CV(dS) = coefficient of variation of the synonymous substitution distribution

the complete genome dataset. Most positively selected sites were detected within the two hyper-variable regions of the G protein ectodomain (marked with an * in Figure 2). The positively selected site 242 is a predicted N-glycosylation site and substitutions can result in loss of glycosylation.

The methods we applied to investigate site-specific selective patterns will generally identify sites that are under positive diversifying selection in most lineages of the tree (pervasive diversifying selection). However, sites may also experience diversifying selection in a restricted number of branches (episodic diversifying selection). To discriminate between these processes, and to identify sites under episodic diversifying selection in particular, we applied a recently developed approach that relaxes the assumption of constant diversifying selection throughout time (MEME,[55]; Table 4). MEME identified 16 sites under positive diversifying selection (P-value ≤ 0.05), six of which were already pinpointed by the other methods (in bold in Table 4). Not surprisingly, there is a large proportion branches estimated to be under diversifying selection (p^+) for these six sites. Most other sites are candidates for sites under episodic diversifying selection, some of which only implicate a small subset of branches but with high nonsynonymous substitution rates (β^+). To contrast the substitution history for sites under pervasive and episodic selection, we mapped the amino acid residues for four G gene sites to the internal MCC tree nodes in Figure S6. This demonstrates that the amino acid patterns for the two sites with the highest RC d_N/d_S estimates (site 274 and 290 under pervasive selection), are the result of a relatively rich substitution history on both external an internal branches, whereas substitutions for the two sites under episodic selection

(site 154 and 255) occur on only three external branches and one external branch respectively. We caution against over-interpreting the latter because a single substitution on an external branch may also reflect a deleterious mutation that has been weeded out by purifying selection [65].

The neutral phylodynamics of RSV-A

To test the impact of diversifying selection in the G gene on virus circulation patterns, we conducted a genealogical test of neutrality using posterior predictive simulation. Comparison of the tree shapes inferred from the complete genomes to those obtained by simulating under the neutral model with the same population dynamics indicated that neutrality could not be rejected at the population level (P

Table 4. Sites under episodic diversifying selection in the G gene data set.

Positive selected site	β^+	p^+	P value
154	2010.12	0.01	0.00
255	19.06	0.09	0.00
274	3.48	1.00	0.00
224	113.89	0.01	0.01
290	2.12	1.00	0.02
256	23.86	0.13	0.02
297	18.83	0.11	0.03
250	21.85	0.12	0.03
286	2.47	1.00	0.03
132	1.67	1.00	0.03
146	1.5	1.00	0.04
247	1.18	1.00	0.04
273	1.37	0.64	0.04
218	252.07	0.01	0.04
258	3.95	0.44	0.05
124	3.96	0.44	0.05

Sites marked in bold are positive selected sites that meet the criteria of all applied methods

= 0.24). The same holds true for the G gene data set although the inclusion of older samples and the resulting longer time scale may have resulted in a lower *P*-value in this case (P = 0.08).

Discussion

Here we present the first study describing RSV-A whole-genomic diversity through time using phylodynamic inference. We mapped the rate and density of nucleotide substitutions in viral gene products to provide an indication of the relative importance of individual protein evolutionary dynamics in RSV replication or immune evasion. The evolutionary and population dynamics of RSV-A were estimated using a Bayesian coalescent approach, based on two different datasets of sequences from time-stamped strains collected between 2001 and 2011. One set was recently collected in the Milwaukee area [36]. Like previous molecular evolution studies [64,66], Rebuffo-Scheer *et al* [35,36] described the epidemiology of these strains according to G gene genotyping. The current study adds another set of thirty-three complete genomic sequences to the literature, enabling in-depth examination of RSV evolutionary dynamics by Bayesian genealogical inference. Although the collective sampling arose from a few specific locations, the general absence of population substructure and the co-circulation of different lineages through time ensure that we analyzed a representative RSV-A genomic sample.

Most nucleotide sequence variability was detected in the non-coding regions, as well as the M2-2 and G genes. The role of intergenic non-coding regions and their high sequence variability is unknown. In the G and M2-2 proteins high nucleotide sequence variability is correlated with elevated amino acid substitution rates. Variability in the M2-2 transcription regulatory protein sequences could largely be attributed to substitutions in reference strains (Long, Line19, and A2), suggesting that this variability may be attributable to adaptation to artificial growth conditions during long-term culturing. Given the role of M2-2 in switching from transcription to RNA replication during the virus infection cycle, together with our observation that none of the clinical strains shows such increased variability, we hypothesize that the absence of immunological constraints on the cultured isolates may have

allowed changes in this switching protein that facilitate more efficient packaging and increased viral yields *in vitro*. We did not detect any difference in the nucleotide sequence of three RSV strains sequenced either directly from patient isolates or after passaging on HEp2 cells for three culture rounds (data not shown). Apparently, short-term culturing does not result in a noticeable accumulation of culture adaptive substitutions and it may therefore be acceptable to use whole genomic sequences from RSV isolates with low culture passage levels in future genetic analyses. The G protein is the virus' most immunogenic protein and it has been suggested that neutralizing antibodies drive positive selection for immune escape variants of the ectodomain [67]. All other RSV genes show high degrees of conservation, suggesting that the impact of immune pressures on the evolution of these genes is minimal. Nevertheless the substitutions observed in these genes might very well affect RSV fitness and its transmission dynamics. The exact mapping of conserved sites throughout the genome might be very useful for both future research into the biological functions of these regions as well as for the development of therapeutic and preventive strategies.

An important prerequisite to performing phylogenetic inference is a thorough check of recombination events occurring in the genomes being analyzed. Recombination in RSV genomes is unprecedented in natural isolates and has only very rarely been detected under controlled laboratory conditions [68]. Therefore, it is not very likely that any of the recombination events detected here are the product of natural RSV evolutionary processes. Indeed, our analyses suggest that in both the case of sequences deposited by Kumaria *et al*. (2011), as well as in seven of our own 454-derived sequences, the recombinant sequences are likely to be either PCR or sequence assembly artifacts. Especially our own findings underscore the potential pitfalls of applying phylogenetics-based analytical approaches to sequences derived through the application of short-read sequencing techniques. To incorporate published strains with recombination signal, we compared a data set where minor recombinant parts were removed and a data set where sequence parts with different phylogenetic clustering were distributed over multiple taxa. These data sets yielded highly similar evolutionary rate and demographic estimates, but we noted that a large number of such distributed strains may complicate the interpretation of demographic inferences. This is because the relationship between

viral coalescent rates and the change in effective number of infections through time only holds when each individual is represented by a single sequence.

To assess whether the genotype grouping is consistent when different genomic regions are used, we compared the phylogenetic trees constructed from all single genes, as well as the whole genome based tree. The analysis of the non-recombinant complete genome strains revealed consistent genotype clustering patterns constructed from all single genes, including the most variable G gene (data not shown). Therefore, we conclude that genotyping based only on G gene sequences, appears to be a suitable method for assigning genotypes to RSV strains. However, in general there are no equidistant clusters corresponding to RSV genotypes and differences in the analyzed strain population, i.e. typing strains against different background populations, might give rise to different classifications as observed for six Milwaukee strains (A/WI/629-4071/98, A/WI/629-3248/98, A/WI/629-4302/98, A/WI/629-4255/98, A/WI/629-3868/98 and A/WI/629-4266/98), which were previously assigned to the GA7 lineage, but are classified as genotype GA2 in our analysis. This indicates that RSV genotyping may be intrinsically volatile.

Changes in amino acid sites that are either N-glycosylated or susceptible to selective pressure correlate with RSV phenotypic differences and variations in RSV infectiousness. This may ultimately affect the outcome of RSV infection and transmission dynamics. We studied the N-glycosylation sites within the two major RSV glycoproteins, F and G. Here we observed differences in the predicted N-glycosylation sites for the F protein when comparing our data to those of Zimmer *et al* [62]. Modifications in asparagine residues and the related alterations in N-glycosylation potential could affect RSV G and F protein folding, possibly resulting in altered virus-host attachment, changes in immune evasion strategies or distorted fusion abilities. Zimmer *et al* used the RSV-Long strain to conduct radio-immunoprecipitations with recombinant RSV F mutants to determine the N-glycan positions. They reported glycosylation at N27, N70 and N500. Here, position N500, which is known to be required for efficient syncytia formation, was not predicted to be glycosylated, underscoring the need for *in vitro* confirmation of predicted N-glycosylation sites. In addition, N-glycosylation is not only determined by sequence composition, but also heavily depends on the type of host cells that are infected [69].

Selective pressure analysis showed that the RSV genes have predominantly negatively selected, or neutrally evolving sites. Only the RSV G gene displays clear evidence of pervasive positive diversifying selection. Using the full genome dataset we detected positively selected sites at G protein positions 101, 237, 244, 250, 258, 274, and 286. In the larger G gene exclusive dataset, additional positively selected sites at positions 146, 242, 255 and 258 were detected by the various detection methods (Table 2). Positively selected site 237 has an asparagine residue with N-glycosylation potential [18] for approximately 50% of the strains included in this study. More positively selected sites were found at predicted N-glycosylation positions 242, 244, 250 and 258. Substitutions at positions 274 and 286 have previously been associated with antibody escape [70]. Less is known about the effects of amino acid replacements at sites 101, 146, and 255, although positions 101 and 146 fall within the immunogenic region of RSV G [57]. Clearly, the exact G gene sites identified as evolving under positive selection are not consistent among different studies [64,71]. The vast majority of previously described sites [59,72] were not predicted here, even when using the G gene only data set. Among the sites predicted in this study, positions 242, 255 and 258 were also predicted by Botosso *et al.* (2009) [71]. Differences in positively selected sites can be attributed to different data sets and/or the use of different analytical methods. For example, unaccounted recombination is known to yield false evidence of positive selection. The fact that the amount of sequence data strongly impacts the power of selection analyses is evident from the positively selected sites in the G gene when comparing the smaller complete genome data set to the larger G gene specific data set. Our codon model comparison also indicates an important degree of variation in synonymous substitution rates among sites, which could result in sites being falsely identified as evolving under positive diversifying selection if not accounted for. The sensitivity to sequence sampling and the amount of data is less of an issue for the recently developed MEME approach [55], which also captures episodic diversifying selection by disentangling lineages under purifying and positive selection for each site. MEME confirmed six sites under pervasive diversifying selection, but also identifies a considerable number of sites under episodic diversifying selection. This confirms the notion that site-specific positive selection may have been largely underestimated because purifying selection in many lineages can mask episodic adaptation.

The degree of phylogenetic intermixing of samples from different geographical regions in the maximum clade credibility tree indicates that there is likely very little, if any, temporal or long-term geographical structuring of RSV populations. Several genotypes may co-circulate and selection leading to the replacement of predominating genotypes is likely to be far more sporadic than that which, for example, occurs with influenza viruses where high host population-wide prevalence's of antibodies targeting the hemagglutunin (HA) and neuraminidase (NA) genes are able to cause extinction of specific clades. In contrast to human Influenza A [50] and Norovirus GII.4 [51], no evidence of selection-driven viral population turnover was uncovered by our posterior predictive analysis of the whole genome (P=0.24), which indicates that substitutions are generally random. This, in combination with the high degree of variability detected for the G protein allows us to conclude that the RSV attachment process, mediated by N- and O-linked sugars of the RSV G attachment glycoprotein, does not put significant evolutionary constraints on the selection of circulating variants. Consequently, the RSV G protein is not expected to induce functional B-cell mediated immunity and is therefore unlikely to be an effective vaccine target. For the RSV F protein the situation is completely different given the fact that antibodies targeting RSV F (e.g. RespiGam®, Synagis®) are clearly protective when administered as RSV prophylactics. Our inability to detect RSV F escape mutants indicates that immune evasion mutations in this gene are either rare, episodic, or have a strongly detrimental effect on virus viability. This reinforces the suitability of RSV F as an effective vaccine target.

In conclusion, our evolutionary analysis indicates that the RSV genome is largely conserved and, with the exception of the G protein, almost exclusively accumulates neutral substitutions. There is very little evidence to suggest that the evasion of adaptive host immunity is a primary driver of the observed RSV diversity. This might reflect a situation in which individuals with immune systems that attempt to acquire immunologic memory of the virus destabilize the immunologic balance of infections, both decreasing their tolerance for RSV and, increasing the likelihood that they will experience an enhanced disease outcome. Based on these findings we suggest firstly that future control strategies should target the RSV F protein and, secondly, that these strategies might be potentiated by co-therapeutic strategies modulating natural host immune responses.

Acknowledgements

We are grateful to Hanneke Hoelen for providing support in 3D-modelling and Marc van Ranst for giving access to his collection of Belgian RSV-A isolates.

References

1. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, et al. (2010) *Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis*. Lancet 375: 1545-1555.

2. Glezen WP, Taber LH, Frank AL, Kasel JA (1986) *Risk of primary infection and reinfection with respiratory syncytial virus*. Am J Dis Child 140: 543-546.

3. Henrickson KJ, Hoover S, Kehl KS, Hua W (2004) *National disease burden of respiratory viruses detected in children by polymerase chain reaction*. Pediatr Infect Dis J 23: S11-18.

4. Falsey AR, Walsh EE (2000) *Respiratory syncytial virus infection in adults*. Clin Microbiol Rev 13: 371-384.

5. Englund JA, Anderson LJ, Rhame FS (1991) *Nosocomial transmission of respiratory syncytial virus in immunocompromised adults*. J Clin Microbiol 29: 115-119.

6. Fleming DM, Cross KW (1993) *Respiratory syncytial virus or influenza?* Lancet 342: 1507-1510.

7. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE (2005) *Respiratory syncytial virus infection in elderly and high-risk adults*. N Engl J Med 352: 1749-1759.

8. White LJ, Mandl JN, Gomes MG, Bodley-Tickell AT, Cane PA, et al. (2007) *Understanding the transmission dynamics of respiratory syncytial virus using multiple time series and nested models*. Math Biosci 209: 222-239.

9. Anderson LJ, Hendry RM, Pierik LT, Tsou C, McIntosh K (1991) *Multicenter study of strains of respiratory syncytial virus*. J Infect Dis 163: 687-692.

10. Cane PA, Matthews DA, Pringle CR (1994) *Analysis of respiratory syncytial virus strain variation in successive epidemics in one city*. J Clin Microbiol 32: 1-4.

11. Peret TC, Hall CB, Hammond GW, Piedra PA, Storch GA, et al. (2000) *Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America*. J Infect Dis 181: 1891-1896.

12. Mufson MA, Orvell C, Rafnar B, Norrby E (1985) *Two distinct subtypes of human respiratory syncytial virus*. J Gen Virol 66 (Pt 10): 2111-2124.

13. Anderson LJ, Hierholzer JC, Tsou C, Hendry RM, Fernie BF, et al. (1985) *Antigenic characterization of*

respiratory syncytial virus strains with monoclonal antibodies. J Infect Dis 151: 626-633.

14. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ (1998) *Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community*. J Gen Virol 79 (Pt 9): 2221-2229.

15. Hall CB, Walsh EE, Schnabel KC, Long CE, McConnochie KM, et al. (1990) *Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children*. J Infect Dis 162: 1283-1290.

16. Mufson MA, Belshe RB, Orvell C, Norrby E (1988) *Respiratory syncytial virus epidemics: variable dominance of subgroups A and B strains among children, 1981-1986*. J Infect Dis 157: 143-148.

17. Hendry RM, Pierik LT, McIntosh K (1989) *Prevalence of respiratory syncytial virus subgroups over six consecutive outbreaks: 1981-1987*. J Infect Dis 160: 185-190.

18. Johnson PR, Spriggs MK, Olmsted RA, Collins PL (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: 5625-5629.

19. Sullender WM (2000) *Respiratory syncytial virus genetic and antigenic diversity*. Clin Microbiol Rev 13: 1-15, table of contents.

20. Chin J, Magoffin RL, Shearer LA, Schieble JH, Lennette EH (1969) *Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population*. Am J Epidemiol 89: 449-463.

21. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE (1969) *An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine*. Am J Epidemiol 89: 405-421.

22. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, et al. (1969) *Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine*. Am J Epidemiol 89: 422-434.

23. Murphy BR, Walsh EE (1988) *Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity*. J Clin Microbiol 26: 1595-1597.

24. Olson MR, Varga SM (2007) *CD8 T cells inhibit respiratory syncytial virus (RSV) vaccine-enhanced disease*. J Immunol 179: 5415-5424.

25. Castilow EM, Olson MR, Meyerholz DK, Varga SM (2008) *Differential role of gamma interferon in inhibiting pulmonary eosinophilia and exacerbating systemic disease in fusion protein-immunized mice undergoing challenge infection with respiratory syncytial virus*. J Virol 82: 2196-2207.

26. Kruijsen D, Schijf MA, Lukens MV, van Uden NO, Kimpen JL, et al. (2011) *Local innate and adaptive*

immune responses regulate inflammatory cell influx into the lungs after vaccination with formalin inactivated RSV. Vaccine 29: 2730-2741.

27. Kruijsen D, Bakkers MJ, van Uden NO, Viveen MC, van der Sluis TC, et al. (2010) *Serum antibodies critically affect virus-specific CD4+/CD8+ T cell balance during respiratory syncytial virus infections*. J Immunol 185: 6489-6498.

28. Friedewald WT, Forsyth BR, Smith CB, Gharpure MA, Chanock RM (1968) *Low-temperature-grown RS virus in adult volunteers*. JAMA 204: 690-694.

29. Kim HW, Arrobio JO, Pyles G, Brandt CD, Camargo E, et al. (1971) *Clinical and immunological response of infants and children to administration of low-temperature adapted respiratory syncytial virus*. Pediatrics 48: 745-755.

30. Wright PF, v Mills J, Chanock RM (1971) *Evaluation of a temperature-sensitive mutant of respiratory syncytial virus in adults*. J Infect Dis 124: 505-511.

31. Murphy BR, Hall SL, Kulkarni AB, Crowe JE, Jr., Collins PL, et al. (1994) *An update on approaches to the development of respiratory syncytial virus (RSV) and parainfluenza virus type 3 (PIV3) vaccines*. Virus Res 32: 13-36.

32. Tang YW, Graham BS (1997) *T cell source of type 1 cytokines determines illness patterns in respiratory syncytial virus-infected mice*. J Clin Invest 99: 2183-2191.

33. Klein Klouwenberg P, Tan L, Werkman W, van Bleek GM, Coenjaerts F (2009) *The role of Toll-like receptors in regulating the immune response against respiratory syncytial virus*. Crit Rev Immunol 29: 531-550.

34. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) *Unifying the epidemiological and evolutionary dynamics of pathogens*. Science 303: 327-332.

35. Kumaria R, Iyer LR, Hibberd ML, Simoes EA, Sugrue RJ (2011) *Whole genome characterization of non-tissue culture adapted HRSV strains in severely infected children*. Virol J 8: 372.

36. Rebuffo-Scheer C, Bose M, He J, Khaja S, Ulatowski M, et al. (2011) *Whole genome sequencing and evolutionary analysis of human respiratory syncytial virus A and B from Milwaukee, WI 1998-2010*. PLoS One 6: e25468.

37. Holmes EC, Grenfell BT (2009) *Discovering the phylodynamics of RNA viruses*. PLoS Comput Biol 5: e1000505.

38. Gupta R, Brunak S (2002) *Prediction of glycosylation across the human proteome and the correlation to protein function*. Pac Symp Biocomput: 310-322.

39. Katoh K, Misawa K, Kuma K, Miyata T (2002) *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform*. Nucleic Acids Res 30: 3059-3066.

40. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) *RDP3: a flexible and fast computer program for analyzing recombination*. Bioinformatics 26: 2462-2463.

41. Martin DP, Lemey P, Posada D (2011) *Analysing recombination in nucleotide sequences*. Mol Ecol Resour 11: 943-955.

42. Drummond A, Pybus OG, Rambaut A (2003) *Inference of viral evolutionary rates from molecular sequences*. Adv Parasitol 54: 331-358.

43. Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) *Estimating maximum likelihood phylogenies with PhyML*. Methods Mol Biol 537: 113-137.

44. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. Mol Biol Evol.

45. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) *Relaxed phylogenetics and dating with confidence*. PLoS Biol 4: e88.

46. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH (2011) *Improving marginal likelihood estimation for Bayesian phylogenetic model selection*. Syst Biol 60: 150-160.

47. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) *Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty*. Mol Biol Evol.

48. Shapiro B, Ho SY, Drummond AJ, Suchard MA, Pybus OG, et al. (2011) *A Bayesian phylogenetic method to estimate unknown sequence ages*. Mol Biol Evol 28: 879-887.

49. Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) *Time dependency of molecular rate estimates and systematic overestimation of recent divergence times*. Mol Biol Evol 22: 1561-1568.

50. Drummond AJ, Suchard MA (2008) *Fully Bayesian tests of neutrality using genealogical summary statistics*. BMC Genet 9: 68.

51. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, et al. (2010) *Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants*. PLoS Pathog 6: e1000884.

52. Kosakovsky Pond SL, Frost SD (2005) *Not so different after all: a comparison of methods for detecting amino acid sites under selection*. Mol Biol Evol 22: 1208-1222.

53. Lemey P, Minin VN, Bielejec F, Kosakovsky Pond SL, Suchard MA (2012) *A counting renaissance: Combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection*. Bioinformatics.

54. Pond SK, Muse SV (2005) *Site-to-site variation of synonymous substitution rates*. Mol Biol Evol 22: 2375-2385.

55. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) *Detecting individual sites subject to episodic diversifying selection*. PLoS Genet 8: e1002764.

56. Wertheim JO, Worobey M (2009) *Relaxed selection and the evolution of RNA virus mucin-like*

pathogenicity factors. J Virol 83: 4690-4694.

57. Simard C, Nadon F, Seguin C, Trudel M (1995) *Evidence that the amino acid region 124-203 of glycoprotein G from the respiratory syncytial virus (RSV) constitutes a major part of the polypeptide domain that is involved in the protection against RSV infection*. Antiviral Res 28: 303-315.

58. Langedijk JP, de Groot BL, Berendsen HJ, van Oirschot JT (1998) *Structural homology of the central conserved region of the attachment protein G of respiratory syncytial virus with the fourth subdomain of 55-kDa tumor necrosis factor receptor*. Virology 243: 293-302.

59. Plotnicky-Gilquin H, Goetsch L, Huss T, Champion T, Beck A, et al. (1999) *Identification of multiple protective epitopes (protectopes) in the central conserved domain of a prototype human respiratory syncytial virus G protein*. J Virol 73: 5637-5645.

60. Martin D, Calder LJ, Garcia-Barreno B, Skehel JJ, Melero JA (2006) *Sequence elements of the fusion peptide of human respiratory syncytial virus fusion protein required for activity*. J Gen Virol 87: 1649-1658.

61. McLellan JS, Yang Y, Graham BS, Kwong PD (2011) *Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes*. J Virol 85: 7788-7796.

62. Zimmer G, Trotz I, Herrler G (2001) *N-glycans of F protein differentially affect fusion activity of human respiratory syncytial virus*. J Virol 75: 4744-4751.

63. Lefeuvre P, Lett JM, Varsani A, Martin DP (2009) *Widely conserved recombination patterns among single-stranded DNA viruses*. J Virol 83: 2697-2707.

64. Zlateva KT, Lemey P, Vandamme AM, Van Ranst M (2004) *Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup a: positively selected sites in the attachment g glycoprotein*. J Virol 78: 4675-4683.

65. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, et al. (2007) *Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics*. PLoS Comput Biol 3: e29.

66. Zlateva KT, Lemey P, Moes E, Vandamme AM, Van Ranst M (2005) *Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein*. J Virol 79: 9157-9167.

67. Woelk CH, Holmes EC (2001) *Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV)*. J Mol Evol 52: 182-192.

68. Spann KM, Collins PL, Teng MN (2003) *Genetic recombination during coinfection of two mutants of human respiratory syncytial virus*. J Virol 77: 11201-11211.

69. Rixon HW, Brown C, Brown G, Sugrue RJ (2002) *Multiple glycosylated forms of the respiratory syncytial virus fusion protein are expressed in virus-infected cells*. J Gen Virol 83: 61-66.

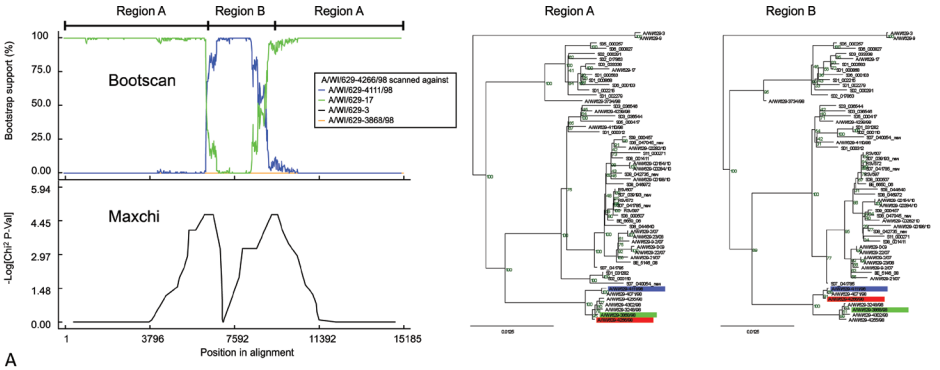
70. Rueda P, Delgado T, Portela A, Melero JA, Garcia-Barreno B (1991) *Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies*. J Virol 65: 3374-3378.

71. Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, et al. (2009) *Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus*. PLoS Pathog 5: e1000254.

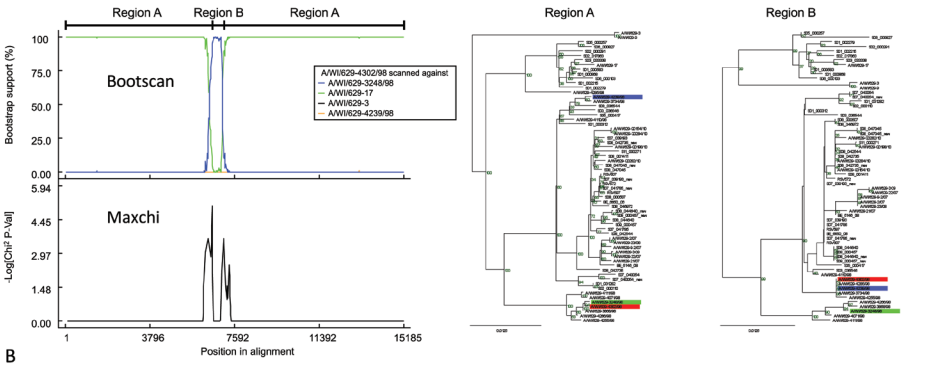
72. Gaunt ER, Jansen RR, Poovorawan Y, Templeton KE, Toms GL, et al. (2011) *Molecular epidemiology and evolution of human respiratory syncytial virus and human metapneumovirus*. PLoS One 6: e17427.

SUPPLEMENTARY INFO

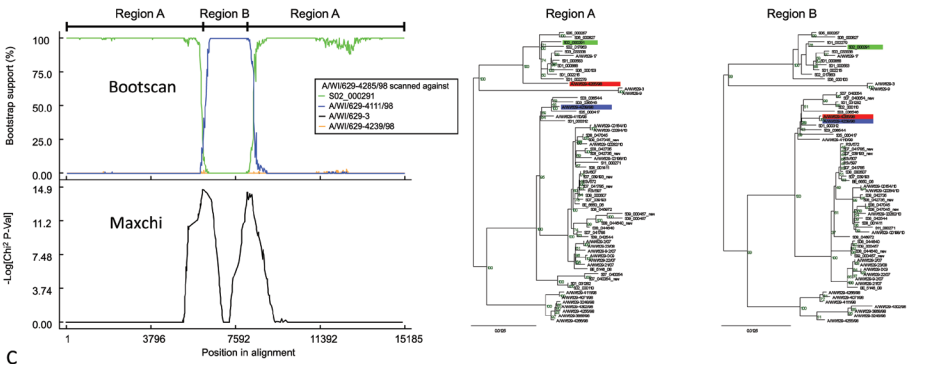
CHAPTER 2


Region A
Region B
Region A
Bootscan
Bootstrap support (%)
A/WI/629-4266/98 scanned against
A/WI/629-4111/98
A/WI/629-17
A/WI/629-3
A/WI/629-3868/98
Maxchi
-Log[Chi² P-Val]
Position in alignment
Region A
Region B

A


Region A
Region B
Region A
Bootscan
Bootstrap support (%)
A/WI/629-4302/98 scanned against
A/WI/629-3248/98
A/WI/629-17
A/WI/629-3
A/WI/629-4239/98
Maxchi
-Log[Chi² P-Val]
Position in alignment
Region A
Region B

B


Region A
Region B
Region A
Bootscan
Bootstrap support (%)
A/WI/629-4285/98 scanned against
S02_000291
A/WI/629-4111/98
A/WI/629-3
A/WI/629-4239/98
Maxchi
-Log[Chi² P-Val]
Position in alignment
Region A
Region B

C

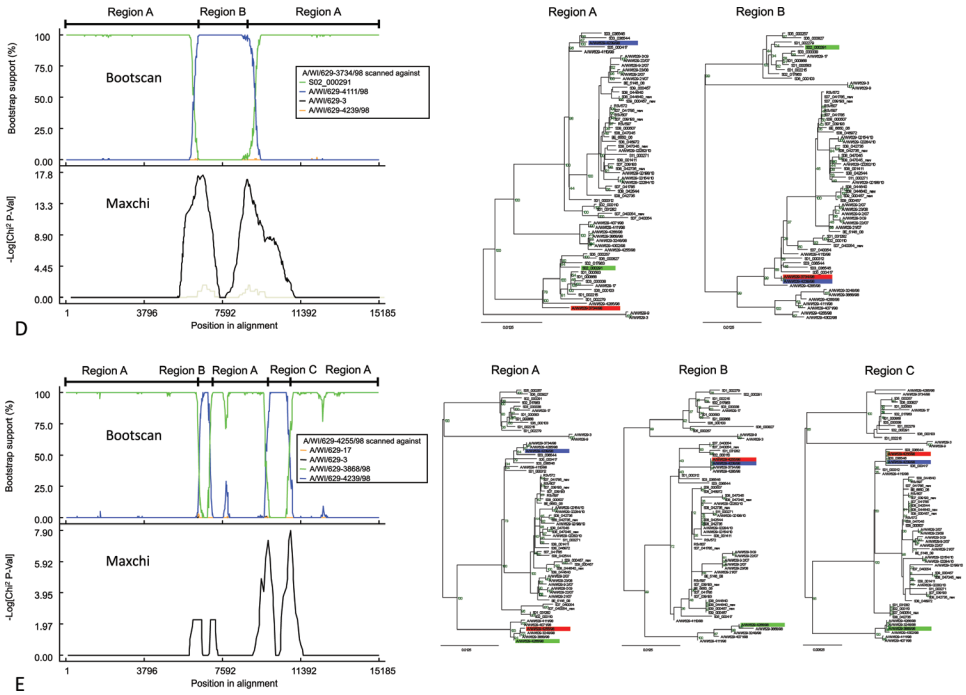


Figure S1. Evidence of recombination events within the genomes of RSV isolates. (A) A/WI/629-4266/98; (B) A/WI/629-4302/98; (C) A/WI/629-4285/98; (D) A/WI/ 629-3734/98; and (E) A/WI/629-4255/98. Whereas bootscan plots indicate regions of sequence where these isolates are differentially most closely related to major (in green) and minor (in blue) parental viruses, peaks in the Maxchi plots indicate the most probable locations of recombination breakpoints. Neighbor joining trees (Jukes Cantor model with 100 bootstrap replicates) constructed from different regions of the RSV full genome alignment (where region A corresponds with the genome region inherited from the recombinant's major parent and regions B and C correspond with the genome regions inherited from the recombinant's minor parent/parents). Within these trees, recombinant sequences are highlighted in red, sequences closely related to the recombinants' major parents are highlighted in green, and sequences closely related to the recombinants' minor parents are highlighted in blue.

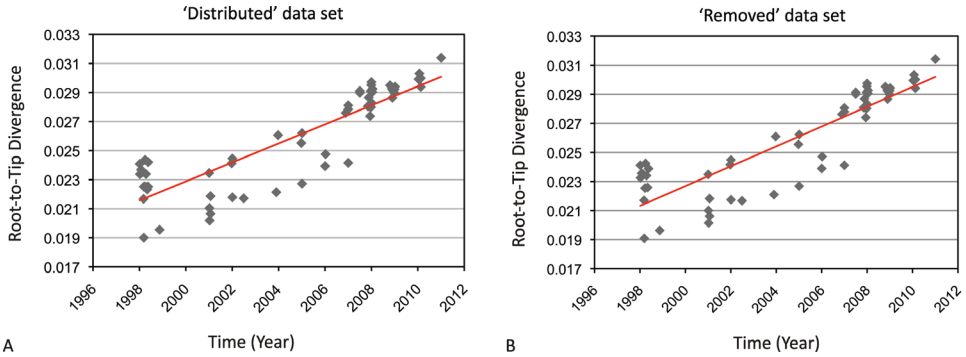


Figure S2. Root-to-tip divergence plot. A) Distributed data set, B) Removed data set

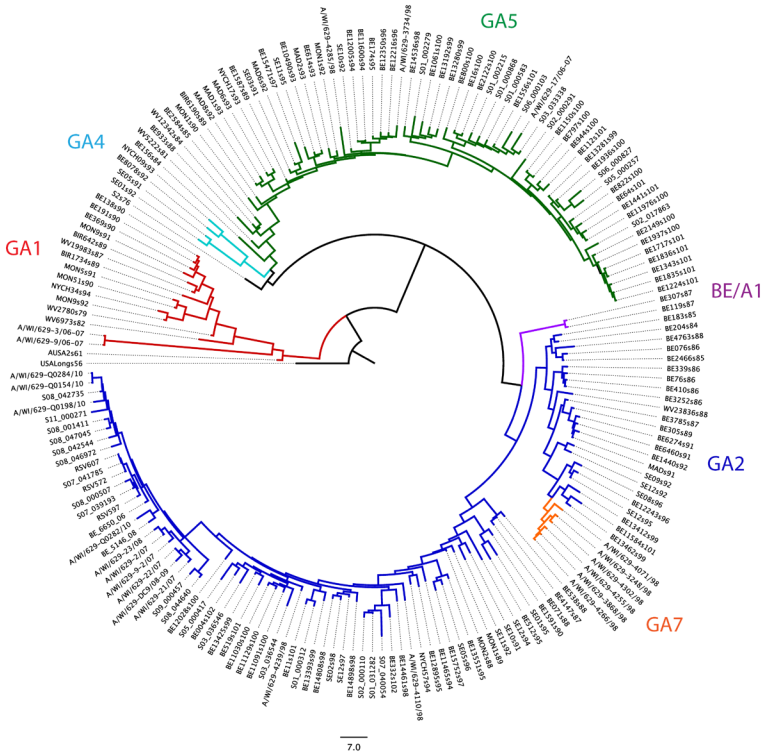


Figure S3. Phylogeny of the RSV G gene. The accession numbers, country of isolation, isolation date and references for the sequences included are listed in Table S3.

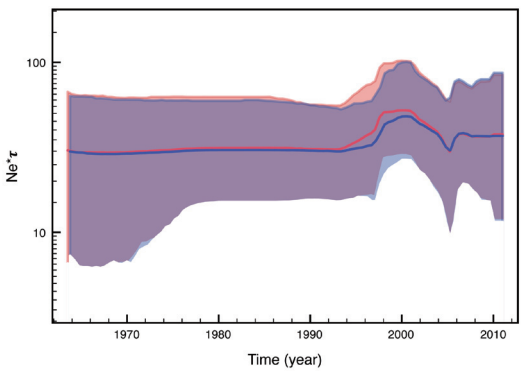


Figure S4. Bayesian Skyline plot reconstructions. The superimposed estimated change in effective population size through time for the distributed (blue) and removed (red) data sets.

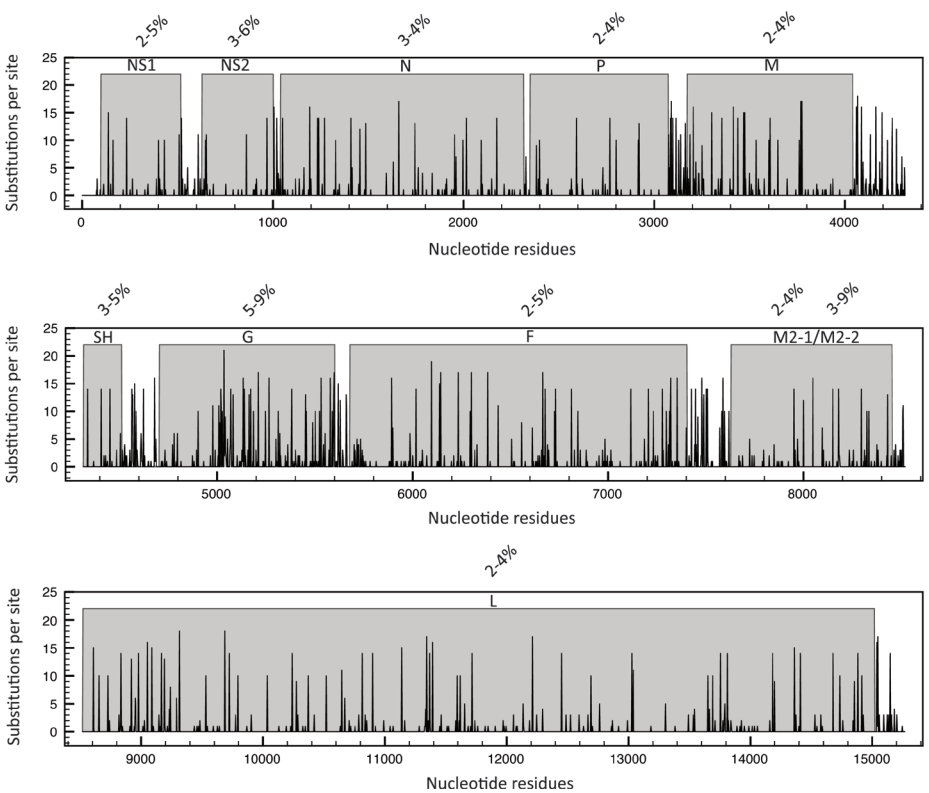
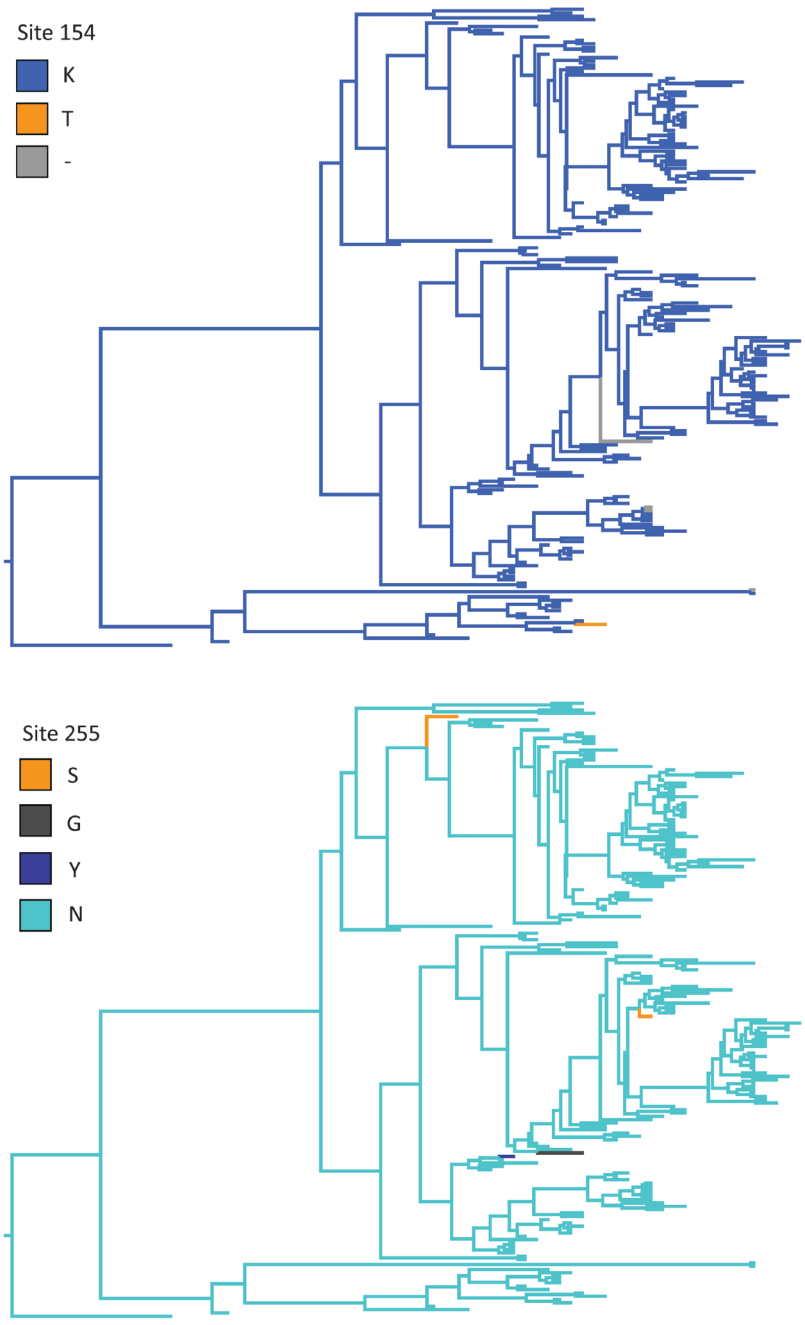


Figure S5. Mutational hotspots and nucleotide sequence variability within the RSV genome. The number of substitutions per site (black bars) and the nucleotide sequence variability (%) in each RSV gene calculated per strain relative to the consensus.


Site 154
K
T
-
Site 255
S
G
Y
N

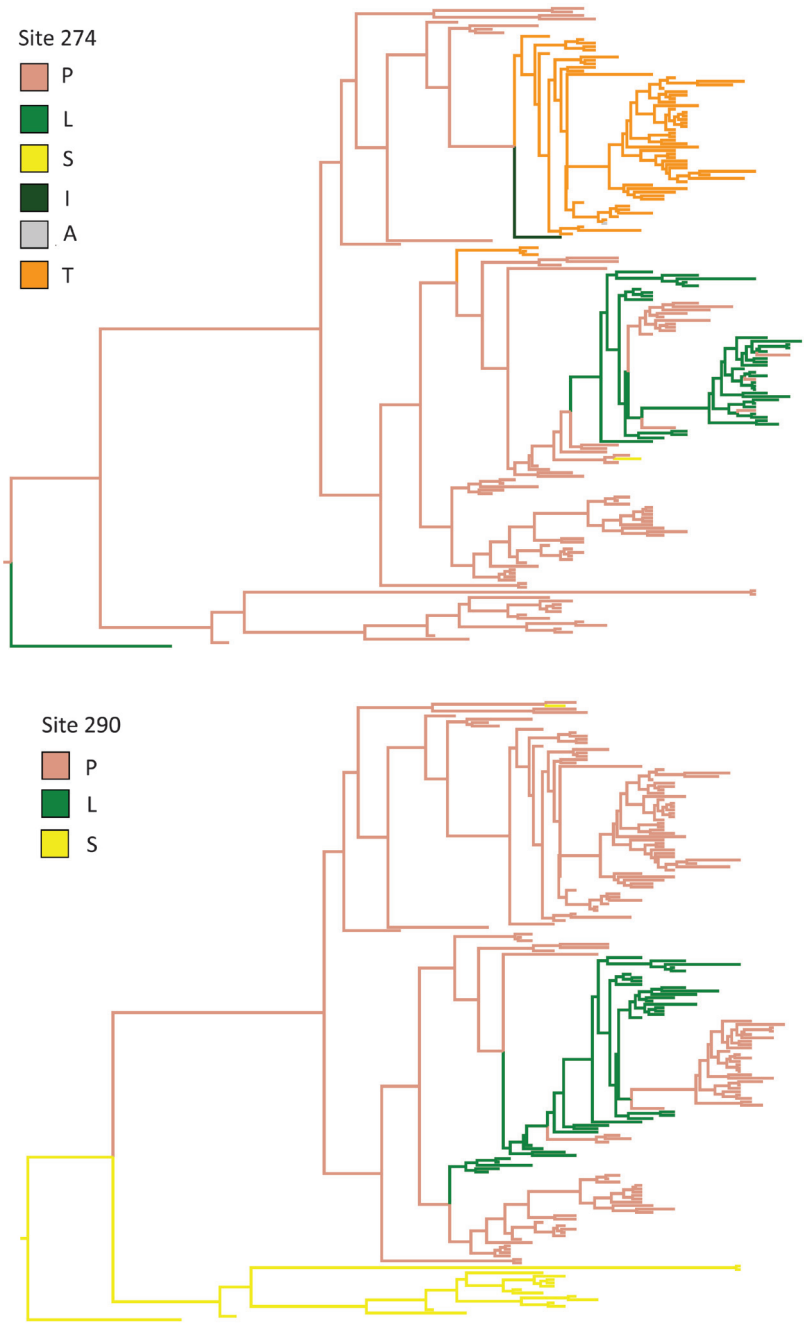


Figure S6. Pervasive and episodic selection sites in the G gene. The G gene based phylogenetic trees show the substitution history for episodic sites 154 and 255 plus the pervasive sites 274 and 290.

Table S1. Patient and sample information

Code	Gender	Age (months)*	Sample date	Material	Sequence Method	Genotype	Accession #
01-000312	M	5.7	1/3/2001	N Asp	Conventional	GA2	JQ901447
01-000583	V	0.8	1/7/2001	N Asp	Conventional	GA5	JQ901448
01-000868	M	3.5	1/10/2001	N Asp	Conventional	GA5	JQ901449
01-002215	M	1.4	1/24/2001	N Asp	Conventional	GA5	JQ901450
01-002279	M	3.9	1/24/2001	N swab	Conventional	GA5	JQ901451
01-031282	V	13.4	12/22/2001	N Asp	Conventional	GA2	JQ901452
02-000110	V	3.7	1/3/2002	N Asp	Conventional	GA2	JQ901453
02-000291	M	4.3	1/4/2002	N Asp	Conventional	GA5	JQ901454
02-017863	M	2.1	6/29/2002	N Asp	Conventional	GA5	JQ901455
03-033338	M	3.1	11/25/2003	Sputum	Conventional	GA5	JQ901456
03-036456	M	1.3	12/22/2003	N Asp	Conventional	GA2	JQ901457
03-036544	M	7.9	12/23/2003	N Asp	Conventional	GA2	JQ901458
05-000257	M	1.2	1/4/2005	N Asp	Conventional	GA5	JX015485
05-000417	M	3	1/5/2005	N Asp	Conventional	GA2	JX015486
06-000103	M	48.9	1/2/2006	N Asp	Conventional	GA5	JX015487
06-000827	V	10.4	1/9/2006	Sputum	Conventional	GA5	JX015488
07-039193	V	0.5	12/3/2007	N Asp	454, Conventional	GA2	JX015489
07-040054	V	4.8	12/10/2007	N Asp	454, Conventional	GA2	JX015480
07-041785	M	3.4	12/24/2007	N Asp	454, Conventional	GA2	JX015492
08-000507	M	1	1/7/2008	N Asp	454	GA2	JX015493
08-001411	V	4.2	1/14/2008	N Asp	454	GA2	JX015494
08-042544	V	1.6	11/21/2008	N Asp	454	GA2	JX015495
08-042735	M	0.4	11/26/2008	N Asp	454, Conventional	GA2	JX015496
08-044640	M	2.7	12/6/2008	N Asp	454, Conventional	GA2	JX015497
08-046972	M	1.6	12/27/2008	N Asp	454	GA2	JX015498
08-047045	M	2.2	12/29/2008	T Asp	454, Conventional	GA2	JX015483
RSV572	M	10.8	11/6/2007	N swab	Conventional	GA2	JX015484
RSV597	F	1	11/19/2007	N swab	Conventional	GA2	JX015490
RSV607	F	8.1	11/23/2007	N swab	Conventional	GA2	JX015491
09-000457	M	2.1	1/7/2009	N Asp	454, Conventional	GA2	JX015481
11-000271	V	1.9	1/3/2011	N Asp	Conventional	GA2	JX015479
BE-6650-06	V	36	11/22/2006	N Asp	Conventional	GA2	JX015482
BE-5146-08	V	12	10/22/2008	N Asp	Conventional	GA2	JX015499

N Asp = Nasopharyngeal Aspirates, T Asp = Tracheal Aspirates, N swab = Nose-throat swab and * Age at time of sampling.

Table S2. Primers used for cDNA preparation and whole genome sequencing

Primer	5'-3' sequence	Start	End	Origin
110984SP-1F	ACGCGAAAAAATGCGTACAAC	1	21	Modified from Kumaria *et al*
110983SP-1R	TGGATGGTGTATTTGCTGGA	1197	1216	Modified from Kumaria et al
110982SP-2F	ATTTCAACACAAKAKTCACACAA	1003	1025	Modified from Kumaria et al
110981SP-2R	TTCAGGAGCAAACTTTTCCAT	2346	2366	Modified from Kumaria et al
111056SP-F2A for	CCATAGTCCAAATGGAGCCTG	1058	1078	This study
111061SP-F2A rev	CAACTCTACTGCCACCTCTGG	1837	1857	This study
110980SP-3F	AAAAATTGGGTGGWGAAGCA	2014	2033	Modified from Kumaria et al
110979SP-3R	CCCTTGGGTGTGGATATTTG	3441	3460	Modified from Kumaria et al
111057SP-F3A for	GCAGAAGAACTAGAGGCTATC	2253	2273	This study
111062SP-F3A rev	GGTTGGATGGGTGAGTTGG	3127	3145	This study
110978SP-4F	AACCTRTTGGAAGGGAATGA	3018	3037	Modified from Kumaria et al
110977SP-4R	AGGCCAGAATTTGCTTGAGA	4331	4350	Modified from Kumaria et al
111175F4b-for	GAAGCTATGGCAAGGCTCAG	2922	2941	This study
111176F4b-rev	CCTTGATTTCGCAGGGTGTG	3581	3600	This study
111058SP-F5A for	CAGATCATCCCAAGTCATTG	4153	4172	This study
111063SP-F5A rev	GCTGCATATGCTGCAGGGTAC	5198	5218	This study
110974SP-6F	AAGTCAACCCTGCAATCCAC	5054	5073	Modified from Kumaria et al
111167SP-F6a for	CACCATACTAGCTTCAACAACACC	5023	5046	This study
110973SP-6R	GCATTAACACTAAATTCCCTGGT	6360	6382	Modified from Kumaria et al
111059SP-F7A for	GTGAACAAGCAAAGCTGC	6279	6296	This study
111064SP-F7A rev	GTGTGACTGGTGTGCTTCTGG	7315	7335	This study
110969SP-8F	CCCATTAGTRTTCCCCTCTG	7097	7116	Modified from Kumaria et al
110968SP-8R	TCCATTAATAATGGGATCCATT	8497	8518	Modified from Kumaria et al
111060SP-F9A for	GATTGCCAGCAGACGTATTGAAG	8057	8079	This study
111065SP-F9A rev	GGCTAATATCTTTCCATGTC	9307	9326	This study
110965SP-10F	CAATGCAACATCCTCCATCA	9084	9103	Modified from Kumaria et al
110964SP-10R	GGTTGCATTGCAAACATTCTA	10375	10395	Modified from Kumaria et al
110963SP-11F	CGTGAGTTTCGGTTGCCTA	10064	10082	Modified from Kumaria et al
110962SP-11R	GGGATCACCACCACCAAATA	11448	11467	Modified from Kumaria et al
110961SP-12F	AGTGGGACCGTGGATAAACA	11164	11183	Modified from Kumaria et al
110960SP-12R	TGACTGTAAGGCGATGCAA	12506	12524	Modified from Kumaria et al
110959SP-13F	TGGACATCAAATATACWACAAGCA	12180	12203	Modified from Kumaria et al
110958SP-13R	TTAACAACCCAAGGGCAAAC	13361	13380	Modified from Kumaria et al
110957SP-14F	AAAAAGATTGGGGAGAGGGATA	13041	13062	Modified from Kumaria et al
110956SP-14R	TGCAYTTTCTTACATGCTTGC	14355	14375	Modified from Kumaria et al

Table S2. Primers used for cDNA preparation and whole genome sequencing (continued)

Primer	5'-3' sequence	Start	End	Origin
110955SP-15F	GGTGAAGGAGCAGGGAATTT	14054	14073	Modified from Kumaria et al
110954SP-15R	ACGAGAAAAAAAAGTGTCAAAAACT	15199	15222	Modified from Kumaria et al
111067F1-seqbeginrev	CTGGCATTGTTGTGAAATTGG	322	342	This study
111066F15-seqendfor	CCCATAGCTACACACTAAC	15022	15040	This study
110951Seq-F:	GCAGTCAACCATATGCCTTG			This study
110952Seq-R	GCTTGCGGAATTCTGACAAC			This study

Table S3. Genbank human RSV-A sequences used in this study

Strain	Accesion No	Country of isolation	Year of isolation	Reference
USALongs56	M17212	U.S.A.	1956	Johnson et al (1987) PNAS
AUSA2s61	M11486	Australia	1961	Wertz et al (1985) PNAS
S2s76	U39662	U.K.	1976	Tolley et al (1996) Vaccine
'A/WI/629-3248/98'	JF920062	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-3734/98'	JF920059	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-3868/98'	JF920063	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4071/98'	JF920065	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4110/98'	JF920056	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4239/98'	JF920057	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4255/98'	JF920068	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4266/98'	JF920064	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4285/98'	JF920061	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-4302/98'	JF920067	U.S.A.	1998	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-3/06-07'	JF920069	U.S.A.	2006-7	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-9/06-07'	JF920070	U.S.A.	2006-7	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-17/06-07'	JF920058	U.S.A.	2006-7	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-2/07'	JF920046	U.S.A.	2007	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-9-2/07'	JF920048	U.S.A.	2007	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-21/07'	JF920051	U.S.A.	2007	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-22/07'	JF920049	U.S.A.	2007	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-23/08'	JF920047	U.S.A.	2008	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-DC9/08-09'	JF920050	U.S.A.	2008-9	Rebuffo-Scheer et al (2011) PLoSONE

Table S3. Genbank human RSV-A sequences used in this study (continued)

Strain	Accesion No	Country of isolation	Year of isolation	Reference
'A/WI/629-Q0154/10'	JF920052	U.S.A.	2010	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-Q0198/10'	JF920055	U.S.A.	2010	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-Q0282/10'	JF920054	U.S.A.	2010	Rebuffo-Scheer et al (2011) PLoSONE
'A/WI/629-Q0284/10'	JF920053	U.S.A.	2010	Rebuffo-Scheer et al (2011) PLoSONE
BE156s84	AY343597	Belgium	1984	Zlateva et al (2004) J Virol
BE204s84	AY343625	Belgium	1984	Zlateva et al (2004) J Virol
BE183s85	AY343626	Belgium	1985	Zlateva et al (2004) J Virol
BE2466s85	AY343631	Belgium	1985	Zlateva et al (2004) J Virol
BE2584s85	AY343596	Belgium	1985	Zlateva et al (2004) J Virol
BE76s86	AY343632	Belgium	1986	Zlateva et al (2004) J Virol
BE339s86	AY343633	Belgium	1986	Zlateva et al (2004) J Virol
BE410s86	AY343640	Belgium	1986	Zlateva et al (2004) J Virol
BE3252s86	AY343641	Belgium	1986	Zlateva et al (2004) J Virol
BE119s87	AY343656	Belgium	1987	Zlateva et al (2004) J Virol
BE307s87	AY343655	Belgium	1987	Zlateva et al (2004) J Virol
BE3785s87	AY343654	Belgium	1987	Zlateva et al (2004) J Virol
BE4147s87	AY343622	Belgium	1987	Zlateva et al (2004) J Virol
BE071s88	AY343623	Belgium	1988	Zlateva et al (2004) J Virol
BE538s88	AY343624	Belgium	1988	Zlateva et al (2004) J Virol
BE933s88	AY343595	Belgium	1988	Zlateva et al (2004) J Virol
BE4763s88	AY343628	Belgium	1988	Zlateva et al (2004) J Virol
BE305s89	AY343653	Belgium	1989	Zlateva et al (2004) J Virol
BE1587s89	AY343587	Belgium	1989	Zlateva et al (2004) J Virol
BE1591s90	AY343620	Belgium	1990	Zlateva et al (2004) J Virol
BE191s90	AY343659	Belgium	1990	Zlateva et al (2004) J Virol
BE138s90	AY343658	Belgium	1990	Zlateva et al (2004) J Virol
BE369s90	AY343660	Belgium	1990	Zlateva et al (2004) J Virol
BE6274s91	AY343649	Belgium	1991	Zlateva et al (2004) J Virol
BE6460s91	AY343652	Belgium	1991	Zlateva et al (2004) J Virol
BE1440s92	AY343651	Belgium	1992	Zlateva et al (2004) J Virol
BE8078s92	AY343657	Belgium	1992	Zlateva et al (2004) J Virol
BE614s93	AY343586	Belgium	1993	Zlateva et al (2004) J Virol
BE10490s93	AY343584	Belgium	1993	Zlateva et al (2004) J Virol
BE11465s94	AY343615	Belgium	1994	Zlateva et al (2004) J Virol

Table S3. Genbank human RSV-A sequences used in this study (continued)

Strain	Accesion No	Country of isolation	Year of isolation	Reference
BE11600s94	AY343588	Belgium	1994	Zlateva et al (2004) J Virol
BE12005s94	AY343590	Belgium	1994	Zlateva et al (2004) J Virol
BE174s95	AY343591	Belgium	1995	Zlateva et al (2004) J Virol
BE521s95	AY343621	Belgium	1995	Zlateva et al (2004) J Virol
BE12895s95	AY343619	Belgium	1995	Zlateva et al (2004) J Virol
BE13551s95	AY343618	Belgium	1995	Zlateva et al (2004) J Virol
BE12216s96	AY343592	Belgium	1996	Zlateva et al (2004) J Virol
BE12243s96	AY343647	Belgium	1996	Zlateva et al (2004) J Virol
BE12350s96	AY343593	Belgium	1996	Zlateva et al (2004) J Virol
BE15471s97	AY343594	Belgium	1997	Zlateva et al (2004) J Virol
BE15752s97	AY343617	Belgium	1997	Zlateva et al (2004) J Virol
BE14461s98	AY343614	Belgium	1998	Zlateva et al (2004) J Virol
BE14536s98	AY343583	Belgium	1998	Zlateva et al (2004) J Virol
BE14808s98	AY343609	Belgium	1998	Zlateva et al (2004) J Virol
BE14898s98	AY343610	Belgium	1998	Zlateva et al (2004) J Virol
BE13192s99	AY343556	Belgium	1999	Zlateva et al (2004) J Virol
BE13280s99	AY343554	Belgium	1999	Zlateva et al (2004) J Virol
BE13281s99	AY343568	Belgium	1999	Zlateva et al (2004) J Virol
BE13393s99	AY343612	Belgium	1999	Zlateva et al (2004) J Virol
BE13412s99	AY343646	Belgium	1999	Zlateva et al (2004) J Virol
BE13425s99	AY343603	Belgium	1999	Zlateva et al (2004) J Virol
BE13462s99	AY343643	Belgium	1999	Zlateva et al (2004) J Virol
BE16s100	AY343558	Belgium	2000	Zlateva et al (2004) J Virol
BE797s100	AY343581	Belgium	2000	Zlateva et al (2004) J Virol
BE800s100	AY343561	Belgium	2000	Zlateva et al (2004) J Virol
BE822s100	AY343567	Belgium	2000	Zlateva et al (2004) J Virol
BE944s100	AY343582	Belgium	2000	Zlateva et al (2004) J Virol
BE1061s100	AY343557	Belgium	2000	Zlateva et al (2004) J Virol
BE1150s100	AY343579	Belgium	2000	Zlateva et al (2004) J Virol
BE1936s100	AY343569	Belgium	2000	Zlateva et al (2004) J Virol
BE1937s100	AY343572	Belgium	2000	Zlateva et al (2004) J Virol
BE2122s100	AY343559	Belgium	2000	Zlateva et al (2004) J Virol
BE2149s100	AY343573	Belgium	2000	Zlateva et al (2004) J Virol
BE11030s100	AY343599	Belgium	2000	Zlateva et al (2004) J Virol

Table S3. Genbank human RSV-A sequences used in this study (continued)

Strain	Accesion No	Country of isolation	Year of isolation	Reference
BE11091s100	AY343601	Belgium	2000	Zlateva et al (2004) J Virol
BE11129s100	AY343600	Belgium	2000	Zlateva et al (2004) J Virol
BE11976s100	AY343565	Belgium	2000	Zlateva et al (2004) J Virol
BE12028s100	AY343608	Belgium	2000	Zlateva et al (2004) J Virol
BE11s101	AY343611	Belgium	2001	Zlateva et al (2004) J Virol
BE64s101	AY343566	Belgium	2001	Zlateva et al (2004) J Virol
BE112s101	AY343578	Belgium	2001	Zlateva et al (2004) J Virol
BE519s101	AY343605	Belgium	2001	Zlateva et al (2004) J Virol
BE1441s101	AY343562	Belgium	2001	Zlateva et al (2004) J Virol
BE1224s101	AY343571	Belgium	2001	Zlateva et al (2004) J Virol
BE1343s101	AY343574	Belgium	2001	Zlateva et al (2004) J Virol
BE1556s101	AY343560	Belgium	2001	Zlateva et al (2004) J Virol
BE1717s101	AY343577	Belgium	2001	Zlateva et al (2004) J Virol
BE1835s101	AY343570	Belgium	2001	Zlateva et al (2004) J Virol
BE1836s101	AY343576	Belgium	2001	Zlateva et al (2004) J Virol
BE11584s101	AY343645	Belgium	2001	Zlateva et al (2004) J Virol
BE004s102	AY343606	Belgium	2002	Zlateva et al (2004) J Virol
BE332s102	AY343613	Belgium	2002	Zlateva et al (2004) J Virol
BIR642s89	X73354	U.K.	1989	Cane et al (1994) J Clin Micro
BIR6190s89	X73350	U.K.	1989	Cane et al (1994) J Clin Micro
BIR1734s89	X73352	U.K.	1989	Cane et al (1994) J Clin Micro
MADs91	Z33420	Spain	1991	Garcia et al (1994) J Virol
MAD6s92	Z33418	Spain	1992	Garcia et al (1994) J Virol
MAD8s92	Z33419	Spain	1992	Garcia et al (1994) J Virol
MAD1s93	Z33414	Spain	1993	Garcia et al (1994) J Virol
MAD2s93	Z33493	Spain	1993	Garcia et al (1994) J Virol
MAD6s93	Z33410	Spain	1993	Garcia et al (1994) J Virol
MON2s88	Z33424	Uruguay	1988	Garcia et al (1994) J Virol
MON1s89	Z33422	Uruguay	1989	Garcia et al (1994) J Virol
MON1s90	Z33494	Uruguay	1990	Garcia et al (1994) J Virol
MON5s90	Z33427	Uruguay	1990	Garcia et al (1994) J Virol
MON5s91	Z33428	Uruguay	1991	Garcia et al (1994) J Virol
MON9s91	Z33431	Uruguay	1991	Garcia et al (1994) J Virol
MON1s92	Z33423	Uruguay	1992	Garcia et al (1994) J Virol

Table S3. Genbank human RSV-A sequences used in this study (continued)

Strain	Accesion No	Country of isolation	Year of isolation	Reference
MON9s92	Z33432	Uruguay	1992	Garcia et al (1994) J Virol
NYCH09s93	AF065254	U.S.A.	1993	Peret et al (1996) J Gen Virol
NYCH17s93	AF065255	U.S.A.	1993	Peret et al (1996) J Gen Virol
NYCH34s94	AF065257	U.S.A.	1994	Peret et al (1996) J Gen Virol
NYCH57s94	AF065258	U.S.A.	1994	Peret et al (1996) J Gen Virol
SE03s91	AF193304	Korea	1991	Choi et al (2000) J Infect Dis
SE05s91	AF193306	Korea	1991	Choi et al (2000) J Infect Dis
SE10s91	AF193307	Korea	1992	Choi et al (2000) J Infect Dis
SE01s92	AF193308	Korea	1992	Choi et al (2000) J Infect Dis
SE09s92	AF193309	Korea	1992	Choi et al (2000) J Infect Dis
SE10s92	AF193310	Korea	1992	Choi et al (2000) J Infect Dis
SE11s92	AF193311	Korea	1992	Choi et al (2000) J Infect Dis
SE12s92	AF193312	Korea	1992	Choi et al (2000) J Infect Dis
SE12s94	AF193316	Korea	1994	Choi *et al* (2000) J Infect Dis
SE01s95	AF193317	Korea	1995	Choi *et al* (2000) J Infect Dis
SE11s95	AF193318	Korea	1995	Choi *et al* (2000) J Infect Dis
SE12s95	AF193319	Korea	1995	Choi *et al* (2000) J Infect Dis
SE05s96	AF193321	Korea	1996	Choi *et al* (2000) J Infect Dis
SE08s96	AF193322	Korea	1996	Choi *et al* (2000) J Infect Dis
SE12s97	AF193323	Korea	1997	Choi *et al* (2000) J Infect Dis
SE02s98	AF193325	Korea	1998	Choi *et al* (2000) J Infect Dis
WV2780s79	AF065405	U.S.A.	1979	Sullender *et al* (1991) J Clin Micro
WV5222s81	AF065406	U.S.A.	1981	Sullender *et al* (1991) J Clin Micro
WV6973s82	AF065407	U.S.A.	1982	Sullender *et al* (1991) J Clin Micro
WV12342s84	AF065409	U.S.A.	1984	Sullender *et al* (1991) J Clin Micro
WV19983s87	AF065408	U.S.A.	1987	Sullender *et al* (1991) J Clin Micro
WV23836s88	AF065410	U.S.A.	1988	Sullender *et al* (1991) J Clin Micro

Table S4. Overview protein sequence variability per RSV strain.

	NS1	NS2	N	P	M	SH	G	F	M2-1	M2-2	L
Consensus	-	-	-	-	-	-	-	-	-	-	-
01-000312	6	8	7	5	6	10	14	6	5	9	5
01-000583	6	8	7	5	6	8	18	7	5	10	5

Table S4. Overview protein sequence variability per RSV strain (continued)

	NS1	NS2	N	P	M	SH	G	F	M2-1	M2-2	L
01-000868	6	8	8	5	6	8	17	7	5	10	5
01-002215	6	8	7	5	6	8	17	7	5	12	5
01-002279	6	8	7	5	6	8	16	7	5	13	5
01-031282	6	7	7	5	6	8	13	7	5	10	5
02-000110	6	7	7	5	6	8	13	7	5	10	5
02-000291	6	8	8	5	6	8	16	7	6	10	5
02-017863	6	8	7	5	6	8	18	7	5	12	5
03-033338	6	8	8	5	6	8	16	7	6	10	5
03-036456	6	8	7	5	6	8	14	6	5	12	5
03-036544	6	7	7	5	6	8	13	7	6	9	5
05-000257	6	8	7	5	6	8	16	7	6	12	5
05-000417	6	7	7	5	6	8	13	7	6	10	5
06-000103	6	8	7	5	6	8	17	7	5	10	5
06-000827	6	8	7	5	6	8	17	7	6	12	5
07-039193	6	7	7	5	6	8	11	6	5	9	5
07-040054	6	7	7	5	6	8	14	7	5	14	5
07-041785	6	7	7	5	6	8	11	6	5	9	5
08-000507	6	7	7	6	6	8	11	6	5	9	5
08-001411	6	7	7	5	6	10	10	6	5	10	5
08-042544	6	7	8	5	6	8	10	6	5	9	5
08-042735	6	7	7	5	6	8	11	7	5	9	5
08-044640	6	7	7	5	6	8	11	6	5	12	5
08-046972	6	8	7	5	6	8	11	6	5	9	5
08-047045	6	8	7	5	6	8	10	7	5	9	5
09-000457	6	7	7	5	6	8	12	6	5	10	5
11-000271	6	9	7	5	6	10	11	7	5	10	5
AY911262.1_Long	6	8	7	6	6	10	18	7	6	20	6
BE-5146-08	6	7	7	5	6	8	11	6	5	9	5
BE-6650-06	6	7	7	5	6	8	10	6	5	10	5
FJ614813.1_Line_19	6	8	7	6	6	10	18	8	6	20	6
M74568.1_A2	6	9	7	5	6	8	18	9	5	20	6
NC_001803.1_RSS-2	6	8	7	6	6	8	13	7	5	9	5
RSV572	6	7	7	5	6	8	11	6	5	9	5
RSV597	6	8	7	5	6	8	12	6	5	9	5
RSV607	6	7	7	5	6	8	11	6	5	9	5

Table S5. N-glycosylation probability report of Asn-Xaa-Ser/Thr sequons in the RSV G protein

Position of Asp	85	103	135	179	237	242	244	250	251	258	273	294
Consensus	---	X	+	-	X	X	X	X	+	X	+	-
01-000312	---	+	+	-	+	X	X	X	X	X	---	+
01-000583	---	++	+	-	+	X	X	+	X	X	---	+
01-000868	---	++	+	-	+	X	X	+	X	X	---	+
01-002215	---	++	+	-	+	X	X	+	X	X	---	+
01-002279	---	++	+	-	+	X	X	+	X	X	---	+
01-031282	---	+	+	-	+	X	X	X	+	X	---	+
02-000110	---	+	+	-	+	X	X	X	+	X	---	+
02-000291	---	++	+	-	+	X	X	+	X	X	---	+
02-017863	---	++	+	-	+	X	X	+	X	X	---	+
03-033338	---	++	+	-	+	X	X	+	X	X	---	+
03-036456	---	+	-	-	+	X	X	X	+	X	---	+
03-036544	---	+	-	-	+	X	X	X	+	X	---	+
05-000257	---	++	+	-	+	X	X	+	X	X	---	+
05-000417	---	+	-	-	+	X	-	X	+	X	---	+
06-000103	---	++	+	-	+	X	X	+	X	X	---	+
06-000827	---	++	+	-	+	X	X	+	X	X	---	+
07-039193	---	+ ProX1	+	-	X	X	X	X	+	X	---	+
07-040054	---	+	---	-	+	X	X	X	++	-	---	+
07-041785	---	+ ProX1	+	-	X	X	X	X	+	X	---	+
08-000507	---	+ ProX1	+	-	X	X	X	X	+	X	---	+
08-001411	---	+	+	-	X	X	X	X	+	X	---	+
08-042544	---	+	+	-	X	X	X	X	+	X	---	+
08-042735	---	+	+	-	X	X	X	X	+	X	---	+
08-044640	---	++	+	-	X	X	X	X	-	X	---	+
08-046972	---	+	+	-	X	X	X	X	X	X	---	+
08-047045	---	+	+	-	X	X	X	X	+	X	---	+
09-000457	---	++	+	-	X	X	X	X	-	X	---	+
11-000271	---	+	+	-	X	X	X	X	+	X	---	X
AY911262.1_Long	---	+	+	-	+	+	X	-	+	X	---	+
BE-5146-08	---	+	+	-	X	X	X	X	X	X	---	+
BE-6650-06	---	+ ProX1	+	-	X	X	X	X	+	X	---	+

Table S5. N-glycosylation probability report of Asn-Xaa-Ser/Thr sequons in the RSV G protein (continued)

Position of Asp	85	103	135	179	237	242	244	250	251	258	273	294
FJ614813.1_Line_19	---	+	+	-	+	+	X	-	+	X	---	+
M74568.1_A2	---	+	-	-	+	X	X	X	+	X	---	X
NC_001803.1_RSS-2	---	+	-	-	+	X	X	-	+	X	---	+
RSV572	---	+ ProX1	+	-	X	X	X	X	+	X	---	+
RSV597	---	+	+	-	X	X	X	X	+	X	---	-
RSV607	---	+ ProX1	+	-	X	X	X	X	+	X	---	+

X shows that there is no Asn-Xaa-Ser/Thr sequon at the indicated position for that specific strain, Xaa = any amino acid except Proline, +ProX1 = Proline at the second position of the sequon precludes N-linked glycosylation, N-glycosylation and Non-glycosylation potential is marked by + and -, respectively.

Table S6. Clock comparison for marginal likelihood estimates

	Model	Strict	Relaxed
Distributed	HME	**-42070.931**	-42074.427
	AICm	**84186.937**	84214.735
	PS	-42371.638	**84043.193**
	SS	-42372.117	**-42363.759**
Removed	HME	-41994.966	**-41983.643**
	AICm	**84021.414**	84043.193
	PS	-42268.322	**-42263.492**
	SS	-42268.735	**-42263.951**

HME = Harmonic Mean Estimator, AICM = Akaike Information Content through MCMC, PS = Path Sampling and SS = Stepping-stone Sampling

CHAPTER THREE

The Comparative Genomics of Human Respiratory Syncytial Virus Subgroup A and B: Genetic Variability and Molecular Evolutionary Dynamics

Lydia Tan[1], Frank E.J. Coenjaerts[1], Lieselot Houspie[2], Marco C. Viveen[1], Grada M. van Bleek[1], Emmanuel J.H.J. Wiertz[1#], Darren P. Martin[3], Philippe Lemey[2]

1 Department of Medical Microbiology, University Medical Center Utrecht, The Netherlands

2 Department of Microbiology and Immunology, Rega Institute, KU Leuven, Belgium

3 Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa

Journal of Virology; 2013, 87(14):8213-26

Abstract

Genomic variation and related evolutionary dynamics of human respiratory syncytial virus (RSV), a common causative agent of severe lower respiratory tract infections, may affect its transmission behavior. RSV evolutionary patterns are likely to be influenced by a precarious interplay between selection favoring variants with higher replicative fitness and variants that evade host immune responses. Studying RSV genetic variation can reveal both genes and individual codons within these genes that are most crucial for RSV survival. This study conducts genetic diversity and evolutionary rate analyses on 36 RSV subgroup B whole genome sequences. The attachment protein, G, was the most variable protein and, accordingly, the G gene had a higher substitution rate than other RSV-B genes. Overall, less genetic variability was found amongst the available RSV-B genome sequences than that amongst a comparable sample of RSV-A genome sequences. The mean substitution rates of both subgroups were, however, similar (A: 6.47×10^{-4} [Credible interval (CI): 5.56, 7.38×10^{-4}], B: 7.76×10^{-4} [CI: 6.89, 8.58×10^{-4}] substitution/site/yr), with the time to their most recent common ancestors (TMRCAs) being much lower for RSV-B (19 yrs) than for RSV-A (46.8 yrs). The more recent RSV-B TMRCA is apparently the result of a genetic bottleneck that, over longer time-scales, is still compatible with neutral population dynamics. Whereas the immunogenic G protein seems to require high substitution rates to ensure immune evasion, strong purifying selection in conserved proteins such as the fusion protein and nucleocapsid protein is likely essential to preserve RSV viability.

Introduction

Human respiratory syncytial virus (RSV) is the leading cause of severe respiratory tract infections in infants under the age of two [1]. Premature neonates and children with chronic pulmonary or congenital heart diseases are especially at risk, but also immuno-compromised adults and the elderly have been reported as high-risk groups [2-4]. Individuals infected by RSV mostly suffer from rhinitis and common cold-like symptoms, but can also develop serious pneumonia or bronchiolitis thereby requiring hospitalization [5]. According to the Centers for Disease Control (CDC), RSV infections annually account for the hospitalization of up to 126,000 children and 62,000 elderly people in the USA alone (http://www.cdc.gov/rsv/about/faq.html). Seasonal epidemic outbreaks of RSV disease in Europe and North America occur during the winter and early spring months with a peak incidence in December [6]. In tropical countries, there is a correlation between increased RSV infection rates and rainy seasons [7]. Relative humidity is an important factor for RSV activity, but the exact influence on the course of infection or on virus transmission dynamics is unknown [8].

RSV is an enveloped, non-segmented negative-sense, single-stranded RNA virus of approximately 15,000 nucleotides, that is classified in the genus *Pneumovirus* belonging to the family of *Paramyxoviridae*. The viral genome, which is surrounded by a nucleocapsid protein complex, encodes eleven proteins that have roles in three different stages of the RSV life cycle. The nucleocapsid protein (N), phosphoprotein (P), large polymerase (L), and the transcription regulatory proteins M2-1 and M2-2 are essential mediators of the RSV transcription and replication processes [9]. Virus entry and assembly is mediated by the structural matrix protein (M), attachment- (G) and fusion glycoprotein (F) [10,11].

Virus attachment occurs via binding of the immunogenic RSV G to the glycosaminoglycans (GAGs) of the host target cell [12,13]. RSV entry into host cells is then mediated via the fusion protein, which requires enzymatic cleavage by furin-like proteases for RSV F activation [14]. RSV F alone can initiate infection as proven with RSV G deletion mutants [15]. Recently, the structural small hydrophobic protein (SH) has been identified as a viroporin, which permeabilizes the host membrane, suggesting facilitation of viral entry [16,17]. This accessory structural protein is not

pivotal for RSV infection and mainly accumulates in the lipid raft structures of the Golgi-complex [18,19]. Finally, the RSV non-structural proteins 1 and 2 (NS1 and -2) are host interference proteins that impair the anti-viral state by impeding interferon activity [20]. In addition, the soluble truncated form of RSV G has been described as an antigen decoy in antibody-mediated neutralization, thereby impairing the antibody mediated antiviral effects of Fc receptor-bearing leukocytes [21].

For RSV, two antigenic subgroups (A and B) have been identified via antibody-cross-reactivity patterns and these were further classified into genotypes according to genetic divergence within the highly variable G gene [22-24]. Both subgroups can co-circulate and RSV re-infections occur frequently throughout life, which indicates that there is only partial cross-immunity against different strains [25]. Initial infection with RSV-A is frequently followed by infection with RSV-B, but this order is not consistently observed [26]. It has been suggested that particularly the antigenic variability of the G protein, both within and between the antigenic subgroups, facilitates evasion of the pre-existing host immune responses [27,28]. Multiple genotypes can be present in one population, although new genotypes may replace older predominating genotypes in time over successive epidemic seasons [24,29]. The heterogeneity in genotypic distribution patterns observed among different populations is most likely influenced by variation in the herd immunity and also by community-specific cultural and behavioral patterns [30,31].

The pathogenesis and disease outcome of RSV infection is most likely determined by the course of the induced immune response [32] and the capacity of RSV to modify this response for its own benefits (e.g through Toll-like receptor-signaling interference [33]). This has been hypothesized on the basis of several vaccine studies, where the critical immune balance was negatively affected in numerous ways. Early vaccination attempts in children in the 1960s with a formalin-inactivated RSV vaccine resulted in enhanced disease severity upon subsequent virus exposure [34-37]. It is thought that inadequate serum neutralizing antibody levels, low $CD8^+$ memory T cell, and enhanced $CD4^+$ memory T cell response were the consequence of this intensified disease profile [35,38-41]. Later vaccination attempts with live, attenuated RSV vaccines did not show similar respiratory disease exacerbation upon natural infection, but nevertheless failed to provide appreciable protection from infection. From clinical-immunological data and the epidemiological profile of RSV, it is clear that this

virus, which lacks clear-cut virulence factors, causes disease by modifying the host immune response in a way that affects pathogenesis and that it has typical seasonal transmission behavior. However, an association between RSV genetic diversity, RSV induced immune responses and the formation of specific epidemiological patterns remains poorly understood.

Understanding fluctuations in the genetic diversity of viral populations that are driven by ecological and evolutionary processes could elucidate the roles of specific genes and proteins in RSV survival. In addition, knowledge of evolutionary conserved domains or substitution hotspots could highlight potentially useful therapeutic targets. Phylodynamic analyses are increasingly used to infer viral population dynamics from phylogenetic patterns [42]. These analyses reveal both the impact of short- and long-term genomic evolutionary changes that enable viruses to escape from host immune responses, and how these changes affect viral epidemic behavior. Although the molecular epidemiology and evolutionary dynamics of RSV have been extensively studied, these analyses have primarily focused on single genes – in particular the G gene [43,44]. Little is known about either the diversity or evolution rates of other genes or the RSV genome as a whole. Importantly, as has recently been demonstrated in many other virus groups, full genome analyses can provide valuable insights into the processes that shape viral epidemiology and evolution [45]. Complete genome sequences from RSV clinical isolates have, however, only recently been generated in sufficiently large numbers to facilitate such analyses [46-48] and this work is further extended with this study.

Here, we conducted phylodynamic analyses on a collection of RSV-B strains and present the first calculations made within the phylodynamic framework of RSV-B genomic diversity through time. In addition to detailed genome-wide genetic variability analyses, we inferred genome-wide evolution rates and estimate the time since the most recent common ancestor (TMRCA) of all the currently available RSV-B genome sequences. Furthermore, we compare these estimates with those previously made for RSV-A [48]. We show that whereas RSV-B genomes evolved with rates similar to RSV-A genomes, there is substantially less protein variability amongst currently sampled RSV-B strains suggesting that these circulating RSV-B strains have a more recent ancestry due either to a chance fixation of a particular variant, or to fixation of a variant harboring advantageous mutations. The observed differences

in G, SH and M2-2 gene variability between RSV-A and -B have the highest impact on this dissimilarity. As has been previously shown for RSV-A, the RSV G gene is evolving considerably faster than the remainder of the RSV genome – a finding that is consistent with evidence of both persistent and episodic diversifying selection in this gene. Whereas positive selection in the G gene likely reflects the impact of host immune responses, more pervasive evidence of purifying selection within the more conserved genes likely reflects selection favoring the optimization of viral replication and transmission.

Materials and Methods

Clinical virus isolates and ethics statement

From 2002 to 2012, nasopharyngeal aspirates and nose-throat swabs were collected from thirty-four patients who were hospitalized in the Wilhelmina Children's Hospital of Utrecht, the Netherlands, or in the Gasthuisberg University Medical Hospital of Leuven, Belgium (Table S1). At the time of sampling, twenty-eight patients were under the age of two and six were adults ranging from twenty-seven to seventy-nine years old. As part of the routine diagnostic process, collected samples were cultured and checked for the formation of syncytia in HEp2 cell cultures, a typical RSV phenotype. The anonymized clinical strains were used in this study according to the guidelines of the institutions' ethical committees (METC for Dutch samples; CME for Belgian samples) and the study was performed in concordance with Dutch privacy legislation. The institutional review board (IRB) confirmed (protocol 12/320) that viral strains are not regarded as patient-owned and that the use of these strains is not restricted in the applicable Dutch law ("Law Medical Scientific Research with People", WMO; art. 1b).

Viral RNA extraction, cDNA synthesis and real-time Taqman PCR

Viral genomic RNA was directly isolated from patient material using MagnaPure LC total nucleic acid kits (Roche Diagnostics, Mannheim, Germany). Multiscribe reverse transcriptase kits and random hexamers (Applied Biosystems, Foster City, CA) were used for reverse transcription of isolated viral RNA. Subtyping of the RSV

patient strains was achieved using real-time PCR on viral cDNA with primers and probes designed on the basis of highly conserved genomic regions of the N gene for both RSV subgroup A (RSV-A) and B (RSV-B) as described by Tan *et al.*; 2012 [48].

Synthesis and sequencing of genomic RSV PCR fragments

Human RSV-B PCR fragments were obtained via fractional amplification of MagNAPure LC genomic RNA isolates using the Superscript III one-step RT-PCR System with the Platinum Taq High Fidelity kit (Invitrogen) according to the manufacturer's protocol and a 9800 Fast thermal cycler (Applied Biosystems). PCR products were purified from 1% agarose gels by a GeneJET Gel Extraction kit (Fermentas) and were sequenced according the conventional Sanger Technique. Fragments ranging between 650 and 1,400 nucleotides were sequenced on an ABI 3730 48-capillary DNA analyzer using Big-Dye Terminator 3.1 (ABI) and sequence-specific primers (Table S2). Whole genome sequences were assembled with these strain specific PCR fragments by aligning them to the reference RSV strain, B1 (AF013254.1) using the computer program Seqman Pro (DNASTAR Lasergene 10, http://www.dnastar.com).

Substitution analysis on the whole genome and individual protein levels

Individual gene sequences were extracted from the whole genome sequences of all individual RSV-B strains *in silico*, using Seqman (34 RSV-B isolates and 2 reference RSV-B strains). Seaview4 was used to translate genes into protein coding sequences that were subsequently aligned with the EMBL-EBI ClustalW2-Multiple Sequence Alignment tool [49]. The sequence variability score per gene and coding protein was calculated relative to the gene and protein consensus, respectively, obtained from the data set of aligned clinical and reference strains. Genomic and protein substitutions per site were mapped using the Plot0.997 program (http://plot.micw.eu/). The NetNGlyc 1.0 server (shown to achieve correct prediction of 86% glycosylated and 61% nonglycosylated sites with an accuracy of 76%) [50] and NetOGlyc 3.1 server (shown to achieve correct prediction of 76% glycosylated and 93% nonglycosylated sites with an accuracy 21%) [51] were used to predict the gain and loss of N-glycosylation and O-glycosylation sites, respectively. Previously described RSV-A genome sequences [48] were included in all these analysis for comparison

with RSV-B specific genetic variations. In addition, we analyzed previously obtained RSV-A data for evidence of O-glycosylation.

Data set assembly and recombination analysis

Analysis of RSV complete genome evolutionary dynamics was conducted on a data set containing a combination of newly obtained RSV-B isolates and strains from a recent study of Rebuffo-Scheer *et al*.; 2011. The latter five full genome sequences (JN032115-JN032117, JN032119-JN032120) were derived from nasopharyngeal and nasal swabs collected from patients in the Milwaukee metro area. A total of 41 RSV-B complete genomes were aligned using Mafft. Manual editing of the alignment was performed using Se-Al (available at http://tree.bio.ed.ac.uk/software/seal). Strain recombination was evaluated using the RDP, GENECONV, RECSCAN, MAXCHI, CHIMAERA, SISCAN and 3SEQ recombination detection methods implemented in RDP3 [52]. Evidence of likely recombination events was only considered robust when these were detected by two or more different recombination detection methods together with evidence that different genome fragments apparently derived from different parental strains clearly fell within different clades of RSV-B phylogenetic trees. Two Dutch-Belgian RSV-B strains showed significant evidence of recombination. Since the analyses carried out in this study relied on accurate phylogenetic inference and a single phylogenetic tree cannot adequately describe the evolutionary history of recombinant sequences, we opted to remove the minor recombinant parts of these two genomes from the data set and replaced them by gaps in the alignment. These two strains were only incorporated in the data set for Bayesian inference, but were not included in the substitution analysis.

The evolutionary dynamics of the G gene were investigated based on an alignment of the G gene partitions from our complete RSV-B genomes together with two other data sets previously evaluated by Zlateva *et al*. (2005) [43] and Baek *et al*. (2012) [53] (overview of strains in table S3). Zlateva *et al*. aligned 204 sequences encompassing the region of amino acid position 58 to 299 of the G protein (according the AF013254 reference strain). Baek *et al.* aligned full-length G proteins from Korean origin. All RSV-B estimates obtained by the molecular evolutionary analyses were compared with recent published RSV-A estimates of molecular evolutionary analyses [48].

Bayesian inference of the RSV-B evolutionary history

In order to determine whether there existed a clear temporal signal of nucleotide divergence within our complete genome data set, exploratory analyses were performed with the computer program Path-O-Gen (available at http://tree.bio.ed.ac.uk/software/pathogen/) [54], which performs a linear regression analysis of root-to-tip divergence of individual samples against their sample dates. For this purpose, we constructed a maximum likelihood tree using PhyML [55] with a general time-reversible (GTR) substitution model and a discretized gamma distribution to model rate variation among sites. These exploratory analyses were performed with the exact sampling date (month + year) for our novel RSV-B genomes. For the five sequences from Rebuffo-Scheer *et al*. (2011) collection months were not available, so we set the sampling date at the midpoint of the reported sampling year.

The RSV evolutionary and demographic histories were reconstructed for both the whole genome sequences and the G gene data set using Bayesian genealogical inference implemented in BEAST v1.7 [56]. Using Markov Chain Monte Carlo (MCMC) analyses, we estimated a posterior distribution of time-measured genealogies based on a full probabilistic model including a nucleotide substitution model, a molecular clock model and a coalescent model as prior distribution for the phylogenetic tree. We used the GTR substitution model with a discretized gamma distribution to model rate variation among sites. Substitution rate variability across the RSV genome was studied using a partition model that allowed for separate substitution rates for all individual genes and a single concatenated non-coding region. The best molecular clock model fit was assessed on the basis of marginal likelihood estimates using path sampling (PS) estimators and stepping stone (SS) estimators [57]. We compared both strict and relaxed clock models that assume homogenous and heterogeneous substitutions rates across phylogenetic branches, respectively [58]. Tip ages of the five sequences with unknown exact sample date were estimated within a one-year time interval with a uniform prior [59]. With the Bayesian skyline plot model as a flexible demographic prior, we analyzed the changes in the effective population size through time [60]. MCMC analyses were performed until convergence could be convincingly assumed according to evaluation with Tracer (available at http://tree.bio.ed.ac.uk/software/tracer/). Marginal posterior distributions for evolutionary rates and times to the most common ancestors (TMRCA) of various sequence

groups were summarized using means and 95% highest posterior density intervals (HPDs). A maximum clade credibility (MCC) tree annotated with divergence time and evolutionary rate summaries was used as a representation of the evolutionary history and visualized using FigTree (available at http://tree.bio.ed.ac.uk/software/figtree) [58].

A genealogical test based on posterior predictive simulation was conducted in order to determine whether the posterior phylogenetic tree shapes deviated from neutral expectations [61]. In short, the tree shapes estimated by the Bayesian coalescent analysis were summarized with the genealogic Fu and Li statistic (D_F), which compares the length of the terminal branches to the total length of the coalescent genealogy and returns negative values for long terminal branch lengths. This in turn indicates an excess of slightly deleterious mutations on these branches and consequently a deviation from neutrality. Trees with the same tip numbers, the same tip ages and under the same neutral coalescent model as applied to the real data were randomly simulated by the posterior predictive simulation analysis. In this study we used the genealogical neutrality test extension that allowed simulation of trees under the Bayesian skyline plot model [62]. On the basis of the frequency at which the D_F of the inferred trees being more extreme than the D_F of the simulated trees, we derived P-values to test departures from neutrality.

Diversifying selection analyses

We identified diversifying selection in the G gene data using a combination of a fixed-effects likelihood (FEL) approach [63], a recently developed renaissance counting (RC) procedure [64] and a random-effects likelihood (REL) approach [65]. The FEL method fits codon models to each site independently and uses a likelihood ratio test to assess whether a model assuming equal nonsynonymous and synonymous rates (d_N=d_S) can be rejected in favor of a model with different d_N and d_S rates. We used a P value < 0.1 to consider sites as diversifying in our consensus approach (see below). Estimates of site-specific d_N/d_S ratios via the RC approach were obtained by combining stochastic mapping under nucleotide substitution models and empirical Bayes regularization. Neutrality was rejected and sites were considered to be under positive diversifying selection only when the 97.5% lower limit of the d_N/d_S cumulative density interval exceeded one. The REL approach fits a codon model to

the entire alignment, but allows the d_N/d_S ratio to vary among sites. We used the Akaike's information criterion to select the best fitting codon substitution model (comparing the Nonsynonymous, Dual and Lineage-Dual model, [65]) and we identify positively selected sites using an empirical Bayes method with a log Bayes factor (ln BF) cut-off of three. The cut-off values for the three different approached (FEL, RC and DUAL REL) are fairly liberal, but we use a consensus approach in considering evidence for diversifying selection as advised by Kosakovsky Pond *et al*. (2005) [63], and analogously to [48], we only list sites that are identified to be under diversifying selection by at least two of these methods.

Whether diversifying selection in the G gene was pervasive or episodic was evaluated using the recently developed mixed-effects model of evolution (MEME) [66]. This model allows the d_N/d_S distribution to vary from site to site, but importantly, also from branch to branch at a specific site, thereby relaxing the assumption that the strength of natural selection is constant among all lineages. MEME extends the FEL approach by modeling two categories of lineage-specific d_N rates for each site; one category has a $d_N \leq d_S$ ($ß^-$) and the other category is characterized by an unrestricted d_N rate ($ß^+$). The codon substitution model including these categories of sites is tested against the same model in which also $ß^+$ is constrained to values below or equal to one. We only reported sites identified as evolving under diversifying selection with an associated P-value ≤ 0.05. For these sites, we list the $ß^+$ estimate and the proportion of branches estimated to be part of this d_N rate class (p^+).

Nucleotide sequence accession numbers

The nucleotide sequences from the Dutch and Belgian RSV-B isolates were deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/index.html) under the accession numbers JX576720 – JX576762. Previous documented prototype RSV-B reference strains B1 (AF013254) and 9320 (AY353550) plus RSV-A strains JQ901447-JQ901458, JX015479-JX015499, A2 (M74568), Long (AY911262), Line19 (FJ614813) and RSS-2 (NC_001803) were included in genomic and protein substitution analysis studies. All sequence alignments used in this study are available upon request.

Results

Protein substitution mapping of RSV-A and RSV-B strains

For each RSV protein, the amino acid residue changes within 34 Dutch-Belgian RSV-B clinical strains (Table S1) and two reference lab-strains were mapped to determine the protein substitution density (Fig.1). Changes were relative to the consensus of all aligned RSV-B strains and the degree of variation to the consensus was indicated as the percentage of sequence variability per protein. The overall variation within RSV-B proteins is much lower than that previously described for RSV-A [48]. This is clearly reflected by the greater amount of sites that are substituted among the RSV-A subgroup and the greater number of RSV-A strains having a substitution at one unique site. The G protein is the most variable RSV-B protein with 2-12% variation compared to less than 5% variation in the other ten proteins. When comparing variation between the RSV subgroup A and B consensus sequences, the structural proteins SH, G and the accessory transcription regulation protein M2-2 displayed the highest sequence variability (35, 53 and 45%, respectively).

Since the SH and G proteins can be glycosylated, a process that might affect their folding and function, variations in glycosylation were studied both for these proteins, and for the more conserved RSV F glycoprotein that is essential for initializing RSV infection. Predicted glycosylation patterns show remarkable differences between the G, F and SH proteins derived from subgroup A and B strains.

Insertions and deletions alter N- and O-glycosylation predictions in the RSV-B G protein

The highly immunogenic RSV G protein binds to glycosaminoglycans (GAGs) of the host target cell to facilitate subsequent virus-host fusion [12,13]. Amino acid residues at sites located in the highly variable domains present in the ectodomain of the G protein are most prone to substitutions in both RSV-A and RSV-B strains (Fig.2A). The high degree of amino acid sequence variation found in these RSV G domains, also designated as the mucin-like regions, indicates that there are more relaxed selective constraints operating on these regions, which may allow molecular adaptation without loss of protein function. These regions act as pathogenicity

factors and are characterized by the high prevalence of serine and threonine residues, which allow the binding of O-linked glycans [67]. The central conserved domain, which contains the immunogenic domain, the TNFr homologous region and the CX3C chemokine motif, only displays substitutions at individual sites that are unique to single RSV-B sequences. This conserved part of the ectodomain is also marked by the absence of N-glycosylation (Fig.2B) or O-glycosylation (Fig.2C) sites. Specifically, N-glycan binding sites were predicted for RSV G asparagine residues 86, 144, 256, 276, 290, 294, 296, 308 and 310 within the RSV-B dataset. The first N residue (N86) is predicted to be glycosylated in all RSV-B strains, while only a single strain was predicted to have glycosylation potential at N144 or 256 due to NKP to NKS sequon changes for N144 and a K256N substitution, respectively.

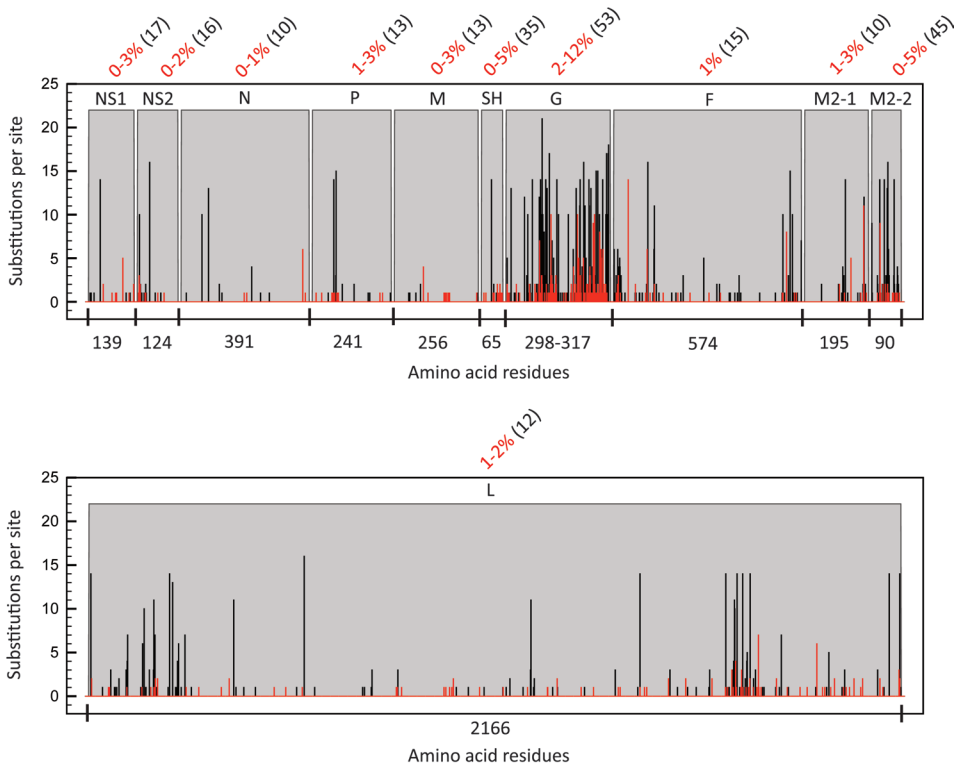


Figure 1. RSV protein sequence variability. The number of substitutions per site for RSV-A (black bars) and RSV-B (red bars) and the protein sequence variability (red %) in each RSV-B protein is calculated per strain relative to the consensus. The number between brackets indicates the percentage of protein sequence variability between the consensus sequences of type A and type B RSV strains.

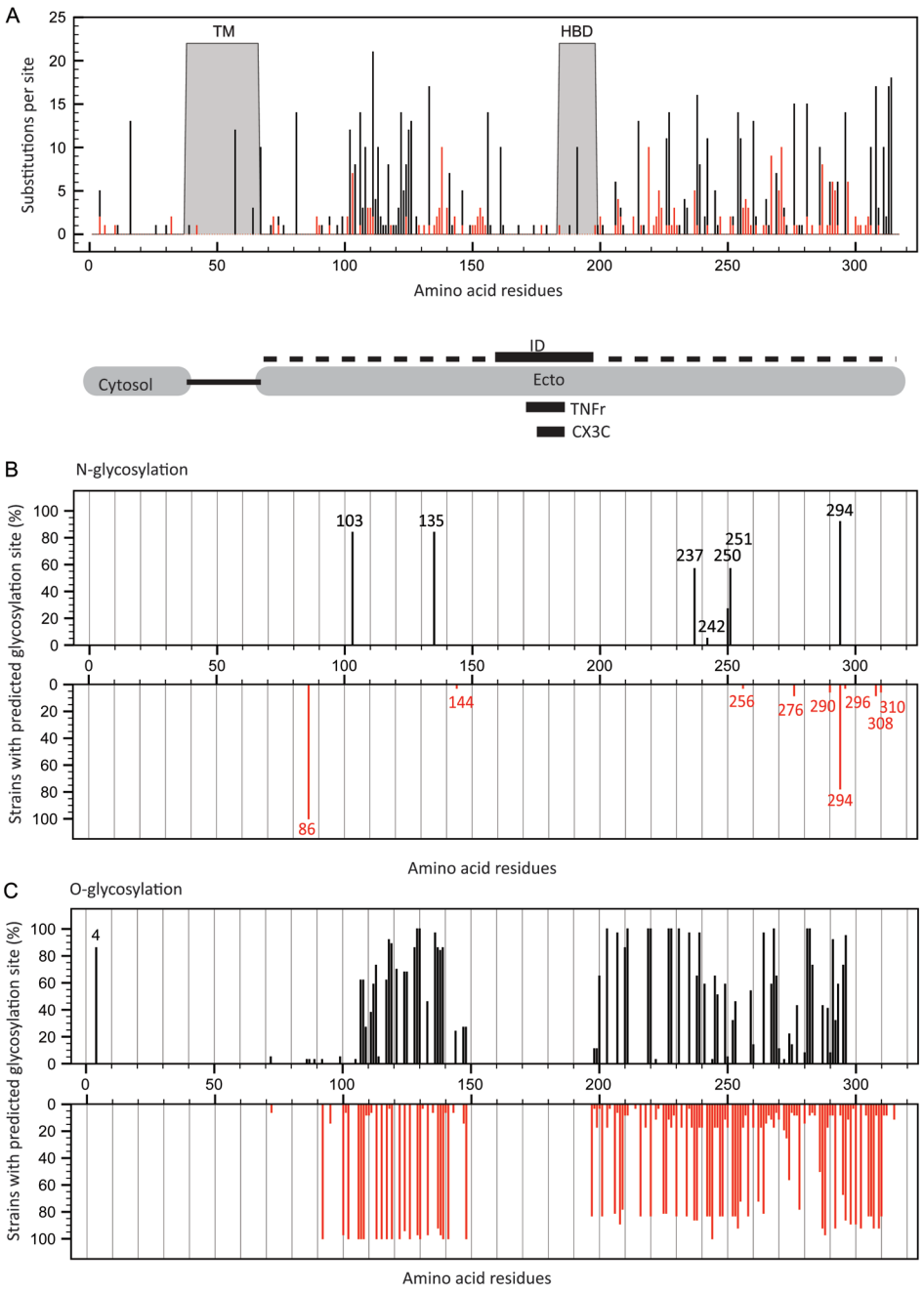

A
Substitutions per site
TM
HBD
Amino acid residues
ID
Cytosol
Ecto
TNFr
CX3C
B
N-glycosylation
Strains with predicted glycosylation site (%)
103
135
237
250
251
242
294
144
86
256
276
290
296
310
308
Amino acid residues
C
O-glycosylation
Strains with predicted glycosylation site (%)
4
Amino acid residues

Figure 2 ***left page*****. Substitution hotspots in specific RSV G protein domains and consequences for glycosylation.** (A) Schematic representation of the RSV G protein and its specific domains: the transmembrane domain (TM), heparin binding domain (HBD), N-terminal cytosolic domain, C-terminal ectodomain, immunogenic domain (ID, aa 159-198), CX3C chemokine motif (CX3C, aa 182-186) and the region homologous to the fourth subdomain of TNFr (TNFr, aa 171-186). Dashed lines represent mucin-like regions. (B,C) The percentage of strains with predicted sites for N-glycosylation (B) and O-glycosylation (C) within the G protein at a certain amino acid position for both RSV-A (black; 37 strains) and RSV-B (red, 36 strains).

The N-glycan binding residues at positions 276, 290, 294 and 296 all contain an asparagine residue (underlined) within the same PENTPNXXQTPASE acceptor sequence context. Variation in the sequence coordinates of this asparagine residue are the consequence of deletions and insertions located upstream of this acceptor sequence in various sequences (Fig.3). The majority (33 out of 36) of the studied RSV-B strains have the duplication region of 240-259 ERDTSTPQSTVLDTTTSKHT and 260-280 ERDTSTSQSIXXDTTTSKHT, but in the two reference strains AF013254, AY353550 and in the clinical strain P03-000005 the second 260 to 280 sequence is missing. On the other hand, the sequences from strains 02-000467, 02-028215, 03-000005, 05-001965, AF013254 and AY353550 have a PK insertion in the 155-164 PPKKPKDDYH sequence. Asparagine residues with similar acceptor sequences were found for N308 and 310 as well, which depict the N-glycosylation site in the STSNST C-terminal ectodomain sequence. This acceptor sequence needs to be followed by additional amino acids, which give alternative RSV G sequence ends (Fig.3) and are necessary for glycosylation potential, because strains lacking these extra residues are not predicted to be N-glycosylated at the STSNST.

The total number of predicted O-glycosylation sites in the G protein is 36% higher for RSV-B strains compared to RSV-A, and these sites are all present in the ectodomain of this protein. The O-glycan binding site at residue T4, is found in more than 80% of known RSV-A strains, and is exceptional as it lies within the cytosol domain of the attachment protein. However,this site is not present in the reference strains.

N-glycan binding sites are exclusively predicted in the F2 domain of the highly conserved fusion protein

Viral-host membrane fusion mediated via the enzymatic activated fusion protein

occurs directly after host attachment via RSV G [14]. RSV F is highly conserved among the RSV-B strains and similar to the other proteins, contains fewer substitutions than its RSV-A counterpart. Variability is particularly restricted to the signal peptide sequence of the F2 domain. In the RSV-B sequences, the asparagine residues at sites 27, 70, 116, 120 and 126 in the F2 domain are predicted to have N-glycosylation potential in all strains sequenced in this study. This is different within the RSV A strain dataset, where some strains lack glycosylation potential at N120 [48]. Despite the fact that none of the RSV-B strains have a predicted N-glycosylation at N500 (N-glycosylation potential <0.5), this specific site, with sequon NQS in both RSV-A and –B, has previously been shown to be glycosylated according to *in vitro* experiments [68].

Although rich in threonine and serine residues, the fusion protein is the only protein among the three RSV structural glycoproteins for which no O-glycosylation sites were predicted. This is in strong contrast with the attachment protein of RSV and

Aa region	Coding sequence	Strain		Domain
155-164	PPKK--DDYH	Consensus		Immunogenic
	PPKKPKDDYH	P02-000467		
		P02-028215		
		P03-000005		
		P05-001965		
		AF013254.1		
		AY353550.1		
257-283	SKHTERDTSTSQSIXXDTTTSKHTIQQ	Consensus		Mucin-like
	SG--------------------HTIQQ	P03-000005		
	LE--------------------HTIQQ	AF013254.1		
	SK--------------------HTIQQ	AY353550.1		
313-End	STSNSTKLQSYA-	Consensus	(316aa)	Mucin-like
	STSNSTQKL----	P02-000467	(315aa)	
		P05-001965	(315aa)	
	STSNSTQKLQSYA	P03-000005	(299aa)	
		P03-034613	(317aa)	
		P05-040058	(317aa)	
		P08-045952	(317aa)	
		P10-051639	(317aa)	
		P12-002874	(317aa)	
	STSNSTQNTQSHA	AF013254.1	(299aa)	

Figure 3. Amino acid sequence divergence in the G protein of RSV-B strains. Indicated are insertions (grey, underlined), deletions (-), and variable C-terminal ends.

may contribute to the low level of molecular variation in RSV F.

Comparison of N- and O-glycosylation in the SH protein of subgroup A and B strains

SH is a small hydrophobic type II integral membrane protein that acts as a viroporin in ion channel formation [16,17], thereby enhancing host membrane permeability. It is not pivotal for RSV infection [18] and the amino acid substitutions observed within this protein are mainly located in the C-terminal ectodomain. The O-glycosylation predicted for the SH protein at threonine position T64 of RSV-A strains is missing in the RSV-B subgroup strains. The difference of one amino acid in protein length, between subgroup A (64 amino acids, aa) and -B (65 aa), is most likely the cause of either O-glycosylation potential loss at the last C-terminal located threonine in RSV-B strains, or the gain of an O-glycan binding site in RSV-A strains. The Asparagine residues at positions 3 and 52 in the SH proteins of RSV-A and RSV-B are predicted to be N-glycan binding sites. However, the reference RSV-B strain, B1, lacks N-glycosylation at N52, which is most likely caused by a K53N substitution at the second position of the NKT glycosylation sequon. The T4 residue is a predicted O-linked glycan site for all RSV strains. For RSV-B strain 08-045420 the three additional O-glycosylation sites, S5, T7 and T8, were predicted to have sugar binding potential as well. It is unclear why the S5 and T7, which are present in all RSV-B strains, are specific potential glycosylation sites in 08-045420. The I8T substitution in this specific strain, however, is predicted to have created a new O-linked glycan site, which could also influence the sugar acceptor potential for the other two sites

M2-2 protein transcription variants differ between RSV subgroups A and B

The transcription regulatory protein, M2-2, acts as a switch from RNA transcription to genome replication during the RSV infection cycle and is tightly regulated by the M2-1 protein [69,70]. M2-2 can be transcribed in different truncated forms via a ribosomal termination-dependent re-initiation mechanism [71]. From the alternative start codons at position 1, 3 and 7 described for RSV-A strains, only 1 and 7 are present in all RSV-B strains included in this study. Nevertheless, the first methionine is in some RSV-A strains also substituted by a threonine or proline residue, which is predicted to result in the transcription of only the truncated M2-2 forms.

The phylodynamic history of RSV-B

Plotting the root-to-tip divergence as a function of sampling time for each RSV-B genome clearly showed a temporal nucleotide divergence signal for the RSV-B complete genomes sampled over the last ten years (Fig.4). Although this analysis is exploratory in nature, a comparison with RSV-A already suggests a number of interesting aspects. First, the regression lines are roughly parallel to each other, which indicates similar rates of evolution. Second, the levels of divergence from the root are considerably higher for RSV-A compared to RSV-B implying a shorter TMRCA for the latter, which offers an explanation for the over-all lower degrees of genomic variability among the analyzed RSV-B genomes. Finally, the regression analyses also suggest a lower variability in evolutionary rates among lineages for the RSV-B genomes.

The root-to tip divergence analysis clearly justified the application of a dated tip molecular clock model during a Bayesian evolutionary reconstruction. Both strict and relaxed clock models were compared using two different approaches

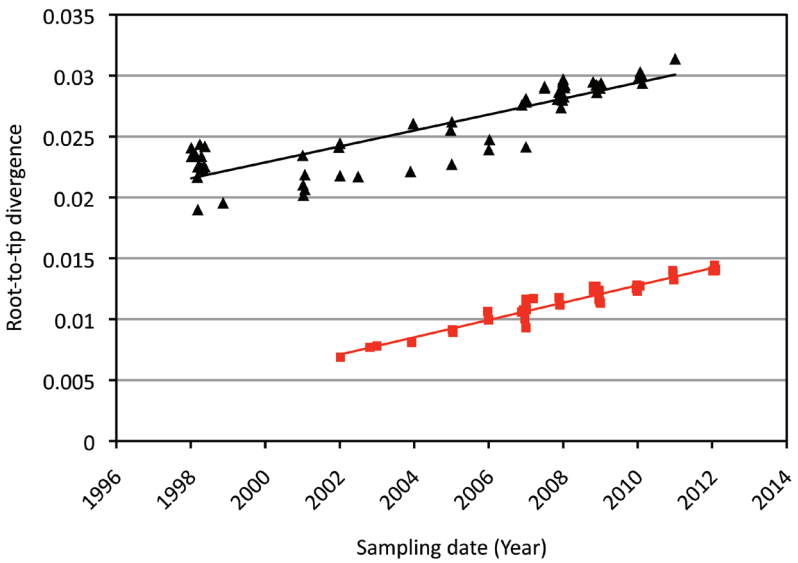


Figure 4. Plot of the root-to-tip divergence as a function of sampling time for the RSV-A and RSV-B genomes. RSV-A is indicated by black triangles and RSV-B is indicated by red squares; the corresponding regression lines are plotted in the same colors.

Table 1. Model-fit of molecular clock models based on log likelihood estimates.

Marginal likelihood estimator	Whole genome		G gene	
	Strict	Relaxed	Strict	Relaxed
Path sampling	-32304.88367	-32307.19482	-7706.969731	-7682.638227
Stepping-stone sampling	-32305.06152	-32307.41622	-7718.684767	-7696.065831

to compare model fit in a Bayesian framework (Table 1). The best model fit was indicated by the highest log marginal likelihood estimate (in bold). Both the path sampling and stepping stone sampling approaches, which have been shown to be the most accurate Bayesian phylogenetic model selection methods [57], agreed with the assumption of a constant substitution rate across phylogenetic tree branches based on genomic sequences. This is in line with the root-to-tip regression exploration and argues for the use of the strict clock model for further analysis of these data. However, for the G gene data set that includes many more RSV-B strains [43,47,53], this assumption had to be rejected and a relaxed clock model was required to accommodate rate variation among lineages.

The estimated posterior tree distribution for the genome-wide RSV-B data set represented by a maximum clade credibility tree did not provide evidence for geographical or strong temporal clustering of the RSV-B genomes (Fig.5). Strains from several different epidemic seasons are present in the same clusters and the Milwaukee strains are phylogenetically interspersed with the Dutch-Belgian genomes. The MRCA of the RSV-B genome phylogeny dated back to approximately 1993 (95% credibility interval between 1991 and 1995; Table 2), which is, as suggested by the linear regression analyses, considerably younger than estimates for a similar sample of RSV-A genomes (1964, CI 95% [1956,1973], [48]). The shorter TMRCA for RSV-B explains why we observed lower genetic diversity over the complete genome but this does not appear to be the result of a lower rate of evolution because the latter was estimated to be 7.76 x 10^{-4} substitutions per site per year (CI 95% [6.89 x 10^{-4} to 8.58 x 10^{-4}]), which is very similar to that of RSV-A (6.47 x 10^{-4}, CI 95% [5.56 x 10^{-4}, 7.38 x 10^{-4}], [48]).

Table 2. Evolutionary rate and TMRCA estimates for RSV-A and RSV-B

Evolutionary parameter	Input	Mean	95% HPD lower	95% HPD upper
Mean rate x 10^{-4} (subst./site/yr)	RSV-A (genomic)	6.47	5.56	7.38
	RSV-B (genomic)	7.76	6.89	8.58
	RSV-A (G gene)	22.2	19.3	25.6
	RSV-B (G gene)	27.8	23.5	32.3
TMRCA (yr)	RSV-A (genomic)	1964	1957	1972
	RSV-B (genomic)	1993	1991	1995
	RSV-A (G gene)	1942	1929	1953
	RSV-B (G gene)	1955	1946	1960

In light of the short TMRCA, we studied the clustering of the strains for which complete genomes are available in the G gene data set (data not shown). This revealed that these strains also coalesce into a single internal node that dates back to 1994 (1992, 1995) while the root of the G gene tree is dated back to 1955 (1946, 1960) (Table 2). So, whereas a recently obtained RSV-A full genome sample appeared to be representative for the strains diversity in G gene phylogeny [48], the RVS-B sample represents the progeny of a more recent ancestor within the complete strain diversity. This could represent a chance fixation in a largely neutral epidemiological scenario or it may reflect the fixation of a variant with advantageous mutations. This also impacts the classification of recent RSV-B strains according to the G gene tree; the GB13 lineage is predominant amongst the Dutch-Belgian strains and only strain 03-000005 belongs to another lineage (GB12).

To further investigate the impact of this fixation event on RSV-B population dynamics, we reconstructed the change in relative genetic diversity through time using a Bayesian skyline plot model in our Bayesian genealogical inference for both the G gene and full genome data set (superimposed in Fig.6). This reveals a noticeable degree of variation in relative genetic diversity through time, including a contraction and expansion between 2002 and 2008. From the full genome sample, we only reconstruct the recovery dynamics due to the absence of samples prior to

2002. The increase in relative genetic diversity from the early nineties in the G gene skyline might also reflect such a bottleneck event because the paucity in samples before this time hampers the reconstruction of the complete ancestral diversity. The timing of such a potential bottleneck event coincides with the MRCA of the strains for which complete genomes are available in the G gene data set.

It is important to point out that, as is the case with RSV-A, the RSV-B attachment protein gene appears to have a significantly higher substitution rate than the RSV-B full genome (Table 2) [43,44]. The variability in evolutionary rates across RSV genomes was further determined for all the various gene partitions and the combined non-

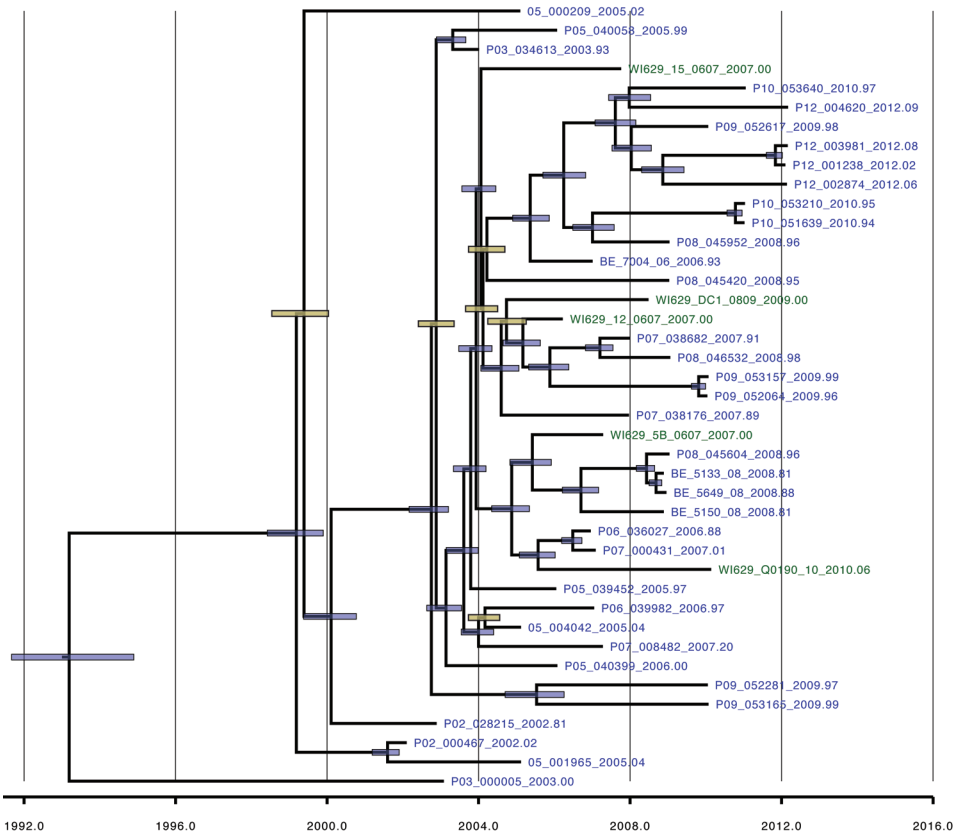


Figure 5. RSV-B whole genome based phylogeny. Distribution of Dutch-Belgian strains (blue) and Milwaukee strains (green). The node bars depict the credibility intervals for nodes showing a posterior probability support >95% (blue) or <95% (yellow).

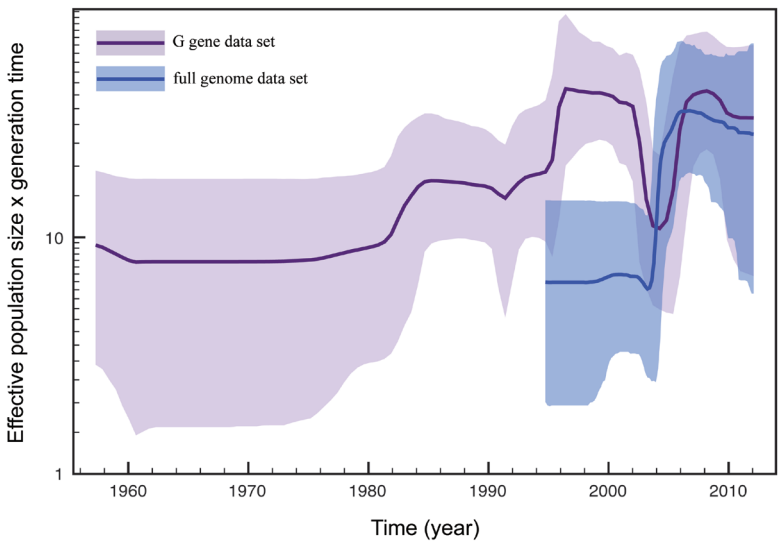


Figure 6. Bayesian skyline plot. Indication of the estimated change in effective population size over time for both the full genome (blue) and G gene (purple) data sets. The thick lines represent the mean estimate whereas the transparent areas represent the 95% highest posterior density intervals.

coding regions via the Bayesian estimation approach in order to individually identify the impacts of different genes on the genome-wide substitution rate (Fig.7). Analogous to the previous RSV-A results, considerably elevated substitution rates were observed in the non-coding regions and G gene, while the other genes were more conserved in both RSV subgroups. Relatively high genetic G gene variability (1-8% variability; Fig.8) and the associated relatively high (almost three-fold higher compared to other genes) substitution rate estimate, which was in line with previous estimates [44], indicate that either there is a general relaxation of negative selection (i.e. selection disfavoring change) acting on this gene, or that there are various codons within this gene that are evolving under positive selection (i.e. selection favoring change).

Patterns of natural selection acting at individual sites

Sites within the RSV G protein that are potentially evolving under diversifying selection were identified by querying nonsynonymous/synonymous substitution

rate ratios (d_N/d_S) at each codon position of the G-protein gene using the FEL, RC and REL methods (Table 3). From the 11 sites identified by at least two of these three methods as displaying evidence of diversifying selection, only site 224 and 230 (marked in bold), were significantly supported by all three of the methods. As has been previously found with RSV-A most of the 11 sites were within the two hyper-variable regions of the G protein ectodomain. However, sites 50 and 53 are part of the transmembrane domain. In general the methods applied here to study site-specific selective patterns will identify sites that are under pervasive diversifying selection (i.e diversifying selection in most lineages of the phylogeny). However, the RC, FEL and REL methods cannot efficiently detect episodic diversifying selection, which involves diversifying selection of specific sites in a restricted number of branches in the phylogeny with either neutral or negative selection occurring at these sites along the remaining branches. Therefore MEME analysis was conducted to discriminate between these two kinds of diversifying selection. In total ten sites were identified to be under diversifying selection by MEME (with an associated P-value ≤ 0.05). Five

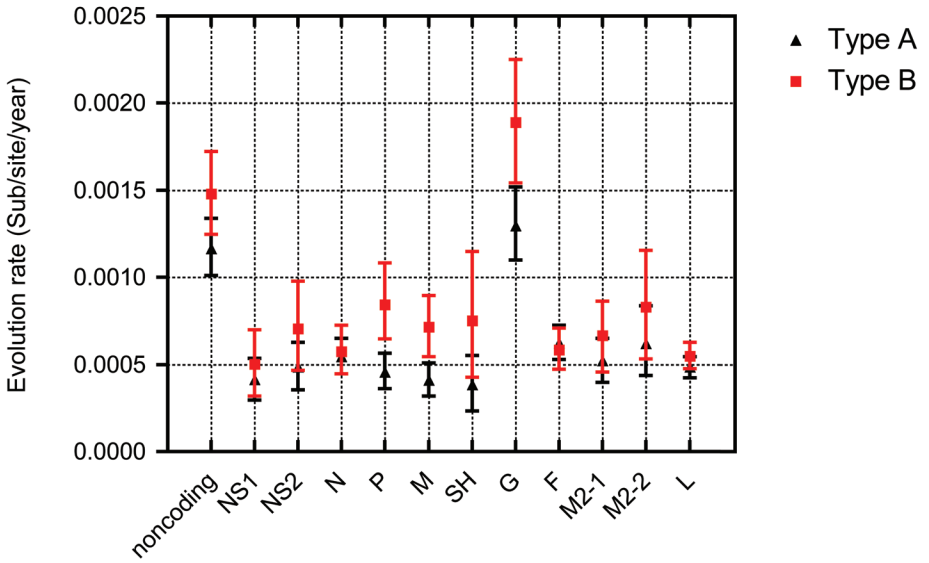


Figure 7. Comparison of RSV-A and RSV-B evolutionary rate partitions. The mean rate of evolutionary changes indicated by the numbers of substitutions per site for each year was estimated per gene and for the combined non-coding sequence parts for the RSV-A (black triangles) and RSV-B (red squares) datasets.

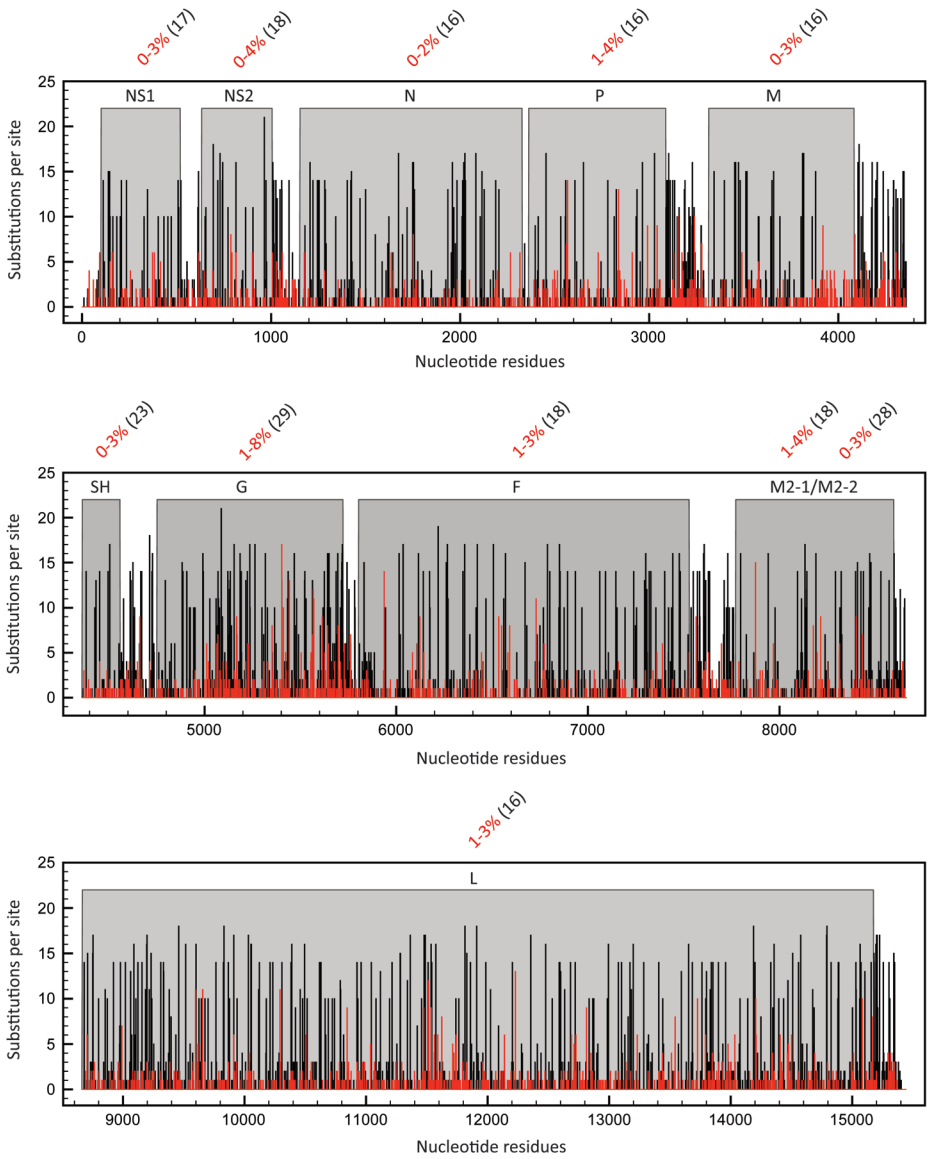


Figure 8. RSV nucleotide sequence variability in whole genome. The number of substitutions per site for the RSV-A (black bars) and RSV-B (red bars) genomes and the nucleotide sequence variability (red %) in each RSV-B gene is calculated per strain relative to the consensus. The number between brackets indicates the nucleotide sequence variability between the consensus of type A and type B RSV strains.

Table 3. Sites under diversifying selection in the G gene data set

	RC	FEL		REL	
Site	d_N/d_S	$d_N - d_S$	P value	E [d_N/d_S]	Log (BF)
50	8.00 (3.59, 15.47)	3.60	0.05	3.21	2.54
53	4.74 (1.50, 9.46)	1.88	0.26	3.67	3.49
81	7.74 (3.62, 15.23)	3.58	0.05	3.20	2.53
85	3.77 (1.20, 7.56)	1.41	0.37	2.53	3.15
160	6.80 (3.17, 13.35)	10.15	0.00	2.90	2.10
187	4.82 (2.12, 9.35)	2.00	0.20	3.79	3.91
210	7.82 (1.70, 18.98)	10.39	0.06	2.15	1.20
224	7.62 (3.29, 15.19)	3.50	0.06	3.84	4.15
226	2.99 (1.40, 5.90)	1.67	0.31	3.52	3.12
230	9.79 (3.90, 19.48)	4.76	0.03	3.75	3.78
239	4.97 (2.14, 9.88)	1.86	0.26	3.72	3.66

Sites marked in bold are positive selected sites that meet the criteria of all applied methods

of these sites, which are apparently evolving under diversifying selection across a considerable proportion of branches in the phylogeny (p^+), were also identified by two of the three other selection detection methods (marked in bold in Table 4) and can therefore be considered as sites under pervasive diversifying selection. All other sites detected via MEME are evolving under diversifying selection across only small numbers of subsets of branches in the phylogeny ($ß^+$) and are thus identified as evolving under episodic diversifying selection.

Testing genealogical departures from neutrality

To assess the impact of diversifying selection in the G gene on RSV-B circulation patterns we performed a genealogical test of neutrality using posterior predictive simulation. Phylogenetic tree shapes deduced from complete genome Bayesian inference were compared to those simulated with the same population dynamics and under a model of neutrality. This test indicated that neutrality could be rejected

Table 4. Sites under episodic diversifying selection in the G gene data set

Site	β^+	p^+	P value
18	58.53	0.01	0.00
138	135.60	0.01	0.04
145	10000.00	0.01	0.04
153	600.31	0.01	0.02
160	3.12	1.00	0.02
187	3.34	0.42	0.04
226	6.35	0.23	0.03
230	4.65	0.61	0.01
238	768.59	0.01	0.04
239	1.84	0.97	0.04

Sites marked in bold are positive selected sites that meet the criteria of one or more of the applied methods. Lineage-specific unrestricted d_N rate estimate for indicated site (β^+) and proportion of branches estimated to be part of this d_N rate class (p^+).

at the population level for the complete genome RSV-B data set (P = 0.01). This is in contrast with the RSV-A population that did not show departures from neutrality (P = 0.24) [48]. Although approaching significance, neutrality cannot be convincingly rejected when the large data set of the highly variable G gene is considered in the predictive simulation (RSV-B P = 0.06; RSV-A P = 0.08).

Discussion

Here we have used 34 RSV-B full genome sequences sampled between 2002 and 2012 to infer genome-wide mutation density maps and nucleotide substitution rates and compared these with those determined previously for RSV-A [48]. Although the identified substitution hotspots closely resemble those detected for RSV-A, particular differences may reflect RSV subtype-specific variation in replication and/or immune evasion strategies and could therefore be helpful in explaining RSV

subtype specific differences in transmission dynamics and epidemiology.

Similar to RSV-A, most sequence variability within RSV-B strains was detected in the intergenic non-coding regions and the G gene; in the G gene, high nucleotide sequence variability correlated with an elevated amino acid substitution rate. In contrast with the RSV-A strains analyzed to date, the G genes from RSV-B strains showed a high frequency of nucleotide insertions and deletions within the G ectodomain. The highly variable and immunogenic G protein has been previously suggested to be susceptible to neutralizing antibodies; this might drive the selection of strains carrying immune escape mutations within the ectodomain [72].

Unlike with RSV-A strains, we did not observe high variability in the M2-2 gene in the RSV-B strains studied. When comparing RSV-A and RSV-B gene products, the highest protein sequence variation between these subtypes was found between the G, SH and M2-2 proteins. For the SH protein, variation is mainly caused by C-terminal polymorphisms in the ectodomain, which were more common among RSV-A strains. Apparently, the impact of immunogenic pressure on the evolution of other RSV genes is far less substantial since these genes show high degrees of conservation. However, even low substitution rates in these genes could still have a large effect on viral fitness and the course of RSV transmission dynamics. It is therefore entirely relevant to map substitutions in conserved regions to both support future research into the biological functions of the various RSV proteins, and to inform the development of preventive and therapeutic approaches. In the latter case it is particularly relevant to compare RSV-A and RSV-B variation in efforts to develop protective strategies that will be effective against all RSV infections.

Amino acid site alterations that are either correlated with changes in N- and O-glycosylation potential or susceptibility to antibody neutralization can both trigger RSV phenotypic differences. Such differences might include increased RSV infectiousness through adaptation to host defenses and altered transmission dynamics. Although we explicitly recognize the urge to experimentally confirm the true nature of glycosylation predictions, we also feel that the *in silico* assessment of N- and O-glycosylation potential on the extended whole-genome database presented here yield an excellent starting point for further in-depth analyses throughout the RSV field. We show that RSV-A and -B strains contain similar numbers of predicted

N-glycan binding sites, although the exact glycosylation sites vary greatly due to substantial sequence variability within and between the two subgroups. Only two sites, N86 and N294, were found predominately in the G protein of RSV-B strains. The latter asparagine position was indicated as N276, N290 and N296 within other strains of the RSV-B subgroup due to deletions or insertions in upstream sequences. Binding sites for N-glycans on the fusion protein of RSV-B strains were largely similar to those predicted for RSV-A strains, with binding potential being predicted for all of the analyzed RSV-B strains. In contrast to the findings of Zimmer *et al*. (2001) [68] neither the asparagine residue at position N500 in the RSV-B strains nor the homologous asparagine in RSV-A strains had any predicted glycosylation potential. Zimmer *et al*. (2001) showed that glycosylation at this site was required for efficient syncytia formation. This repeatedly emphasizes the requirement for *in vitro* analysis to verify predicted glycosylation sites. O-glycosylation was only predicted for the RSV attachment protein where it potentially occurs mainly in the mucin-like regions of this protein. Surprisingly, the T4 O-glycan binding site in the cytosolic domain of the G protein was predicted for 80% of the RSV-A strains, but it was not detected in either the reference strains or any of the analyzed RSV-B strains. Relative to the RSV-B strains, the RSV-A strains contained fewer sites with O-glycan binding potential. Furthermore, the RSV F protein, which is rich in threonine and serine residues, also appears to lack the potential to be O-glycosylated. Since the extensive glycosylation of the G protein may shield these viruses from immune recognition, the conserved F protein, with its minimal degree of glycosylation, is possibly more suitable as a target for intervention strategies.

RSV-B evolutionary rates and demography were analyzed using Bayesian phylogenetic techniques based on a recombination-free whole genome sequence dataset. Two of the RSV genomes in our initial data set showed some evidence of potential recombination. Since genomic recombination in RSV is believed to be extremely rare [73], it is most likely that these recombinants arose as a result of PCR or sequencing artifacts. We removed the genomic parts that appeared to be inherited from minor parental strains from these mosaic genomes and treated them as unobserved sequence for further phylogenetic, population genetic, molecular clock and selection analyses.

For both the RSV-B strains analyzed here and the RSV-A strains analyzed previously it is evident that strains isolated from different geographical regions tend to have a phylogenetically mixed distribution within the RSV phylogeny. This indicates that there is no strong long-term geographical or temporal structuring of RSV populations.

Nevertheless, the RSV-B strains we sampled coalesced into a relatively recent root of the complete genome phylogeny (dated back to 1993 [1991,1995]) as well as into an internal node with similar age in the G gene phylogeny. As a consequence of the relatively recent TMRCA, all except one of the RSV-B strains (03-000005) in this study belong to the GB13 clade. The relatively recent ancestry was not the case for the RSV-A genome sample, which appeared to be representative for the complete diversity in the corresponding G gene tree. This recent TMRCA may be the result of a chance fixation of a variant or a variant driven to fixation by some advantageous mutation in its genome. The Bayesian skyline plot reconstructions indicated that such fixations could impose population genetic bottlenecks, which can result in relatively constant levels of genetic diversity over longer time scales. The fact that we were able to detect these dynamics for RSV-B and not for RSV-A may be due to differences in sampling. Whereas the RVS-B complete genome sample was restricted to 2002-2012, the RSV-A data set contained a considerable amount of strains sampled in 1998, which increases the probability of capturing more ancestral diversity. Also, despite the fact that there does not seem to be strong geographical clustering of RSV-B diversity, additional sampling from different locations may be useful to further elucidate RSV population dynamics.

Similar to the RSV-A strain data set, the majority of RSV-B genes are apparently evolving under pervasive negative selection against a background of neutrally evolving sites. Both pervasive positive selection and episodic diversifying selection were, however, readily detectable within the G gene. Of the 11 positively selected sites identified by the RC, FEL and REL methods, only five sites were confirmed by the MEME method to be evolving under pervasive positive selection. In addition, MEME identified episodic selection at four other sites. This novel method for selective pressure analyses is less sensitive to the influences of sequence sampling and to false identification of diversifying selection. Surprisingly, besides the fact that

most diversifying selection was detected in the highly variable mucin-like regions of the RSV-B G gene, some positively selected sites were also found in the cytosolic domain (site 18), transmembrane domain (sites 50 and 53) and the more conserved region of the ectodomain (sites 160 and 187). This contrasts with the patterns of positively selected sites previously observed for RSV-A strains for which positive selection was primarily found in the mucin-like regions. Also in contrast to RSV-A strains, we did not detect any positively selected sites in RSV-B that might directly influence N-glycosylation.

According to the genealogical tests of neutrality, the genetic bottlenecks are still compatible with population turnover under neutral expectations for the G gene evolutionary history even though the test approaches significance (P = 0.06). This was similar to the situation observed with the RSV-A dataset [48]. Although it would be useful to explore other genealogical test statistics in the posterior predictive simulation, such as a recently proposed temporal clustering statistic [74], which may prove more sensitive in detection of departures from neutrality, it is clear that population turnover is not as pervasive and selection-driven as that observed for human Influenza A [62] and Norovirus GII.4 [61]. Neutrality was however rejected for the complete genome RSV-B data set, but since this focuses on a smaller time scale that accommodates one of bottleneck events observed over longer time scales (Fig.6), it remains difficult to make strong conclusions based on this. The exclusive accumulation of neutral substitutions in the G gene and the diversifying selection detected in multiple RSV G domains suggests that there are relaxed evolutionary constrains on this gene. Therefore, the RSV G protein may not be the best target for effective vaccines as it is not expected to induce functional B-cell mediated immunity.

For the determination of the evolutionary rate, as well as for constructing a phylogenetic tree for the genotyping of RSV strains, the whole-genomic approach followed both in this paper, and that of Tan *et al*.; 2012 [48], yielded results that were broadly consistent with those obtained from analyses of the G gene partitions. From that perspective, a conclusion that G gene analysis is an appropriate choice for molecular epidemiology studies might be justified. This is due to the fact that the G gene accounts for the majority of genomic variation, which in turn is mainly situated in regions of the gene encoding the highly variable mucin-like regions of

the G protein. Most likely, the only structural and functional requirements for these regions are the presence of heavily glycosylated amino acids. However, the whole genome analyses presented in this paper and that of Tan *et al.*; 2012 [48] revealed potentially important aspects of M2-2, F and SH gene diversity, information that would have remained undisclosed when focusing solely on the G gene.

In summary, RSV-A and RSV-B whole genome evolutionary analyses have revealed that these subgroups have similar evolutionary rates and display comparable degrees of conservation. In both subgroups, the G gene is exceptionally variable and displays elevated substitution rates compared to other genes. In both subgroups the high substitution rates observed in this gene are at least in part due to the fact that a substantial number of residues are evolving under diversifying selection. Despite sporadic instances of positive selection, the majority of accumulating substitutions are likely neutral in both subgroups. This implies that the high degrees of RSV G gene diversity are not entirely driven by the evasion of adaptive host immunity. Individuals attempting to acquire immunological memory only weaken their immunological balance, ironically decreasing their tolerance to RSV and increasing the probability of experiencing more severe disease. Overall, the high degrees of G protein variability, its likely extensive glycosylation, and evidence that various of its amino acids are evolving under diversifying selection imply that the G protein might not be as good a target for prevention/intervention strategies as the more conserved F protein, which displays far less evidence of glycosylation and diversifying selection.

At present, this study represents the broadest conceivable RSV genome analysis, which provides us with unique information on the degree of genetic conservation and the existence of natural genomic variants. These data constitute a key source for further extensive research in the field of RSV protein structure-function analyses, immune regulatory onset studies and are highly valuable for specific peptide selection in therapeutic development studies.

Acknowledgements

We are grateful to Drs. Anton M. van Loon and Marc van Ranst for allowing access to their collection of Dutch and Belgian RSV-B isolates, respectively.

References

1. Henrickson KJ, Hoover S, Kehl KS, Hua W (2004) *National disease burden of respiratory viruses detected in children by polymerase chain reaction*. Pediatr Infect Dis J 23: S11-18.

2. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE (2005) *Respiratory syncytial virus infection in elderly and high-risk adults*. N Engl J Med 352: 1749-1759.

3. Falsey AR, Walsh EE (2000) *Respiratory syncytial virus infection in adults*. Clin Microbiol Rev 13: 371-384.

4. Han LL, Alexander JP, Anderson LJ (1999) *Respiratory syncytial virus pneumonia among the elderly: an assessment of disease burden*. J Infect Dis 179: 25-30.

5. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, et al. (2010) *Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis*. Lancet 375: 1545-1555.

6. Zlateva KT, Vijgen L, Dekeersmaeker N, Naranjo C, Van Ranst M (2007) *Subgroup prevalence and genotype circulation patterns of human respiratory syncytial virus in Belgium during ten successive epidemic seasons*. J Clin Microbiol 45: 3022-3030.

7. White LJ, Mandl JN, Gomes MG, Bodley-Tickell AT, Cane PA, et al. (2007) *Understanding the transmission dynamics of respiratory syncytial virus using multiple time series and nested models*. Math Biosci 209: 222-239.

8. Meerhoff TJ, Paget JW, Kimpen JL, Schellevis F (2009) *Variation of respiratory syncytial virus and the relation with meteorological factors in different winter seasons*. Pediatr Infect Dis J 28: 860-866.

9. Cowton VM, McGivern DR, Fearns R (2006) *Unravelling the complexities of respiratory syncytial virus RNA synthesis*. J Gen Virol 87: 1805-1821.

10. Batonick M, Wertz GW (2011) *Requirements for Human Respiratory Syncytial Virus Glycoproteins in Assembly and Egress from Infected Cells*. Adv Virol 2011.

11. Mitra R, Baviskar P, Duncan-Decocq RR, Patel D, Oomens AG (2012) *The Human Respiratory Syncytial Virus Matrix Protein is Required for Maturation of Viral Filaments*. J Virol.

12. Hallak LK, Kwilas SA, Peeples ME (2007) *Interaction between respiratory syncytial virus and glycosaminoglycans, including heparan sulfate*. Methods Mol Biol 379: 15-34.

13. Hallak LK, Spillmann D, Collins PL, Peeples ME (2000) *Glycosaminoglycan sulfation requirements for respiratory syncytial virus infection*. J Virol 74: 10508-10513.

14. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, et al. (2001) *Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion*. Proc Natl Acad Sci U S A 98: 9859-9864.

15. Techaarpornkul S, Barretto N, Peeples ME (2001) *Functional analysis of recombinant respiratory syncytial virus deletion mutants lacking the small hydrophobic and/or attachment glycoprotein gene*. J Virol 75: 6825-6834.

16. Gan SW, Tan E, Lin X, Yu D, Wang J, et al. (2012) *The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels*. J Biol Chem 287: 24671-24689.

17. Gan SW, Ng L, Lin X, Gong X, Torres J (2008) *Structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain*. Protein Sci 17: 813-820.

18. Collins PL, Mottet G (1993) *Membrane orientation and oligomerization of the small hydrophobic protein of human respiratory syncytial virus*. J Gen Virol 74 (Pt 7): 1445-1450.

19. Rixon HW, Brown G, Aitken J, McDonald T, Graham S, et al. (2004) *The small hydrophobic (SH) protein accumulates within lipid-raft structures of the Golgi complex during respiratory syncytial virus infection*. J Gen Virol 85: 1153-1165.

20. Lo MS, Brazas RM, Holtzman MJ (2005) *Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness*. J Virol 79: 9315-9319.

21. Bukreyev A, Yang L, Fricke J, Cheng L, Ward JM, et al. (2008) *The secreted form of respiratory syncytial virus G glycoprotein helps the virus evade antibody-mediated restriction of replication by acting as an antigen decoy and through effects on Fc receptor-bearing leukocytes*. J Virol 82: 12191-12204.

22. Anderson LJ, Hierholzer JC, Tsou C, Hendry RM, Fernie BF, et al. (1985) *Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies*. J Infect Dis 151: 626-633.

23. Mufson MA, Orvell C, Rafnar B, Norrby E (1985) *Two distinct subtypes of human respiratory syncytial virus*. J Gen Virol 66 (Pt 10): 2111-2124.

24. Peret TC, Hall CB, Schnabel KC, Golub JA, Anderson LJ (1998) *Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community*. J Gen Virol 79 (Pt 9): 2221-2229.

25. Hall CB, Walsh EE, Long CE, Schnabel KC (1991) *Immunity to and frequency of reinfection with respiratory syncytial virus*. J Infect Dis 163: 693-698.

26. Mufson MA, Belshe RB, Orvell C, Norrby E (1987) *Subgroup characteristics of respiratory syncytial virus strains recovered from children with two consecutive infections*. J Clin Microbiol 25: 1535-1539.

27. Johnson PR, Spriggs MK, Olmsted RA, Collins PL (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: 5625-5629.

28. Sullender WM (2000) *Respiratory syncytial virus genetic and antigenic diversity*. Clin Microbiol Rev 13: 1-15, table of contents.

29. Cane PA, Matthews DA, Pringle CR (1994) *Analysis of respiratory syncytial virus strain variation in successive epidemics in one city*. J Clin Microbiol 32: 1-4.

30. Anderson LJ, Hendry RM, Pierik LT, Tsou C, McIntosh K (1991) *Multicenter study of strains of respiratory syncytial virus*. J Infect Dis 163: 687-692.

31. Peret TC, Hall CB, Hammond GW, Piedra PA, Storch GA, et al. (2000) *Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America*. J Infect Dis 181: 1891-1896.

32. Tang YW, Graham BS (1997) *T cell source of type 1 cytokines determines illness patterns in respiratory syncytial virus-infected mice*. J Clin Invest 99: 2183-2191.

33. Klein Klouwenberg P, Tan L, Werkman W, van Bleek GM, Coenjaerts F (2009) *The role of Toll-like receptors in regulating the immune response against respiratory syncytial virus*. Crit Rev Immunol 29: 531-550.

34. Chin J, Magoffin RL, Shearer LA, Schieble JH, Lennette EH (1969) *Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population*. Am J Epidemiol 89: 449-463.

35. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE (1969) *An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine*. Am J Epidemiol 89: 405-421.

36. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, et al. (1969) *Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine*. Am J Epidemiol 89: 422-434.

37. Murphy BR, Walsh EE (1988) *Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity*. J Clin Microbiol 26: 1595-1597.

38. Castilow EM, Olson MR, Meyerholz DK, Varga SM (2008) *Differential role of gamma interferon in inhibiting pulmonary eosinophilia and exacerbating systemic disease in fusion protein-immunized mice undergoing challenge infection with respiratory syncytial virus*. J Virol 82: 2196-2207.

39. Kruijsen D, Bakkers MJ, van Uden NO, Viveen MC, van der Sluis TC, et al. (2010) *Serum antibodies critically affect virus-specific CD4+/CD8+ T cell balance during respiratory syncytial virus infections*. J Immunol 185: 6489-6498.

40. Kruijsen D, Schijf MA, Lukens MV, van Uden NO, Kimpen JL, et al. (2011) *Local innate and adaptive immune responses regulate inflammatory cell influx into the lungs after vaccination with formalin inactivated RSV*. Vaccine 29: 2730-2741.

41. Olson MR, Varga SM (2007) *CD8 T cells inhibit respiratory syncytial virus (RSV) vaccine-enhanced disease*. J Immunol 179: 5415-5424.

42. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) *Unifying the epidemiological and evolutionary dynamics of pathogens*. Science 303: 327-332.

43. Zlateva KT, Lemey P, Moes E, Vandamme AM, Van Ranst M (2005) *Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein*. J Virol 79: 9157-9167.

44. Zlateva KT, Lemey P, Vandamme AM, Van Ranst M (2004) *Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup a: positively selected sites in the attachment g glycoprotein*. J Virol 78: 4675-4683.

45. Holmes EC, Grenfell BT (2009) *Discovering the phylodynamics of RNA viruses*. PLoS Comput Biol 5: e1000505.

46. Kumaria R, Iyer LR, Hibberd ML, Simoes EA, Sugrue RJ (2011) *Whole genome characterization of non-tissue culture adapted HRSV strains in severely infected children*. Virol J 8: 372.

47. Rebuffo-Scheer C, Bose M, He J, Khaja S, Ulatowski M, et al. (2011) *Whole genome sequencing and evolutionary analysis of human respiratory syncytial virus A and B from Milwaukee, WI 1998-2010*. PLoS One 6: e25468.

48. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJG, et al. (In press) *Genetic Variability among Complete Human Respiratory Syncytial Virus Subgroup A*

Genomes: Bridging Molecular Evolutionary Dynamics and Epidemiology. PLoS One.

49. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) *Clustal W and Clustal X version 2.0*. Bioinformatics 23: 2947-2948.

50. Gupta R, Brunak S (2002) *Prediction of glycosylation across the human proteome and the correlation to protein function*. Pac Symp Biocomput: 310-322.

51. Julenius K, Molgaard A, Gupta R, Brunak S (2005) *Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites*. Glycobiology 15: 153-164.

52. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) *RDP3: a flexible and fast computer program for analyzing recombination*. Bioinformatics 26: 2462-2463.

53. Baek YH, Choi EH, Song MS, Pascua PN, Kwon HI, et al. (2012) *Prevalence and genetic characterization of respiratory syncytial virus (RSV) in hospitalized children in Korea*. Arch Virol 157: 1039-1050.

54. Drummond A, Pybus OG, Rambaut A (2003) *Inference of viral evolutionary rates from molecular sequences*. Adv Parasitol 54: 331-358.

55. Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) *Estimating maximum likelihood phylogenies with PhyML*. Methods Mol Biol 537: 113-137.

56. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. Mol Biol Evol.

57. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) *Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty*. Mol Biol Evol.

58. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) *Relaxed phylogenetics and dating with confidence*. PLoS Biol 4: e88.

59. Shapiro B, Ho SY, Drummond AJ, Suchard MA, Pybus OG, et al. (2011) *A Bayesian*

phylogenetic method to estimate unknown sequence ages. Mol Biol Evol 28: 879-887.

60. Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) *Time dependency of molecular rate estimates and systematic overestimation of recent divergence times*. Mol Biol Evol 22: 1561-1568.

61. Drummond AJ, Suchard MA (2008) *Fully Bayesian tests of neutrality using genealogical summary statistics*. BMC Genet 9: 68.

62. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, et al. (2010) *Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants*. PLoS Pathog 6: e1000884.

63. Kosakovsky Pond SL, Frost SD (2005) *Not so different after all: a comparison of methods for detecting amino acid sites under selection*. Mol Biol Evol 22: 1208-1222.

64. Lemey P, Minin VN, Bielejec F, Kosakovsky Pond SL, Suchard MA (2012) *A counting renaissance: Combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection*. Bioinformatics.

65. Pond SK, Muse SV (2005) *Site-to-site variation of synonymous substitution rates*. Mol Biol Evol 22: 2375-2385.

66. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) *Detecting individual sites subject to episodic diversifying selection*. PLoS Genet 8: e1002764.

67. Wertheim JO, Worobey M (2009) *Relaxed selection and the evolution of RNA virus mucin-like pathogenicity factors*. J Virol 83: 4690-4694.

68. Zimmer G, Trotz I, Herrler G (2001) *N-glycans of F protein differentially affect fusion activity of human respiratory syncytial virus*. J Virol 75: 4744-4751.

69. Bermingham A, Collins PL (1999) *The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription*. Proc Natl Acad Sci U S A 96: 11259-11264.

70. Cheng X, Park H, Zhou H, Jin H (2005) *Overexpression of the M2-2 protein of respiratory syncytial virus inhibits viral replication*. J Virol 79: 13943-13952.

71. Ahmadian G, Randhawa JS, Easton AJ (2000) *Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism*. EMBO J 19: 2681-2689.

72. Woelk CH, Holmes EC (2001) *Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV)*. J Mol Evol 52: 182-192.

73. Spann KM, Collins PL, Teng MN (2003) *Genetic recombination during coinfection of two mutants of human respiratory syncytial virus*. J Virol 77: 11201-11211.

74. Gray RR, Pybus OG, Salemi M (2011) *Measuring the Temporal Structure in Serially-Sampled Phylogenies*. Methods Ecol Evol 2: 437-445.

SUPPLEMENTARY INFO

CHAPTER 3

Table S1. Patient and RSV-B strain information

Code	Gender	Age (months)*	Sample date	Patientmaterial	Genotype	Accession #
12-004620	F	1.7	05-02-12	N Asp	GB13	JX576733
12-003981	M	951.0	29-01-12	N swab	GB13	JX576734
12-002874	F	590.8	23-01-12	N swab	GB13	JX576735
12-001238	F	900.2	07-01-12	N swab	GB13	JX576736
10-053640	M	3.2	21-12-10	N Asp	GB13	JX576737
10-053210	F	8.4	12-12-10	N Asp	GB13	JX576738
09-053165	M	5.1	28-12-09	N Asp	GB13	JX576739
09-053157	M	20.2	28-12-09	N Asp	GB13	JX576740
09-052617	M	1.5	22-12-09	N Asp	GB13	JX576741
09-052281	F	0.8	20-12-09	N Asp	GB13	JX576742
09-052064	M	12.5	18-12-09	N Asp	GB13	JX576743
08-046532	F	4.2	22-12-08	N Asp	GB13	JX576744
08-045952	M	8.1	17-12-08	N Asp	GB13	JX576745
08-045604	M	13.6	15-12-08	N Asp	GB13	JX576746
08-045420	M	6.9	12-12-08	N Asp	GB13	JX576747
07-038682	M	5.8	27-11-07	N Asp	GB13	JX576748
07-038176	F	2.6	23-11-07	N Asp	GB13	JX576749
07-008482	M	30.4	14-03-07	BAL	GB13	JX576750
07-000431	M	0.8	05-01-07	N Asp	GB13	JX576751
06-039982	F	6.5	21-12-06	N Asp	GB13	JX576752
06-036027	M	12.1	17-11-06	N Asp	GB13	JX576753
05-040399	F	1.0	31-12-05	N Asp	GB13	JX576754
05-040058	M	11.8	26-12-05	N Asp	GB13	JX576755
05-039452	M	41.6	20-12-05	N Asp	GB13	JX576756
05-004042	F	333.4	14-01-05	N swab	GB13	JX576757
05-001965	F	657.5	14-01-05	N swab	GB13	JX576758
03-034613	M	8.6	05-12-03	N Asp	GB13	JX576759
03-000005	F	6.2	01-01-03	N Asp	**GB12**	JX576760
02-028215	M	0.4	21-10-02	N Asp	GB13	JX576761
02-000467	F	22.6	08-01-02	N Asp	GB13	JX576762
BE/5133/08	F	8.0	23-10-08	N Asp	GB13	JX576729
BE/5150/08	F	4.0	24-10-08	N Asp	GB13	JX576730
BE/5649/08	M	0.5	19-11-08	N Asp	GB13	JX576731
BE/7004/06	F	7.0	05-12-06	N Asp	GB13	JX576732

N Asp = Nasopharyngeal Aspirates, N swab = Nose-throat swab, BAL = Bronchoalveolar lavage and * Age at time of sampling.

Table S2. Primers applied for cDNA preparation and whole genome sequencing

General primers for RSV-B whole genome sequencing			
Primer name	5'-3' sequence	Start	End
RSV B F1 For	ACGCGAAAAAATGCGTACTAC	1	21
RSV B F1 Rev	GAGCCATCTTTGTATTTGCC	1147	1128
RSV B F2 For	CAATCATGGCGGGTTTCTAG	916	935
RSV B F2 Rev	GCTCTGCATATGCTTTGGC	2206	2188
RSV B F3 For	GCACAATCATCCACAAGAGG	1825	1844
RSV B F3 Rev	GTTGATTGATTGGTTGGTGG	3195	3175
RSV B F4 For	GATAGGTTAGAGGCTATGGCAAG	2916	2938
RSV B F4 Rev	GCGGTATAAATGGTTGTTGAG	4273	4253
RSV B F5 For	GGTGCCTACCTAGAAAAAGAGAG	3942	3964
RSV B F5 Rev	GGTCTCTTTTGTTTGTGGTTTTG	5337	5314
RSV B F6 For	GTCTCACCAGAAAGGGTTAGC	4970	4990
RSV B F6 Rev	CTTTGGATACAGCTATACCACTTG	6195	6172
RSV B F7 For	CCAAATGCAATGGAACTGAC	5926	5945
RSV B F7 Rev	CAGAAGGAAACACTAGAGGGTC	7182	7161
RSV B F8 For	CTCTTGGAGCTATAGTGTCATGC	6952	6974
RSV B F8 Rev	GATGGATGGTTTGCTTGTTG	8121	8102
RSV B F9 For	GCTCTTGGTATAGTTGGAGTGC	7896	7917
RSV B F9 Rev	GGAGGATGTTGCATTGAACAC	9172	9152
RSV B F10 For	GTAAAGGTGTACGCCATCTTG	8934	8954
RSV B F10 Rev	GCAGATGAAACTCACGATAGAAC	10152	10130
RSV B F11 For	GATGAAAGACAAGCAATGGATG	9873	9894
RSV B F11 Rev	GTCTGACCCTCTATAAGTCTAACTGG	11053	11028
RSV B F12 For	GTGGTATTGAGGGCTGGTG	10894	10912
RSV B F12 Rev	AATACTTGGCGATGTTACTCCTAC	12242	12219
RSV B F13 For	CAAAATATAGAACCAACTTACCCTC	11904	11928
RSV B F13 Rev	AATCAGAATCCAATGTCCAGC	13079	13059
RSV B F14 For	AAAACAGCATATGTTCCTACCAG	12893	12915
RSV B F14 Rev	TCTGGATGAAGTTCTACTACCGTAC	14179	14155
RSV B F15 For	CACCACCACTTCACATCAGAC	13949	13969
RSV B F15 Rev	ACGAGAAAAAAGTGTCAAAAACT	15288	15265
RSV B F1 Begin3 (rev)	GTTATCAGGGCACACTTCAC	286	305
RSV B Seq End (for)	TACCTTTCCTTTGTTACCCTATAAC	15113	15089

Table S2. Primers applied for cDNA preparation and whole genome sequencing (continued)

Additional primers for RSV-B whole genome sequencing			
Primer name	5'-3' sequence	Start	End
RSV B F3a For	GAGGGGGTAGTAGAGTTGAAGG	1873	1894
RSV B F5a For	GTGTTCTCCCATTATGCTGTG	4500	4520
RSV B F5 A Rev	GTTGCCACATATACTACAGGGAAC	5221	5200
RSV B F8a For	CACACCAGTTACACTAAGCAAAGAC	7327	7351
RSV B F8a Rev	TGTCTTCTTCAGCACGTCTGC	8097	8077
RSV B F11a For	AAAGCAGGAATAAGCAACAAGTC	10528	10550
RSV B F11a Rev	CACCACCACCAAACAGCATAG	11473	11453

Nucleotide position relative to consensus RSV B [AF013254.1].

Attached to the 5' end of the forward primer is the following sequence primer site: GCA GTC AAC CAT ATG CCT TG

Attached to the 5' end of the reverse primer is the following sequence primer site: GCT TGC GGA ATT CTG ACA AC

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*.

Strain	Accesion No	Year of isolation	Country of isolation	Reference
BA3833_1999.50	AY333362	1999	Buenos Aires (Argentina)	Trento et al. (2003) J Gen Virol
BA3859_1999.50	AY333363	1999	Buenos Aires (Argentina)	Trento et al. (2003) J Gen Virol
BE10495_1993.50	AY751215	1993	Belgium	Zlateva et al. (2005) J Virol
BE10557_2000.50	AY751266	2000	Belgium	Zlateva et al. (2005) J Virol
BE1066_2003.50	AY751123	2003	Belgium	Zlateva et al. (2005) J Virol
BE10701_2000.50	AY751163	2000	Belgium	Zlateva et al. (2005) J Virol
BE10760_2000.50	AY751267	2000	Belgium	Zlateva et al. (2005) J Virol
BE1091_1996.50	AY751222	1996	Belgium	Zlateva et al. (2005) J Virol
BE11120_1993.50	AY751230	1993	Belgium	Zlateva et al. (2005) J Virol
BE1134_2003.50	AY751159	2003	Belgium	Zlateva et al. (2005) J Virol
BE11500_2001.50	AY751105	2001	Belgium	Zlateva et al. (2005) J Virol
BE11508_2001.50	AY751116	2001	Belgium	Zlateva et al. (2005) J Virol
BE11535_2001.50	AY751126	2001	Belgium	Zlateva et al. (2005) J Virol
BE11583_2001.50	AY751201	2001	Belgium	Zlateva et al. (2005) J Virol
BE11609_2001.50	AY751102	2001	Belgium	Zlateva et al. (2005) J Virol
BE1162_2002.50	AY751104	2002	Belgium	Zlateva et al. (2005) J Virol
BE11712_1993.50	AY751244	1993	Belgium	Zlateva et al. (2005) J Virol
BE11718_1996.50	AY751197	1996	Belgium	Zlateva et al. (2005) J Virol

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*. (continued)

Strain	Accesion No	Year of isolation	Country of isolation	Reference
BE11754_2000.50	AY751268	2000	Belgium	Zlateva et al. (2005) J Virol
BE1176_1988.50	AY751256	1988	Belgium	Zlateva et al. (2005) J Virol
BE11842_2001.50	AY751206	2001	Belgium	Zlateva et al. (2005) J Virol
BE11875_2001.50	AY751196	2001	Belgium	Zlateva et al. (2005) J Virol
BE11900_1996.50	AY751239	1996	Belgium	Zlateva et al. (2005) J Virol
BE11972_1996.50	AY751226	1996	Belgium	Zlateva et al. (2005) J Virol
BE12015_1996.50	AY751132	1996	Belgium	Zlateva et al. (2005) J Virol
BE12024_2000.50	AY751265	2000	Belgium	Zlateva et al. (2005) J Virol
BE12031_1999.50	AY751270	1999	Belgium	Zlateva et al. (2005) J Virol
BE12059_1999.50	AY751223	1999	Belgium	Zlateva et al. (2005) J Virol
BE12160_2002.50	AY751271	2002	Belgium	Zlateva et al. (2005) J Virol
BE12186_1996.50	AY751225	1996	Belgium	Zlateva et al. (2005) J Virol
BE12228_1998.50	AY751246	1998	Belgium	Zlateva et al. (2005) J Virol
BE12237_2002.50	AY751175	2002	Belgium	Zlateva et al. (2005) J Virol
BE12252_1996.50	AY751241	1996	Belgium	Zlateva et al. (2005) J Virol
BE12308_1996.50	AY751272	1996	Belgium	Zlateva et al. (2005) J Virol
BE12321_1998.50	AY751134	1998	Belgium	Zlateva et al. (2005) J Virol
BE12349_2002.50	AY751121	2002	Belgium	Zlateva et al. (2005) J Virol
BE12350_2002.50	AY751185	2002	Belgium	Zlateva et al. (2005) J Virol
BE12353_2001.50	AY751162	2001	Belgium	Zlateva et al. (2005) J Virol
BE12357_1998.50	AY751182	1998	Belgium	Zlateva et al. (2005) J Virol
BE12358_2002.50	AY751084	2002	Belgium	Zlateva et al. (2005) J Virol
BE12370_2001.50	AY751118	2001	Belgium	Zlateva et al. (2005) J Virol
BE12445_1999.50	AY751094	1999	Belgium	Zlateva et al. (2005) J Virol
BE12446_2001.50	AY751125	2001	Belgium	Zlateva et al. (2005) J Virol
BE12522_2001.50	AY751119	2001	Belgium	Zlateva et al. (2005) J Virol
BE12543_1996.50	AY751157	1996	Belgium	Zlateva et al. (2005) J Virol
BE12562_1999.50	AY751213	1999	Belgium	Zlateva et al. (2005) J Virol
BE12595_2001.50	AY751107	2001	Belgium	Zlateva et al. (2005) J Virol
BE12598_2001.50	AY751108	2001	Belgium	Zlateva et al. (2005) J Virol
BE12670_2001.50	AY751086	2001	Belgium	Zlateva et al. (2005) J Virol
BE128_1992.50	AY751248	1992	Belgium	Zlateva et al. (2005) J Virol
BE12817_2003.50	AY751091	2003	Belgium	Zlateva et al. (2005) J Virol
BE12894_1999.50	AY751150	1999	Belgium	Zlateva et al. (2005) J Virol

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*. (continued)

Strain	Accesion No	Year of isolation	Country of isolation	Reference
BE12973_2003.50	AY751089	2003	Belgium	Zlateva et al. (2005) J Virol
BE13159_2002.50	AY751117	2002	Belgium	Zlateva et al. (2005) J Virol
BE13182_1999.50	AY751210	1999	Belgium	Zlateva et al. (2005) J Virol
BE13236_1995.50	AY751221	1995	Belgium	Zlateva et al. (2005) J Virol
BE1334_1996.50	AY751218	1996	Belgium	Zlateva et al. (2005) J Virol
BE13350_1999.50	AY751207	1999	Belgium	Zlateva et al. (2005) J Virol
BE13415_2002.50	AY751174	2002	Belgium	Zlateva et al. (2005) J Virol
BE13417_1999.50	AY751131	1999	Belgium	Zlateva et al. (2005) J Virol
BE13418_2002.50	AY751172	2002	Belgium	Zlateva et al. (2005) J Virol
BE13436_2002.50	AY751151	2002	Belgium	Zlateva et al. (2005) J Virol
BE13457_2003.50	AY751087	2003	Belgium	Zlateva et al. (2005) J Virol
BE13491_2002.50	AY751165	2002	Belgium	Zlateva et al. (2005) J Virol
BE13768_1997.50	AY751198	1997	Belgium	Zlateva et al. (2005) J Virol
BE13833_1997.50	AY751154	1997	Belgium	Zlateva et al. (2005) J Virol
BE14273_1995.50	AY751227	1995	Belgium	Zlateva et al. (2005) J Virol
BE143_1994.50	AY751233	1994	Belgium	Zlateva et al. (2005) J Virol
BE14361_1998.50	AY751181	1998	Belgium	Zlateva et al. (2005) J Virol
BE14364_1998.50	AY751245	1998	Belgium	Zlateva et al. (2005) J Virol
BE14516_1997.50	AY751192	1997	Belgium	Zlateva et al. (2005) J Virol
BE14517_1997.50	AY751147	1997	Belgium	Zlateva et al. (2005) J Virol
BE14548_1998.50	AY751140	1998	Belgium	Zlateva et al. (2005) J Virol
BE14610_2003.50	AY751093	2003	Belgium	Zlateva et al. (2005) J Virol
BE14819_1997.50	AY751152	1997	Belgium	Zlateva et al. (2005) J Virol
BE15_1994.50	AY751231	1994	Belgium	Zlateva et al. (2005) J Virol
BE1509_1996.50	AY751190	1996	Belgium	Zlateva et al. (2005) J Virol
BE15116_1998.50	AY751148	1998	Belgium	Zlateva et al. (2005) J Virol
BE15760_1997.50	AY751158	1997	Belgium	Zlateva et al. (2005) J Virol
BE1732_1982.50	AY751260	1982	Belgium	Zlateva et al. (2005) J Virol
BE1802_2003.50	AY751109	2003	Belgium	Zlateva et al. (2005) J Virol
BE1932_1984.50	AY751281	1984	Belgium	Zlateva et al. (2005) J Virol
BE1950_2001.50	AY751149	2001	Belgium	Zlateva et al. (2005) J Virol
BE196_2001.50	AY751212	2001	Belgium	Zlateva et al. (2005) J Virol
BE2003_1984.50	AY751276	1984	Belgium	Zlateva et al. (2005) J Virol
BE2098_2000.50	AY751238	2000	Belgium	Zlateva et al. (2005) J Virol

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*. (continued)

Strain	Accesion No	Year of isolation	Country of isolation	Reference
BE210_2003.50	AY751111	2003	Belgium	Zlateva et al. (2005) J Virol
BE23_1991.50	AY751216	1991	Belgium	Zlateva et al. (2005) J Virol
BE259_1996.50	AY751187	1996	Belgium	Zlateva et al. (2005) J Virol
BE2756_1985.50	AY751282	1985	Belgium	Zlateva et al. (2005) J Virol
BE277_1996.50	AY751195	1996	Belgium	Zlateva et al. (2005) J Virol
BE2803_1998.50	AY751191	1998	Belgium	Zlateva et al. (2005) J Virol
BE283_2000.50	AY751242	2000	Belgium	Zlateva et al. (2005) J Virol
BE2968_1985.50	AY751257	1985	Belgium	Zlateva et al. (2005) J Virol
BE302_2004.50	AY751124	2004	Belgium	Zlateva et al. (2005) J Virol
BE3280_1987.50	AY751251	1987	Belgium	Zlateva et al. (2005) J Virol
BE3455_1986.50	AY751259	1986	Belgium	Zlateva et al. (2005) J Virol
BE352_2000.50	AY751208	2000	Belgium	Zlateva et al. (2005) J Virol
BE361_1994.50	AY751235	1994	Belgium	Zlateva et al. (2005) J Virol
BE381_2001.50	AY751211	2001	Belgium	Zlateva et al. (2005) J Virol
BE382_2000.50	AY751145	2000	Belgium	Zlateva et al. (2005) J Virol
BE3837_2002.50	AY751186	2002	Belgium	Zlateva et al. (2005) J Virol
BE391_1989.50	AY751254	1989	Belgium	Zlateva et al. (2005) J Virol
BE395_1985.50	AY751280	1985	Belgium	Zlateva et al. (2005) J Virol
BE4383_1989.50	AY751253	1989	Belgium	Zlateva et al. (2005) J Virol
BE45_1991.50	AY751274	1991	Belgium	Zlateva et al. (2005) J Virol
BE450_1983.50	AY751262	1983	Belgium	Zlateva et al. (2005) J Virol
BE456_1987.50	AY751252	1987	Belgium	Zlateva et al. (2005) J Virol
BE46_2003.50	AY751110	2003	Belgium	Zlateva et al. (2005) J Virol
BE4618_1988.50	AY751255	1988	Belgium	Zlateva et al. (2005) J Virol
BE4929_1988.50	AY751250	1988	Belgium	Zlateva et al. (2005) J Virol
BE5222_1988.50	AY751249	1988	Belgium	Zlateva et al. (2005) J Virol
BE552_1993.50	AY751236	1993	Belgium	Zlateva et al. (2005) J Virol
BE552_2003.50	AY751173	2003	Belgium	Zlateva et al. (2005) J Virol
BE553_2003.50	AY751122	2003	Belgium	Zlateva et al. (2005) J Virol
BE612_1996.50	AY751228	1996	Belgium	Zlateva et al. (2005) J Virol
BE618_2000.50	AY751237	2000	Belgium	Zlateva et al. (2005) J Virol
BE705_2001.50	AY751180	2001	Belgium	Zlateva et al. (2005) J Virol
BE7374_1992.50	AY751243	1992	Belgium	Zlateva et al. (2005) J Virol
BE860_1990.50	AY751264	1990	Belgium	Zlateva et al. (2005) J Virol

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*. (continued)

Strain	Accesion No	Year of isolation	Country of isolation	Reference
BE90_1992.50	AY751247	1992	Belgium	Zlateva et al. (2005) J Virol
BE91_2003.50	AY751161	2003	Belgium	Zlateva et al. (2005) J Virol
BE922_2003.50	AY751176	2003	Belgium	Zlateva et al. (2005) J Virol
BE946_1987.50	AY751258	1987	Belgium	Zlateva et al. (2005) J Virol
BE96_1996.50	AY751219	1996	Belgium	Zlateva et al. (2005) J Virol
BE975_2002.50	AY751269	2002	Belgium	Zlateva et al. (2005) J Virol
CH10b_1990.50	AF065250	1990	New York (USA)	Peret et al. (1998) J Gen Virol
CH18537_1962.50	M17213	1962	District of Colombia	Johnson et al. (1987) PNAS
CH18b_1993.50	AF065252	1993	New York (USA)	Peret et al. (1998) J Gen Virol
CH53b_1993.50	AF065253	1993	New York (USA)	Peret et al. (1998) J Gen Virol
CH9b_1993.50	AF065251	1993	New York (USA)	Peret et al. (1998) J Gen Virol
HQ699287_2010.21	HQ699287	2010	South Korea	Baek et al. (2012) Arch Virol
HQ699288_2010.21	HQ699288	2010	South Korea	Baek et al. (2012) Arch Virol
HQ699289_2010.29	HQ699289	2010	South Korea	Baek et al. (2012) Arch Virol
HQ699290_2010.29	HQ699290	2010	South Korea	Baek et al. (2012) Arch Virol
HQ699291_2008.87	HQ699291	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699292_2008.87	HQ699292	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699293_2008.87	HQ699293	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699294_2008.96	HQ699294	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699295_2008.96	HQ699295	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699296_2009.04	HQ699296	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699297_2009.79	HQ699297	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699298_2008.46	HQ699298	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699299_2008.71	HQ699299	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699300_2009.79	HQ699300	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699301_2008.71	HQ699301	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699302_2009.87	HQ699302	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699303_2009.87	HQ699303	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699304_2009.87	HQ699304	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699305_2009.87	HQ699305	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699306_2009.87	HQ699306	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699307_2009.96	HQ699307	2009	South Korea	Baek et al. (2012) Arch Virol
HQ699308_2008.79	HQ699308	2008	South Korea	Baek et al. (2012) Arch Virol
HQ699309_2010.04	HQ699309	2010	South Korea	Baek et al. (2012) Arch Virol

Table S3. Strain information on G gene sequences derived from the studies of Zlateva *et al*. and Baek *et al*. (continued)

Strain	Accesion No	Year of isolation	Country of isolation	Reference
HQ699310_2008.50	HQ699310	2008	South Korea	Baek et al. (2012) Arch Virol
MON15_1990.50	AY333361	1990	Montevideo (Uruguay)	Trento et al. (2003) J Gen Virol
NM1355_1989.50	M73543	1989	New Mexico (USA)	Sullender et al. (1991) J Virol
Sap4_2000.50	AB117522	2000	Sapporo (Japan)	Nagai et al. (2004) J Med Virol
SW8_1960.50	M73545	1960	Sweden	Sullender et al. (1991) J Virol
WV10010_1983.50	M73541	1983	West Virginia (USA)	Sullender et al. (1991) J Virol
WV15291_1985.50	M73542	1985	West Virginia (USA)	Sullender et al. (1991) J Virol
WV4843_1980.50	M73540	1980	West Virginia (USA)	Sullender et al. (1991) J Virol
WVB1_1985.50	AF013254	1985	West Virginia (USA)	Karron et al. (1997) PNAS

CHAPTER FOUR

Human antibodies specific for the RSV F prefusion protein are prevalent *in vivo* and efficiently reduce respiratory tract infection by RSV

Michaël V. Lukens[4*], Lydia Tan[1*], Etsuko Yasuda[2*], Arjen Q. Bakker[2], Myra N. Widjojoatmodjo[4], Frank E. Coenjaerts[1], Hergen Spits[2], Grada M. van Bleek[3] and Tim Beaumont[2]

[1]Department of Virology, University Medical Center, Utrecht

[2]AIMM Therapeutics, Academic Medical Center, Amsterdam

[3]Department of Pediatrics, The Wilhelmina Children's Hospital

[4] Department of Vaccine Research, National Institute for Public Health and the Environment, RIVM, The Netherlands

* Authors contributed equally to the paper

Submitted

Abstract

Severe lower respiratory tract infections caused by Respiratory Syncytial Virus (RSV) form a major disease burden in young infants. Currently a vaccine is not available and high-risk infants can only receive prophylactic treatment with Palivizumab, a humanized mouse monoclonal antibody recognizing the fusion (F) protein of RSV. Using a novel method to immortalize antibody-producing B cells, we studied the total RSV memory B cell response in three healthy donors and found that approximately 50% of the prefusion specific antibodies are neutralizing. In addition we selected a series of broadly reacting highly neutralizing antibodies of the IgG1 subclass. Most antibodies recognize non-linear epitopes on the RSV F prefusion protein and these antibodies are more potent in reducing viral load, pathology and the cellular influx in mouse and cotton rat models than antibodies, such as Palivizumab, that recognize epitopes also expressed in the post-fusion structure of RSV-F. One of these antibodies, D25, is superior to the others and therefore holds great promise for treatment in RSV disease.

Introduction

RSV infection is the main cause of lower respiratory tract disease in children [1]. Hall *et al*. estimated the total hospitalization rate for RSV-related respiratory tract infections in the U.S.A. at 17 cases per 1000 infants under six months and three cases per 1000 children younger than 5 years [2]. Especially premature neonates and children with bronchopulmonary dysplasia or congenital heart disease are at risk for pathological consequences of RSV infection [3-5]. Previously it has been observed that there is an association between maternally derived RSV neutralizing IgGs and protection in infants younger than 6 months of age [6-8]. Based on this survey, a humanized monoclonal antibody specific for the RSV F protein (Palivizumab) is prophylactically administered to children at risk [9,10]. The observations that maternally derived antibodies, treatment with RespiGam (a polyclonal mixture of human IgG) or treatment with Palivizumab are protective, indicate that prophylaxis with a good neutralizing and cost-effective antibody against RSV is a suitable method to protect high-risk patients [7,9,11-13]. This is important since lower respiratory tract infection early in life, especially in high-risk infants, is associated with wheezing and asthma during infancy. RSV, but also other viral etiologies may cause or enhance airway sensitization that could result in hyper-responsive airways and asthma [14-19]. Furthermore, RSV induced disease may also be more severe in predisposed individuals. Debate about a causal relation and mechanisms by which lower respiratory tract infection cause asthma is still ongoing [15,20-23]. Prophylactic treatment of children may therefore result in a significant reduction in wheezing days during the first year of life [24,25]. The high-dose requirements of Palivizumab are less convenient for this approach and the search for alternative more potent neutralizing antibodies continues.

We previously described the isolation and immortalization of human memory B cells from peripheral blood by introducing the transcriptional repressor BCL6 and the anti-apoptotic gene Bcl-xL. These cells resemble germinal center B cells and produce relatively high amounts of antigen specific immunoglobulin while still expressing the B cell receptor (BCR) [26-29]. In this study we isolated a series of B cell clones from a healthy individual and show that several of these B cell clones produced RSV specific IgGs directed against the RSV fusion protein, which possess

high neutralizing activity *in vitro* and *in vivo*. The majority of the highly neutralizing human RSV specific antibodies solely recognize the RSV F prefusion protein. Furthermore, we determined that approximately 20% of the memory RSV specific B cell response is directed against prefusion F structures but only 50% of these antibodies are neutralizing, suggesting that prefusion specificity is not a prerequisite for neutralization. During *in vitro* analyses, these antibodies neutralize a broad selection of subgroup A and subgroup B RSV clinical isolates. Neutralization efficacy of the IgG1 antibodies in a mouse and cotton rat model was higher compared to Palivizumab. In addition, the reduced viral load resulted in reduced innate- and RSV specific adaptive immune responses, which prevented severe immune pathology.

Materials and Methods

Viruses and cells

RSV reference strains, A2 (VR-1540, ATCC, subgroup A), RSV-X (GenBank FJ948820.1, subgroup A) of isolate 98-25147-X, [30] and 9320 (VR-955, ATCC, subgroup B) and RSV subgroup A and B clinical isolates (Table S1) obtained from RSV infected patients [31,32] were grown on HEp-2 cells in DMEM (Dulbecco's Minimal Essential Medium; Gibco) with 5% FCS and 100 µg/ml Normocin. For animal experiments, RSV A2 was cultured in IMDM (Iscove's Modified Dulbecco's Medium; Gibco) supplemented with 5% FCS, penicillin/streptomycin, purified by PEG-6000 precipitation and stored in liquid nitrogen. Wild type and recombinant vesicular stomatitis virus (VSV) expressing RSV-G protein (VSV-G) and VSV expressing RSV F protein (VSV-F) (kindly provided by Dr. J.S. Kahn and Prof. J.K Rose, Yale University School of Medicine) were cultured on BHK cells in DMEM containing 5% FCS, penicillin/streptomycin and 50 µM 2-mercapto-ethanol [33]. Culture supernatant of infected BHK cells was used to infect EL-4 cells that were cultured in IMDM supplemented with 5% FCS, penicillin/streptomycin and 50 µM 2-mercapto-ethanol. The murine H-2^b dendritic cell line D1 was cultured in IMDM with 10% FCS, penicillin/streptomycin, 50 µM 2-mercapto-ethanol and GM-CSF [34].

Stable cell line expressing protein G membrane form of RSV-X strain

Vero cells were transfected with a pcDNA3-vector (Invitrogen) expressing the

membrane bound form of RSV-X G behind a chicken βactin promoter. Three days after transfection Vero cells were harvested, stained with a mouse monoclonal antibody against RSV-G (Mab 858-2, Chemicon), followed by staining with a FITC conjugated goat anti mouse IgG. RSV G positive cells were sorted on a FACSVantage cell sorter (Becton Dickinson) and collected in a 96 round-bottom plate at one cell per well in 100 µl of medium. After two-to-three weeks single colonies were counted (>50) by visual microscopy and checked for G expression. In parallel, single colonies were harvested and grown in complete medium with neomycine. To ensure clonality, single colonies were subjected to an additional round of FACS sorting and culture.

B cell culture and immortalization

B cells were immortalized and cultured as described before [27-29,35]. In brief, we isolated B cells from peripheral blood by Ficoll separation, CD22 MACS microbeads (Miltenyi Biotech) and subsequently cell sorting for CD19+CD3–CD27+IgM–IgA– (IgG memory cells) on a FACSAria (BD). The use of peripheral blood samples was approved by the medical ethical committees of the institution and was contingent on informed consent. Retroviral transduced B cells, were maintained at $2x10^5$ cells/ml in IMDM supplemented with recombinant mouse IL-21 (25 ng/ml, R&D systems) and co-cultured with g-irradiated (50Gy) mouse L cell fibroblasts stably expressing CD40L (CD40L-L cells, 10^5 cells/ml) for 36 hrs. The BCL6 and Bcl-xL retroviral constructs were described previously [26,36] and were cloned into the LZRS retroviral vector and transfected in Phoenix packaging cells as described before and added to the stimulated B cells [26,27,35]. Transduced B cells were maintained in IMDM, in the presence of recombinant IL-21 and CD40L-L cells for prolonged periods of time and culture supernatants containing antibodies were tested for anti-RSV activity. To analyze BCR specificity, B cell clones of AM14, AM16, AM22, AM23 and D25 were incubated with non-purified culture supernatant of recombinant His-tagged RSV F protein (subtype Long strain [37]) expressed in 293T cells. Binding of the antigen to the BCR was detected with an anti-penta His Alexa fluor 647 antibody (Qiagen) by FACS analysis.

Cloning of selected antibodies

To obtain antibody heavy (VH) and light (VL) chain sequences for subsequent recombinant protein expression, total RNA was isolated with the RNeasy® mini kit (Qiagen), cDNA was generated using Superscript 3 (Invitrogen) and a Variable Heavy (VH) and Light (VL) chain PCR were performed. VH and VL variable regions were cloned into the pCR2.1 TA cloning vector (Invitrogen) [35]. To rule out reverse transcriptase or DNA polymerase induced mutations, several independent cloning experiments were performed. To produce recombinant antibodies the VH and VL regions were cloned in frame with human IgG1 and Kappa constant regions into a pcDNA3.1 (Invitrogen) based vector. For protein production the vectors were transiently transfected into 293T cells. IgG1 antibodies were subsequently purified from 293T culture supernatant using HiTrap Protein A or G columns on an ÄKTA instrument (GE healthcare).

Animal models

Pathogen-free 6 week-old female C57BL/6cjo mice were purchased from Charles River Nederland (Maastricht, The Netherlands) and pathogen-free 6 week-old female Balb/c mice were purchased from Janvier (Le Genest-St-Isle, France). Mice were lightly anesthetized with isoflurane and intranasally infected with 10^7 $TCID_{50}$ RSV A2 in a volume of 50 µl diluted in PBS with 10% sucrose. One day (day -1) before infection mice were intraperitoneal (i.p.) injected either with 10, 1 or 0.1 mg/kg Palivizumab (MedImmune, Gaithersburg, MD) dissolved in PBS or with 1, 0.1 or 0.01 mg/kg D25 (AIMM Therapeutics, Amsterdam, The Netherlands) or with 1, 0.1 mg/kg hIgG1 isotype control (Eureka therapeutics, Emeryville, CA). Five days post infection mice were sacrificed by i.p. injection of pentobarbital. For determination of the viral load, the lungs were removed after performing Bronchial Alveolar Lavage (BAL) and lung tissue was homogenized in 1 ml of IMDM and then centrifuged at 1,000 x g for 10 min. For measurement of T cell responses, the lungs were cut in pieces and the fragments were digested for 25 min at 37°C and 5% CO_2 with 3.2 mg/ml collagenase A and 1 mg/ml DNAse (Roche Applied Science, Basel, Switzerland). For the last 5 min 1 mM EDTA was added. Single-cell suspensions were prepared from the pre-treated lungs by processing the tissue trough cell strainers (BD Falcon, Franklin Lakes, NJ). The mouse study protocol was approved by the Animal Ethics

Committee of the University Medical Center Utrecht and the Utrecht University.

Pathogen free 7–9 week-old Cotton rats (*Sigmodon hispidus*, Charles River Nederland) were anesthetized with isoflurane and given 0.1 ml of purified antibody by intramuscular (i.m.) injection at doses of 2.0 or 0.4 mg/kg for the control antibody, Palivizumab, AM22 and D25, while AM14 was administered at 0.4 and 0.1 mg/kg. Twenty-four hrs later, animals were anesthetized, bled for serum hIgG determination and challenged by intranasal instillation of 10^6 $TCID_{50}$ RSV-X (100 µl). Five days later animals were sacrificed and their lungs were harvested. Lung virus titers were determined using standard TCID50 culture for which the lowest limit of detection (LD) was 2.1 $\log_{10}$ $TCID_{50}$ /g. The Animal Experiment Committee of the Netherlands Vaccine Institute approved all procedures involving cotton rats.

RSV viral load and virus neutralization analysis

Plaque assays were performed in 24-well culture plates (Corning Costar) to determine the neutralization efficacy of antibodies *in vitro*. Three-step serial antibody dilutions, prepared in DMEM containing 2% FCS and 100 µg/ml Normocin, were pre-incubated for two hrs with 60 pfu virus. These virus-antibody mixtures were subsequently applied to Hep2 cells that were seeded $8x10^4$ cells/well 48 hrs before infection. After two hrs of viral adherence at 37°C and 5% CO_2 the Hep2 monolayers were covered by a 0.5% Seakam LE agarose overlay. Cells were fixed with 10% formalin on day 5 post-infection and plaque formation was visualized using 0.5% crystal violet/ 20% ethanol (w/v) staining solution. Plaques were counted for each antibody dilution and neutralizing titers, responsible for 50% plaque-reduction, were calculated from the plaque-reduction curves.

The *in vivo* infectious viral load present in lungs and BAL was determined by standard $TCID_{50}$ culture analysis. Serial dilutions of the clarified lung supernatants or BAL fluid obtained 5 days after challenge were cultured on HEp-2 cells in flat bottom 96 well plates (Costar, Netherlands) with IMDM, containing penicillin/streptomycin and 1% FCS.

Antibody neutralization of RSV *in vitro* was performed by using a method similar to the method described by Zielinska et al. [38]. For this, 100 $TCID_{50}$ of RSV A2 was mixed and pre-incubated with B cell culture supernatants for 30 min at 37°C and added in

triplicate to HEp-2 cells. After 96 hrs of culture at 37°C with 5% CO_2, the monolayers were fixed with 80% acetone followed by a washing step with PBS/0.05% Tween 20. Virus plaques were visualized by incubating the cells for 2 hrs at 37°C with polyclonal goat anti-RSV HRP (Biodesign, Kennebunk, ME, USA), and subsequent addition of the 3-amino-9-ethylcarbazole (AEC) substrate. Plaques were counted under a light microscope.

RSV specific RT-PCR

Total viral load present in mouse lungs and BAL was analyzed by quantitative RT-PCR. Viral RNA was isolated using the RNeasy mini kit (Qiagen, Valencia, CA) according to the manufacturer instructions. Then cDNA was prepared by reverse transcriptase of the viral RNA using MultiScribe RT and 2.5 µM random hexamers (Applied Biosystems, Foster City, CA) for 5 min at 25°C, 30 min at 48°C and 10 min at 95°C. Real-time PCR analysis based on the amplification and detection of the RSV N gene was performed using an ABI Prism 7700 (Applied Biosystems, Foster City, CA) with the following run conditions; 2 min at 50°C, 10 min at 95°C, 45 cycles of 15 seconds at 95°C and 1 min at 60°C. Samples were assayed in a 25 µl reaction mixture containing 10 µl cDNA, PCR master mix, RSV-A primers (forward: AGA TCA ACT TCT GTC ATC CAG CAA and reverse TTC TGC ACA TCA TAA TTA GGA GTA TCA AT) and FAM-labeled probe (CAC CAT CCA ACG GAG CAC AGG AGA T) as described in Tan et al, 2012 [31]. Known concentrations of RSV A2 were used to form a standard RT-PCR cycli-curve for the calculation of viral RNA copies/ml in samples with unknown viral load. Standards and negative controls were run together with each PCR.

Intra-cellular Cytokine Staining (ICS) and flow cytometry

ICS analysis was conducted to study RSV specific T cell responses in mouse lungs. Lung single-cell suspensions (10^6 cells) were stimulated for 6 hours at 37°C, 5% CO_2 either with 1 µg/ml $M_{187\text{-}195}$ peptide (NAITNAKII) [39,40] or with $2x10^5$ RSV A2 infected D1 cells (multiplicity of infection (MOI) of 2) in 200 µl IMDM containing 10% FCS, penicillin/streptomycin and 50 µM 2-mercapto-ethanol and 25 U/ml recombinant human IL-2 (Roche Applied Science, Basel, Switzerland). Brefeldin-A (10 µg/ml; Sigma, St. Louis, Mo.) was added throughout the stimulation to facilitate intracellular accumulation of cytokines. After cytokine stimulation, the cells were

washed with PBS containing 2% FCS, 2 mM EDTA and 0.02% NaN_3 (FACS buffer) and surface stained for flow cytometric analysis. Cells were fixed and permeabilized using CytoFix/CytoPerm solution and Perm/Wash buffer (BD Pharmingen, San Diego, CA) according to the guidelines of the manufacturer. All cell suspensions were pre-incubated with 5 µg/ml blocking antibody against CD16/CD32 (2.4G2) before staining to reduce non-specific binding. Freshly isolated lung lymphocytes and BAL cells were stained in FACS buffer using fluorchrome-conjugated antibodies and the H-$2D^b$ $M_{187-195}$ tetramer. This tetramer was produced as described by Haanen *et al.* [41]. The antibodies, anti-CD3 (145-2C11), anti-CD4 (L3T4), anti-CD8 (clone 53-6.7), anti-CD11a (2D7), anti-CD11c (clone HL3), anti-CD62L (clone MEL-14), anti-NKG2a (20D5), MHC-II (I-A^b/I-E^b) (clone M5/114.15.2), anti-B220 (RA3-6B2) and Streptavidin were used for cell surface staining. For intra-cellular IFN-γ staining cells were treated with rat-anti-mouse IFN-γ FITC (clone XMG 1.2). All antibodies were purchased from BD biosciences (San Diego, CA) except anti-CCR3 (clone 83101), which was obtained from R&D systems (Minneapolis, MN). Stained samples were assessed on a FACS Calibur flowcytometer (BD). Data were analyzed using CellQuest software (BD) or Flow Jo (Treestar).

In another setting, antibody binding was tested against F proteins that most likely were in the prefusion conformation. This was done by staining of intra-cellular expressed RSV F trimer constructs in 293T cells transfected with a construct expressing the RSV F protein of RSV A2 with a isoleucine zipper trimerization domain (IZ) and His-tag with eight histidine residues (RSV F/IZ/8xHis) [42,43]. Transfected cells were stained with the anti-RSV F antibody panel (Palivizumab, AM14, AM16, AM22, AM23 and D25) and binding was detected with a PE-labeled goat anti-human F(ab')2 antibody (Southern Biotech).

ELISA

Binding of the human anti-RSV antibody panel (Palivizumab, AM14, AM16, AM22, AM23 and D25) to denatured RSV and monomeric RSV F was studied in an ELISA based experimental setup. In one setting regular ELISA plates (Corning) were coated with lysate of RSV infected HEp-2 cells in PBS for 1 hour at 37°C or o/n at 4°C and washed with PBS, 0.5% Tween-20. In the second setting purified His-tagged RSV F protein was captured on HisSorp ELISA plates (Qiagen). Plates were blocked

by incubation with PBS containing 4% milk, before incubation with the anti-RSV antibodies or polyclonal goat anti-RSV (Biodesign) and the HRP-conjugated anti-IgG antibodies (Jackson; diluted 1:2500). Antibody binding was visualized after addition of TMB substrate and stop solution (Biosource).

Results

Human RSV specific neutralizing IgGs dominantly target a prefusion specific domain in the F protein

We previously described the isolation and immortalization of memory B cells from a healthy donor, producing RSV specific human IgGs [28]. A panel of five monoclonal cell lines (D25, AM14, AM16, AM22 and AM23), obtained via this novel technology, was selected to study binding characteristics of RSV antigens. Sequences of the variable heavy chain (VH) and variable light chain (VL) regions of the new antibody set revealed that they were all different from each other and contained many somatic hyper mutations, an indication that these clones have gone through several rounds of selection and affinity maturation during Germinal Center reactions (Table 1 and Figure S1) [44].

The antibody specificity for either the RSV F or the attachment- (G) protein was determined by using recombinant Vesicular Stomatitis Virus (VSV) expressing either the RSV F- or G- protein [33]. EL-4 cells were infected with VSV-Wt, VSV-F_{rsv} or VSV-G_{rsv} and stained with Palivizumab, control IgG or one of the human antibodies. Similar to Palivizumab, D25 and the AM antibodies recognized the RSV F protein when expressed by recombinant VSV (Figure 1A) and these antibodies were also able to bind RSV infected HEp-2 cells (data not shown). However, except for Palivizumab and AM16, which also recognize RSV F in the postfusion conformation, recognition was dependent on the presence of the RSV F prefusion complex since D25, AM14, AM22 and AM23 did not recognize RSV infected HEp-2 cell lysate in ELISA, assuming that in such an ELISA the majority of the proteins is linearized and in a postfusion conformation (Figure 1B left panel). In agreement with this observation, no binding was detected to a monomeric His-tagged recombinant RSV F (Figure 1B right panel) or to RSV in a commercial ELISA (from r-Biopharm, data not shown). However, when the original stable B cell receptor (BCR) expressing B cell lines were incubated with

Table 1. Composition of antibody VDJ heavy and light chain recombination

Clone	Isotype	IGHV	IGHD (frame)	IGHJ	CDR3H	aa replacement
						VH
D25	IgG1	1-60*01 F	5-18*01 F(1)	4*02	CARHSTDYYYYYGMDVW	24
AM14	IgG1	3-30*04 F	1-26*01 F(1)	6*02	CARDRIVDDYYYYGMDVW	13
AM16	IgG1	3-21*01 F	3-10*01 F(2)	5*02	CAREDYGPGNYYSPNWFDPW	15
AM22	IgG1	1-24*01 F	6-19*01 F(2)	4*02	CGTLGVTVTEAGLGIDDYW	31
AM23	IgG1	3-33*01F	1-26*01 F(1)	4*02	CVRDKVGPTPYFDSW	17
Clone	Isotype	IGKV/LV	IGK/LJ		CDR3L	VL
D25	κ	1-33*01 F	4*01 F		CQQYDNLPLTF	6
AM14	κ	1-33*01 F	4*01 F		CQQYDNLPPLTF	9
AM16	λ	1-40*02 F	210*01 F		CHSYDRSLSGSVF	5
AM22	κ	3-20*01 F	3*0151 F		CLSSDSSIFTF	13
AM23	λ	3-21*02 F	2*0131 F		CQVWDRSNYHQVF	12

IMGT V-quest was used to match framework and CDR antibody sequences with corresponding genes (IMGT, the international ImMunoGeneTics information system, www.imgt.org)

His-tagged RSV F protein, binding could be observed for B cell clones AM16, AM23 and to a lesser extent D25. This indicates that the preparation of soluble F protein does contain epitopes that can be recognized by prefusion F specific antibodies like D25. In addition it shows that assays in which the antigen is coated, or immobilized will yield different results compared to protein assays in which the target is soluble and the receptor cell bound. No RSV F protein capture was observed to the BCRs of the AM14 or AM22 B cell lines (Figure 1C). Binding of the antibodies to the RSV F prefusion structure was confirmed by intracellular staining of 293T cells expressing the RSV F construct, RSV F/IZ/8xHis (Figure 1D). Except for AM14, all antibodies efficiently recognized the non-linear prefusion complex of the RSV F protein.

Competition of the AM antibody panel against D25 and Palivizumab revealed that D25 binding was reduced when cells were pre-incubated with AM22 or AM23, but none of the antibodies was interfering with Palivizumab binding (Figure 2). The absence of AM14 binding to the RSV trimer (Figure 1D) suggests that AM14

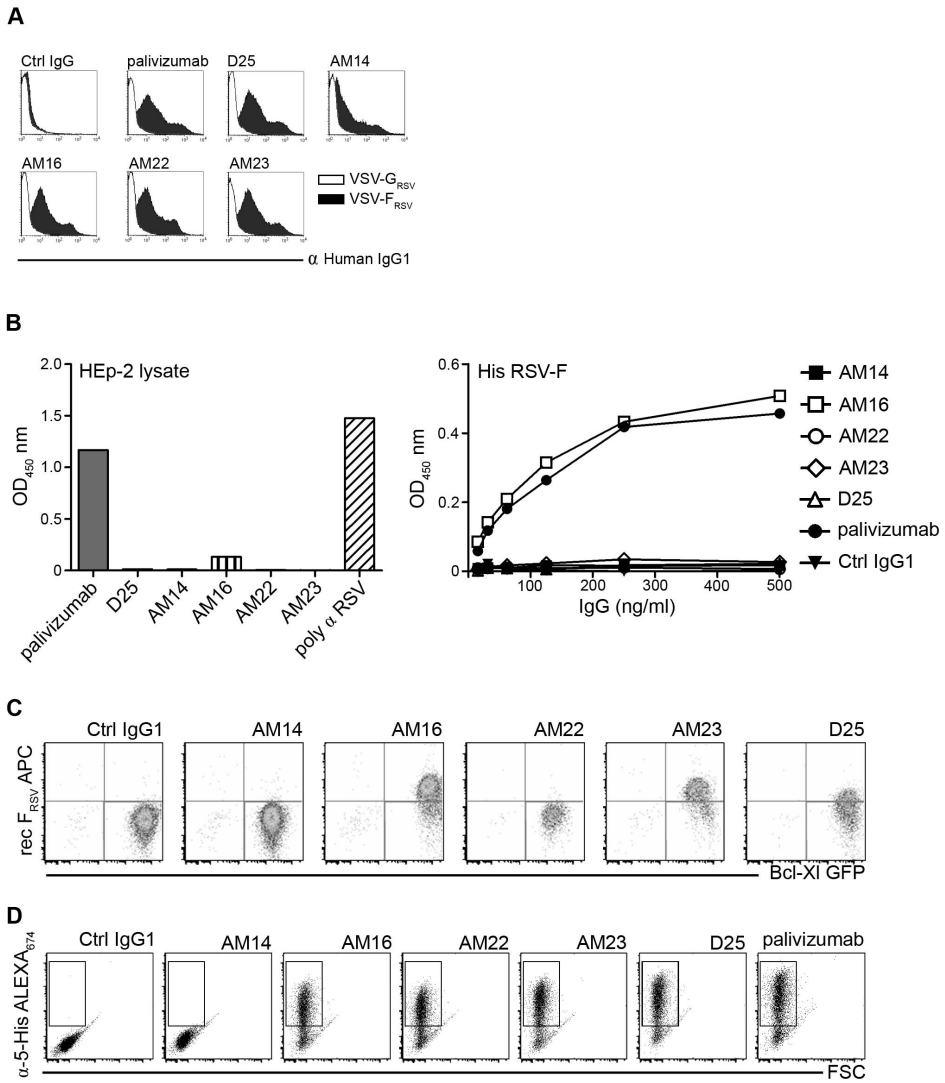


Figure 1. Determination of antibody target and epitope on the RSV virus. A) Binding of human anti-RSV IgGs to EL-4 cells infected with VSV-F_{RSV} or VSV-G_{RSV} was detected by goat anti human IgG-PE (Jackson Immunoresearch) and measured by flow cytometry. (B left panel) Binding of human anti RSV IgGs to a RSV A2 infected HEp-2 cell lysate measured by ELISA or (B right panel) capture ELISA with a HIS tagged RSV-A2 F protein. C) Binding of recombinant HIS tagged RSV-A2 F protein to GFP positive B cell clones. Bound antigen was detected using an APC labeled anti-His antibody and measured by FACS. D) Intra-cellular staining of 293T cells transfected with the RSV F/IZ/8xHis construct. Antibody binding was detected with a PE labeled goat anti-human F(ab')2 (Southern Biotech).

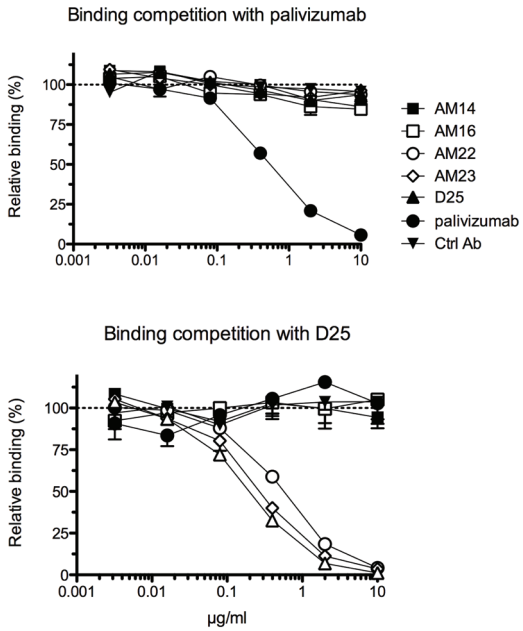


Figure 2. Binding competition for RSV F epitopes. RSV A2 infected HEp-2 cells were pre-incubated with increasing concentrations of the indicated unlabeled antibodies. The HEp-2 cells were then stained with (A) $Alexa_{633}$-labeled palivizumab or (B) $Alexa_{555}$-labeled D25. Relative binding was measured by flow cytometry. As a control IgG1 antibody rituximab (anti-CD20) was used.

recognizes a different prefusion specific epitope compared to the other antibodies (most probably C terminal since that part is absent in the trimer). These data on antibody competition can be largely explained by the crystallography data recently published by McLellan et.al. [45] in which they propose the existence of epitope Ø, an epitope specifically recognized by RSV F prefusion specific antibodies like D25 and AM22. AM14 and Palivizumab do not recognize this site Ø.

Frequency of RSV F prefusion specific B cell clones in vivo

Neutralizing antibodies against the prefusion form of RSV F offer unique possibilities for clinical intervention. Magro et al. suggested that these prefusion specific antibodies are present in sera at relative high titers and more recently Corti et al.

isolated an RSV-directed antibody which cross-reacted with other paramyxoviruses like hMPV [46,47]. Since the prefusion specific antibodies described here were isolated from a donor who by means of work was frequently challenged with RSV we choose to study the frequency of these B cells in the memory IgG^+ B cell population of healthy individuals. Therefore we screened the memory B cell repertoire of 3 healthy donors (Table 2). In the primary screening, RSV A2 infected and fluorescent (CFSE) labeled cells were mixed with Vero cells stably expressing G protein derived from the RSV-X strain (Figure S2). Antibodies present in B cell supernatants recognizing the RSV A2 infected cells (population 1) recognize RSV F, while antibodies recognizing both RSV A2 infected and Vero G cells (population 2) most probably are specific for G (around 30% of the repertoire). Around 10% of the cultures were neutralizing and prefusion specific, while the frequency of B cells that neutralize and are postfusion specific was 6% in two donors and 23% in the other donor (e.g. these antibodies also bind the prefusion form). This relative high frequency was confirmed by screening memory IgG+ B cells from a fourth donor in whom we found 25 unique clones of which 20 neutralized RSV A2 and 9 neutralizing clones were prefusion F specific (data not

Table 2. Frequency of RSV specific B cells in vivo

Donor	RSV specific cultures	Number of clones generated	% RSV G	% RSV F			
				Neutralizing		Non-neutralizing	
				Prefusion	Postfusion	Prefusion	Postfusion
R	0.36%	18	22	11	6	17	44
J	0.41%	39	33	8	23	10	26
V	0.31%	16	38	13	6	6	38

BCL6 Bcl-xL transduced IgG+CD27+ memory B cells of each donor were cultured in ten 96 well plates at 20 cells per well (19,200 cells per donor). Using FACS, culture supernatants were tested for binding to RSV A2 infected HEp2 cells and binding to Vero cells stably expressing RSV G from the X strain (Figure S2). Subsequently the specific clones were retrieved by single cell culture and culture supernatants were tested for neutralization against RSV A2, intracellular binding to 293T cells expressing a RSV F construct containing a trimerization domain and were tested in ELISA for binding to RSV A2 infected HEp2 cell lysate or a commercial kit (RIDASCREEN® RSV IgG, r-Biopharm GmbH, Darmstadt, Germany).

shown). In addition we did find a relative high frequency (between 6 and 17%) of prefusion F specific B cells, which do not seem to neutralize. These antibodies thus bind epitopes that do not interfere with attachment or fusion.

D25 efficiently neutralizes a panel of genetically distinctive primary RSV strains

Neutralization efficacy of the B cell derived antibodies AM14, AM22 and D25 was studied *in vitro* and on a panel of genetically divergent RSV strains (Table 3). The antibodies efficiently neutralized a panel of five RSV A and three RSV B clinical strains that originated from the epidemic years 2001 till 2008 (Figure 3). These strains show 1% and 9% amino acid difference in the fusion protein of RSV subgroup B and subgroup A, respectively, and differ 15% between the two subgroups (data not shown) [31,32]. Viral loads of reference strains RSV A2 and RSV 9320 were also

Table 3. RSV neutralizing activity of purified IgGs (ng/ml)

RSV A Strains	Palivizumab	D25	AM14	AM22
06-000827	284	7	183	75
01-000583	282	9	296	223
02-000110	66	16	556	284
03-036544	433	9	1096	238
RSV572	137	8	390	610
08-001411	168	4	170	271
A2	504	4	229	130
Mutual difference (Log 10)	0.88	1.40	0.80	0.90
RSV B strains	**Palivizumab**	**D25**	**AM14**	**AM22**
07-038176	205	7	163	197
03-000005	97	13	74	297
06-039982	30	4	75	182
9320	897	77	293	166
Mutual difference (Log 10)	1.48	1.34	0.59	0.26

The RSV neutralizing antibody titers of palivizumab, D25, AM14 and AM22 were determined with plaque assay on a panel of subgroup A and B clinical strains, RSV A2, and RSV 9320. Titers represent the concentration of antibody required for 50% reduction of the infectious viral load.

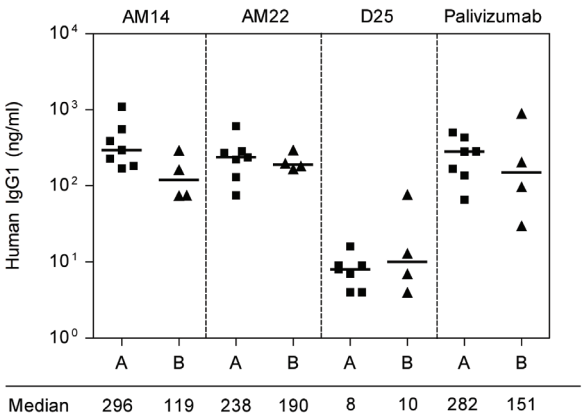


Figure 3. Neutralization of primary clinical isolates of RSV subgroup A and B *in vitro*. Neutralization efficacy of AM14, AM22 and D25 and palivizumab was determined by plaque reduction experiments with subgroup A and B clinical RSV strains on HEp-2 cells. Each dot represents the neutralizing antibody concentration required for 50% plaque reduction of each RSV strain type.

efficiently reduced by the B cell derived antibodies. A comparable neutralization efficacy range as with Palivizumab was observed for AM14 and AM22 (ranging from 30 to 1096 ng/ml). In contrast, D25 required only 4 to 77 ng/ml IgG to reduce the viral load with 50% (RSV A, 1.55 log; RSV B, 1.17 log compared to Palivizumab). All antibodies show similar efficacy in virus reduction between the two RSV subgroups with a maximum median difference of 0.35 log for AM14. However, neutralization of RSV B strain 9320 required up to one log higher doses of D25 and Palivizumab compared to other RSV subgroup B strain variants (Figure 3). Although up to 1.48 log maximum mutual strain differences in neutralization efficacy was observed for Palivizumab (Table 3), no natural antibody escape mutants for any of the antibodies were found in our panel of RSV strains.

Prefusion specific antibodies neutralize RSV efficiently in Cotton rats

Three RSV F prefusion specific antibodies, AM14, AM22 and, D25 were studied *in vivo* in cotton rats, which are known to support RSV replication [48]. Animals received prophylactic treatment with 2.0 or 0.4 mg/kg of monoclonal antibodies before intranasal challenge with RSV-X [30], a primary RSV A isolate. Due to relatively low

antibody production, the AM14 antibody was administered at 0.4 and 0.1 mg/kg.

The retrieval of RSV virus from the lungs of sacrificed animals was strongly reduced in all animal groups treated with 2.0 mg/kg of immunoglobulin compared to the control group (Figure 4A). Animals treated with 0.4 mg/kg Palivizumab and AM23 showed significant virus replication. No virus could be retrieved from animals treated with AM22 while in the AM14 and D25 groups only one out of 6 animals showed detectable virus replication. Surprisingly, animals treated with 0.1 mg/kg AM14 still showed reduced virus titers. These results indicated that the antibodies that specifically recognize non-linear, prefusion RSV F specific epitopes harbor strong *in vivo* neutralizing capacities.

The level of recovered human IgG from cotton rat sera at day 1 (day of virus inoculation) and day 6 (day of sacrifice) was in the same range for all antibodies and the decline of antibodies in time was comparable (data not shown). At day 5 post-infection, the left lung from each cotton rat was removed and fixed with formalin. Lung damage was classified and the severity of pathology was calculated according to the pathology index score. Lung pathology after RSV infection was significantly reduced in animal groups treated with high doses of immunoglobulin (2 mg/kg, Figure 4B). However, a complete absence of pathology was only seen in 3 out of 5 animals treated with AM22 at 2 mg/kg (Table 4). In cotton rats treated with AM14 and AM22 the pathology was also significantly reduced at lower concentrations (0.4 mg/kg). For these antibodies, complete protection was detected for 2 or 1 out of 6 animals in the AM14 and AM22 groups, respectively.

D25 lowers RSV viral load in Balb/c mice

The effect of RSV challenge in Balb/c mice was tested in the presence of D25. One day before intra-nasal challenge with RSV A2 mice received D25 at concentrations of 1.0, 0.1 and 0.01 mg/kg, Palivizumab at 10, 1 and 0.1 mg/kg or an isotype control IgG1 at 1 mg/kg. In contrast to D25, where weight loss was absent at 1 and 0.1 mg/kg, the control and Palivizumab treated mice lost weight between day 1 and 2 post-infection when pre-treated with 1 mg/kg antibody (Figure 5A). At lower antibody concentrations there were no differences in relative weight between mice treated with D25, Palivizumab or Ctrl IgG1.

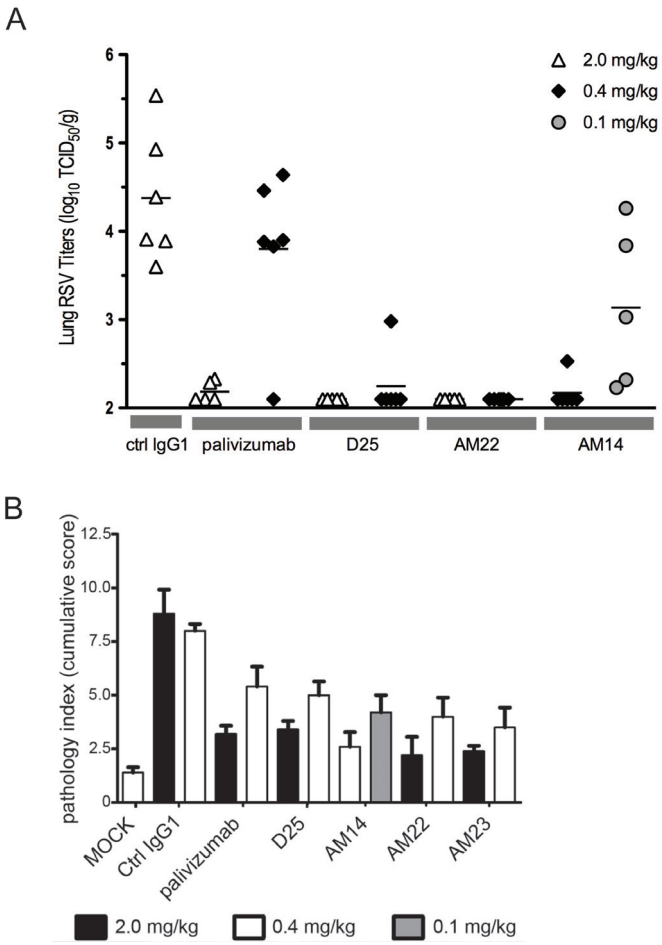


Figure 4. Antibody efficacy in Cotton rats. A) The RSV load in the lungs of cotton rats treated with the indicated antibodies 24 hours before infection with the RSV-X strain, was determined 5 days after infection by $TCID_{50}$ culture. Open symbols indicate animals treated with 2.0 mg/kg, closed symbols 0.4 mg/kg and gray symbols 0.1 mg/kg. Experiments were performed twice with four to six individual animals per treatment group. B) Cumulative pathology score of the lungs of cotton rats, treated 24 hour before RSV infection with the indicated antibodies. Lung specimens obtained 5 days after infection were evaluated randomly by a pathologist. Lung pathology was classified with scores between 0-5 for three individual markers: 1) hyperthropy of bronchus and bronchioli epithelia, 2) peribronchiolitis and 3) alveolitis. Average pathology score (maximum score 15) with standard error of the mean (SEM). Experiments were performed twice with four to six individual animals per treatment group.

Table 4. Lung pathology in cotton rats after RSV challenge

Antibody	mg/kg	Pathology reduction	No pathology
Palivizumab	2	5/5	1/5
Palivizumab	0.4	3/5	0/6
D25	2	5/5	0/5
D25	0.4	2/6	0/6
AM14	0.4	5/6	2/6
AM14	0.1	4/5	0/5
AM22	2	5/5	3/5
AM22	0.4	5/6	1/6

Number of animals with significant reduction in lung pathology /or complete absence of pathology at day 5 post-infection with RSV-X (accession number FJ948820.1). Experiments were performed twice with five to six individual animals per treatment group.

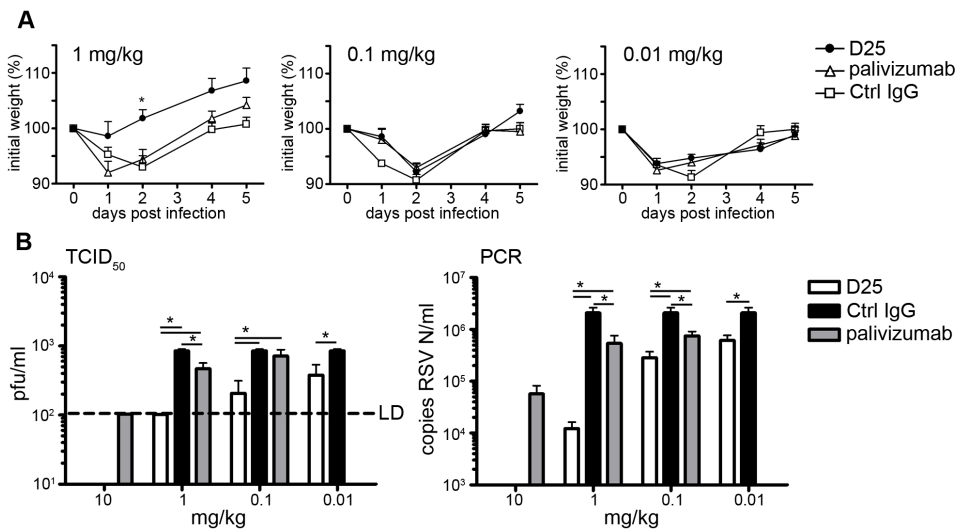


Figure 5. D25 efficiently neutralizes RSV A2 in mice. A) Percentage weight change of mice treated with different concentrations of control IgG, palivizumab and D25. Weight on the day of RSV A2 infection is set to 100%. The RSV load in the BAL (B) of mice treated with the indicated antibody, was determined 5 days after infection with $TCID_{50}$ culture (left panel) or quantitative PCR for the RSV N gene (right panel). The error bars represent the standard error of the mean (SEM), with * $P < 0.05$. The experiment was performed twice with 5 individual mice per group. LD is limit of detection.

The effect of antibody doses on viral replication in lungs and BAL was determined five days after RSV infection. Using a standard $TCID_{50}$ culture assay, the infectious viral load in the BAL was determined (Figure 5B, left panel). The highest concentration of both Palivizumab (10 mg/kg) and D25 (1.0 mg/kg) reduced viral loads below the limit of detection of $2.1x10^2$ $TCID_{50}$/ml. In contrast, Palivizumab only partially reduced viral load when administered at 1.0 mg/kg and no reduction could be observed at lower concentration. D25 treated mice, however, showed partial protection at a concentration of 0.01 mg/kg compared to Ctrl IgG1 treated mice. These data were further supported by quantitative RT-PCR of the RSV N gene, which detects the total viral load (non-viable and infectious virus; Figure 5B, right panel). The same efficient reduction in viral load by D25 compared to Palivizumab was also found in the lung tissue compared to the BAL samples (data not shown). Since there was no difference in the decay of serum IgG levels measured at day 6 (data not shown), the effects seen with D25 compared to Palivizumab must be attributed to the more potent neutralization capacity of D25.

Innate and adaptive immune responses are diminished in D25 treated mice

There is much debate whether the pathology of RSV infection is directly virus mediated or mediated by exacerbation of the anti-viral immune response [49-51]. Similar to the gain of weight and the lower virus levels in the lungs of D25 pre-treated mice, total cellular infiltrate in the airways was lower in BAL of mice treated with D25 at 1 and 0.1 mg/kg compared to Palivizumab treated mice. Palivizumab only reduced the influx of cells at 10 mg/kg (Figure 6A). The majority of the cellular infiltrate in the airways consisted of neutrophils and T cells based on the expression patterns of CD3, CD11c, B220 and CCR3 [52]. Mice treated with 1.0 mg/kg D25 had significantly lower absolute numbers of neutrophils in the airways compared to control IgG1 treated mice (Figure 6B). There was no difference in neutrophil influx between mice treated with Palivizumab or control IgG1, despite the lower viral load in the first group (Figure 5B). The same pattern was found for the influx of T cells in the airways. Treatment with D25, at a concentration of 1.0 mg/kg, lowered the migration of T cells into the airways compared to Palivizumab or control IgG1 (Figure 6C). No difference in the recruitment of T cells in the BAL between Palivizumab and control IgG treated mice was found. This suggested that D25 was not only capable

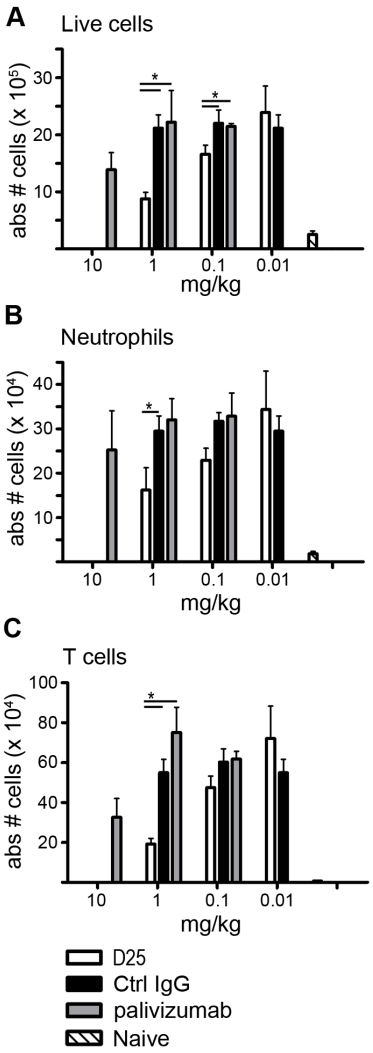


Figure 6. D25 treated mice show reduced cellular influx in the airways during RSV infection. A) Absolute live cell counts of BAL were obtained 5 days after RSV A2 infection. B) Absolute numbers of neutrophils in the airways of mice was determined 5 days after RSV A2 infection and was based on the FSC/SSC pattern and the absence of CD3/B220/CCR3/ MHC II expression as determined by flow cytometry. C) Absolute numbers of T cells in the airways of mice was determined 5 days after RSV A2 infection and was based on the FSC/SSC pattern and expression of CD3 as determined by flow cytometry, The error bars represent the (SEM) and * indicates $P < 0.05$ The experiment was performed twice with 5 individual mice per group.

of lowering the viral load, but concurrently reduced the virus induced inflammatory cell recruitment initiated after recognition of RSV by the innate immune system.

After the innate immune response, the adaptive immune response develops to eliminate the virus and generate long-term memory responses. Antibodies affect antigen uptake and influence the generation and nature of the T cell response [53,54]. Here we determined whether reduced virus titers or formation of immune complexes affected the subsequent T cell response. Eight days post intra-nasal RSV infection, at the peak of the T cell response of mice pre-treated with 0.1, 1.0 and 10 mg/kg of antibodies, lungs were removed and the RSV specific $CD4^+$ and $CD8^+$ T cell responses were determined. D25, at a concentration of 1.0 mg/kg, significantly reduced the absolute number of activated $CD11a^{hi}CD62L^{lo}CD8^+$ T cells compared to controls. These low $CD8^+$ T cell numbers were only found at 10 mg/kg for Palivizumab (Figure 7A). The RSV specific response was studied with tetramer staining against the dominant $M_{187\text{-}195}$ H-$2D^b$ epitope from the RSV matrix protein [39,40]. Again D25 was 10 times more efficient then Palivizumab and significantly reduced the absolute number of $M_{187\text{-}195}$ specific $CD8^+$ T cells found in the lung of RSV challenged mice (Figure 7B). In addition, the functional RSV specific responses of lung $CD4^+$ and $CD8^+$ T cells were determined *ex vivo*. Lung $CD4^+$ and $CD8^+$ T cells isolated 8 days after infection were stimulated with RSV A2 infected dendritic cells (D1) and IFN-γ production was measured by intra cellular cytokine staining. The absolute number of virus specific IFN-γ producing $CD8^+$ and $CD4^+$ T cells recovered from lungs of RSV A2 infected mice was significantly reduced when mice were pretreated with 1.0 mg/kg of D25 compared to mice receiving Ctrl IgG or Palivizumab (Figure 7C and 7D). The later required a ten-fold higher concentration to significantly reduce the absolute number of RSV specific IFN-γ producing $CD8^+$ T cells, while no significant reduction in absolute numbers of RSV specific $CD4^+$ T cells was found (Figure 7D). These data correlated with the ability of D25 to reduce virus replication at 1.0 mg/kg and Palivizumab at 10 mg/kg. The virus specific $CD4^+$ T cells producing the Th2 cytokine IL-5 were also enumerated. The absolute numbers of IL-5 producing $CD4^+$ T cells were low in all mice, with no differences between the different antibody treated groups (data not shown). Thus both Palivizumab and D25 lowered viral loads, cell recruitment and the level of RSV specific T cell responses in the lungs of RSV infected mice, but D25 was in all aspects more potent than Palivizumab.

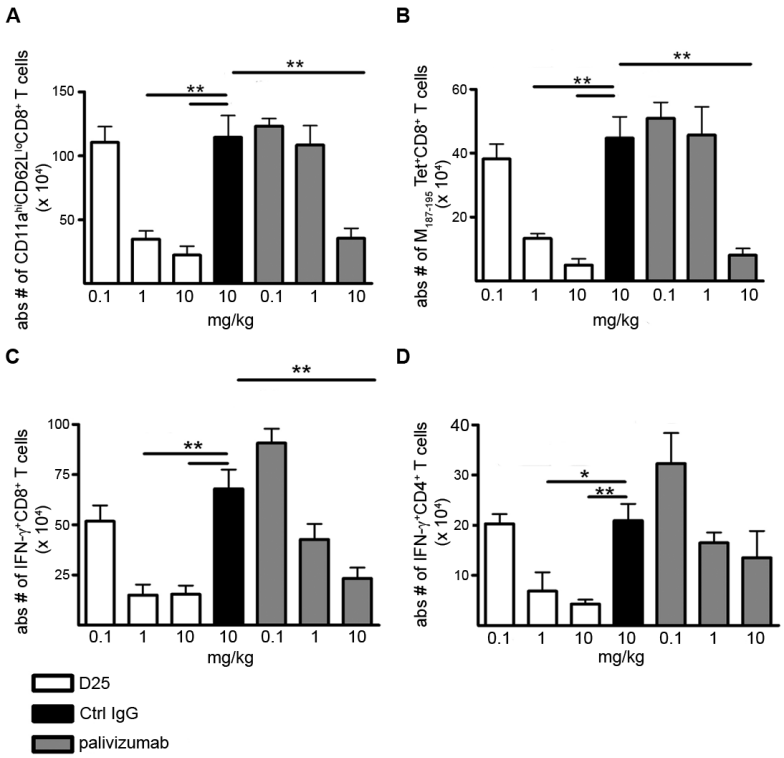


Figure 7. D25 and palivizumab treated mice show reduced RSV specific T cell responses in the lungs. Lung derived T cells of mice treated with the designated antibody at 0.1, 1 or 10 mg/kg were obtained 8 days after primary RSV infection. A) Total number of activated $CD8^+$ T cells was determined by measuring the $CD11a^{hi}CD62L^{lo}$ $CD8^+$ T cells by flow cytometry. B) Total number of $M_{187\text{-}195}$ $CD8^+$ T cells was determined by H-$2D^b$ $M_{187\text{-}195}$ tetramer staining in combination with anti-CD8. C) RSV specific $CD8^+$ T cell numbers or (D) $CD4^+$ T cell numbers were measured by stimulating lung T cells with RSV infected or uninfected LPS matured D1 cells (a H-2^{b+} dendritic cell line) and staining for intracellular IFN-γ. The error bars represent the (SEM), with * $P < 0.05$ and ** $P < 0.01$, the experiment was performed twice with 5 individual mice per group.

Discussion

In this study we determined the prevalence, binding characteristics, neutralization efficacy and impact on immune responses of a panel of fully human, highly potent RSV neutralizing IgG1 antibodies derived from germinal center-like B cells. All

antibodies are derived from different VH and VL families and contain multiple somatic hyper mutations suggesting that the original B cell clones have undergone multiple rounds of selection in the host. Interestingly, they all react with the RSV F protein, but they have the highest specificity for the prefusion form of RSV F. It has been shown for the related para-influenza subtype 5 and Newcastle disease virus that the F protein undergoes profound conformation changes from the pre- to postfusion state upon infection [55,56]. Also the transition of RSV F protein from a cone to lollypop shaped form has been implicated to affect membrane fusion and infection [57]. Recently, McLellan *et al.* reported that D25 binds specifically to the metastable prefusion form of RSV F, thereby preventing the conformational change towards the more stable postfusion structure [45]. Similar to D25, the AM antibodies did not recognize denatured RSV virus or monomeric forms of the F protein, suggesting that they are specific for the prefusion form of RSV F and interfere with changes in conformation towards a postfusion state. The presence of B cells producing only prefusion specific antibodies seems not unique as we found relatively high frequencies (approximately 10%) of these B cells in the memory IgG+ population. However, also B cells producing non-neutralizing prefusion specific antibodies were found, indicating that prefusion specificity is not a prerequisite for neutralization. Whether in the non-neutralizing postfusion specific antibody population (26 to 44%) only postfusion specific B cells are present cannot be determined by this study. In addition, we found that prefusion specificity is not restricted to a single epitope since we found several antibodies that did not compete with each other for binding to the prefusion RSV F. This indicated that there are several important conformational epitopes on the RSV F protein accessible by antibodies, a finding also observed by Corti et al. [58] who described an antibody directed to a prefusion-specific, Palivizumab-like binding site.

The conformation dependent antibodies effectively decreased the viral load of a selection of genetic divergent RSV strains and with similar efficacy for both subgroup A and subgroup B variants. This suggests that strain variance does not affect the neutralization epitope of these antibodies. Among the novel IgG1s, D25 is the most potent neutralizing antibody tested with average *in vitro* IC_{50} (half-maximum inhibitory concentrations) values around 10 ng/ml, A study by Magro *et al* [47] also showed that natural derived antibodies directed against the pre-active form of RSV

F were potent RSV neutralizing agents. Further *in vivo* analyses in cotton rats and mice showed that the human RSV specific IgG1s diminished RSV viral load in the lungs at relatively low concentrations. Compared to Palivizumab, the novel antibody panel was at least ten-fold more efficient in RSV neutralization *in vivo*.

The main issues in RSV disease that remain to be clarified are the cause of illness and mechanisms to prevent RSV related illnesses like wheezing and asthma. The association between early-life wheezing and the development of asthma later in life is strong [14,59]. Viral respiratory tract infections with RSV or human rhinovirus [17,21] have been associated with the development of wheezing and childhood asthma, suggesting that prevention of infection has the potential to reduce the risk of developing these diseases. Several studies have now shown that reduction of viral load using Palivizumab or high levels of maternal antibodies as a prophylactic can reduce virus related disease [8,13,22,25,60]. The virus can cause serious damage to lung epithelial layers and induce bronchial hyper responsiveness. Disease severity may be enhanced by strong innate and adaptive immune responses [13]. High numbers of neutrophils have been observed in the lungs of children admitted to the intensive care unit with severe RSV disease [61]. These cells can be major contributors to RSV-induced inflammation and disease severity. Using mouse models, it has been suggested that innate cell influx is in part mediated by the fracktalkine chemokine motif (CX_3C) in the RSV G protein [62]. Early therapeutic administration of an anti-RSV G specific antibody directed against this CX_3C motif reduced cellular infiltrate, lung pathology and viral load in RSV challenged mice [62,63]. In addition, the F protein of RSV has been implicated in TLR4 mediated signaling leading to NFκB activation facilitating innate immune responses during RSV infection [64]. TLR4 triggering and subsequent production of pro-inflammatory cytokines could be reduced by I) direct decrease of the viral load, II) RSV F antibody binding competition that blocks TLR4 activation [65]. Both actions will lead to less epithelial cell damage, less induction of innate immune responses and less antigen presentation or T cell activation.

Here we showed that both the innate and adaptive immune responses in mice were diminished, when D25 was prophylactically administered. Innate immune responses in mice, as measured by lower inflammatory cell influx into the airways were decreased. In addition, neutralizing antibodies also reduced the induction of adaptive immune response as measured by reduced RSV specific $CD4^+$ and $CD8^+$ T

cell numbers 8 days post infection. These T cells still mainly produced IFN-γ, but there was no shift towards a Th2 type response. Th2 type T cell responses with increased IL-4, IL-5 and Il-13 production are often observed after vaccination with inactivated or subunit vaccines, and have been associated with enhanced disease [65,66]. Therefore, lowered inflammatory cell infiltrates most likely reduce pathology and on the long term reduce the risk of wheezing and asthma exacerbations. This change in pathology was observed in cotton rats treated with D25, AM14 and AM22.

This study has shown that human B cell derived IgG1s directed against specific conformational epitopes are highly neutralizing antibodies both *in vitro* and *in vivo*. Furthermore, it proves that the novel IgG selection technology is a powerful tool to specifically isolate antibodies that hold potential for the development of antibody therapeutics against RSV and many other infectious diseases. Our findings suggest that antibodies with a higher neutralizing efficacy compared to Palivizumab and capacity to reduce inflammation are expected to improve the RSV clinical outcome in young infants and in immuno compromised adults with RSV disease.

Acknowledgements

H.S, E.Y., A.Q.B and T.B. are inventors on an international patent application describing the technology used to isolate D25 and AM22 (Means and methods for influencing the stability of antibody producing cells, WO2007067046A1). A.Q.B, E.Y. and T.B. are inventors on international patent applications for antibody D25 (RSV-specific binding molecules and means for producing them, WO2008147196A2) and antibody AM22 (RSV-specific binding molecule, WO2011043643A1). The work performed by M.V.L. and G.M.v.B was supported by grant no. OZF-02-004 from the Wilhelmina Research Fund.

We thank Yvonne van Remmerden for virus titrations, Paul Roholl for histopathology and Debby Kruijsen for assistance with RSV culture work and Rogier Sanders for helpful discussions in using the isoleucine zipper trimerization domain.

References

1. Glezen P, Denny FW (1973) *Epidemiology of acute lower respiratory disease in children*. N Engl J Med 288: 498-505.

2. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, et al. (2009) *The burden of respiratory syncytial virus infection in young children*. N Engl J Med 360: 588-598.

3. MacDonald NE, Hall CB, Suffin SC, Alexson C, Harris PJ, et al. (1982) *Respiratory syncytial viral infection in infants with congenital heart disease*. N Engl J Med 307: 397-400.

4. Meissner HC (2003) *Selected populations at increased risk from respiratory syncytial virus infection*. Pediatr Infect Dis J 22: S40-44; discussion S44-45.

5. Thorburn K (2009) *Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection*. Arch Dis Child 94: 99-103.

6. Glezen WP, Paredes A, Allison JE, Taber LH, Frank AL (1981) *Risk of respiratory syncytial virus infection for infants from low-income families in relationship to age, sex, ethnic group, and maternal antibody level*. J Pediatr 98: 708-715.

7. Stensballe LG, Ravn H, Kristensen K, Meakins T, Aaby P, et al. (2009) *Seasonal variation of maternally derived respiratory syncytial virus antibodies and association with infant hospitalizations for respiratory syncytial virus*. J Pediatr 154: 296-298.

8. Stensballe LG, Ravn H, Kristensen K, Agerskov K, Meakins T, et al. (2009) *Respiratory syncytial virus neutralizing antibodies in cord blood, respiratory syncytial virus hospitalization, and recurrent wheeze*. J Allergy Clin Immunol 123: 398-403.

9. Group TI-RS (1998) *Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants.* . Pediatrics 102: 531-537.

10. Johnson S, Oliver C, Prince GA, Hemming VG, Pfarr DS, et al. (1997) *Development of a humanized monoclonal antibody (MEDI-493) with potent in vitro and in vivo activity against respiratory syncytial virus*. J Infect Dis 176: 1215-1224.

11. Groothuis JR, Simoes EA, Hemming VG (1995) *Respiratory syncytial virus (RSV) infection in preterm infants and the protective effects of RSV immune globulin (RSVIG). Respiratory Syncytial Virus Immune Globulin Study Group*. Pediatrics 95: 463-467.

12. Meissner HC, Long SS (2003) *Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus*

infections. Pediatrics 112: 1447-1452.

13. Feltes TF, Cabalka AK, Meissner HC, Piazza FM, Carlin DA, et al. (2003) *Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease*. J Pediatr 143: 532-540.

14. Stensballe LG, Simonsen JB, Thomsen SF, Larsen AM, Lysdal SH, et al. (2009) *The causal direction in the association between respiratory syncytial virus hospitalization and asthma*. J Allergy Clin Immunol 123: 131-137 e131.

15. Kusel MM, de Klerk NH, Kebadze T, Vohma V, Holt PG, et al. (2007) *Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma*. J Allergy Clin Immunol 119: 1105-1110.

16. Sigurs N, Bjarnason R, Sigurbergsson F, Kjellman B (2000) *Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7*. Am J Respir Crit Care Med 161: 1501-1507.

17. Jackson DJ, Gangnon RE, Evans MD, Roberg KA, Anderson EL, et al. (2008) *Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children*. Am J Respir Crit Care Med 178: 667-672.

18. Garcia-Garcia ML, Calvo C, Casas I, Bracamonte T, Rellan A, et al. (2007) *Human metapneumovirus bronchiolitis in infancy is an important risk factor for asthma at age 5*. Pediatr Pulmonol 42: 458-464.

19. Caliskan M, Bochkov YA, Kreiner-Moller E, Bonnelykke K, Stein MM, et al. (2013) *Rhinovirus wheezing illness and genetic risk of childhood-onset asthma*. N Engl J Med 368: 1398-1407.

20. Sly PD, Kusel M, Holt PG (2010) *Do early-life viral infections cause asthma?* J Allergy Clin Immunol 125: 1202-1205.

21. Wu P, Hartert TV (2011) *Evidence for a causal relationship between respiratory syncytial virus infection and asthma*. Expert Rev Anti Infect Ther 9: 731-745.

22. Lemanske RF, Jr. (2013) *Early-life wheezing and respiratory syncytial virus prevention*. N Engl J Med 368: 1839-1841.

23. Jackson DJ, Lemanske RF, Jr. (2010) *The role of respiratory virus infections in childhood asthma inception*. Immunol Allergy Clin North Am 30: 513-522, vi.

24. Blanken MO, Rovers MM, Molenaar JM, Winkler-Seinstra PL, Meijer A, et al. (2013) *Respiratory syncytial virus and recurrent wheeze in healthy preterm infants*. N Engl J Med

368: 1791-1799.

25. Simoes EA, Groothuis JR, Carbonell-Estrany X, Rieger CH, Mitchell I, et al. (2007) *Palivizumab prophylaxis, respiratory syncytial virus, and subsequent recurrent wheezing*. J Pediatr 151: 34-42, 42 e31.

26. Shvarts A, Brummelkamp TR, Scheeren F, Koh E, Daley GQ, et al. (2002) *A senescence rescue screen identifies BCL6 as an inhibitor of anti-proliferative p19(ARF)-p53 signaling*. Genes Dev 16: 681-686.

27. Scheeren FA, Naspetti M, Diehl S, Schotte R, Nagasawa M, et al. (2005) *STAT5 regulates the self-renewal capacity and differentiation of human memory B cells and controls Bcl-6 expression*. Nat Immunol 6: 303-313.

28. Kwakkenbos MJ, Diehl SA, Yasuda E, Bakker AQ, van Geelen CM, et al. (2010) *Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming*. Nat Med 16: 123-128.

29. Diehl SA, Schmidlin H, Nagasawa M, van Haren SD, Kwakkenbos MJ, et al. (2008) *STAT3-mediated up-regulation of BLIMP1 Is coordinated with BCL6 down-regulation to control human plasma cell differentiation*. J Immunol 180: 4805-4815.

30. Widjojoatmodjo MN, Boes J, van Bers M, van Remmerden Y, Roholl PJ, et al. (2010) *A highly attenuated recombinant human respiratory syncytial virus lacking the G protein induces long-lasting protection in cotton rats*. Virol J 7: 114.

31. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, et al. (2012) *Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology*. PLoS One 7: e51439.

32. Tan L, Coenjaerts FE, Houspie L, Viveen MC, van Bleek GM, et al. (2013) *The Comparative Genomics of Human Respiratory Syncytial Virus Subgroup A and B: Genetic Variability and Molecular Evolutionary Dynamics*. J Virol.

33. Kahn JS, Schnell MJ, Buonocore L, Rose JK (1999) *Recombinant vesicular stomatitis virus expressing respiratory syncytial virus (RSV) glycoproteins: RSV fusion protein can mediate infection and cell fusion*. Virology 254: 81-91.

34. Winzler C, Rovere P, Rescigno M, Granucci F, Penna G, et al. (1997) *Maturation stages of mouse dendritic cells in growth factor-dependent long-term cultures*. J Exp Med 185: 317-328.

35. Scheeren FA, van Geelen CM, Yasuda E, Spits H, Beaumont T (2011) *Antigen-specific*

monoclonal antibodies isolated from B cells expressing constitutively active STAT5. PLoS One 6: e17189.

36. Jaleco AC, Stegmann AP, Heemskerk MH, Couwenberg F, Bakker AQ, et al. (1999) *Genetic modification of human B-cell development: B-cell development is inhibited by the dominant negative helix loop helix factor Id3*. Blood 94: 2637-2646.

37. Ternette N, Tippler B, Uberla K, Grunwald T (2007) *Immunogenicity and efficacy of codon optimized DNA vaccines encoding the F-protein of respiratory syncytial virus*. Vaccine 25: 7271-7279.

38. Zielinska E, Liu D, Wu HY, Quiroz J, Rappaport R, et al. (2005) *Development of an improved microneutralization assay for respiratory syncytial virus by automated plaque counting using imaging analysis*. Virol J 2: 84.

39. Lukens MV, Claassen EA, de Graaff PM, van Dijk ME, Hoogerhout P, et al. (2006) *Characterization of the CD8+ T cell responses directed against respiratory syncytial virus during primary and secondary infection in C57BL/6 mice*. Virology 352: 157-168.

40. Rutigliano JA, Rock MT, Johnson AK, Crowe JE, Jr., Graham BS (2005) *Identification of an H-2D(b)-restricted CD8+ cytotoxic T lymphocyte epitope in the matrix protein of respiratory syncytial virus*. Virology 337: 335-343.

41. Haanen JB, Toebes M, Cordaro TA, Wolkers MC, Kruisbeek AM, et al. (1999) *Systemic T cell expansion during localized viral infection*. Eur J Immunol 29: 1168-1174.

42. Harbury PB, Kim PS, Alber T (1994) *Crystal structure of an isoleucine-zipper trimer*. Nature 371: 80-83.

43. Melchers M, Matthews K, de Vries RP, Eggink D, van Montfort T, et al. (2011) *A stabilized HIV-1 envelope glycoprotein trimer fused to CD40 ligand targets and activates dendritic cells*. Retrovirology 8: 48.

44. Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, et al. (1999) *IMGT, the international ImMunoGeneTics database*. Nucleic Acids Res 27: 209-212.

45. McLellan JS, Chen M, Leung S, Graepel KW, Du X, et al. (2013) *Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody*. Science 340: 1113-1117.

46. Corti D, Bianchi S, Vanzetta F, Minola A, Perez L, et al. (2013) *Cross-neutralization of four paramyxoviruses by a human monoclonal antibody*. Nature 501: 439-443.

47. Magro M, Mas V, Chappell K, Vazquez M, Cano O, et al. (2012) *Neutralizing antibodies*

against the preactive form of respiratory syncytial virus fusion protein offer unique possibilities for clinical intervention. Proc Natl Acad Sci U S A 109: 3089-3094.

48. Prince GA, Jenson AB, Horswood RL, Camargo E, Chanock RM (1978) *The pathogenesis of respiratory syncytial virus infection in cotton rats*. Am J Pathol 93: 771-791.

49. Graham BS, Bunton LA, Rowland J, Wright PF, Karzon DT (1991) *Respiratory syncytial virus infection in anti-mu-treated mice*. J Virol 65: 4936-4942.

50. Cannon MJ, Openshaw PJ, Askonas BA (1988) *Cytotoxic T cells clear virus but augment lung pathology in mice infected with respiratory syncytial virus*. J Exp Med 168: 1163-1168.

51. DeVincenzo JP, El Saleeby CM, Bush AJ (2005) *Respiratory syncytial virus load predicts disease severity in previously healthy infants*. J Infect Dis 191: 1861-1868.

52. van Rijt LS, Kuipers H, Vos N, Hijdra D, Hoogsteden HC, et al. (2004) *A rapid flow cytometric method for determining the cellular composition of bronchoalveolar lavage fluid cells in mouse models of asthma*. J Immunol Methods 288: 111-121.

53. Kruijsen D, Bakkers MJ, van Uden NO, Viveen MC, van der Sluis TC, et al. (2010) *Serum antibodies critically affect virus-specific CD4+/CD8+ T cell balance during respiratory syncytial virus infections*. J Immunol 185: 6489-6498.

54. Kruijsen D, Einarsdottir HK, Schijf MA, Coenjaerts FE, van der Schoot EC, et al. (2013) *Intranasal administration of antibody-bound respiratory syncytial virus particles efficiently primes virus-specific immune responses in mice*. J Virol 87: 7550-7557.

55. Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS (2006) *Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation*. Nature 439: 38-44.

56. Chen L, Gorman JJ, McKimm-Breschkin J, Lawrence LJ, Tulloch PA, et al. (2001) *The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion*. Structure 9: 255-266.

57. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, et al. (2001) *Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion*. Proc Natl Acad Sci U S A 98: 9859-9864.

58. Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, et al. (2011) *A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins*. Science 333: 850-856.

59. Stein RT, Sherrill D, Morgan WJ, Holberg CJ, Halonen M, et al. (1999) *Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years*. Lancet 354: 541-545.

60. Groothuis JR, Simoes EA, Levin MJ, Hall CB, Long CE, et al. (1993) *Prophylactic administration of respiratory syncytial virus immune globulin to high-risk infants and young children. The Respiratory Syncytial Virus Immune Globulin Study Group*. N Engl J Med 329: 1524-1530.

61. Heidema J, Lukens MV, van Maren WW, van Dijk ME, Otten HG, et al. (2007) *CD8+ T cell responses in bronchoalveolar lavage fluid and peripheral blood mononuclear cells of infants with severe primary respiratory syncytial virus infections*. J Immunol 179: 8410-8417.

62. Tripp RA, Jones LP, Haynes LM, Zheng H, Murphy PM, et al. (2001) *CX3C chemokine mimicry by respiratory syncytial virus G glycoprotein*. Nat Immunol 2: 732-738.

63. Haynes LM, Caidi H, Radu GU, Miao C, Harcourt JL, et al. (2009) *Therapeutic monoclonal antibody treatment targeting respiratory syncytial virus (RSV) G protein mediates viral clearance and reduces the pathogenesis of RSV infection in BALB/c mice*. J Infect Dis 200: 439-447.

64. Kurt-Jones EA, Popova L, Kwinn L, Haynes LM, Jones LP, et al. (2000) *Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus*. Nat Immunol 1: 398-401.

65. Openshaw PJ, Tregoning JS (2005) *Immune responses and disease enhancement during respiratory syncytial virus infection*. Clin Microbiol Rev 18: 541-555.

66. Openshaw PJ, Culley FJ, Olszewska W (2001) *Immunopathogenesis of vaccine-enhanced RSV disease*. Vaccine 20 Suppl 1: S27-31.

SUPPLEMENTARY INFO

CHAPTER 4

Table S1. Primary RSV strains

Code	Sample date	Genotype	Accession #
01-000583	07-01-01	GA5	JQ901448
02-000110	03-01-02	GA2	JQ901453
03-036544	23-12-03	GA2	JQ901458
06-000827	09-01-06	GA5	JX015488
08-001411	14-01-08	GA2	JX015494
RSV572	06-11-07	GA2	JX015484
07-038176	23-11-07	GB13	JX576749
06-039982	21-12-06	GB13	JX576752
03-000005	01-01-03	GB12	JX576760

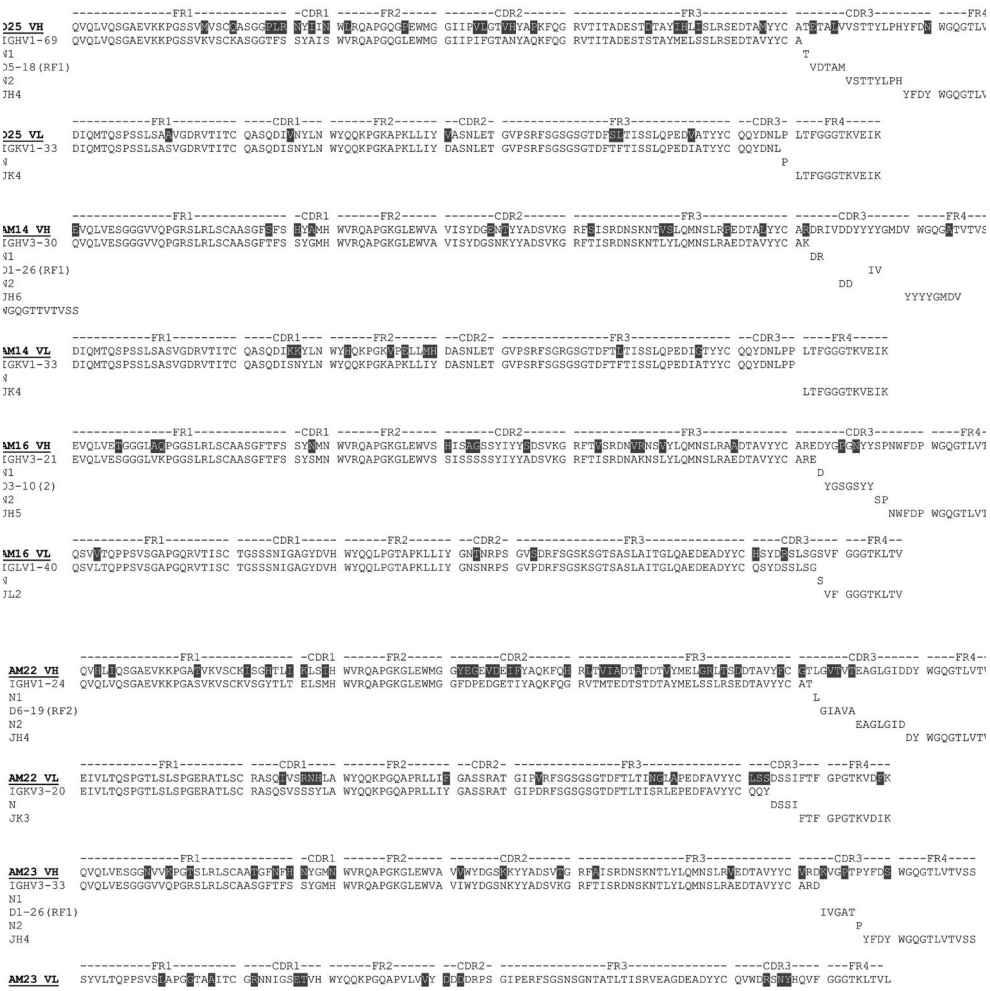


Figure S1. Composition of antibody VDJ heavy and light chain recombination. Framework and CDR regions are numbered according to Kabat and IMGT V-quest was used to match antibody sequences with corresponding genes (IMGT, the international ImMunoGeneTics information system www.imgt.org). N indicates insertion, and amino acids highlighted gray are replacements compared to germline.

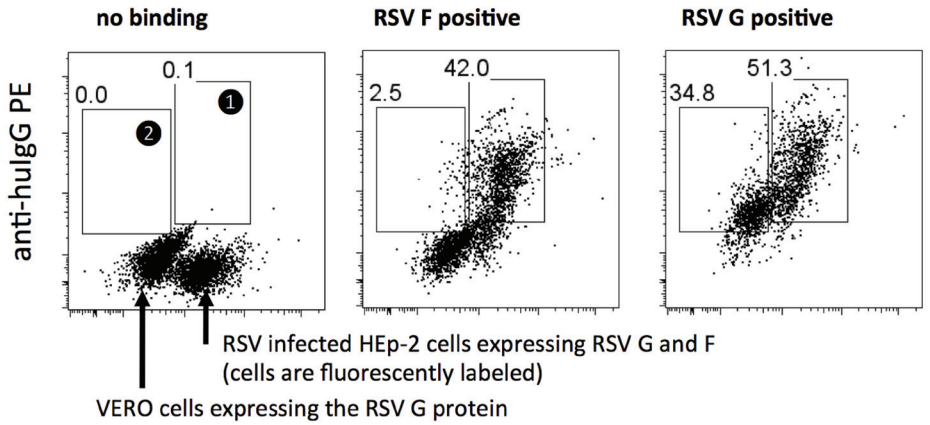


Figure S2. Screening of the IgG+CD27+ memory B cell repertoire of 3 healthy donors. BCL6 Bcl-xL transduced memory B cells of each donor were cultured in ten 96 well plates at 20 cells per well (19,200 cells per donor). Culture supernatants were tested for binding to CFSE labeled RSV A2 infected HEp2 cells (population 1) and Vero cells stably expressing RSV G from the X strain (population 2) and detected with anti-human IgG-PE by FACS.

CHAPTER FIVE

RSV Antibody Neutralization Efficacy in Multi-epitope Antibody Target Strategies

Lydia Tan[1], Barry Benaissa-Trouw[1], Marco C. Viveen[1], Emmanuel J.H.J. Wiertz[1], Frank E.J. Coenjaerts[1,], Mark Throsby[2]

[1] Department of Medical Microbiology, University Medical Center Utrecht, The Netherlands

[2] Merus Biopharmaceuticals, Utrecht, The Netherlands

Paper in preparation

Abstract

Human Respiratory Syncytial Virus (RSV) is an important causative agent of severe pneumonia and bronchiolitis in premature neonates and patients with weakened immunologic protection. Prophylactic antibody therapy is currently the only effective method to counteract RSV infections. This therapy comprises Palivizumab, a virus-neutralizing antibody directed against the fusion protein of RSV. Antibody binding impedes fusion of the viral envelope with the host cell membrane, thereby preventing infection. However, the existence of Palivizumab escape mutants and the high therapy costs urge the search for alternative antibody therapy strategies. Here, we studied the virus neutralization effect of multi-epitope antibody targeting on a panel of genetically divergent RSV strains by using a bi-specific antibody and combinatorial antibodies against two different proteins: the fusion (F) protein and attachment (G) protein. Virus reduction efficacy of these antibodies was higher than that observed for Palivizumab, whereas less IgG was required to gain 50% neutralization. The analyzed antibodies displayed broad efficacy against RSV strains with high genetic diversity. We conclude that multi-epitope targeting strategies offer a promising direction for the development of cost-effective RSV antibody therapies that cover a broad spectrum of RSV strain variants.

Introduction

Human respiratory syncytial virus (RSV) is worldwide the primary cause of severe lower respiratory tract infections during infancy. The virus affects premature neonates, but also high-risk groups such as immuno-compromised patients and elderly people [1-3]. Compared to patients with influenza or parainfluenza infections, a three times higher hospitalization rate has been observed for young infants infected with RSV [4]. By the age of two, the majority of children has been infected by RSV at least once [5].

RSV is an enveloped RNA virus belonging to the genus *Pneumovirus* of the *Paramyxoviridae* family. The two glycoproteins incorporated in the viral envelope, attachment protein (G) and fusion protein (F), are responsible for viral attachment and fusion with the host cell, respectively. These two proteins form the most essential antigenic targets for neutralizing antibodies. Important in this respect, antibody cross-reactivity assays showed that there are two distinct RSV antigenic subgroups (A and B) that have been further classified into genotypes based on genetic divergence in the G gene [6,7]. This genetic diversity plays a crucial role in immune escape [8,9].

RSV neutralizing antibodies are acquired during primary infection and play a major role in the protection against RSV. However, the balance in antibody concentration, specific antibody target and the pre-existing immunologic conditions of the patient strongly determine the outcome of RSV-induced disease. High antibody levels were shown to provide protection against RSV disease, but their protection is incomplete as re-infections in seropositive patients still occur [10,11]. In addition, low levels of nasal IgA and serum IgG to RSV F and G are significant risk factors for acquiring RSV infection [12]. For neonates, not the immunologic immaturity, but substantial pre-existing neutralizing maternal antibodies at time of RSV infection may negatively affect the development of a specific humoral response in RSV-naive infants [13]. Young adults and elderly have similar baseline serum antibody levels, but upon re-infection older patients have significantly increased humoral immune responses against the fusion protein of RSV [14,15]. These increased antibody levels correlated with protection against RSV infection in frail elderly [16].

Since the 1960's, vaccines and antibody therapies have been developed to protect patients from severe RSV disease. Formalin-inactivated RSV was the first candidate vaccine, but administration resulted in augmented cell-mediated immune responses and enhanced disease in children undergoing natural RSV infection [17,18]. Antibodies induced by this vaccine were directed against the F glycoprotein of RSV, although they were deficient in fusion-inhibiting activity [19]. Other vaccine design attempts with either RSV subunits or live attenuated virus showed to be ineffective in inducing adequate host immune responses or resulted in immunopathogenic damage of the host [20-22].

Identification of altered amino acid residues in antibody escape mutants and the evaluation of antibody binding to RSV F peptides revealed specific neutralizing epitopes, designated II and IV, on the F glycoprotein [23], which were later specified as site A (255-275) and site C (422-438) [24,25], respectively. These sites have been further studied for the development of therapeutic prophylactic antibody treatment. RSV-F site A led to the development of the humanized antibody Palivizumab (Synagis®) that is currently administered to high-risk patients as passive prophylaxis [26,27]. Although this monoclonal antibody binds to a highly conserved epitope [28], the existence of several Palivizumab escape mutants [29] makes efficient treatment of RSV-infected patients difficult. These mutants are induced when prolonged pulmonary RSV replication takes place in the presence of Palivizumab [30]. The affinity-optimized Motavizumab antibody that targets the same antigenic site as Palivizumab showed a 44-fold increase in neutralization capacity *in vitro*, but lacks bioavailability *in vivo* due to excessive aspecific binding to lung tissue [31]. Therefore, Motavizumab did not show any significant benefit over Palivizumab and is also unable to provide protection against Palivizumab escape mutants. In addition to the existence of escape mutants, the high costs of prophylactic treatment with Palivizumab [32] justify further search for alternative antibody treatment strategies.

It is therefore of great importance to continue research on Palivizumab substitutes and explore the correlation between antigenic diversity and the neutralization efficacy of RSV-specific antibodies. Characterization of the neutralization capacity of antibodies directed against different regions of the F protein towards a broad selection of clinical isolates [33,34] may yield important insights on the relative importance of various antibody recognition sites. Furthermore, the effect of

multi-epitope targeting on reduction of viral loads will provide us insight in the effectiveness with which escape mutants are counteracted.

Here, we studied the binding of mono-specific monoclonal antibodies and one bi-specific antibody directed against the RSV fusion protein. Furthermore, we have studied the effect in virus reduction with combinations of these antibodies and a monoclonal antibody directed against the attachment protein.

Materials and Methods

Viruses

RSV subgroup A and B clinical strains were isolated from nasopharyngeal swabs and aspirates of patients who were hospitalized between 2001 and 2012 in the University Medical Center of Utrecht, The Netherlands (Table 1) [34]. The isolates were grown at 37˚C, 5% CO_2 on Hep2 cells (CCL-23, American Type Culture Collection [ATCC]) in DMEM (Dulbecco's Minimal Essential Medium) with 5% fetal calf serum (FCS) and Normocin (100μg/ml, Invivogen). Viruses with typical RSV syncytia formation were stored at -80°C when 70% of monolayer cells displayed cytopathological effects (CPE). Virus titers of the clinical isolates and RSV reference strains Long (VR-26, ATCC) and 9320 (VR-955, ATCC) were determined by 24-well based plaque-purification on Hep2 cells.

RSV-specific human IgGs

The Fab arms of monoclonal human antibodies PG2729, PG2845 and PG3082 were initially obtained from donor-derived human Fab-phage display libraries constructed as described by Throsby *et al*. [35], but with the modification that all libraries were constructed with a single common human kappa light chain instead of the donor light chain repertoire. The Fabs were either selected against RSV fusion protein recombinantly expressed on the surface of 293T cells, or against a 15-mer peptide comprising the conserved central domain of RSV-G protein. Specific binding Fabs were recloned into IgG1 expression vectors, expressed and purified as described previously [36]. The IgG1 molecules PG2729 and PG2845 were shown to bind specifically to RSV-F and PG3082 to RSV-G (Table 2). To generate

Table 1. Selection of RSV clinical strains

Strain	Subgroup	Year of isolation	Genotype	Accession #
01-000583	A	2001	GA5	JQ901448
02-000110	A	2002	GA2	JQ901453
03-036544	A	2003	GA2	JQ901458
05-000417	A	2005	GA2	JX015486
06-000103	A	2006	GA5	JX015487
06-000827	A	2006	GA5	JX015488
RSV572	A	2007	GA2	JX015484
08-001411	A	2008	GA2	JX015494
11-000271	A	2011	GA2	JX015479
03-000005	B	2003	GB12	JX576760
05-001965	B	2005	GB13	JX576758
05-040058	B	2005	GB13	JX576755
06-039982	B	2006	GB13	JX576752
07-000431	B	2007	GB13	JX576751
07-038176	B	2007	GB13	JX576749

Table 2. Overview of RSV-specific human antibodies

Antibody	Type	Target	Site	Structure	Reference
PG2729	Monoclonal	F	M2	Prefusion	N.A
PG2845	Monoclonal	F	M1	Postfusion	N.A
Ch101F	Monoclonal	F	C	Postfusion	Mclellan et al. 2010 J Virology
D25	Monoclonal	F	X	Prefusion	Kwakkenbos et al. 2010 Nature Medicine
Palivizumab	Monoclonal	F	A	Pre-/postfusion	Mclellan et al. 2010 Nat struct Mol Biol
PB2172	Bispecific	F	M1/M2	Pre-/postfusion	N.A
PG3082	Monoclonal	G	Conserved region in ectodomain	N.A	N.A

the next generation bispecific antibody PB2172 (Biclonics™, http://www.merus.nl/products/ cobra.html), complementary modifications were made in the CH3 region of PG2729 and PG2845 to favor heterodimerization of heavy chain pairs over normal homodimerization as described [37]. PB2172 was expressed and purified as described above.

Plaque-reduction neutralization assay

The complete panel of novel RSV-specific human antibodies developed by Merus Biopharmaceuticals, the RSV reference sera Ch101F, D25, and Palivizumab [25,38-40] were tested in plaque reduction neutralization assays. PG1207, a nonspecific IgG directed against tetanus toxoid was used as a negative control. Monoclonal antibodies were diluted by three-step serial dilutions in DMEM with 2% FCS and 100μg/ml Normocin pre-incubated with 60 pfu virus for two hours at 37°C. Virus-antibody mixtures were then applied to Hep2 monolayers in 24-well cell culture plates (Corning Costar) that were seeded at a density of $8x10^4$ cells/well 48 hours pre-infection. After two hours of viral adherence at 37°C and 5% CO_2, monolayers were covered by a 0.5% Seakam LE agarose overlay diluted in DMEM FCS 5% with Normocin. On day 5 post-infection, when virus plaque formation was clearly visible, the cells were fixed with 10% formalin and stained using a 0.5% crystal violet/ 20% ethanol (w/v) staining solution. Plaques were counted for each dilution and depicted in plaque-reduction curves for each antibody separately. Neutralizing antibody titers, responsible for 50% plaque-reduction relative to the mock-treated virus-infected cells, were calculated from these curves.

ELISA-based antibody binding test

Hep2 cells were seeded at $8x10^4$ cells/well in 96-well cell culture plates (Corning Costar) and infected with RSV clinical isolates and reference strains 48 hours later. Two days post-infection, cells were fixed with 0.025% glutaraldehyde in Dulbecco's phosphate buffered saline (DPBS, Lonza) at room temperature for 20 minutes. Cells were then washed twice with tap water and one time with PBS. Ten-fold serial antibody dilutions ranging from 10-10,000ng/ml were prepared in PBS antibody diluent supplemented with Tween-20 (0.05%) and 4% ELK. Antibody dilutions were applied to the fixed cells at 37˚C for two hours. Four washing steps with tap

water and one with PBS removed all unbound antibodies. The mouse anti-human IgG_1 (Fc) – HRPO conjugate (Millipore) was diluted 1:10,000 and incubated at 37˚C for one hour. Similar washing steps as for the primary unbound antibodies removed all unbound conjugate. Antibody binding was visualized after addition of the HRPO enzyme substrate solution (0.11M sodium-acetate pH 5.5, 0.025M 3,3',5,5'-tetramethylbenzidine (TMB), 2mM urea peroxide) for approximately 20 minutes. The oxidation reaction was terminated by adding 0.18M H_2SO_4, thereby changing the color from yellow to blue. The color intensity, indicating the degree of antibody binding, was measured at OD 450nm.

Results

Minor differences in virus neutralization between RSV subgroups

Nine subgroup A and six subgroup B clinical strains from different epidemic years and with different genotypes [33,34] were exposed to RSV-specific human IgGs directed against non-overlapping epitopes of the fusion protein. The virus neutralization efficacy of these antibodies was measured by plaque-reduction assays. Neutralizing antibody titers represent the antibody concentration required to gain 50% virus reduction relative to the non-treated virus-infected cells. The neutralization results for the clinical strains were compared to the efficacies of antibodies on the commonly used Long and 9320 reference strains. Antibody PG2845 has a similar range of virus reduction efficacy as Palivizumab (Figure 1A). The PG2729 antibody, however, shows increased efficacy for subgroup A strains compared to this clinically used Palivizumab (0.85 log less IgG required), but is not as potent as the two reference antibodies, Ch101F and D25. These antibodies require one (1.05 log, RSV A; 1.24 log, RSV B) to two (1.98 log, RSV A; 1.74 RSV B) logs less antibody for 50% virus reduction, respectively. Palivizumab and D25 have similar neutralizing efficacy for both RSV subgroup A and B strains. The antibodies PG2729, PG2845 and to some extend Ch101F, however, show minor differences in neutralization efficacy between the two RSV subgroups ranging from 0.74 to 0.33 logs, respectively. The maximum mutual strain difference in neutralization efficacy shows the broadness in strain coverage for each antibody. For subgroup A strains the mutual strain difference is 1.27 log when using PG2845 in the plaque-reduction assays (Figure

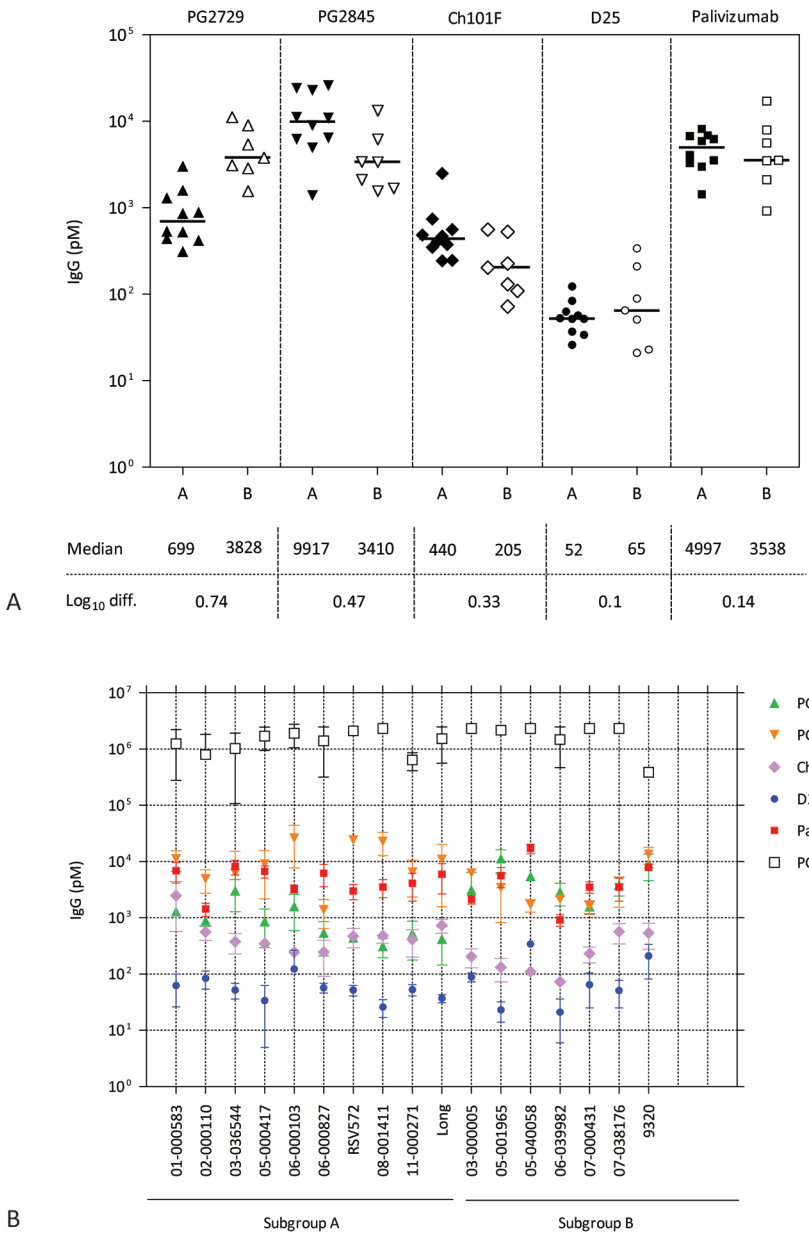


Figure 1. Virus neutralization by RSV-F directed monoclonal antibodies. A) Comparison of the virus reduction efficacy for all monoclonal antibodies on each RSV subgroup and B) on all individual virus strains. Each dot represents the neutralizing antibody titer, responsible for 50% virus reduction relative to the non-treated RSV-infected cells.

1B and Supplementary Table 1). Other antibodies have similar, but slightly lower maximum mutual strain differences than PG2845. Palivizumab shows the highest mutual difference for subgroup B strains (1.27 log). The antibody binding affinity of each monoclonal determined by an ELISA-based binding assay showed a diversity of affinity among strains (Figure 2). However, this binding divergence does not correlate to the neutralization efficacy determined by the plaque-reduction assays, assuming that epitope conformation could play a role in neutralization activity.

Increased virus neutralization efficacy using bispecific antibody PB2172

Antibody monomers with variable Fab fragments of PG2729 and PG2845 were combined via CH3 hetero-dimerization to create a single bi-specific antibody, PB2172. Comparison of the neutralization efficacy between mono-specific and bi-specific antibodies shows that PB2172 reduces the viral load for both RSV subgroups more efficiently than mono-specific PG2729 and PG2845 (Figure 3). Particularly, this increased effect is stronger among subgroup A strains when comparing the neutralization efficacy of PB2172 to the virus reduction capacity of

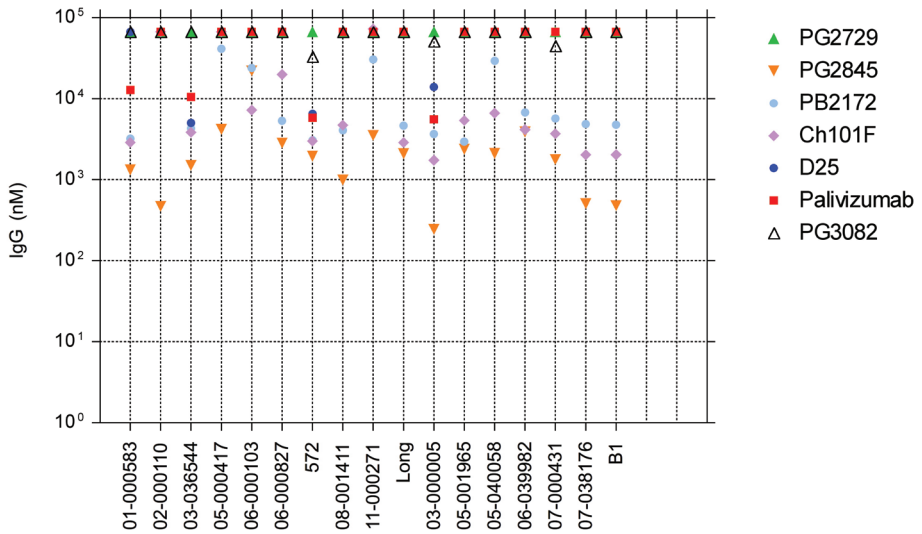


Figure 2. Antibody binding potential of the bispecific and monoclonal antibodies. Each dot represents the antibody titer that shows two times more binding to the virus-infected cells than the mock-treated cells.

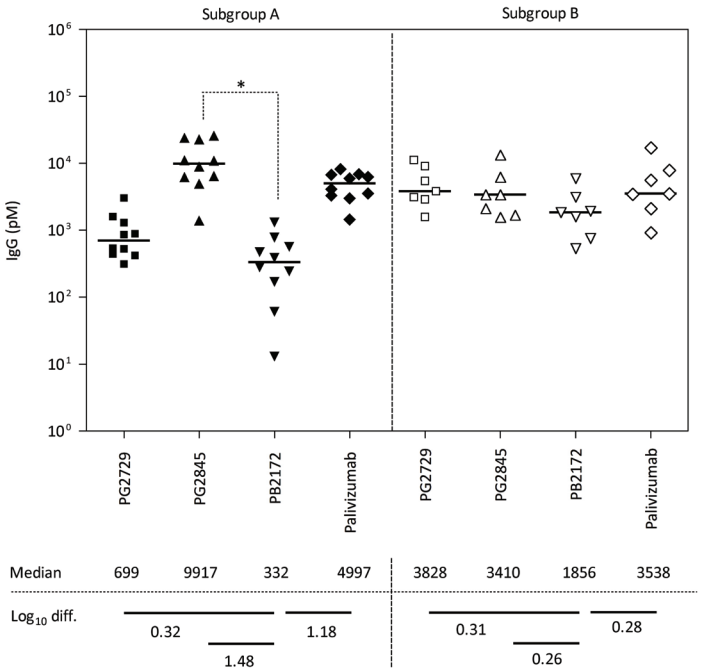


Figure 3. Neutralization efficacy of bispecific antibody PB2172. Virus reduction effect per RSV subgroup. Each dot represents the neutralizing antibody titer, responsible for 50% virus reduction relative to the non-treated RSV-infected cells. Significant improvement of efficacy with P = <0.05 is marked by an asteriks.

Palivizumab (1.18 log, RSV A; 0.28 log, RSV B). PB2172 is especially effective on the Long strain and clinical strain 11-000271 (up to 2.5 log compared to palivizumab, Supplementary Figure 1). The viral load of RSV strain 08-001411, however, is less well reduced by the bi-specific antibody when compared to mono-specific PG2729. The maximum mutual difference in RSV subgroup A strain neutralization is increased by approximately 1 log when using PB2172 instead of PG2729 or PG2845 (Supplementary Table 1). Also among the subgroup B strains the maximum mutual difference in efficacy is marginally increased, albeit non-significantly.

Additional RSV-G antibody genotype-independently improves efficacy of virus neutralization

We analyzed antibody combinations made by combining each of the four RSV

F-directed antibodies with the RSV G-directed antibody PG3082. The neutralization efficacy of these antibody mixtures was compared to the virus-reducting capacity of PG2729, 2845, Ch101F, bi-specific antibody PB2172 and PG3082 alone. Final antibody concentrations for pre-incubation with RSV were equal for both the single and combinatorial antibodies. Addition of PG3082 increased the median neutralization efficacy in all cases, but the strongest increase was observed for the combination with PB2172 on subgroup B strains (0.97 log; Figure 4A). Strikingly, the combination of RSV F-directed antibodies with PG3082 had the largest effect on reducing the viral load of the Long and 9320 reference strains (Supplementary figure

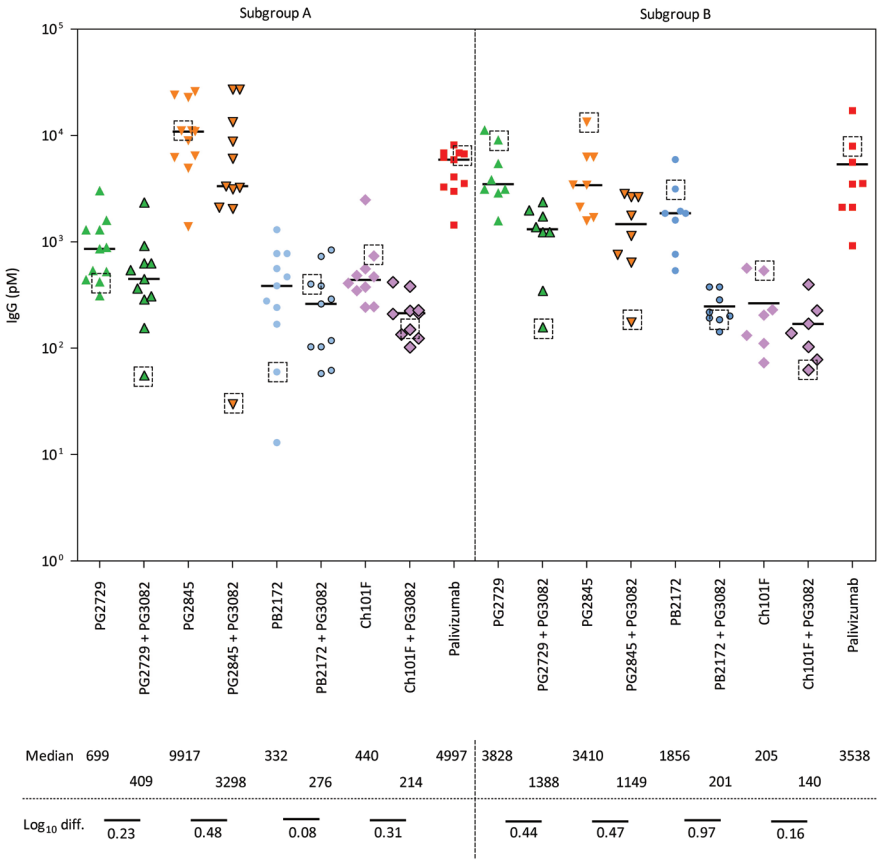


Figure 4. RSV-F and -G antibody combinations for improved neutralization efficacy. Virus reduction effect on each RSV subgroup. Each dot represents the neutralizing antibody titer, responsible for 50% virus reduction relative to the non-treated RSV-infected cells. Dashed boxes represent Long strain (RSV-B) and 9320 strain (RSV-B)

2A-D and marked with dashed boxes in Figure 4A). Generally, much lower antibody concentrations are required to gain 50% virus reduction for these two strains in combination therapy. However, the combination of PB2172 with PG3082 on the Long strain is an exception as addition of the RSV-G directed antibody decreases the neutralization efficacy compared to PB2172 alone. Although PG3082 increases the efficacy of RSV-F antibody treatment during RSV infection, this antibody did not show any intrinsic neutralization potential against the RSV strain panel. The mutual difference in neutralization efficacy between RSV subgroup A strains is strongly increased when adding PG3082 to PG2845 (1.69 log increase) and is therefore responsible for the most diverse effect in virus reduction (Supplementary Table 1). All RSV F and PG3082 combinations, except for the combination with PB2172 (0.42 log), showed similar maximum mutual differences among subgroup B strains (approximately 1 log).

Discussion

In this study we analyzed the virus neutralization effects of combinatorial antibody strategies and bi-specific antibody targeting. The effectiveness of various antibody preparations was assessed on both RSV subgroups comprising a selection of genetically divergent strains that were sampled over ten years in the Netherlands. Thereafter, we compared the results obtained with the selected patient strains to the results obtained with the Long and 9320 reference strains.

All mono-specific, bi-specific and combinatorial antibodies could neutralize each strain in the subgroup A and B strain selection, although with different efficacies. We did not detect natural escape mutants for any of the antibodies. Considerable differences in virus reduction were found between RSV subgroups A and B, as observed with PG2729, PG2845, Ch101F and PB2172. However, no correlation could be found between the neutralization efficacy of the antibodies and their binding capacity as determined by ELISA binding tests. The binding affinities of the analyzed antibodies were generally low to, in some cases, undetectable. This could be the consequence of epitope inaccessibility.

Bi-specific antibody PB2172 had a highly neutralizing capacity on subgroup A and

B strains. The increase in virus reduction by PB2172 was 0.32 log and 0.26 log compared to the best-contributing mono-specific parent PG2729 (for subgroup A) and PG2845 (for subgroup B), respectively. The combinatorial antibody strategy with epitope targets on both the RSV F and G proteins resulted in enhanced neutralization efficacy as well. RSV G-directed PG3082 did not neutralize, but could enhance the virus reduction efficacy when combined with RSV F-directed antibodies, albeit non-significantly. This suggests a stronger role for RSV F as neutralizing antibody target and might be correlated to the fact that RSV G deletion mutants were also able to establish infection *in vitro* [41]. This study showed that both the bi-specific antibody and combinatorial antibody strategy required less IgG for 50% reduction of the viral load compared to Palivizumab, but the efficacy was lower compared to reference sera Ch101F and D25.

The major advantage of IgG bi-specificity is that it allows to target more strain variants that normally escape from mono-specific antibodies. The effectiveness, however, strongly differs as some strains have both epitopes available and other strains might have only one free epitope. Alternative combinations of Fab fragments may increase neutralization efficacies even more than what we have found for the PG2729 and PG2845 Fab fragment combination. A second putative advantage of bi-specific antibodies compared to monoclonal antibody mixtures is the reduced steric hindrance when targeting multiple epitopes on one protein. However, steric hindrance with monoclonal antibody mixtures seems to be less of an issue when different proteins are targeted. In addition, the multi-epitope targeting therapy strategies may form a higher genetic barrier that restricts the development of antibody escape mutants. Both therapeutic approaches presented in this study look promising and will help the field to find the optimal antibody combination that covers all strain variants and subgroups with a similar neutralization efficacy. Finding the ideal Fab fragment or monoclonal antibody combination with the lowest required effective concentration will ultimately lead to a highly cost-effective antibody therapy with broad strain coverage.

References

1. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, et al. (2010) *Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis*. Lancet 375: 1545-1555.

2. Englund JA, Anderson LJ, Rhame FS (1991) *Nosocomial transmission of respiratory syncytial virus in immunocompromised adults*. J Clin Microbiol 29: 115-119.

3. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE (2005) *Respiratory syncytial virus infection in elderly and high-risk adults*. N Engl J Med 352: 1749-1759.

4. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, et al. (2009) *The burden of respiratory syncytial virus infection in young children*. N Engl J Med 360: 588-598.

5. Henrickson KJ, Hoover S, Kehl KS, Hua W (2004) *National disease burden of respiratory viruses detected in children by polymerase chain reaction*. Pediatr Infect Dis J 23: S11-18.

6. Mufson MA, Orvell C, Rafnar B, Norrby E (1985) *Two distinct subtypes of human respiratory syncytial virus*. J Gen Virol 66 (Pt 10): 2111-2124.

7. Anderson LJ, Hierholzer JC, Tsou C, Hendry RM, Fernie BF, et al. (1985) *Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies*. J Infect Dis 151: 626-633.

8. Zhu Q, Patel NK, McAuliffe JM, Zhu W, Wachter L, et al. (2012) *Natural Polymorphisms and Resistance-Associated Mutations in the Fusion Protein of Respiratory Syncytial Virus (RSV): Effects on RSV Susceptibility to Palivizumab*. J Infect Dis 205: 635-638.

9. Walsh EE, Falsey AR, Sullender WM (1998) *Monoclonal antibody neutralization escape mutants of respiratory syncytial virus with unique alterations in the attachment (G) protein*. J Gen Virol 79 (Pt 3): 479-487.

10. Kruijsen D, Bakkers MJ, van Uden NO, Viveen MC, van der Sluis TC, et al. (2010) *Serum antibodies critically affect virus-specific CD4+/CD8+ T cell balance during respiratory syncytial virus infections*. J Immunol 185: 6489-6498.

11. Mufson MA, Belshe RB, Orvell C, Norrby E (1987) *Subgroup characteristics of respiratory syncytial virus strains recovered from children with two consecutive infections*. J Clin Microbiol 25: 1535-1539.

12. Walsh EE, Falsey AR (2004) *Humoral and mucosal immunity in protection from natural respiratory syncytial virus infection in adults*. J Infect Dis 190: 373-378.

13. Shinoff JJ, O'Brien KL, Thumar B, Shaw JB, Reid R, et al. (2008) *Young infants can develop protective levels of neutralizing antibody after infection with respiratory syncytial virus*. J Infect Dis 198: 1007-1015.

14. Walsh EE, Falsey AR (2004) *Age related differences in humoral immune response to respiratory syncytial virus infection in adults*. J Med Virol 73: 295-299.

15. Falsey AR, Walsh EE, Looney RJ, Kolassa JE, Formica MA, et al. (1999) *Comparison of respiratory syncytial virus humoral immunity and response to infection in young and elderly adults*. J Med Virol 59: 221-226.

16. Falsey AR, Walsh EE (1998) *Relationship of serum antibody to risk of respiratory syncytial virus infection in elderly adults*. J Infect Dis 177: 463-466.

17. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE (1969) *An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine*. Am J Epidemiol 89: 405-421.

18. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, et al. (1969) *Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine*. Am J Epidemiol 89: 422-434.

19. Murphy BR, Walsh EE (1988) *Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity*. J Clin Microbiol 26: 1595-1597.

20. Tristram DA, Welliver RC, Mohar CK, Hogerman DA, Hildreth SW, et al. (1993) *Immunogenicity and safety of respiratory syncytial virus subunit vaccine in seropositive children 18-36 months old*. J Infect Dis 167: 191-195.

21. Falsey AR, Walsh EE (1997) *Safety and immunogenicity of a respiratory syncytial virus subunit vaccine (PFP-2) in the institutionalized elderly*. Vaccine 15: 1130-1132.

22. Murphy BR, Hall SL, Kulkarni AB, Crowe JE, Jr., Collins PL, et al. (1994) *An update on approaches to the development of respiratory syncytial virus (RSV) and parainfluenza virus type 3 (PIV3) vaccines*. Virus Res 32: 13-36.

23. Arbiza J, Taylor G, Lopez JA, Furze J, Wyld S, et al. (1992) *Characterization of two antigenic sites recognized by neutralizing monoclonal antibodies directed against the fusion glycoprotein of human respiratory syncytial virus*. J Gen Virol 73 (Pt 9): 2225-2234.

24. Beeler JA, van Wyke Coelingh K (1989) *Neutralization epitopes of the F glycoprotein of respiratory syncytial virus: effect of mutation upon fusion function*. J Virol 63: 2941-2950.

25. McLellan JS, Chen M, Chang JS, Yang Y, Kim A, et al. (2010) *Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F*. J Virol 84: 12236-12244.

26. Meissner HC, Long SS (2003) *Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus infections*. Pediatrics 112: 1447-1452.

27. Feltes TF, Cabalka AK, Meissner HC, Piazza FM, Carlin DA, et al. (2003) *Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease*. J Pediatr 143: 532-540.

28. DeVincenzo JP, Hall CB, Kimberlin DW, Sanchez PJ, Rodriguez WJ, et al. (2004) *Surveillance of clinical isolates of respiratory syncytial virus for palivizumab (Synagis)-resistant mutants*. J Infect Dis 190: 975-978.

29. Zhao X, Chen FP, Megaw AG, Sullender WM (2004) *Variable resistance to palivizumab in cotton rats by respiratory syncytial virus mutants*. J Infect Dis 190: 1941-1946.

30. Zhao X, Sullender WM (2005) *In vivo selection of respiratory syncytial viruses resistant to palivizumab*. J Virol 79: 3962-3968.

31. Wu H, Pfarr DS, Johnson S, Brewah YA, Woods RM, et al. (2007) *Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract*. J Mol Biol 368: 652-665.

32. Kamal-Bahl S, Doshi J, Campbell J (2002) *Economic analyses of respiratory syncytial virus immunoprophylaxis in high-risk infants: a systematic review*. Arch Pediatr Adolesc Med 156: 1034-1041.

33. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, et al. (2012) *Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology*. PLoS One 7: e51439.

34. Tan L, Coenjaerts FE, Houspie L, Viveen MC, van Bleek GM, et al. (2013) *The Comparative Genomics of Human Respiratory Syncytial Virus Subgroup A and B: Genetic Variability and Molecular Evolutionary Dynamics*. J Virol.

35. Throsby M, de Kruif J (2009) *Construction of phage antibody repertoires from the blood of West Nile virus-infected donors*. Methods Mol Biol 562: 45-60.

36. Throsby M, Geuijen C, Goudsmit J, Bakker AQ, Korimbocus J, et al. (2006) *Isolation and characterization of human monoclonal antibodies from individuals infected with West Nile*

Virus. J Virol 80: 6982-6992.

37. Gunasekaran K, Pentony M, Shen M, Garrett L, Forte C, et al. (2010) *Enhancing antibody Fc heterodimer formation through electrostatic steering effects: applications to bispecific molecules and monovalent IgG*. J Biol Chem 285: 19637-19646.

38. Kwakkenbos MJ, Diehl SA, Yasuda E, Bakker AQ, van Geelen CM, et al. (2010) *Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming*. Nat Med 16: 123-128.

39. McLellan JS, Chen M, Kim A, Yang Y, Graham BS, et al. (2010) *Structural basis of respiratory syncytial virus neutralization by motavizumab*. Nat Struct Mol Biol 17: 248-250.

40. Swanson KA, Settembre EC, Shaw CA, Dey AK, Rappuoli R, et al. (2011) *Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers*. Proc Natl Acad Sci U S A 108: 9619-9624.

41. Karron RA, Buonagurio DA, Georgiu AF, Whitehead SS, Adamus JE, et al. (1997) *Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant*. Proc Natl Acad Sci U S A 94: 13961-13966.

SUPPLEMENTARY INFO

CHAPTER 5

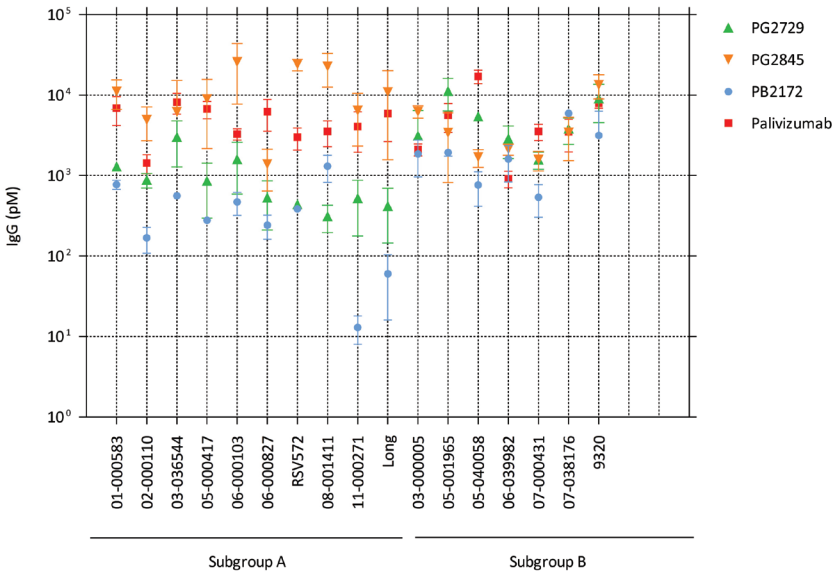


Figure S1. Neutralization efficacy of bispecific antibody PB2172 per strain. Each dot represents the neutralizing antibody titer, responsible for 50% virus reduction relative to the non-treated RSV-infected cells.

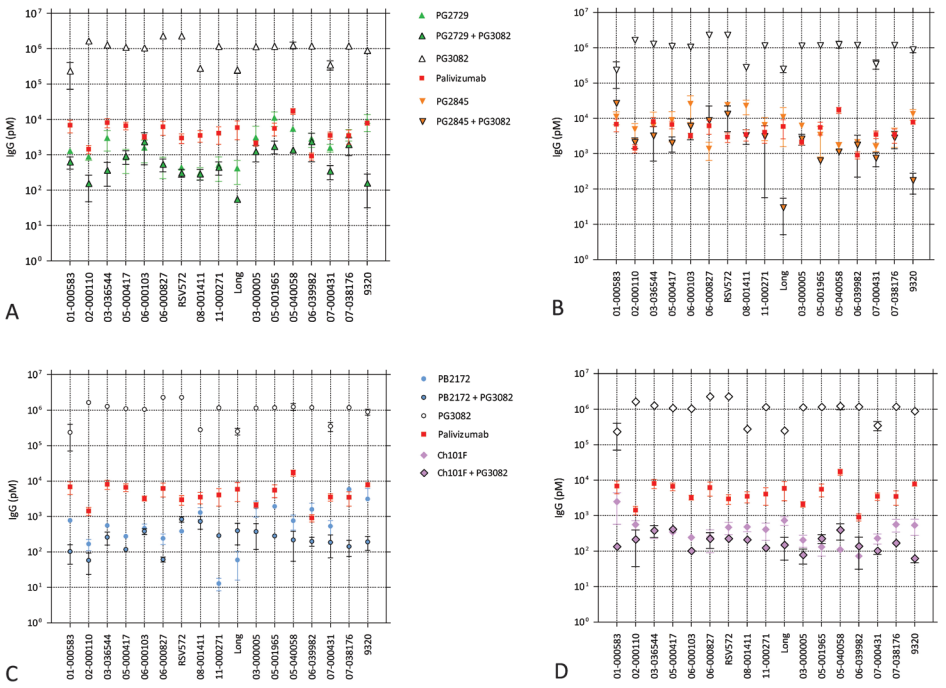


Figure S2. Viral load reduction per strain with combinatorial antibodies. Neutralization efficacy per RSV strain with antibody combinations with PG3082 and A) PG2729, B) PG2845, C) PB2172 and D) Ch101F. Each dot represents the neutralizing antibody titer, responsible for 50% virus reduction relative to the non-treated RSV-infected cells.

Table S1. Exact antibody concentrations required for 50% virus reduction

	PG2729		PG2845		PB2172	
Strain	pM	µg/ml	pM	µg/ml	pM	µg/ml
01-000583	1296	0.193	11066	1.652	776	0.116
02-000110	884	0.132	4923	0.735	168	0.025
03-036544	3029	0.452	6206	0.926	562	0.084
05-000417	861	0.129	8933	1.333	278	0.041
06-000103	1595	0.238	25818	3.853	469	0.070
06-000827	536	0.080	1386	0.207	242	0.036
RSV572	440	0.066	24002	3.582	386	0.058
08-001411	312	0.047	22736	3.394	1305	0.195
11-000271	526	0.079	6434	0.960	13	0.002
Long (ATCC VR-26)	420	0.063	10901	1.627	60	0.009
Mutual difference (Log_{10})	0.99		1.27		2.01	
03-000005	3129	0.467	6250	0.933	1856	0.277
05-001965	11249	1.679	3410	0.509	1942	0.290
05-040058	5447	0.813	1687	0.252	764	0.114
06-039982	2879	0.430	2121	0.317	1602	0.239
07-000431	1580	0.236	1573	0.235	537	0.080
07-038176	3828	0.571	3423	0.511	5943	0.887
9320 (ATCC VR-955)	9085	1.356	13383	1.998	3149	0.470
Mutual difference (Log_{10})	0.85		0.93		1.04	

Table S1. Exact antibody concentrations required for 50% virus reduction (continued)

	Ch101F		D25		Palivizumab	
Strain	pM	µg/ml	pM	µg/ml	pM	µg/ml
01-000583	2486	0.371	63	0.009	6868	1.025
02-000110	560	0.084	84	0.013	1436	0.214
03-036544	377	0.056	52	0.008	8165	1.219
05-000417	348	0.052	34	0.005	6722	1.003
06-000103	244	0.036	123	0.018	3277	0.489
06-000827	246	0.037	57	0.009	6219	0.928
RSV572	471	0.070	52	0.008	2990	0.446
08-001411	484	0.072	26	0.004	3536	0.528
11-000271	409	0.061	53	0.008	4081	0.609
Long (ATCC VR-26)	737	0.110	37	0.005	5913	0.882
Mutual difference (Log_{10})	0.86		0.68		0.68	
03-000005	205	0.031	89	0.013	2115	0.316
05-001965	132	0.020	23	0.003	5622	0.839
05-040058	111	0.017	341	0.051	17130	2.557
06-039982	73	0.011	21	0.003	921	0.137
07-000431	230	0.034	65	0.010	3538	0.528
07-038176	566	0.084	51	0.008	3495	0.522
9320 (ATCC VR-955)	536	0.080	210	0.031	7951	1.187
Mutual difference (Log_{10})	0.89		1.21		1.27	

Table S1. Exact antibody concentrations required for 50% virus reduction (continued)

	PG2729 + PG3082		PG2845 + PG3082		PB2172 + PG3082	
Strain	pM	µg/ml	pM	µg/ml	pM	µg/ml
01-000583	631	0.094	27243	4.066	103	0.015
02-000110	156	0.023	2119	0.316	58	0.009
03-036544	368	0.055	3253	0.486	261	0.039
05-000417	922	0.138	2054	0.307	118	0.018
06-000103	2352	0.351	6081	0.908	386	0.058
06-000827	545	0.081	8821	1.317	62	0.009
RSV572	311	0.046	13487	2.013	840	0.125
08-001411	290	0.043	3344	0.499	731	0.109
11-000271	449	0.067	3140	0.469	290	0.043
Long (ATCC VR-26)	56	0.008	30	0.005	400	0.060
Mutual difference (Log_{10})	1.62		2.96		1.16	
03-000005	1243	0.185	2650	0.396	376	0.056
05-001965	1745	0.260	644	0.096	285	0.043
05-040058	1388	0.207	1149	0.171	219	0.033
06-039982	2373	0.354	1784	0.266	201	0.030
07-000431	350	0.052	762	0.114	186	0.028
07-038176	1988	0.297	2848	0.425	143	0.021
9320 (ATCC VR-955)	159	0.024	177	0.026	193	0.029
Mutual difference (Log_{10})	1.00		1.17		0.42	

Table S1. Exact antibody concentrations required for 50% virus reduction (continued)

	Ch101F + PG3082		PG3082	
Strain	pM	µg/ml	pM	µg/ml
01-000583	136	0.020	237189	35
02-000110	215	0.032	1668652	249
03-036544	385	0.058	1293910	193
05-000417	421	0.063	1104442	165
06-000103	103	0.015	1063814	159
06-000827	226	0.034	2331600	348
RSV572	229	0.034	2331600	348
08-001411	212	0.032	282903	42
11-000271	125	0.019	1170148	175
Long (ATCC VR-26)	151	0.023	253892	38
Mutual difference (Log_{10})	0.61		0.99	
03-000005	79	0.012	1157849	173
05-001965	228	0.034	1180017	176
05-040058	400	0.060	1258534	188
06-039982	140	0.021	1207711	180
07-000431	104	0.015	353352	53
07-038176	172	0.026	1200332	179
9320 (ATCC VR-955)	63	0.009	898638	134
Mutual difference (Log_{10})	0.80		0.55	

CHAPTER SIX

Summarizing Discussion

Summarizing discussion

In this thesis, the RSV evolutionary dynamics and genetic variability among whole genome sequences have been described in detail for a representative panel of RSV clinical strains. In addition, the virus neutralization efficacies of different antibody therapies were studied on clinical strains selected from the evolutionary dynamics studies.

RSV-directed immunologic support and its drawbacks

Human Respiratory Syncytial Virus (RSV) is worldwide the main causative agent of airway infections among neonates, but is also more frequently found in elderly people and immuno-compromised patients with serious lower respiratory tract infections [1,2]. These patients either have a naïve immunologic status (neonates) or lack the specific naturally acquired protection against RSV and therefore require protective support via passive antibody prophylaxis or active immunization through vaccination. The latter has shown to be difficult as has been previously observed in clinical studies with the formalin-inactivated RSV vaccine [3]. Vaccinees experienced adverse effects upon subsequent RSV infections due to immune exacerbation. The vaccination disturbed the immunologic balance, resulting in immune disease that even exceeded the burden of RSV infection. Support like antibody prophylaxis, on the other hand, will only temporarily protect patients, but can give strong protection within that period and does not disturb the immunologic balance as observed with vaccination. Although still under debate, RSV may cause or enhance airway sensitization that could result in hyper-responsive airways and asthma on the long term [4-6]. Prophylactic treatment with RSV neutralizing antibodies may therefore reduce the amount of wheezing days and intervene with asthma development as well[7,8].

After decennia of extensive research, many protein functions, protein structures and protein-host cell interactions in relation to RSV infection have been revealed [9,10]. However, in our goal to efficiently battle RSV infection and suppress the consequent burden, we have not succeeded to develop cost-effective and safe therapeutic strategies, nor an effective and safe vaccine. Therapeutic support is specifically focused on the intervention of RSV entry into host cells. Although the

exact mechanisms by which viral entry occurs are still not completely understood, we know that the RSV fusion protein plays a crucial role in this process, in which RSV G is an important mediator [11,12]. Both viral glycoproteins are key antigens for natural acquirement of antibodies and are used for the development of subunit vaccines and selection of prophylactic antibodies [13,14]. The RSV fusion protein-directed Palivizumab is one of the antibodies that is currently used in the clinics for prophylaxis[15]. However, the existence of escape variants [16-18], the high-dose requirements and related high costs [19,20] drives the search for alternative therapeutic strategies. A general overview of RSV and the current RSV-directed strategies has been described in chapter 1 of this thesis.

New insights into RSV pathogenesis via whole genome analysis

Information regarding strain variation is essential in vaccine development and antibody prophylaxis. Nevertheless, strain variability has primarily been studied for only specific RSV genes, i.e. those encoding fusion protein (F) and attachment protein (G) [21,22]. Data regarding variation in the whole genome context are lacking. Two molecular studies by Kumaria et al. and Rebuffo-Scheer et al. investigated a discrete set of complete genome sequences and elaborated on the overall gene and protein sequence variability [23,24]. The latter study also included evolutionary rate analyses on the G gene. However, research into evolutionary dynamics of other genes or the whole genome was not included.

In chapter 2 of this thesis we further extended RSV whole genome analyses with detailed evolutionary dynamics based on Bayesian genealogical inference [25] and applied this on a broader set of RSV strain sequences. The rate of evolution, which is the footprint of mutational changes over time, gives us an indication on the need for viral adaptation in order to "survive". Chapter 2 describes the whole genome variability within RSV subgroup A clinical strains and includes the degree of evolution for whole genome sequences and gene partition sequences [26]. Substitutions in RSV-A genes and proteins that have been plotted in order to determine mutational hotspots showed that the G and M2-2 gene were most variable together with non-coding regions. This emphasizes the importance of variation in these genes for virus endurance. However, the M2-2 variability was primarily related to reference strains and not to clinical strain variation. This might be the result of strain adaptation to lab

culture conditions or caused by the fact that these reference strains were sampled long before the sample time period of the clinical strains. In addition, evolutionary analyses demonstrates that RSV-A has a low degree of evolution compared to e.g. influenza and that directional selection is primarily found in the highly variable regions of the RSV attachment protein G. Several changes on these spots influence glycosylation potential and could therefore influence antibody recognition, but also protein functionality and even the course of infection.

In chapter 3 the evolutionary dynamics study was expanded by comparing subgroup A data with data from RSV subgroup B clinical strains [27]. This study showed high variability in G, SH and M2-2 genes between the two RSV subgroups. For both RSV-A and RSV-B the highest level of inter subgroup variability was found in the G gene, which was primarily the result of G gene sequence length variation among RSV-B strains. Although both subgroups have similar evolutionary rates, the time to the most recent common ancestor (TMRCA) for RSV-B strains was related to a more recent phylogeny root than the TMRCA for RSV-A strains. This recent TMRCA may be the consequence of a specific strain variant that is driven to fixation by some advantageous mutation in its genome or it could be the result of a chance fixation of a specific strain variant. The latter refers to the mutation frequency in a certain population that gives a certain chance that a mutant allele takes over the entire population.

Selective pressure analysis on RSV-A and RSV-B strain panels revealed that the majority of mutational sites in RSV genes are under neutral selection. This indicates that RSV does not require a strong selection of advantageous genomic changes for optimal infection, virus transmission or adaptation. Still a small selection of sites in the G gene evolves under pervasive selection and is located in the hyper variable regions. These sites are therefore thought to be involved in RSV evasion from antibody recognition.

Potential pitfalls when performing RSV phylodynamic analyses

Performing RSV phylodynamic analyses provides us with valuable data regarding gene evolutionary rates, strains common ancestors and genetic regions under selective pressure. However, the selection and preparation of the dataset and the choice of analytical methods can influence the outcome of the investigation. Here

we listed the most crucial pitfalls that can affect phylodynamic results.

First, putative biases of phylodynamic studies would definitely include A. a lack of viral strain samples from human individuals with asymptomatic infections and, B. a small sample size and origin of samples from limited geographical different places. With respect to A, we exclusively looked at RSV strain variants from patients with symptomatic infections. Therefore we do not know whether asymptomatic patients are infected by similar strains or completely different strain variants that might act as an additional RSV reservoir. From our evolutionary data we know that RSV is not highly variable and we could therefore assume that both symptomatic and asymptomatic patients are exposed to similar strains from the circulating variants. However, it could also be that a specific group of strain variants always induces symptomatic infections and we only isolated these specific variants. Although this seems unlikely, a future well-defined study including asymptomatic patients should exclude these assumptions. Remark B points out that we only included strains from two closely related geographical regions (Netherlands and Belgium) and one distant geographical region (Milwaukee). Although this study provided sufficient strain sequences for reliable phylodynamic analyses, we do not know what the influence of strain sequences from other distant geographical regions will have upon the overall evolutionary rate and selective pressure analyses. Future additional whole RSV genome sequence data sets can improve this image.

An additional pitfall in phylodynamic analysis is re-assortment events, i.e. recombination, in genomic sequences that complicates the interpretation of demographic inferences. Each strain must be represented by a single sequence in order to determine the relationship between viral coalescent rates and the change in effective number of infections through time. Mosaic sequences such as in recombinant strain variants complicate these analyses. Thorough checking of recombination events before performing phylodynamic analysis is therefore a prerequisite. In addition, unaccountable recombination can yield false evidence for positive selection. The novel mixed-effects model of evolution (MEME) method that can distinct pervasive selection from episodic selection is less sensitive to this issue [28]. This method captures episodic diversifying selection by disentangling lineages under purifying and positive selection per site.

Advantages for future continuous work in RSV phylodynamics might be the use

of improved high-throughput sequencing, which would enable us to study larger sample sizes in a shorter amount of time. However, in our analyses in silico assembly of the short random fragments (<500 nucleotides in size) in the current deep-sequence sequencing method strongly increased the amount of PCR artifacts and artificial recombinants. This is important and should be taken into account for future experimental setups. The elaborative conventional Sanger sequence method that we used in our genomic studies contained larger specific size fragments (up to 1000 nucleotides), which minimized the occurrence of artificial recombinants as discussed in chapter 2.

Putative strategies for future prophylactic treatment

A selection of RSV strains from different clades was tested in neutralization assays to determine the impact of genetic diversity on antibody recognition in two different prophylactic antibody strategy studies. The first study is based on single antibody therapy with antibodies derived from human B-cells that are directed towards the fusion protein of RSV (Chapter 4). The B-cell IgG selection technology described by Scheeren et al. [29] proved to be a powerful tool for the isolation of highly neutralizing antibodies and the majority of these antibodies were directed towards conformational epitopes of the F protein. This was previous also described for anti-RSV F antibody D25 that is directed against the pre-fusion form of RSV-F [30]. Here we describe that the monoclonal antibody D25 performs like a possible successor of Palivizumab during in vitro and in vivo experiments and would therefore be a promising candidate for future prophylactic treatment. Although antibody neutralization efficacy differed among strain variants within the two RSV subgroups, none of the strains escaped from antibody recognition.

The second antibody strategy study described in chapter 5 is based on antibody combination therapy with a human bi-specific antibody and antibody mixtures including RSV F and G protein-directed antibodies. The bi-specific antibody PB2172 consisting of differential Fab fragments PG2729 and PG2845 directed against RSV-F showed improved neutralization efficacy for subgroup A strains compared to Palivizumab. However, this Fab fragment combination was not much better than the mono-specific version of PG2729. Selection of alternative Fab fragments might improve the efficacy in this type of antibody therapy strategy. Furthermore,

the addition of RSV G protein-directed monoclonal antibody PG3082 enhanced neutralization efficacy of the mono-specific and bi-specific antibodies in both RSV subgroups. This suggests that multi-epitope targeting by means of bi-specific antibody and/or antibody combination therapy could improve prophylactic treatment by targeting multiple strain variants. Despite the strain variation in the RSV F protein (9% RSV-A; 1% RSV-B) and RSV G protein (18% RSV-A; 12% RSV-B), all strains were neutralized via the multi-epitope targeting strategy.

Concluding remarks

The RSV whole genome analyses described in this thesis provide us with a considerable amount of data regarding the importance of genomic variations, evolution and selective pressure in RSV genes and proteins. These findings are not only important for the development of an effective immunologic therapy, but are also valuable for future structure-function studies in gene and protein specific roles relating viral gene products to virus viability, adaptability and putative immune escape. Our prophylactic antibody support strategy studies show promising results regarding neutralization efficacy of newly developed antibodies. Both monoclonal human B-cell-derived antibodies directed against the pre-fusion form of RSV F and multi-epitope targeting via bi-specific and combinatorial antibodies covered all strain variants tested and were more efficient than one of the currently used clinical antibodies, Palivizumab. The combination of strong neutralizing antibodies and multi-epitope targeting will not only effectively dampen RSV infections in future antibody prophylaxis, but also most likely diminish the occurrence of antibody escape mutants. In addition, the smaller amount of inflammatory cell infiltrates upon antibody treatment may reduce pathology and on the long term decrease the risk of developing wheezing and/or asthma.

References

1. Thorburn K (2009) *Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection*. Arch Dis Child 94: 99-103.

2. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE (2005) *Respiratory syncytial virus infection in elderly and high-risk adults*. N Engl J Med 352: 1749-1759.

3. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE (1969) *An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine*. Am J Epidemiol 89: 405-421.

4. Stensballe LG, Simonsen JB, Thomsen SF, Larsen AM, Lysdal SH, et al. (2009) *The causal direction in the association between respiratory syncytial virus hospitalization and asthma*. J Allergy Clin Immunol 123: 131-137 e131.

5. Kusel MM, de Klerk NH, Kebadze T, Vohma V, Holt PG, et al. (2007) *Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma*. J Allergy Clin Immunol 119: 1105-1110.

6. Sigurs N, Bjarnason R, Sigurbergsson F, Kjellman B (2000) *Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7*. Am J Respir Crit Care Med 161: 1501-1507.

7. Blanken MO, Rovers MM, Molenaar JM, Winkler-Seinstra PL, Meijer A, et al. (2013) *Respiratory syncytial virus and recurrent wheeze in healthy preterm infants*. N Engl J Med 368: 1791-1799.

8. Simoes EA, Groothuis JR, Carbonell-Estrany X, Rieger CH, Mitchell I, et al. (2007) *Palivizumab prophylaxis, respiratory syncytial virus, and subsequent recurrent wheezing*. J Pediatr 151: 34-42, 42 e31.

9. Klein Klouwenberg P, Tan L, Werkman W, van Bleek GM, Coenjaerts F (2009) *The role of Toll-like receptors in regulating the immune response against respiratory syncytial virus*. Crit Rev Immunol 29: 531-550.

10. Collins PL, Melero JA (2011) *Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years*. Virus Res 162: 80-99.

11. Teng MN, Whitehead SS, Collins PL (2001) *Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication in vitro and in vivo*. Virology 289: 283-296.

12. Heminway BR, Yu Y, Tanaka Y, Perrine KG, Gustafson E, et al. (1994) *Analysis of respiratory syncytial virus F, G, and SH proteins in cell fusion*. Virology 200: 801-805.

13. Tristram DA, Welliver RC, Hogerman DA, Hildreth SW, Paradiso P (1994) *Second-year surveillance of recipients of a respiratory syncytial virus (RSV) F protein subunit vaccine, PFP-1: evaluation of antibody persistence and possible disease enhancement*. Vaccine 12: 551-556.

14. Gonzalez IM, Karron RA, Eichelberger M, Walsh EE, Delagarza VW, et al. (2000) *Evaluation of the live attenuated cpts 248/404 RSV vaccine in combination with a subunit RSV vaccine (PFP-2) in healthy young and older adults*. Vaccine 18: 1763-1772.

15. (1998) *Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. The IMpact-RSV Study Group*. Pediatrics 102: 531-537.

16. Adams O, Bonzel L, Kovacevic A, Mayatepek E, Hoehn T, et al. (2010) *Palivizumab-resistant human respiratory syncytial virus infection in infancy*. Clin Infect Dis 51: 185-188.

17. Zhu Q, Patel NK, McAuliffe JM, Zhu W, Wachter L, et al. (2012) *Natural polymorphisms and resistance-associated mutations in the fusion protein of respiratory syncytial virus (RSV): effects on RSV susceptibility to palivizumab*. J Infect Dis 205: 635-638.

18. Zhu Q, McAuliffe JM, Patel NK, Palmer-Hill FJ, Yang CF, et al. (2011) *Analysis of respiratory syncytial virus preclinical and clinical variants resistant to neutralization by monoclonal antibodies palivizumab and/or motavizumab*. J Infect Dis 203: 674-682.

19. Vogel AM, Lennon DR, Broadbent R, Byrnes CA, Grimwood K, et al. (2002) *Palivizumab prophylaxis of respiratory syncytial virus infection in high-risk infants*. J Paediatr Child Health 38: 550-554.

20. Nuijten MJ, Wittenberg W, Lebmeier M (2007) *Cost effectiveness of palivizumab for respiratory syncytial virus prophylaxis in high-risk children: a UK analysis*. Pharmacoeconomics 25: 55-71.

21. Zlateva KT, Lemey P, Moes E, Vandamme AM, Van Ranst M (2005) *Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein*. J Virol 79: 9157-9167.

22. Zlateva KT, Lemey P, Vandamme AM, Van Ranst M (2004) *Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup a: positively selected sites in the attachment g glycoprotein*. J Virol 78: 4675-4683.

23. Kumaria R, Iyer LR, Hibberd ML, Simoes EA, Sugrue RJ (2011) *Whole genome characterization of non-tissue culture adapted HRSV strains in severely infected children*. Virol J 8: 372.

24. Rebuffo-Scheer C, Bose M, He J, Khaja S, Ulatowski M, et al. (2011) *Whole genome sequencing and evolutionary analysis of human respiratory syncytial virus A and B from Milwaukee, WI 1998-2010.* PLoS One 6: e25468.

25. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. Mol Biol Evol.

26. Tan L, Lemey P, Houspie L, Viveen MC, Jansen NJ, et al. (2012) *Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology*. PLoS One 7: e51439.

27. Tan L, Coenjaerts FE, Houspie L, Viveen MC, van Bleek GM, et al. (2013) *The Comparative Genomics of Human Respiratory Syncytial Virus Subgroup A and B: Genetic Variability and Molecular Evolutionary Dynamics*. J Virol.

28. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. (2012) *Detecting individual sites subject to episodic diversifying selection*. PLoS Genet 8: e1002764.

29. Scheeren FA, van Geelen CM, Yasuda E, Spits H, Beaumont T (2011) *Antigen-specific monoclonal antibodies isolated from B cells expressing constitutively active STAT5*. PLoS One 6: e17189.

30. McLellan JS, Chen M, Leung S, Graepel KW, Du X, et al. (2013) *Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody*. Science 340: 1113-1117.

Nederlandse Samenvatting

Nederlandse samenvatting

Humaan respiratoir syncytieel virus (RSV) infecteert respiratoire cellen en vormt daarbij karakteristieke syncytia (grote reuzencellen met meerdere celkernen) waarvan ook de naam van het virus is afgeleid. Het is wereldwijd de voornaamste veroorzaker van luchtweginfecties bij pasgeborenen en wordt geassocieerd met het ontstaan van astma. Het virus wordt ook in toenemende mate gevonden bij ouderen en immuun-deficiënte patiënten die lijden aan zeer ernstige lagere luchtweginfecties. Deze patiënten hebben een gebrek aan RSV-specifieke natuurlijk verkregen immuniteit. Ze hebben daardoor immunologische ondersteuning nodig in de vorm van antilichaamprofylaxe (behandeling met antilichamen om ziekte te voorkomen) of doormiddel van actieve immunisatie met een vaccin. Een studie uit de jaren zestig met een formaline-geïnactiveerd RSV vaccin heeft laten zien dat bescherming doormiddel van een vaccin niet gemakkelijk is. Tijdens opvolgende RSV-infecties ondervonden gevaccineerde patiënten nadelige effecten die veroorzaakt werden door een te sterke immuunreactie. Deze vaccinatie verstoorde de immunologische balans en veroorzaakte daarmee een immunologische last die groter was dan de hinder die men ondervond van de RSV-infectie zelf. Tegenwoordig ontvangen patiënten uit risicogroepen antilichaamprofylaxe. Deze bescherming is slechts tijdelijk, maar geeft een goede effectieve bescherming en verstoort de immunologische balans niet zoals dat gebeurd tijdens vaccinatie.

Na decennia van intensief onderzoek zijn er veel eiwitfuncties, eiwitstructuren en eiwit-celinteracties in relatie tot RSV-infectie ontrafeld. Echter, in het streven naar het effectief behandelen van RSV-infecties en het tegenhouden van extreme pathogenese (ziekteverloop), is het tot op heden niet gelukt om een kosten-effectieve en veilige therapeutische strategie, of een effectief en veilig vaccin, te ontwikkelen. Therapeutische ondersteuning is vooral gericht op het voorkomen van het binnendringen van RSV in de gastheercel. Hoewel de exacte mechanismen die daarbij betrokken zijn nog niet volledig duidelijk zijn, weten we dat het RSV fusie eiwit (RSV F) hierin een cruciale rol speelt en dat het aanhechtingseiwit (RSV G) dit proces makkelijker maakt. Beide virale glycoproteïnen (versuikerde eiwitten) zijn cruciale doelwitten voor natuurlijke verwerving van antilichamen en worden vaak gebruikt bij de ontwikkeling van subunitvaccins en de selectie van profylactische antilichamen. Het RSV F eiwit-bindende Palivizumab is één van de antilichamen

die momenteel gebruikt wordt voor profylaxe in de kliniek. Echter, het bestaan van ontsnappingsvarianten, de vereiste hoge dosis voor efficiënte behandeling en de daarmee gepaard gaande hoge kosten, vormen belangrijke redenen voor verder onderzoek naar alternatieve therapeutische strategieën. Een algemeen overzicht over RSV en de huidige behandel methodes wordt beschreven in hoofdstuk 1 van dit proefschrift.

Er bestaan twee RSV subgroepen (A en B) die onderverdeeld zijn in subtypen op basis van variatie in het RSV G eiwit. Variaties in de overige RSV eiwitten maken elke virusstam uniek. Informatie betreffende virusstamvariatie speelt een belangrijke rol tijdens vaccinontwikkeling en profylactische antilichaamtherapie. Stamvariatie is voornamelijk bestudeerd voor individuele genen, met name die coderen voor de RSV F en RSV G eiwitten. Dit echter, sluit belangrijke informatie omtrent variatie in overige delen van het virale genoom (totale set van genen die een sequentie vormen die alle virale erfelijke informatie bevat) uit. In hoofdstuk 2 en 3 van dit proefschrift hebben wij in detail de stamvariaties en de evolutionaire dynamiek per gen, alsook van volledige RSV genoom sequenties bestudeerd voor respectievelijk subgroep A en subgroep B RSV stammen. Substituties (mutaties) in genen en de gerelateerde eiwitten zijn in kaart gebracht om stamvariaties te bepalen en mutatiehotspots te identificeren die mogelijk van cruciaal belang zijn voor virusoverleving. Bepaalde mutaties kunnen mogelijk de potentie voor glycosylering (koppeling van suikergroepen aan eiwitten) beïnvloeden en daarmee antilichaamherkenning, eiwitfunctionaliteit en zelfs het verloop van infectie beïnvloeden. Fylodynamische analyses, zoals de Bayesiaanse genealogische analysemethode die wij hebben gebruikt voor het bepalen van de evolutionaire dynamiek, verschaffen ons waardevolle data omtrent evolutionaire snelheid van genen of volledige genoomsequenties, de tijd tot de meest recente gezamenlijke voorouder (time to the most recent common ancestor, TMRCA) van stammen en genetische regio's die onder selectieve druk staan. De evolutionaire relatie die stammen onderling hebben wordt weergegeven in een fylogenetische boom en door het koppelen van genetische sequenties aan de virus isolatie datum kan de evolutie snelheid worden berekend. De evolutiesnelheid karakteriseert het aantal mutatie veranderingen in de tijd. Dit geeft ons een indicatie van de noodzaak voor virale adaptatie om te kunnen "overleven". De tijdsperiode tussen de meest recente voorouder en stammen per fylogenetische tak kan bepaald worden door

fixatie van een bevorderlijke mutatie in het genoom van één specifieke stamvariant of kan het resultaat zijn van een kansfixatie van een specifieke stam variant (selectieve druk). Dit laatste refereert naar de mutatie frequentie binnen een bepaalde populatie die een bepaalde kans geeft dat een mutant allel (bepaalde variant van een gen) de gehele populatie over neemt. De evolutie snelheid van RSV blijkt relatief laag in vergelijking tot bijvoorbeeld Influenza virus en selectieve druk analyses op RSV-A en RSV-B stammen hebben aangetoond dat het merendeel van de mutaties in RSV-genen onder neutrale selectie staan. Dat betekent dat RSV geen sterke selectie van bevorderlijke genoomveranderingen nodig heeft voor optimale infectie, virustransmissie of adaptatie. Toch is er een kleine selectie van mutaties in de hypervariabele regio's van het G gen die evolueren onder pervasieve selectie (selectie voorkeur voor een virus fenotype die de frequentie van die specifieke virus variant verhoogd binnen de virus populatie). Deze mutaties worden verwacht betrokken te zijn bij het ontwijken van antilichaam herkenning.

Een selectie van RSV stammen van verschillende clades (subtypen) is getest in virus neutralisatie experimenten om de invloed te bepalen van genetische diversiteit op antilichaam herkenning. Dit is gedaan in twee verschillende profylactische antilichaam strategie studies. De eerste studie is gebaseerd op monoklonale antilichaam therapie met antilichamen geproduceerd door humane B cellen en die gericht zijn op de prefusie vorm van het RSV fusie eiwit (hoofdstuk 4). De tweede strategie staat beschreven in hoofdstuk 5 en is gebaseerd op antilichaam combinatie therapie met humane bi-specifieke antilichamen en antilichaam mixen bestaande uit antilichamen gericht tegen RSV F en G . Beide antilichaam sets bleken in staat een breed scala aan RSV stammen te neutraliseren. Meerdere antilichamen bleken daarbij efficiënter dan het momenteel gebruikte monoklonale Palivizumab.

De volledige genoomanalyses zoals beschreven in dit proefschrift voorzien ons van een aanzienlijke hoeveelheid data omtrent het belang van genomische variatie, evolutie en selectieve druk in RSV genen en corresponderende eiwitten. Deze bevindingen zijn niet alleen belangrijk voor de ontwikkeling van immunologische therapieën, maar zijn ook waardevol voor toekomstige eiwitstructuur-functie studies in relatie tot virus levensvatbaarheid, adapterend vermogen en mogelijke ontsnapping aan het immuunsysteem van de gastheer. Onze profylactische antilichaamstrategie studies laten veelbelovende resultaten zien omtrent de

neutralisatie effectiviteit van nieuw ontwikkelde antilichamen. De combinatie van sterk neutraliserende multi-epitoop gerichte antilichamen zal niet alleen RSV-infecties efficiënt kunnen onderdrukken, maar verlaagt naar alle waarschijnlijkheid ook het ontstaan van RSV-mutanten die ontsnappen aan antilichaamherkenning. Bovendien kan vroegtijdige antilichaambehandeling mogelijk op de lange termijn de kans op astma ontwikkeling reduceren doordat de RSV-geïnduceerde pathologie wordt tegen gehouden.

Dankwoord

“Good soil and pure water provide the basis for a seed to sprout

and all environmental conditions affect its development to become a tree.

Ultimately, the tree represents the life cycle it has gone through.

Look here at the result and thank the gardeners for their good care”

Lydia Tan

Dankwoord

Voor u ligt het proefschrift waar ik de afgelopen vijf jaar met hart en ziel aan heb gewerkt en die mede tot stand is gekomen met de hulp en ondersteuning van vele collega's, familie en vrienden. Ieder van hen heeft op zijn of haar manier een bijdrage geleverd. Alsmede direct door het uitvoeren/voorbereiden van experimenten, schrijven/reviewen van de hoofdstukken/artikelen of indirect doormiddel van mentale support naar mij toe. Ik wil graag eenieder van hen hartelijk danken hiervoor en neem de gelegenheid om enkelen van hen hier in het bijzonder te bedanken voor hun steun.

Als eerste wil ik beginnen met mijn co-promotor (Frank) en promotor (Emmanuel) te bedanken voor de mogelijkheid die ze mij hebben geboden om dit traject te doorlopen en de adviezen en ideeën die ze met mij hebben gedeeld. Frank we hebben de nodige tegenslagen gehad in dit project, maar onze gedeelde vastberadenheid en optimisme hebben alles tot een goed einde kunnen brengen.

Het RSV core team bestaande uit Barry (Mrs. golden hands), de altijd kritische Marco (Keen is my middle name) en verzorgende Henny (Indische gewoontes). In alle opzichten hebben jullie mij geweldig geholpen en ik mis onze samenwerking. Jullie zijn mijn toppers. Ik hoop dat dit boekje een mooie bekroning is geworden op het werk dat jullie verricht hebben voor dit onderzoek.

Philippe, Mark en Tim bedank ik voor de goede samenwerking. Philippe ik heb een hoop geleerd op het gebied van phylodynamische analyses en weet dat er nog veel meer valt te leren. Bedankt voor de goede uitleg en altijd super snelle antwoorden op mijn lange mails. Mark en Tim het was leuk om de diverse RSV stammen te kunnen gebruiken in de ontwikkeling van nieuwe therapeutische strategieën.

Onze stagiaires Archena, Rend, Suzanne en Domenique. Leuk dat jullie deel hebben uitgemaakt van het RSV onderzoek team. Wens jullie veel succes met jullie verdere studie/loopbaan.

Hanneke, bedankt voor de kritische vragen, de super goeie adviezen en de gezellige koffie momenten. Jovanka, voor het inlichten van Frank dat ik geïnteresseerd was in een promotietraject.

Mijn oude kamergenoten Nicole , Alex en Bart, maar ook mijn laatste kamergenoten Amin (water), Guido (vuur) en Laurens (de bemiddelaar). Vaak gezellig, soms te rumoerig, maar altijd met plezier gewerkt in jullie aanwezigheid. Nicole ik vond het fijn dat je mij de taak als paranimf toe vertrouwde en ik heb het met plezier gedaan.

De “Olivier” stamkroeg groep (Piet, Ellen, Bertie, Nicole + Harm) die elke dag met een tegenslag in de wetenschap weer tot een vrolijk einde bracht. Voor mij een “hete” thee of één glas rode wijn terwijl ik een toeziend oog hield op de collega’s die heerlijk van de Belgische biertjes genoten en soms de god Bachus ook hadden uitgenodigd.

Vrienden en familie wil ik bedanken voor de onvoorwaardelijke steun en begrip tijdens deze drukke jaren. Eindelijk is hier dan de afronding van datgene wat alle afspraken in de agenda deed opschuiven.

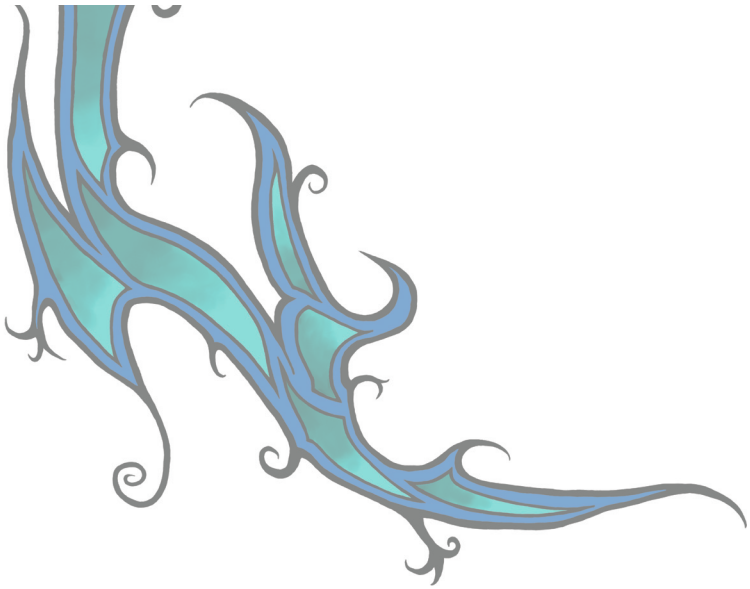
Odile en Malou fijn dat jullie mijn paranimfen willen zijn en leuk om met jullie deze ervaring te kunnen delen.

Oma Dora, onze gedeelde passie voor creativiteit heeft ons vroeger al veel plezier bezorgd tijdens het knutselen en tekenen. Deze keer heb ik de omslag van het boekje als project gebruikt om creatief bezig te zijn.

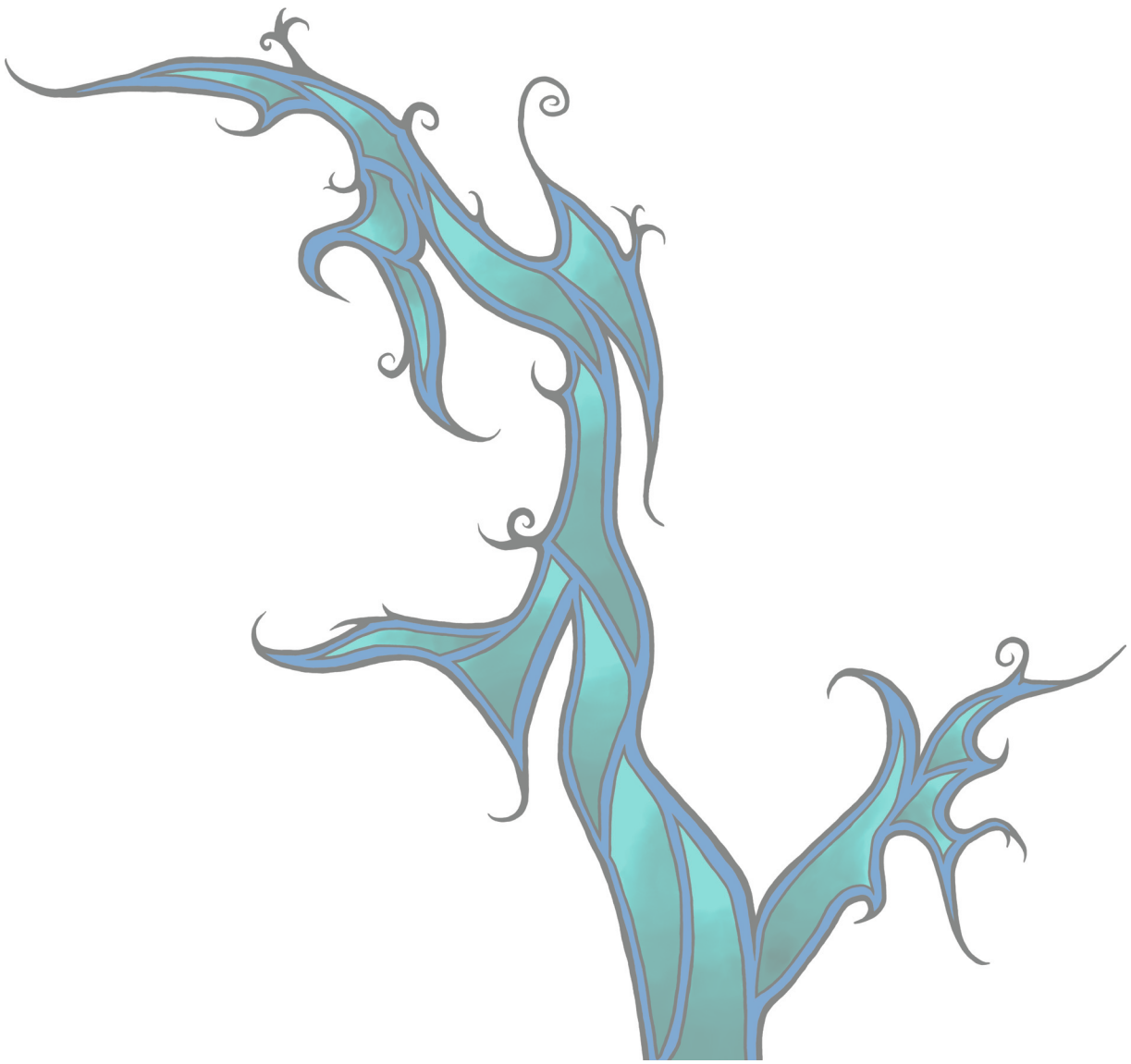
Opa en oma, ik mis jullie enorm, maar in gedachten zijn jullie altijd bij mij. Jullie zijn mijn inspiratiebron geweest tijdens het schrijven van dit boekje, omdat jullie als geen ander hebben laten zien hoe men kan aanpassen en overleven.

Lieve Henk, je bent op het meest chaotische tijdstip in mijn leven gestapt. Ik ben je dankbaar voor je steun, begrip en nuchtere kijk op momenten dat ik weer eens last had van perfectionisme, tegenslagen en frustraties tijdens mijn promotie. Gelukkig hebben wij ook fijne rustige momenten kunnen “inplannen” om samen te zijn en plezier te hebben. Op naar meer plezier en gelukkige momenten.

Als laatste, wil ik jullie bedanken lieve pap en mam. Dit boekje is mede het resultaat van al jullie inspanningen. Alle jaren van goede zorg, aandacht, adviezen, stimulatie om het beste uit jezelf te halen en het gevoel van vertrouwen zijn de sterke basis geweest voor mijn groei en ontwikkeling tijdens dit hele project. Dank jullie wel! Ik ben trots dat jullie mijn ouders zijn.



Curriculum Vitae



Curriculum Vitae

Lydia Tan was born on December 25, 1981 in Oosterhout, The Netherlands. She graduated from high school, MAVO ten Strijen, Oosterhout in 1998. In the same year she started her senior secondary vocational education in Biology and Medical laboratory research, which she completed with a diagnostic microbiology related internship at the St. Elisabeth hospital in Tilburg.

This study was in 2002 followed up with the higher vocational education in the same study direction. Here she majored in Immunology and Medical Microbiology after performing a vaccinology related internship at the National Institute for Public Health and the Environment (RIVM) under supervision of Dr. P.J.M. Roholl and Dr. A.C.M. van Els. The topic of this internship was "Screening of Meningococcal B Proteins for Developing an Effective Vaccine".

From January 2005 till June 2006 she worked as biology technical education assistant at the R.S.G. 't Rijks. Then after a four month sabbatical in Granada, Spain, where she followed intensive courses for Spanish, she continued studying at the University of Utrecht.

During her master study "Infection and immunity" she performed her first internship under the supervision of Prof. J.A.G. van Strijp at the department of Experimental Bacteriology at University Medical Hospital Utrecht (UMCU). Here she searched for a *Staphylococcus Aureus* bacterial Complement Receptor 1 antagonist. During the second internship at the Virology department of the faculty of Veterinary Medicine, University of Utrecht, she studied the interaction between Coronavirus proteins and host proteins. This was done under the supervision of Prof. J.M. Rottier and Dr. C.A.M de Haan.

In November 2008 she started as graduate candidate at the Medical Microbiology department of the UMCU under supervision of Prof. E.J.H.J. Wiertz and Dr. F.E.J. Coenjaerts. The results of this graduate research are described in this thesis.

From June 2013 she pursued her career into the pharmaceutical industry and started as biotech operations engineer at Janssen Biologics BV in Leiden. Here she is responsible for the support and optimization of the purification process of active pharmaceutical ingredients.

Photography by Malou Tan

Lydia Tan
Diamantlaan 244
2332 GS, Leiden The Netherlands

06-11236678
lydiatan1981@gmail.com