USING COMPUTATIONAL AND STATISTICAL STRATEGIES TO STUDY THE EVOLUTION AND EPIDEMIOLOGY OF HUMAN RESPIRATORY SYNCYTIAL VIRUS

by

JIANI CHEN

(Under the Direction of Justin Bahl)

Abstract

Human respiratory syncytial virus (RSV) is one of the leading causes of respiratory infections, especially in infants and young children. Despite the need to easily recognize RSV genotypes for molecular epidemiology, vaccine design, and control efforts, RSV classification criteria are not agreed upon and the potential association of RSV genotype with disease severity or immune response is almost unknown. In addition, RSV co-circulates with seasonal influenza in the U.S. every year causing significant epidemiological and economical burdens. However, very few studies have explored their potential interaction at the population level. To meet the vaccine and disease prevention needs of RSV and other respiratory pathogens, current global genomic data sharing, increased computing capability, and advanced molecular epidemiology methods have become powerful tools to understand rapidly evolving pathogens and have the potential to help us achieve the goal of effective prevention. In this dissertation, I describe the studies of RSV evolutionary and epidemiological dynamics of RSV using computational and statistical approaches to improve disease control. Chapter 2 of this dissertation focuses on developing a novel nomenclature system to better characterize the genetic diversity and evolution of RSV. In Chapter 3, I relate the observed RSV genetic diversity to potential T cell immune profiles. Using computational T cell epitope prediction

approaches, we provide a T cell epitope landscape visualization that shows the co-circulation of three RSV-A T cell epitope groups and two RSV-B T-cell epitope groups, suggesting potentially distinct T cell immunity of different RSV circulating strains. In Chapter 4, we demonstrate RSV may have different evolutionary dynamics compared to seasonal influenza, in that local persistence that may play a role in underlying annual epidemics. We also provide evidence for the potential negative interaction of RSV and seasonal influenza at a population level. Taken together, the findings of this dissertation are important for understanding the evolution of RSV and can greatly enhance our ability to forecast future epidemics, vaccine design, and control strategy developments.

INDEX WORDS: Respiratory Syncytial Virus, Phylogenetic analysis, Genotypic classification, T cell epitope, Epidemiological model

USING COMPUTATIONAL AND STATISTICAL STRATEGIES TO STUDY THE EVOLUTION AND EPIDEMIOLOGY OF HUMAN RESPIRATORY SYNCYTIAL VIRUS

by

JIANI CHEN

BM, Guilin Medical University, China, 2016

MS, University of Kansas, 2018

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2022

USING COMPUTATIONAL AND STATISTICAL STRATEGIES TO STUDY THE EVOLUTION AND EPIDEMIOLOGY OF HUMAN RESPIRATORY SYNCYTIAL VIRUS

by

JIANI CHEN

Approved:

Major Professor: Justin Bahl

Committee: Liliana Salvador
Liang Liu
Mark Ebell

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2022

ACKNOWLEDGEMENTS

I want to express my immense gratitude to Dr. Justin Bahl, who provided me with exceptional guidance on molecular epidemiology research. He is a great mentor to provide me with so many exciting learning and research opportunities and always encourages my thoughts and work, which means a lot to me. Joining Dr. Bahl's research group and working on RSV is one of the best decisions I have made.

I want to thank all members of Bahl lab that I have worked with over the past years. I am especially indebted to Lambodhar Damodaran, one of the most patient and intelligent people I have met. We joined the group at the same time and help each other along the journey. I would also like to thank Swan Tan, who helped me with my T cell epitope project and was always willing to listen to my complaints. I am also very lucky to meet Dr. Xueting Qiu. She served as another mentor not only in my Ph.D. research but also in my life. My Ph.D. would be insane without these three amazing people. I also want to thank my roommate Jingxuan Chen and my friends at IOB, Ruijie Xu, and Hanxia Li. Ph.D. life is challenging but turns out to be fun with you all.

I always thought I am very lucky to meet Dr. Liliana Salvador. She is not only a great scientist but also an amazing faculty that you can always reach out to when in need. I want to thank Dr. Liang Liu for his support of my job search and his constructive feedback on this thesis work. I also want to thank Dr. Mark Ebell for his valuable comments on my dissertation in clinical.

I appreciate all the dedicated collaborators that helped me build my thesis work. For Chapter 2, I want to thank Dr. Samuel Shepard for his training on the LABEL software. I also want to thank Drs. Pedro Piedra and Vasanthi Avadhanula for providing valuable suggestions on virology and immunity for Chapter 2 and Chapter 3 of this thesis work. Chapter 3 cannot

be done without the support of Dr. Annie De Groot, Dr. Leonard Moise, and the EpiVax team on T cell epitope prediction. For Chapter 4, I want to thank Dr. Pejman Rohani for introducing multiple pathogen modeling to me and for his important feedback. I want to, especially thanks to Dr. Deven Gokhale. I could not finish the infectious disease modeling work without his guidance and advice.

Last but not least, my Ph.D. would not have been possible without the full support of my parents.

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION AND LITERATURE REVIEW

Human respiratory syncytial virus (RSV) was first isolated in 1956 in Chimpanzees by Dr. J.A. Morris and colleagues [1]. Since its isolation and discovery, RSV has become the leading cause of hospitalization in infants and children due to bronchiolitis and pneumonia globally [2]. An important challenge in understanding the epidemiology of RSV is a lack of publicly available data [3]. Due to the increasing availability and potency of viral diagnostic tools, many countries and regions have launched RSV surveillance programs in recent years. In the era of genetic sequencing, efforts to study the transmission and severity of RSV through phylogenetics originally focused on partial sequencing of the G gene, but the focus has recently shifted to collecting and using full genome data in analyses to better understand the evolutionary and transmission dynamics not readily apparent in traditional epidemiologic and G gene data. Today we have more information on the epidemiology of the disease and its ubiquitous nature.

In this chapter, I first present an overview of the RSV disease burden and provide background on RSV surveillance and genomics. The increasing availability of viral surveillance and genetic sequences together with the development of advanced algorithms and rapid improvement of computing capability has driven the use of statistical and computational methods to study the epidemiology, evolution, and ecology of rapidly evolving viral pathogens. Following these backgrounds, I then give a literature review of the current research status on RSV genotyping classification systems, recent vaccine development, and the studies of co-circulation of RSV and seasonal influenza with a focus on the computational approaches

that are used in these studies. This chapter aims to lead to the knowledge gaps and research questions I intend to answer in the remainder of this dissertation.

1.1 RSV: GLOBAL BURDEN OF DISEASE

RSV infections usually cause mild, cold-like signs and symptoms, but high-risk populations can develop more severe complications, including bronchiolitis, pneumonia, and/or death [4]. It has a propensity for causing bronchiolitis and often produces a form of the disease that is longer in duration and more severe than bronchiolitis from other causes, while re-infection with RSV occurs regularly throughout life although infants are unlikely to get recurrent bronchiolitis [4].

RSV almost universally affects children worldwide before the age of 2. Each year in the U.S., RSV was estimated to lead to 2.1 million outpatient visits and 58,000 hospitalizations among children younger than 5 years old [5, 6]. The original Global Burden of Disease Study estimated that 6.7% of all deaths in children between 1 month to 1 year old were due to RSV, as were 1.6% of all deaths in children ages between 1 to 4 [7]. Although often characterized as a pediatric disease, RSV infection in adults also represents a substantial health burden. Previous studies have suggested people with compromised immune systems, people with chronic heart or lung disease, and the elderly, are at significant risk of severe infection and death from RSV infection [8, 9]. However, little is known about the global burden of RSV-induced acute lower respiratory tract infections in the elderly or those with underlying health conditions. Current researchers often estimate the number of in-hospital deaths due to RSV by combining in-hospital case-fatality ratio data with hospital admission estimates from hospital-based studies. A review looking at RSV disease in those $>$ 65 years of age found there were an estimated 1.5 million (95% confidence interval [CI], 0.3 million–6.9 million) episodes in high-income countries with 214,000 (95% CI, 100,000–459,000) hospitalizations and globally 336,000 [186,000-614,000] hospitalizations and 14,000 [5,000-50,000] in-hospital death in 2015 [10]. Moreover, there is even less information on RSV in non-hospitalized elderly

persons, in both high- and low-income settings, though some studies suggest the burden of RSV may be comparable to that of influenza in the adult population [11].

In temperate regions, RSV infections show a distinct seasonality with onset in late fall or early winter, a peak between mid-December and early February, and a season offset in late spring. Some areas, in particular in northern Europe, report yearly alternations between an early large outbreak and a late small outbreak. In tropical regions, the patterns are less predictable and can include two yearly peaks in spring and fall or fairly constant infection rates throughout the year [12] and the reasons for the near disappearance of RSV between epidemics remain unclear [13]. To investigate the reasons for the different seasonality of RSV, several studies have been conducted to assess the impact of geography and climate on RSV seasonality. While there is still a knowledge gap in RSV seasonality, RSV incidence was found to be higher when relative humidity was higher and the temperature was lower than the seasonal average [14].

Like influenza, RSV can spread through aerosolized droplets generated by coughs or sneezes, and the virus is able to live on hard surfaces for several hours. According to the US Centers for Disease Control and Prevention (CDC), children are most often exposed to the virus at school or daycare, which then causes household transmission. Although Historical data are largely focused on the United States, more recent studies on the epidemiology of RSV have been conducted in a large number of countries around the world [13, 15]. With the increasing number of countries and regions that are represented in the publicly available dataset, we will have a better understanding of RSV circulation patterns of disease transmission in the near future.

1.2 RSV surveillance system

Previous data analysis indicates that we lack information on RSV-associated diseases both at the community and country levels. The long-term impact of RSV infection on future wheeze and lung function, and the economic burden of RSV are also unknown. Better RSV

surveillance has the potential to provide evidence based on seasonality, healthcare burden (such as hospitalizations), and risk groups to guide immunization practices once the RSV vaccine becomes available and may provide a platform for special studies to address some of these research questions.

The Global Influenza Surveillance and Response System (GISRS) coordinated by WHO and endorsed by national governments tests more than two million respiratory specimens annually to monitor the spread and evolution of influenza viruses through a network of about 150 well-established laboratories in 114 countries representing 91% of the world's population. The GISRS network is ideally placed as a platform for the introduction of more systematic testing for RSV associated with respiratory illness as it already uses molecular diagnostics on respiratory specimens and is conditioned to report regular and disciplined reporting to a global platform. As RSV molecular diagnostics became widely available, many countries within the GISRS started testing for RSV and other respiratory viruses as a by-product of influenza surveillance using the WHO-recommended influenza-like illness (ILI), acute respiratory infection (ARI), and severe acute respiratory infection (SARI) case definitions [16].

In the U.S., CDC analyzes data on RSV activity at the national, regional, and state levels, collected by a surveillance system called the National Respiratory and Enteric Virus Surveillance System (NREVSS). Participating clinical and public health laboratories voluntarily report the number of aggregate and positive RSV tests to NREVSS each week. In previous years, the RSV season was defined by consecutive weeks when RSV antigen-based tests exceeded 10% positivity. Since 2014, the majority of tests and RSV detections among consistently reporting laboratories are determined by PCR [15].

With RSV global and local surveillance systems, more data on the distribution of RSV will be available at the global level as well as in the U.S. We believe better surveillance of RSV in the future will aid efforts to better understand the transmission patterns, control its spread, and eventually to design effective RSV vaccines.

1.3 RSV genome and genetic diversity

RSV is an enveloped, negative-sense, single-stranded RNA virus, with a non-segmented genome that is about 15,000 nucleotides long. It belongs to the family of Paramyxoviridae, genus Pneumovirus, and subfamily Pneumovirinae [13]. There is a single serotype with two major antigenic subgroups in RSV, RSV-A and RSV-B. Strains of both subtypes often co-circulate, but generally, one of the subtypes predominates [17, 18]. The results of molecular analyses show that several genotypes are present simultaneously in any given season and region, but even in neighboring regions, the circulating strains may differ [19].

The RSV genome has 10 genes that code for 11 proteins since two overlapping open reading frames in the M2 yield two distinct matrix proteins, M2-1, and M2-2. Among them, the G protein functions in host cell attachment, and the F protein is responsible for fusion and cell entry, whereas the SH protein is not required in either of these processes, it also play an important role during virus infection. The remaining genes code for nonstructural proteins (NS1 and NS2), nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), transcription regulators (M2-1 and M2-2), and large polymerase (L) (Figure 1.1A).

The G gene produces the key surface glycoprotein in viral binding to host cells and is often called the attachment protein. This gene has traditionally been the focus of studies on the evolutionary history of RSV because there is a hypervariable region at the C-terminus that contains most of the genetic variation in the genome. Before whole-genome sequencing for RSV was widespread, it was easier to base analyses on this section of the gene since it was thought to be the location demonstrating the most evolutionary signal [21]. It has also been suggested that the lack of cross-immunity between RSV subtypes and genotypes is due to variation in the G gene [21]. Therefore, most previous genotypes of RSV are currently classified by this hypervariable region in the G gene (200-300 bp). 14 genotypes among RSV-A (GA1–7, SAA1, CB-A, NA1–4, and ON1) and 24 genotypes in RSV-B (GB1-GB5, SAB1-SAB4, URU1–2, BA1–12, and CB1) have been identified and reported in previous studies [22]. Although there has been disagreement on the classification and naming of RSV genotypes, there are

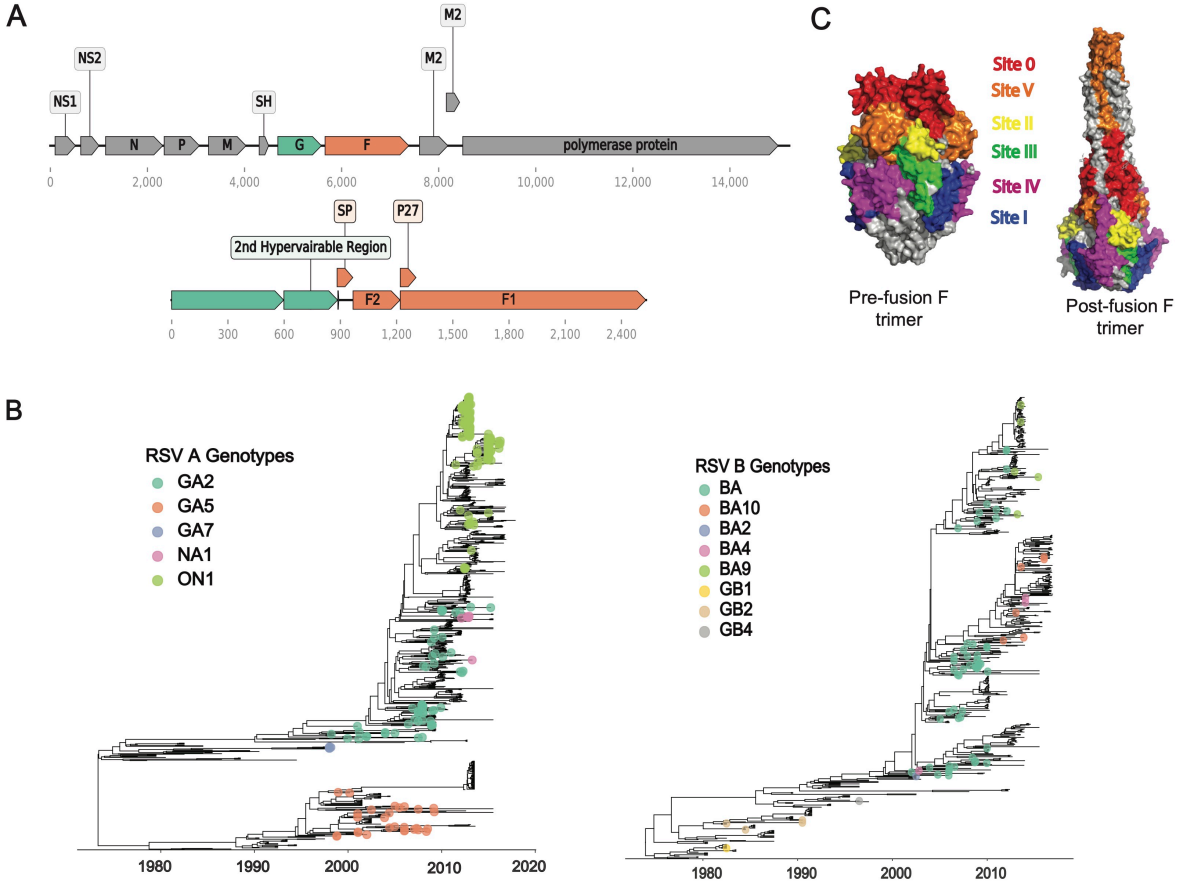

Figure 1.1: **RSV genome and genotypes.** (A) Scheme of the RSV genome. RSV has a negative-stranded RNA genome which is approximately 15 kb long with 10 gene transcripts encoding 11 proteins. Two of these are the major surface glycoproteins, the attachment or G glycoprotein (green) and the fusion (F, orange) glycoprotein. (B) Phylogenetic tree of representative circulating RSV-A and B genotypes. The conventional classifications are shown in the legend. Genotypes are conventionally classified based on genetic variation in the 2nd hypervariable region of the G gene. These include a 72-nucleotide duplication, referred to as the ON1 genotype, which in RSV-A distinguishes this group from the ancestral NA1 genotype. In RSV-B, a 60-nucleotide duplication and a short upstream deletion distinguish what is referred to as the BA clade from the ancestral GB1 genotype. (C) Antigenic structure of the RSV F protein. Surface representation of the 3D structures of RSV F trimer folded in its prefusion and post-fusion conformation, showing six main antigenic sites as defined by Gilman et al [20].

two that are easily identified due to insertions in the G gene: RSV-A genotype ON1 and RSV-B genotype BA. RSV-A genotype ON1 has a 72-nucleotide (nt) duplication that is not present in other common A genotypes, such as NA1 and GA2, RSV-B genotype BA has a similar duplication in the same region that is 60-nt long, this duplication is also missing in other genotypes of RSV-B [21]. These two genotypes, ON1 and BA have been the dominant genotypes in recent outbreaks [23, 24] (Figure 1.1B). Another important RSV gene, the F gene, which is sometimes analyzed in conjunction with the G gene, codes for another surface glycoprotein that is key to viral entry to host cells. This protein has a pre-fusion form and a post-fusion form, and critical antigenic sites in the protein have been used in vaccine design (Figure 1.1C). Compared to the G gene, the F gene is well-conserved across all genotypes of both RSV-A and B [25]. However, there are recent studies conducted that found more genetic variability in the RSV F gene than previously thought, especially in the pre-fusion antigenic sites, and more variations in RSV-B sequences than RSV-A sequences [26]. Another gene that plays a role in viral entry and is often not included in epidemiologic analyses is the SH gene. SH codes for the small hydrophobic protein that changes the permeability of the host membrane and aids in viral entry to host cells, although it is not necessary for entry [27]. Considering the important functions of these genes, some researchers focus on sequencing the three surface glycoproteins genes in their studies and there are recent studies that attempt to build a new RSV genotyping system with SH-G-F concatenated [28].

1.4 Current research status and knowledge gaps of RSV

1.4.1 RSV genotype classification

RSV genotypes proposals Accurate and consistent classification into genotypes is critical in understanding the evolution of divergent viruses like RSV. Genotyping systems of viruses usually rely on monophyletic groups on a phylogenetic tree, which requires a robust and comprehensive virus genomic dataset. Under the classification level of genotype, further subtyping

classification could provide useful epidemiological information regarding the transmission of pathogens.

In a parallel effort, several research groups have been working on RSV genotyping proposals to provide a consensus on RSV uniform genotype designation. The first classification system of RSV genotypes, proposed in 1998, relied on sequencing information of the second hypervariable region (HVR) of the G gene. Based on a visual inspection of a phylogenetic tree, seven genotypes could be distinguished for RSV-A and four genotypes for RSV-B. Bootstrap support (BS) values of 78% or higher were observed for the relevant clusters. The genotypes were named based on the gene used for classification (G), followed by the RSV subtype (A or B) and an ascending number: GA1–GA7 and GB1–GB4 [19, 29]. One year later, Venter et al. used a similar approach to expand the number of RSV genotypes and the method of classification was refined by including genetic distance as a metric to define clusters. If a group of sequences would cluster together with BS values of 70% or more and if characterized with a pairwise distance of ≤ 0.07 nt substitutions per site to all other members part of the same phylogenetic cluster, a genotype was distinguished. In addition to introducing a new genotype definition, the nomenclature system was altered, including now the country of discovery (i.e., SA for South Africa) when naming RSV genotypes [30]. This method resulted in five new genotypes: SAA1 within subtype RSV-A and SAB1–SAB4 within subtype RSV-B [30]. Over several years, this genotype definition method was used in many different studies to distinguish additional genotypes. However, the naming of the genotypes did not adhere to the same subtype- and country-based nomenclature system [31, 32, 33, 23, 34]. Alternatively, Agoti et al. proposed another RSV genotyping system with a complete RSV G gene using BS value greater than 60% and an average genetic distance cutoff of 1.5% [35]. In 2020, a systemic RSV genotyping proposal was published which extends to all G gene sequences that are currently publicly available and classify all RSV sequences into 3 genotypes in RSV-A, and 5 genotypes in RSV-B. The subclade level of classification was proposed with node distance criteria under the level of genotype [28]. The most recent RSV genotyping method

was proposed by Ramaekers, et al [36]. RSV whole genome sequence (WGS) was suggested to be used for RSV genotyping. Additionally, they proposed to use patristic distance rather than genetic distance as a parameter to define RSV genotypes. Patristic distance is a tree-based estimation of the genetic distance, measured as the shortest distance over the branch lengths between two tips of the phylogenetic tree, which result in a better estimation of the true genetic distances represented in the data set compared to genetic distance [36].

RSV viral characteristics and disease severity The severity of RSV infections is multifactorial as it depends both on the host and viral factors. Although much attention has been paid to the severity of RSV infection to host factors, such as immunosuppression or chronic health conditions, the influence of viral factors, such as viral load and RSV genotype on the severity of the infection is less well-defined [37]. Previous studies have shown that RSV-infected infants with severe disease had higher nasopharyngeal viral loads compared to infants with non-severe disease [38]. However, studies focusing on the relationship between disease severity and RSV subtype and genotypes produce controversial results.

In Cyprus, researchers found that a larger proportion of patients that are infected with BA required oxygen, suggesting BA causes severe outcomes more frequently than other genotypes of RSV, which is in contrast to common results indicating that RSV-A is more likely to cause severe disease [37]. In addition, one such study conducted over three epidemic seasons in Japan found that 35.6% of those infected with RSV-A genotype ON1 were hospitalized, and this proportion was greater than all other measured genotype hospitalization rates, with an odds ratio of 6.92:1 of hospitalization for those infected with ON1 to those infected with its ancestral genotype NA1. However, this is opposed by results from a study in Northern Italy, in which NA1 was more likely to cause upper respiratory tract infections and require hospitalizations than ON1 [39]. These discrepancies could be attributed to differences in study design, which in most cases were retrospective, disease definition, inclusion criteria, or inconsistency in genotyping methodology. It is also possible that strain virulence may be variable between epidemics. Furthermore, current RSV disease severity has been mostly

studied at the genotype level, but strain-specific genetic characteristics can also be important determinants of viral pathogenicity. Molecular characterization of ON1 genotype found five amino acid residue substitutions combination in 68% of other ON1 strains in their studies and was associated with decreased disease severity in patients, and one was within the antigenic region, which might affect the antigenic and immunogenic of viral. These results may provide valuable information on the pathogenic mechanism of RSV [40]. Therefore, to have a better understanding of RSV viral characteristics and disease severity, not only a consensus on RSV genotype is needed, but we also need to provide a nomenclature system to describe RSV strains under genotype levels.

Circulation pattern of RSV genotypes Many studies have focused on local circulation patterns of RSV within in single season or across several seasons using phylogenetic approaches. They found the dominant RSV genotypes can alternate between subtype A and subtype B, and within each RSV subtype, the dominant genotype also changes across seasons. For example, RSV-A genotype NA1 was the dominant type in Japan before the introduction of ON1, which then became the dominant genotype [41, 33]. In addition, Multiple genotypes from both RSV-A and RSV-B subtypes can circulate concurrently within a single season, increasing the complexity of circulation patterns [21, 42]. In addition, the geographical clustering observed in these studies were evident in strains from both states with multiple distinct sub-lineages observed and relatively low mixing across jurisdictions, suggesting that endemic transmission was likely seeded from imported, unsampled location [33, 43].

There are a few studies that connect local RSV genomic data to publicly available global data and place local outbreaks in the context of the larger phylogenic tree. This would allow investigators to make inferences about the source of the disease in their country or locality and global transmission patterns. For example, the study analyzing sequences from Guangdong, China, in conjunction with sequences from publicly available strains indicated the dominant strains GA2 and ON1originated in the Americas before spreading to other regions [44]. But the investigators have noted that this result could be due to bias in the dataset since most

early samples are from the United States and much of the available data is also from this country. They also found that many clades contained viral strains from multiple geographic areas in the phylogenetic tree generated by the study, but several clades represented only Guangdong [44]. This indicates that there were introductions of RSV from other regions, but seasonal epidemics may also have been seeded by locally persistent strains of the virus. Another limitation of current molecular epidemiology analysis is previous RSV sequencing studies have largely focused on sequencing only complete or partial G gene sequences because the C-terminal, second hypervariable portion of G is sufficient and required for distinguishing the two RSV subtypes and the various genotypes within each subtype [21]. However, the phylogenetic signal of other genes has not been recapitulated by G gene phylogeny and incomplete lineage sorting could be a confounding variable in relationships that has been proposed by most of these studies [27, 45]. With the advancement of sequencing efforts and phylogenetic approaches, many researchers have suggested using WGS for future RSV studies.

Current molecular epidemiology and transmission studies of RSV are limited by the sampling bias and lack of whole-genome data. As more whole-genome data has become available and more advanced phylogeographic methods are developed, we believe both local and global circulation dynamics of RSV can be better explored in the future which will reveal the relative importance of international introductions in these seasonal epidemics and inform surveillance programs and control measures.

1.4.2 RSV vaccine development

Current RSV vaccine candidates Despite the significant clinical burden of RSV, there is no licensed vaccine currently available for RSV infection. RSV vaccine candidates aim to protect at least three target populations that are at risk for severe RSV disease: (1) young infants (0–6 months), (2) older infants and young children (2 months or older) through active immunization, and (3) older adults (65 years or older). Attempts in the 1960s to develop a formalin-inactivated RSV (FI-RSV) vaccine candidate were hampered by several factors,

including a lack of protection against RSV infection in infants and young children, and an association with a vaccine-enhanced disease that resulted in two deaths upon natural RSV infection following vaccinations [46]. Currently, the only approved RSV prophylactic is palivizumab, which is used for high-risk patients, but such treatment has limited applicability due to cost and treatment logistics. Thus, there is a critical need to develop a vaccine for the vast majority of the RSV-vulnerable population. There is a recent surge in the number of RSV vaccine candidates undergoing clinical evaluation including live-attenuated, chemical mutagenesis, recombinant vector-based, subunit vaccines, and particle-based vaccine [26] (Table 1.1). These candidates that have reached clinical development are nicely reviewed elsewhere, but none are FDA-approved, which is related to our incomplete understanding of the host immune response to RSV [47, 48].

Recombinant live-attenuated vaccines, which mimic natural RSV infection, face the challenge of achieving sufficient attenuation to be safe and remaining immunogenic enough to induce a protective immune response. Therefore, live attenuated vaccines are not ideal for the protection of infants less than 4 months of age whose immune systems are immature and are susceptible to respiratory distress from minor respiratory illness, but this strategy is suited for direct vaccination of older RSV naïve infants [48]. Two main modifications to the RSV genome have been engineered through reverse genetics: the ΔM2-2 deletion which attenuates viral replication and upregulates antigen expression and the ΔNS2 deletion, which reduces viral suppression of host interferon thereby boosting the innate immune response [57]. Further results from clinical trials with the other live-attenuated vaccine candidates are expected. The only chimeric vaccine candidate in clinical development, rBCG-N-hRSV2, allows for combined vaccination against two major respiratory pathogens: Mycobacterium tuberculosis and RSV via T cell immunity [58]. The antigen presented by this vaccine candidate is the RSV N protein. So far, this candidate is the only vaccine candidate intended for administration to newborn babies [59]. Vector-based vaccines are another platform being pursued. These vaccines generally utilize a viral vector to express one or more RSV antigens on the cell

Table 1.1: **RSV vaccine candidates and monoclonal antibodies.**

Vaccine strategy	**Target population**	**Immuno responses**	**Represantative vaccine candidates**
Live-attenuated vaccine	Pediatric population under 2 years of age	Broad humoral and cellular response	RSV MEDI ΔM2-2 [48], RSV ΔNS2 [48]
Chimeric vaccine	Intended for administration to newborn babies	Th1 cellular response	rBCG-N-hRSV [46]
Vector-based vaccine	Children and adults	Broad humoral and cellular response	MVA-BN-RSV [49], VXA-RSV-f 0RSV, Ad26-RSV-preF [50], ChAd155-RSV [51]
Subunit vaccine	Pregnant women and older adult populations	Mainly humoral response	GSK RSF [52], DPX-RSV(SH) [53]
Particle-based vaccine	Multiple target population groups	Broad humoral response	RSV F nanoparticle-based vaccine platform [54], SynGEM (needle free) [55]
Monoclonal antibodies	Infants	NA	Palivizumab, MEDI 8897 (nirsevimab) [56]

surface, allowing for the natural presentation of the antigen of interest. Most RSV vector-based candidates utilize a replication-incompetent adenoviral vector, but a modified vaccinia virus, modified vaccinia Ankara (MVA), is showing promise in late phase II trials [49]. All recombinant vector-based vaccines make use of the F protein, and some include additional viral antigens such as the G protein or the structural proteins, N and M2. Protein-based vaccine candidates focus primarily on the F protein, particularly the pre-fusion conformation, but some candidates also include the two other surface glycoproteins, G and SH. This approach is generally indicated for use in the vaccination of the elderly and maternal immunization rather than direct immunization of RSV-naive infants because of the potential risk for VED [60]. The RSV F nanoparticle-based vaccine platform is being evaluated for protection in multiple target populations and positive results have been shown in both children and adults [54]. SynGEM is a particle-based needle-free vaccine candidate containing the RSV F protein attached to empty bacterial particles made from Lactococcus lactis. An influenza vaccine candidate in clinical trials that uses the same vaccine platform has shown both local and systemic antibody responses but further optimization is needed for RSV vaccination [55]. In addition to the vaccine development, the success of palivizumab in preventing severe disease in high-risk infants has spurred the development of additional monoclonal antibodies with higher affinity and longer half-lives [61].

Antigenic variability of RSV While these RSV vaccine candidates and antibody treatments hold promise, there is the possibility of viral strains developing escape mutations. Palivizumab-resistant strains have been isolated from both cotton rats and humans. This risk is further evidenced by the failure of the Regeneron monoclonal antibody in late-stage trials when circulating RSV-B strains underwent spontaneous point mutations independent of pressure from treatment led to the ineffectiveness of the mAb [32]. This failure emphasizes the need for surveillance of circulating strains and indicates the use of monoclonal antibodies that target critical fitness sites on the F protein or antibody cocktails targeting multiple independent epitopes within the F protein may aid in limiting escape mutants.

To date, mAb-resistant mutants have not been thoroughly studied worldwide and little is known about the prevalence of naturally occurring resistant RSV strains either. Therefore, Respiratory Syncytial Virus Network (ReSViNET; www.resvinet.org) has led an International Network For Optimal Resistance Monitoring of RSV (INFORM RSV) study to describe the molecular epidemiology of RSV by monitoring the temporal and geographic distribution of the whole viral genome sequences [62]. The polymorphisms in the F protein binding regions of RSV mAb are determined and the amino acid changes are recorded. These sequencing and functional have the potential to provide valuable information for vaccines, monoclonal antibodies, and therapeutic drugs in development.

There are limited data currently available about T-cell epitopes to RSV in humans, and to date, no study has looked at the sequence variation of RSV in T-cell epitopes [63]. Amino-acid variation of RSV at the T cell epitope level and the emergence of novel T cell epitopes have been reported, but further studies are needed to illustrate the effect of these variations on T cell recognition [64]. Various immuno-informatics tools have been developed to predict whether a region of the virus genome, usually a protein, can generate a T cell immune response by itself [65], which makes it possible to examine the T cell immune profile changes among a large number of RSV sequences. This refers to many T epitope prediction tools including TEpredict, CTLPred, NetMHC, and Epitopemap. Some tools have employed deep learning and machine learning algorithms to predict potential immunogenic subunits from the viral genome sequences. These tools reduce the time required to identify immunogenic targets and enhance the development of potentially safe vaccine candidates and pertain to the current RSV vaccine development [66].

1.4.3 Co-circulation of RSV and other respiratory viruses

Co-circulation of human respiratory virus Human respiratory viruses include a broad range of viruses that infect cells of the respiratory tract, elicit respiratory and other symptoms, and are transmitted mainly by respiratory secretions of infected persons. They belong to diverse virus

families that differ in viral and genomic structures, populations susceptible to infection, disease severity, seasonality of circulation, transmissibility, and modes of transmission. Epidemiological studies have shown that infants, young children, and the elderly are especially at risk of infection by both subtypes of RSV. The most recognized is seasonal influenza virus infection, which is responsible for about 290,000 to 650,000 deaths each year. Moreover, many other respiratory viruses have also been found to be capable of participating in simultaneous circulation, including human rhinovirus (hRV), human enterovirus (hEV), human metapneumovirus (hMPV), coronavirus (CoV), parainfluenza virus (PIV), adenovirus (AdV), and human bocavirus (hBoV) and some SARS-associated coronaviruses (SARS-CoV-2) [14, 67]. Together, they contribute to substantial morbidity, mortality and concomitant economic losses annually worldwide. In addition, occasional pandemics cause extreme disruption to societies and economies as exemplified by the current COVID-19 pandemic. In addition to the threat from single virus infections, infections with multiple respiratory viruses in the same patient have been reported in many studies. Several respiratory viruses are found to be capable of participating in simultaneous infections.

Potential interaction of RSV and seasonal influenza Among human respiratory viruses, RSV and seasonal influenza viruses cause large burdens of respiratory disease, including in young children. Surveillance systems for RSV are lacking, most influenza-endemic countries have an influenza surveillance system publicly available data and therefore RSV cases are mostly captured through influenza surveillance systems in many countries. Previously, an overwhelming amount of studies reported similar rates of hospitalization and mortality of influenza virus A and B compared to RSV [3]. RSV and seasonal influenza activity consistently peak during the winter months in many regions. Moreover, both viruses can be transmitted by aerosol droplets or direct contact with the virus, such as through fomites. In light of these, it is reasonable to expect the co-infection of RSV and seasonal influenza. However, the observed incidence of co-infectivity of RSV and influenza was significantly less than the expected incidence even when both were co-circulating [68]. Previous biological studies suggest this

might result in a negative association between these two viruses. Several studies have reported evidence of interference between these two respiratory viruses. In vitro, infection with RSV is blocked by competitive infection of influenza A if the host is not infected with the two viruses simultaneously. Similarly, ferret models have shown that influenza A infection may prevent successive infection with RSV and that coinfection with different influenza subtypes is dependent upon the order in which the viruses infect the host [69]. The exact nature of interactions between different respiratory viruses remains unclear, although they are proposed to be driven by the innate immune system.

However, this potentially negative association between infection with RSV and co-infection with influenza cannot be directly observed from epidemiology data. Shretha et al. previously developed a transmission model that clarifies the effect of influenza on pneumococcal pneumonia, which bridges the gap between individual animal experiments and human epidemiolocal data [70]. In their work, they take a mathematical approach by using a mechanics transmission model within a Bayesian likelihood-based inference framework to determine the role of within-host coinfection dynamics. This approach is based on a well-known adapted SIRS model (where S= susceptible, I = infected, and R= recently recovered). This model has been applied in a new way to address questions about pathogen interaction at the population level [71, 72]. Recently, a similar mathematic modeling approach was used to test the potential competition interaction between RSV and influenza in the UK [73].

1.5 Dissertation outline

The overall theme of this dissertation is to develop novel computational frameworks to have a better understanding RSV virus and therefore contribute to the development of RSV preventative strategy and vaccine design.

In Chapter 2, I developed a novel whole genome-based RSV genotyping system using phylogenetic approaches. By comparing different gene regions, I highlight the importance of using whole-genome for RSV genotyping and we assign RSV strains as clade and subclade

with all publicly available RSV whole-genome. I also developed an interactive visualization under the platform of "nextstrain". In addition, I provided a toolbox to perform genotype assignments from RSV raw sequencing data using machine learning instead of time-consuming sequence alignment and phylogenetic tree reconstruction with the "LABEL" toolbox.

In Chapter 3, I computationally examine the putative T cell epitopes of RSV major surface protein and identified T epitopes that might be valuable for future vaccine design. More importantly, I attempted to address the knowledge gaps of virus genetic diversity that I described in chapter 1 for the current RSV vaccine design. I quantitively describe the evolution of T cell epitopes of RSV strains by calculating the T epitope distance between different RSV strains and generating T cell epitope landscapes using a multiple dimensional scaling (MDS) approach.

In Chapter 4, I focus on the co-circulation of Influenza viruses and RSV in the U.S., which are the two most important causes of lower respiratory infections-associated disease. I characterized the co-circulation pattern of these two viruses using genetic data as well as surveillance reports. In addition, I developed a seasonal forced, two-pathogen mechanistic transmission model to evaluate the interaction of two viruses and their potential competitive mechanism at the HHS regional level in the U.S. Our findings suggest a negative interaction of these two pathogens and their competition can be mainly explained by a short period of cross-immunity to the second virus after initial infection.

Chapter 2

Novel and extendable genotyping system for Human Respiratory Syncytial Virus based on whole-genome sequence analysis[1]

[1]Jiani Chen, Xueting Qiu, Vasanthi Avadhanula, Samuel S. Shepard, Do-Kyun Kim, James Hixson, Pedro A. Piedra, Justin Bahl, Novel and extendable genotyping system for Human Respiratory Syncytial Virus based on whole-genome sequence analysis, *Influenza and Other Respiratory Viruses*, 2022.

2.1 ABSTRACT

Human respiratory syncytial virus (RSV) is one of the leading causes of respiratory infections, especially in infants and young children. Previous RSV sequencing studies have primarily focused on partial sequencing of G gene (200-300 nucleotides) for genotype characterization or diagnostics. However, the genotype assignment with G gene has not recapitulated the phylogenetic signal of other genes and there is no consensus on RSV genotype definition. We conducted Maximum Likelihood phylogenetic analysis with 10 RSV individual genes and whole-genome sequence (WGS) that are published in GenBank. RSV genotypes were determined by using phylogenetic analysis and pairwise node distances. In this study, we first statistically examined the phylogenetic incongruence and rate variation for each RSV gene sequence and WGS. We then proposed a new RSV genotyping system based on a comparative analysis of WGS and the temporal distribution of strains. We also provide an RSV classification tool to perform RSV genotype assignment and a publicly accessible up-to-date instance of Nextstrain where the phylogenetic relationship of all genotypes can be explored. This revised RSV genotyping system will provide important information for disease surveillance, epidemiology, and vaccine development.

2.2 Introduction

Human respiratory syncytial virus (RSV) is a major cause of acute lower respiratory tract infection worldwide in infants and young children ($\leq$5 years of age), as well as in the elderly and patients who are immunocompromised [22]. Despite the clinical significance and the burden of RSV infection, we lack an understanding of the patterns of virus emergence, evolution, and spread. Phylogenetic studies of RSV evolution are in need, especially on a global scale due to the limited availability of whole-genome sequence (WGS) data and strongly asynchronous sampling in time and space [33].

RSV is an enveloped virus with a negative-sense, single-stranded, non-segmented RNA genome of 15,200 nucleotides (nt) in length and belongs to the family Pneumoviridae. This genome encodes for 11 proteins, including the polymerase (L), nucleocapsid (N), phosphoprotein (P), transcriptional regulators (M2-1 and M2-2), matrix (M), small hydrophobic surface protein (SH), non-structural proteins (NS1, NS2) and two major surface glycoproteins (F and G) [21]. This virus has been classified as subtype A (RSV-A) or subtype B (RSV-B) according to reactivity with monoclonal antibodies [74]. Both subtypes typically co-circulate during epidemic seasons. Within the RSV-A and RSV-B subtypes, different genotypes have been further classified mainly based on genetic differences in the second hypervariable region (HR) located at the G glycoprotein [29]. Like other respiratory viruses, RSV has diverse circulation patterns. Several genotypes can co-circulate within the same community, while novel RSV genotypes with high genomic diversity may arise and potentially replace the previous dominant genotypes [75]. Fourteen genotypes among RSV-A (GA1-7, SAA1, CB-A, NA1-4 and ON1) and twenty-four genotypes in RSV-B (GB1-GB5, SAB1-SAB4, URU1-2, BA1-12 and CB1) have been identified [22]. The most notable genotype change observed in recent years is the emergence of RSV-A (ON) and RSV-B (BA) strains with a partial duplication of the distal third of the G gene, and have since become the dominant strains in many regions [76, 77].

With the emergence of novel genotypes, a potential association of RSV genotype with disease severity or geographic and temporal restriction of virus circulation has been reported [13, 78]. Moreover, RSV genetic diversity has been considered as an important factor that allows for reinfections to occur and needs to be considered in vaccine development [78]. Therefore, a genotyping system that could reflect RSV genetic diversity is needed. Previous RSV sequencing has largely focused on complete [19] or partial G gene (200-300 nt) [21, 79] for genotype characterization or diagnostics. However, the evolutionary signals from other gene regions should also be taken into account as the phylogeny inferred from other genes might conflict with the phylogeny inferred from complete or partial G gene alone. Furthermore, the

novel identified G gene duplication signature should be considered as a single evolutionary event, whereas current widely used phylogenetic models only account for residue substitution events. In addition, most of the current RSV genotyping methods are based on the pairwise distance (p-distance) matrix by specifying a cutoff value below which individuals are assigned to the same cluster [80, 35]. It is important to note that several factors affect p-distance calculation, including the length of the sequences and the number of sequences used in the analysis. The p-distance defined genotype system also needs to be updated frequently due to the accumulated viral diversity within genotypes over an increased circulation period and new genotypes are likely to be defined within the previously defined genotypes. Despite the need to easily recognize RSV genotypes for molecular epidemiology, vaccine design and control efforts, the delineating criteria are not agreed upon [19, 35, 28, 36]. There is a need for a robust system to define RSV genotypes and to resolve inconsistencies present in the literature arising from previous genotyping methods.

Our study proposed a novel and extendable RSV genotyping system based on a more complete RSV phylogeny. After evaluating the phylogeny inferred from different RSV gene datasets, we concluded that the WGS is the most informative and desirable dataset for RSV genotyping purpose. We categorized RSV into two classification levels, the genotype and subclade, mainly with phylogenetic analysis and detection year of sequences, which provides a convenient and sustainable way to refer to the emergence of RSV strains.

2.3 Materials and Methods

2.3.1 Data Management

RSV sequences from human clinical samples were retrieved from NCBI's GenBank nucleotide database using the search term "HRSV" on April 20, 2019. For all these sequences, metadata including the collection date, isolation country, and previously determined genotype were extracted from the GenBank records using the program gbmunge (https://github.com/sdwfrost/gbmunge). For spatial distribution estimation, the isolation country of each sequence

Table 2.1: **Recombination events in the RSV WGS dataset.**

Subtype	Recombinant	Minor Parent	Major Parent	Start in Alignment	End in Alignment	Best P-Value	Method
A	JX015495	JX015481	KT285064	6374	12817	1.48E-06	MaxChi
	JX069800	KP119746	MG642026	12247	12368	3.47E-02	RDP
	JX627336	KF826830	KX765941	543	1064	3.23E-02	RDP
	KJ672480	MG642074	KU950480	4222	4302	2.75E-08	GENECONV
	KJ672482	JX069800	MH760612	12254	12294	3.05E-02	GENECONV
	MF001054	JQ901456	MF001052	1951	2024	8.31E-14	RDP
B	KJ627251	KJ627342	KJ627254	9522	13303	8.82E-10	RDP
	KJ672473	LC385000	KJ672481	9522	13303	3.38E-04	RDP
	KJ939932	Unknown	KJ939934	3742	6407	6.63E-07	RDP
	KJ939933	KJ939931	MH760701	3959	6516	5.08E-09	SiSscan
	KY249663	KY249674	KY249669	3490	4262	1.33E-10	RDP
	KY924878	MH760677	Unknown	5544	13586	1.04E-09	Maxchi

has been further grouped into 6 WHO regions [81]. These sequences were then assigned to a subtype based on the best match in a nucleotide BLAST alignment against RSV-A and RSV-B reference sequences GenBank acc. no: $NC\_038235.1$ and $NC\_001781.1$). Sequence alignment was generated using MAFFT.v7 [82] and subsequently manually edited in Seqotron to accommodate the open reading frames of all genes [83]. The following inclusion and exclusion criteria were applied: a) each sequence must include collection date (at least year); b) the sequence length for each gene region should be longer than 70% length in the reference sequence; c) sequences with unexpected spurious frame-shift indel in the alignment were removed; d) the recombinant sequences that could interfere the phylogenetic inferences, were identified using the detection methods RDP, GENECONV, MaxChi, BootScan, and SiScan as implemented in the Recombination Detection Program RDP4 were removed [84] (Table 2.1). The final datasets consisted of 860 RSV-A sequences and 591 RSV-B sequences, respectively. The open reading frames of 10 RSV genes (NS1, NS2, N, P, M, SH, G, F, M2, and L) were extracted and the whole-genome sequence (WGS) was generated with a concatenation of each gene.

2.3.2 Phylogenetic Inference

Phylogenetic analysis for different gene datasets was conducted with maximum-likelihood (ML) approaches using RAxML v8.0 [85], which has the advantage of partitioned analysis. We applied the autoMRE option embedded in RAxML for an efficient convergence of bootstrapping process, where the bootstrapping value is one of the criteria for the genotype assignment. We implemented an indel coding method to code gene duplication and deletion region of the G gene as separate binary partitions. The rest of the nucleotides were set as a separate partition with the GTR + Gamma substitution model. The temporal signal of the WGS datasets was diagnosed using TempEst and temporal outliers were removed [86]. Tree topology tests for the phylogenies inferred from different RSV gene datasets were performed using IQ-TREE with the Shimodaira-Hasegawa (SH) test and approximately unbiased (AU) test [87]. The evolutionary rates for different genes were estimated using the program TreeTime [88].

2.3.3 Genotypes and subclades assignment for RSV

We aim to classify RSV strains into two levels in our analysis. The groups of strains that have potential to circulate are further defined as subclades under the classification level of genotypes. The genotype assignment is based on pair-wise node distance. Pair-wise node distances, which are the distances between the most common ancestors of groups of sequences in a phylogeny, were calculated between all nodes in a phylogenetic tree based on the alignment of the RSV whole-genome sequences, using the RRphylo v2.5.7 package [89]. We further employed time (years) of detection for sequences within each clade as another criterion for subclade assignment, which is calculated by the oldest and the latest date of the sequences within clade using personal scripts in R v4.0.2. The R package ape v2.3 [90] were used to define genotypes and subclade, whereas visualizations were created with ggtree v1.16.0 [91].

2.3.4 Genetic Distance Analysis

To characterize the genetic diversity within and between genotypes, the average genetic distance within and between genotypes as well as subgroups were estimated from the alignments with the software MEGA X [92] using the most simplified method, p-distance, which is the proportion of nucleotide sites at which two sequences being compared were different.

2.3.5 RSV genotype classification tool

RSV genotype classification tool was built with Lineage Assignment by Extended Learning (LABEL https://wonder.cdc.gov/amd/flu/label/) pipeline [93], which rapidly determines cladistic information for sequences using support vector machines (SVM) without the need for time-consuming sequence alignment, phylogenetic tree construction or manual annotation. Sequences with an annotated genotype or subclade were used to create a training data library. Training data for each genotype was sub-sampled in an ad hoc manner using PDA v1.0.3 [94] (Table 2.3). The classification module (available at https://github.com/JianiC/rsv-genotype/tree/master/LABEL/RSV) was then built with training sequences and the custom scripts that are implemented in the LABEL program. Both WGS and partial sequences of RSV can be automated to a genotype or subclade and no further information is required.

2.4 Results

2.4.1 Phylogenetic analysis with different RSV gene datasets

We first characterized the indels of RSV sequences in our dataset. In addition to the previously defined RSV-A genotype ON with a 72-nt duplication in the second HR of G gene and RSV-B genotype BA with a 60-nt duplication in a similar region, we also observed a 6-nt deletion in the recent circulating RSV-B strains at the G gene region (Figure 2.1A).

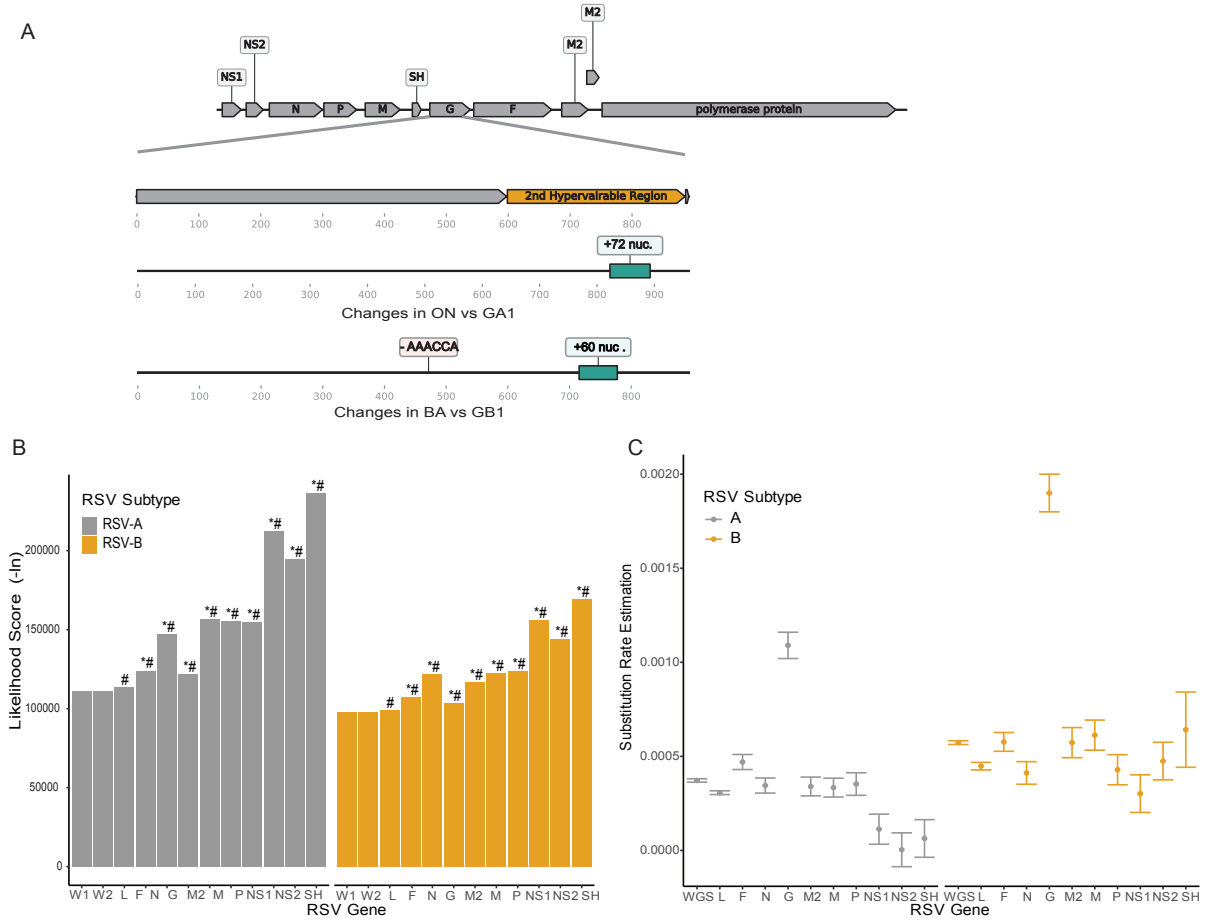

Figure 2.1: **Scheme of RSV genome and comparison of RSV phylogenies inferred from different gene datasets.** (A) RSV genome organization with G gene duplication and indels. (B) Likelihood scores of phylogenies inferred from different gene sequences given to the WGS dataset. W1, the phylogeny inferred from WGS with G gene indels implemented as a single evolutionary event; W2, the phylogeny inferred from WGS with G gene indels implemented as multiple substitution events. *, p$\leq$ 0.005 in Shimodaira-Hasegawa (SH) test compared with W1; , $p \leq 0.005$ in approximately unbiased (AU) test compared with W1. (C) Comparison of evolutionary rates that are estimated from different gene regions. Error bars indicate the confidence intervals of the estimation.

The RSV phylogenies were constructed for each gene as well as WGS using the ML approach and the G gene duplication and deletion region have been further coded as a separate binary partition in our analysis. We also scored the likelihood of phylogenies inferred from different gene datasets given by the WGS dataset to compare the topologies of different phylogenetic trees. The SH test and AU test suggested the phylogenetic trees inferred from individual RSV genes have significantly different likelihood scores compared with WGS, and we did not observe the significant difference with the phylogeny where G gene indels were simply considered as multiple substitution events (Figure 2.1B). The tree topology differences inferred from the WGS and individual gene sequences (L, G and F which have close likelihood scores with that of WGS) have been further demonstrated using a tanglegram approach as seen in Figure 2.2. The mean nucleotide substitution rates for RSV-A and RSV-B estimated from WGS are 3.72×104 and 5.73×104 substitutions/site/year, respectively. The rate estimation of the G gene was approximately 2.5-fold faster than other genes (Figure 2.1C). Overall, our results indicate the phylogenetic analysis based on whole-genome sequences can provide more valuable insight on RSV genetic diversity and evolution.

2.4.2 Novel RSV genotype system with whole-genome sequences

The following criteria were used to build a standardized RSV genotyping system:

1) Genotype and subclade designations are based on the phylogeny derived from WGS.

a. A supported monophyletic clade is defined with 70% bootstrap value at the node.

b. Genotypes are assigned by a maximal pair-wise node distance within the clade, 0.018 for RSV-A and 0.024 for RSV-B (we simulate genotype assignment with different cutoff values in Figure 2.3).

c. Each genotype must contain at least 3 isolates.

d. Under the genotype level, the supported monophyletic clades with a detection time of at least 5 years are assigned as subclades. Time (years) of detection for each monophyletic clade is calculated by the oldest and the latest year of isolation of the sequences.

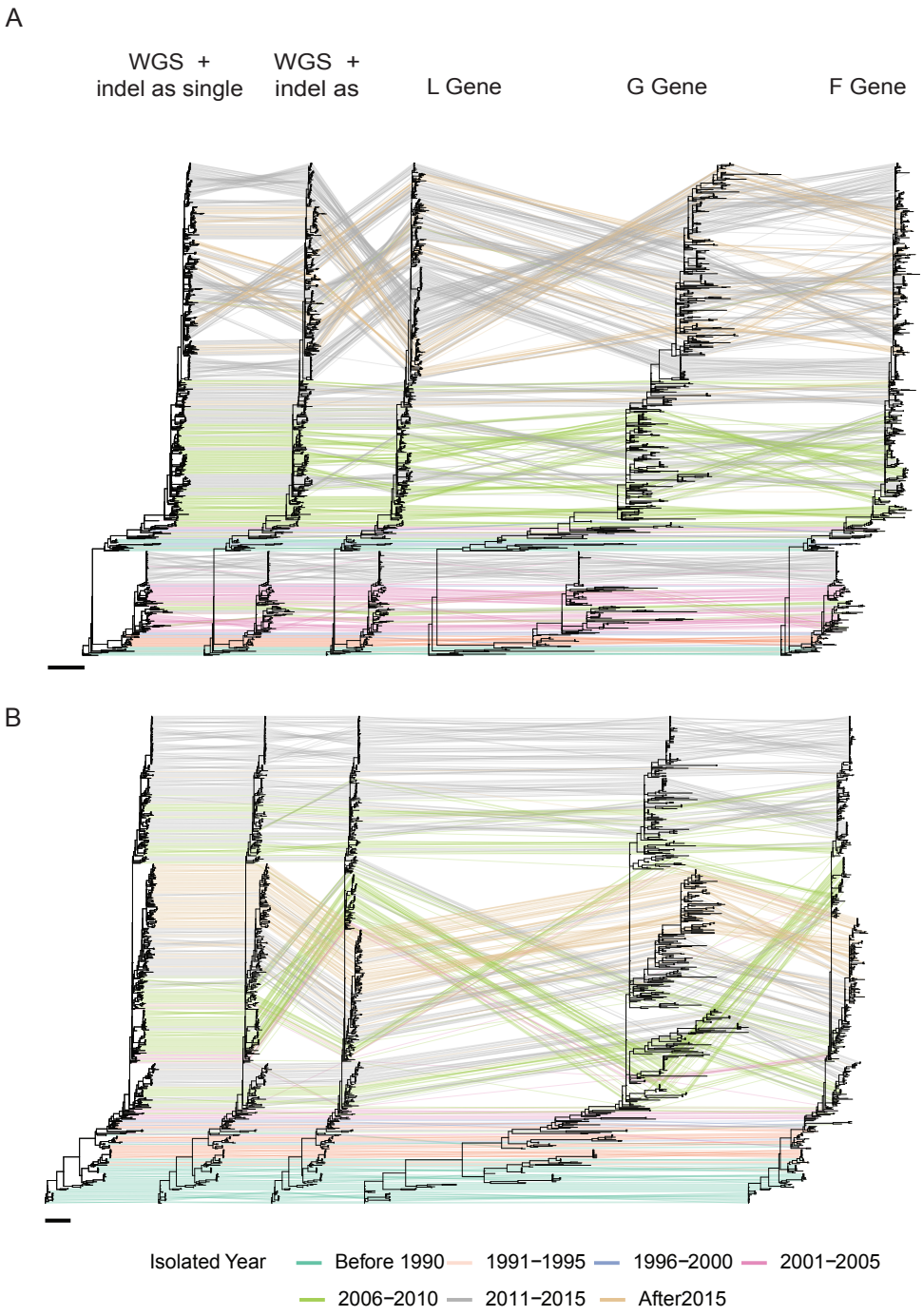

Figure 2.2: **Maximum likelihood phylogeny of RSV-A (A) and RSV-B (B) phylogeny inferred from WGS, L, G and F genes (from Left to right).** The color of the connected line between taxa indicates the isolated year for each strain. Scale bars indicate 0.01 nucleotide substitution per site.

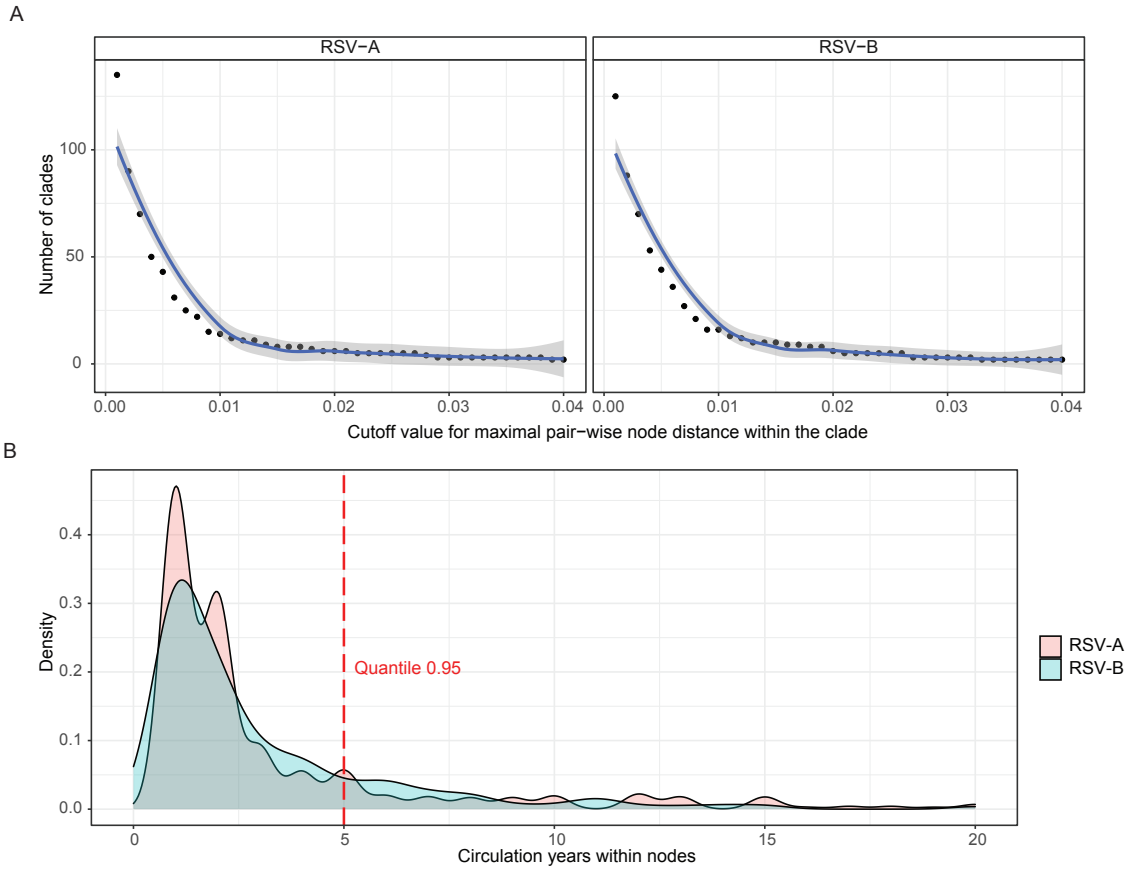

Figure 2.3: **Criteria to assign genotypes and subgroups in RSV whole-genome sequence phylogeny.** (A) Number of genotypes to be assigned with different cut-off value of pair-wise node distance. (B) Density distribution of clade circulation time (year) in RSV whole-genome sequence phylogeny. Red dashed line indicates the 0.95 quantile of the distribution.

2) The genotype or subclade containing the oldest isolated sequence for RSV-A or RSV-B is named as A.1 or B.1, following the same logic for naming the next genotype or subclade.

Using this scheme, we identified 5 genotypes in RSV-A (Figure 2.4A, Table 2.2). A.1 mainly contains the uncharacterized RSV-A strains that circulated in the past, A.2 contains the previously defined GA5 lineage and 4 subclades denoted A.2.1- A.2.4 were identified. A.3

is mainly composed of viruses that are previously described as GA7 genotype. A.4 and A.5 contain the predominantly known global genotype GA2. A.5 has been subdivided into 11 descendant subclades (denoted A.5.1- A.5.11) and strains associated with the 72-nt G gene duplication (ON) are found within genotype A.5.7- A.5.11. We categorized RSV-B into 5 genotypes (Figure 2.4B, Table 2.2). B.1, B.2, and B.3 contain previously described GB1, GB2 and GB4 genotype, respectively. B.4 genotype contains a relatively small group of sequences that have not been characterized before. B.5 contains the BA genotype, which contains 60-nt duplication event in the G gene and is currently divided into 10 subclades.

To understand the genetic diversity of RSV, we computed the intra-genotypic and inter-genotypic p-distance for the genotypes and subgroups assignment (Figure 2.5). The inter-genotypic p-distance for genotypes or subgroups is generally higher than the value of intra-genotypic p-distance. Some discrepancies were observed among subgroups of A.5 in which inter-genotypic p-distances were lower than the threshold compared to the intra-genotype p-distance of other subgroups.

2.4.3 Spatial and temporal distribution of RSV genotypes

We provide a description of the spatial and temporal distribution of RSV genotypes even though the bias in the samples sequenced do not provide a complete resolution of the past distribution of RSV variants. According to our revised genotyping system, RSV-A had at least two important shifts in the dominant genotypes (Figure 2.6A). Before 1990, genotype A.1 and A.2 other old strains that were isolated from the region of the Americas were the dominant genotypes. Since then, A.2 replaced these strains, became the dominant genotype, and co-circulated with A.3 and A.4 in the American and European regions. Recently, a new genotype A.5 emerged and transmitted globally, but the previous dominant strains assigned as genotype A.2 were still circulating in some regions. Genotype B.1, B.2 and B.3 were the dominant genotypes for RSV-B in the past (Figure 2.6B). Post-1995, the novel genotype B.5 emerged and became fixed in the population and has been circulating globally. B.4 is a group

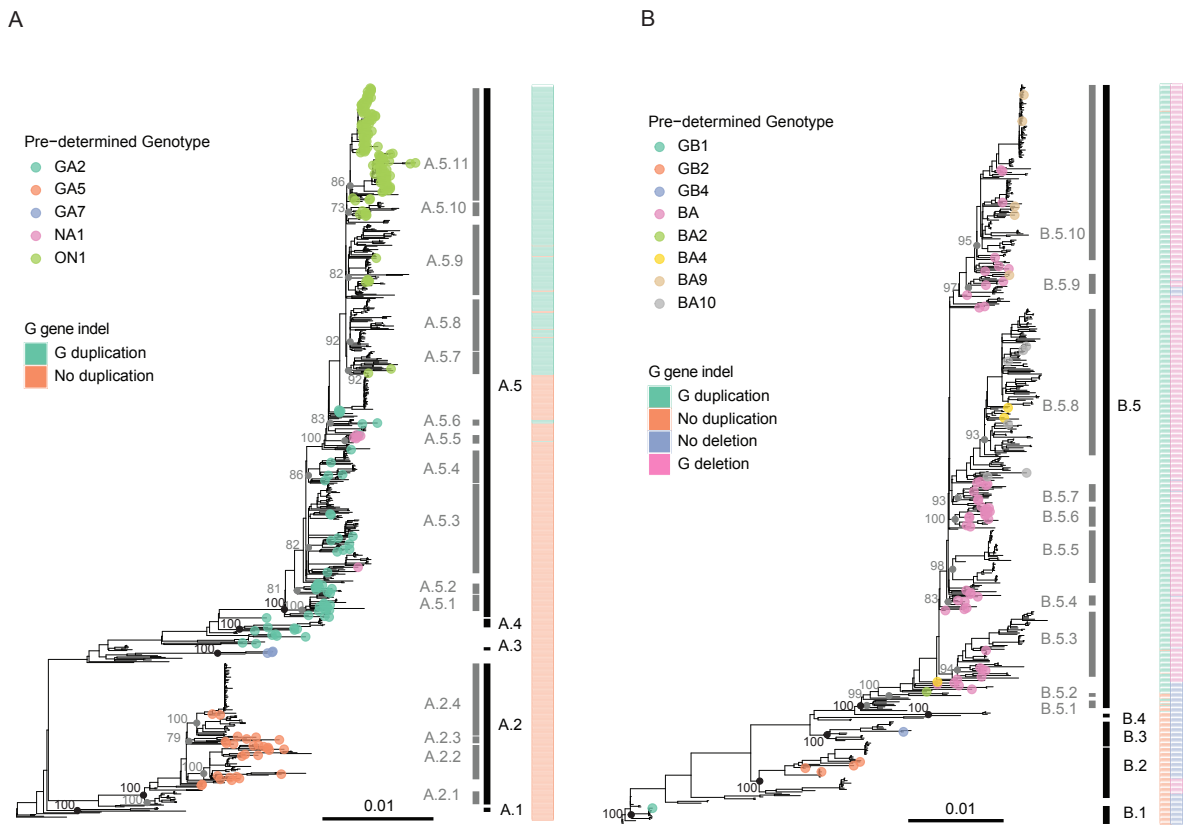

Figure 2.4: **Maximum likelihood phylogeny of RSV-A (A) and RSV-B (B) inferred from WGS with the genotyping assignment.** The genotype assignments are indicated with vertical black bars and are labeled on the right. The subclade assignments are indicated with vertical gray bars and are labeled on the left. Tip point colors represent the previously defined genotype names based on complete or partial G gene sequences. The nodes that define the genotype and subclade are indicated with black and gray node points, respectively. Bootstrap of each ancestral genotype/ subclade node is detailed. Colored columns on the right side represent G gene duplication and indels. Scale bars indicate 0.01 nucleotide substitution per site.

Table 2.2: **List of previously defined genotype name and detection time of new RSV genotype assignment.** a. Periods were detected up to 2017 and may underestimate the circulation time due to bias in GenBank deposition practices. b. The previously defined genotype name was collected from GenBank.

Subtype	Genotype	Subclade	Detection time	Previously defined genotype name
A	A.1		1978-1998	
	A.2			
		A.2.1	1990-1994	
		A.2.2	2001-2015	GA5
		A.2.3	2001-2015	GA5
		A.2.4	1998-2013	GA5
	A.3		1984-1998	GA7
	A.4		1998-2009	GA2
	A.5	A.5.1	2007-2015	GA2
		A.5.2	2006-2010	GA2
		A.5.3	2008-2015	GA2, NA1
		A.5.4	2008-2015	GA2
		A.5.5	2011-2015	NA1
		A.5.6	2011-2015	GA2
		A.5.7	2012-2016	ON1
		A.5.8	2012-2016	
		A.5.9	2008-2016	
		A.5.10	2012-2017	ON1
		A.5.11	2011-2016	ON1
B	B.1		1979-1987	GB1
	B.2		1979-1991	GB2
	B.3		1989-2002	GB4
	B.4		2008-2012	
	B.5	B.5.1	1992-1996	
		B.5.2	1997-2013	
		B.5.3	2008-2015	BA
		B.5.4	2004-2009	BA
		B.5.5	2006-2012	
		B.5.6	2006-2015	BA
		B.5.7	2006-2013	BA
		B.5.8	2012-2016	BA
		B.5.9	2008-2013	BA
		B.5.10	2009-2016	BA

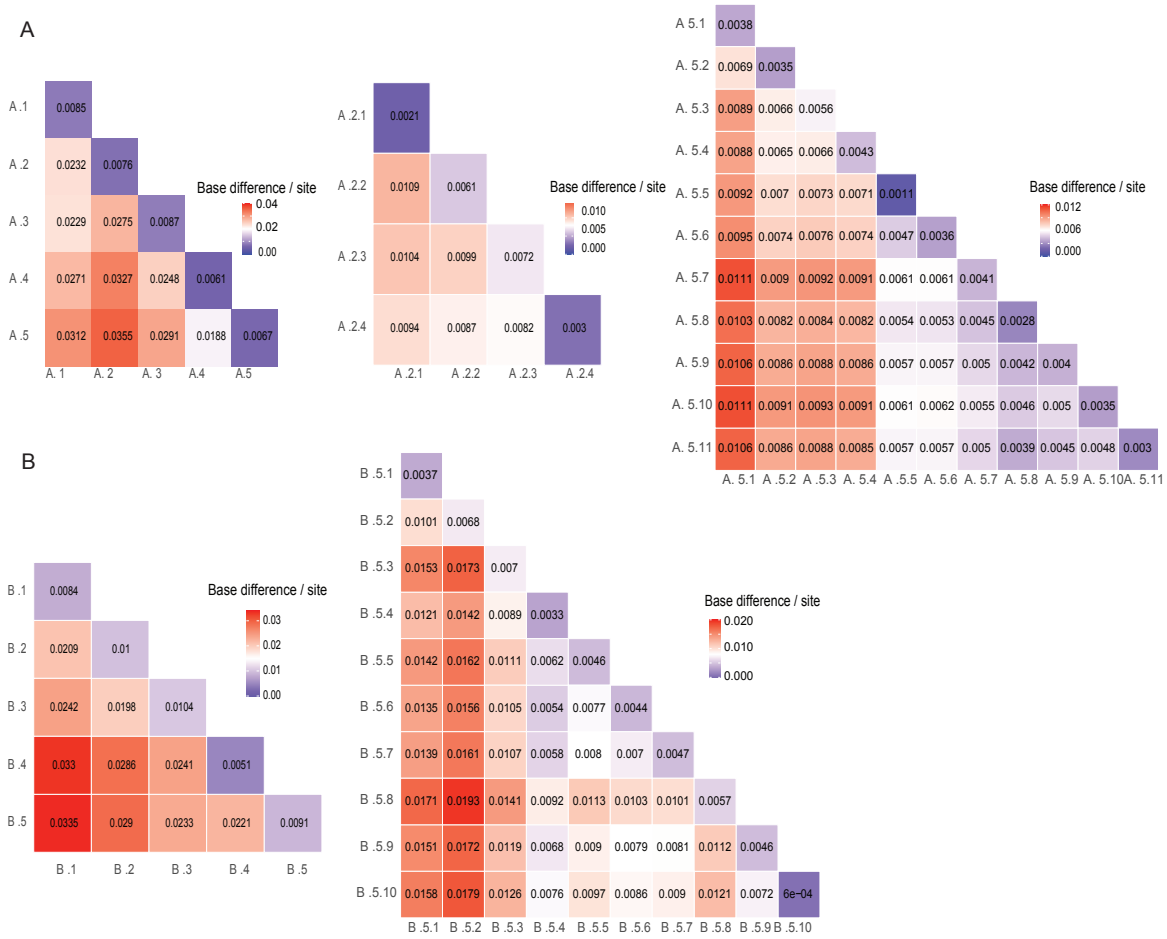

Figure 2.5: **P-distance calculation within and between RSV genotypes.** (A) RSV-A intra-genotypic and inter-genotypic p-distance for genotypes (left), subclades within genotype A.2 (middle) and genotype A.5 (right). (B) RSV-B intra-genotypic and inter-genotypic p-distance for genotypes (left), subclades within genotype B.5 (right).

of strains that are detected after 2006. We have also deployed our genotyping assignment using the open-source tools Nextstrain, which provides a graphical demonstration of the global transmission events and genomic diversity over time with our new RSV genotype assignment (https://nextstrain.org/community/JianiC/rsv-genotype) [95].

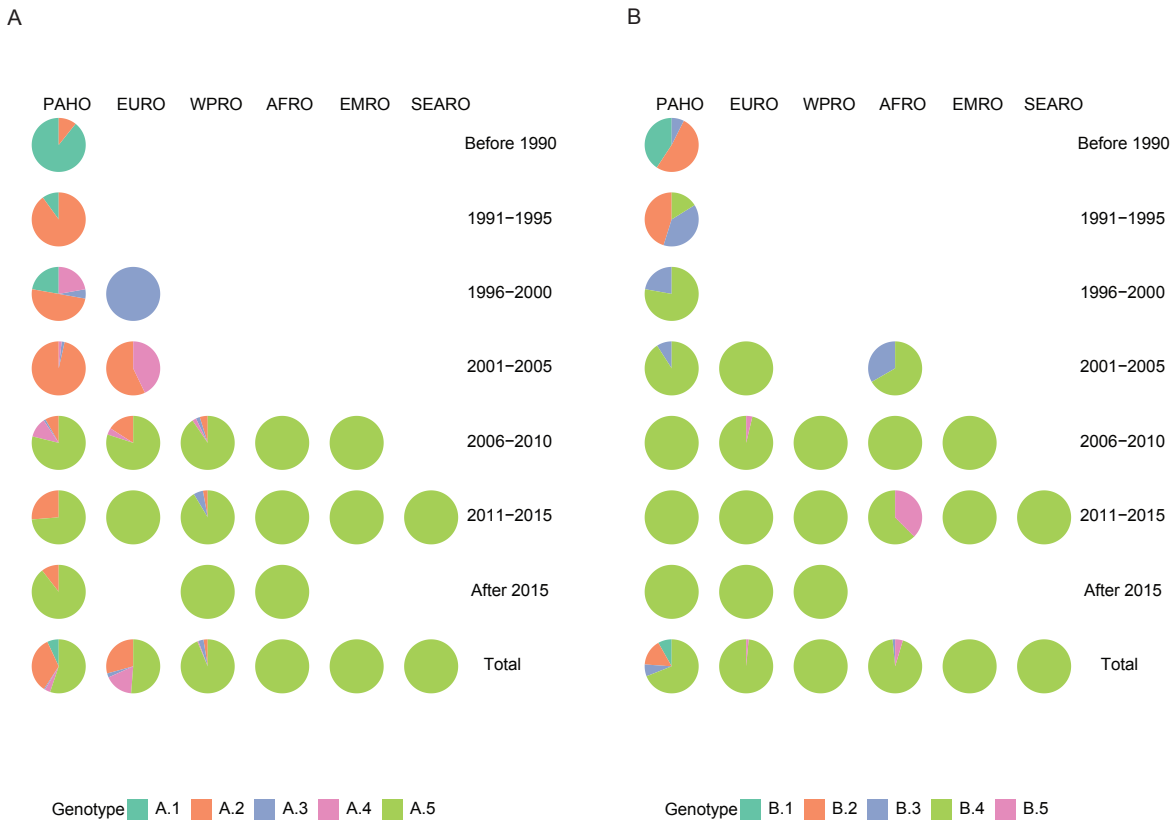

Figure 2.6: **Spatial and temporal distribution of RSV-A (A) and RSV-B (B) genotypes.** The temporal and spatial distribution of RSV genotypes is based on the detection year and isolated WHO region of sequence for each assigned genotype. African Region (AFRO), Region of the Americas (PAHO), South-East Asia Region (SEARO), European Region (EURO), Eastern Mediterranean Region (EMRO), and Western Pacific Region (WPRO).

2.4.4 Automated RSV classification tool

Representative RSV genotypes were used to build a custom RSV classification module within LABEL [93]. The classifier ascribed the correct genotypes and subclades in all sequences but 49 instances, with 95.4% accuracy. Of these 49 sequences, 32 sequences from genotype B.2 were incorrectly assigned as genotype B.1, which are ancient RSV-B strains circulating before 2000. The remaining 17 sequences were assigned to a sister clade (subclade) that shared ancestry with the correct genotype assignment (Table 2.3 and Table 2.4). Overall, this classifier is fast and accurate in capturing our RSV classifications without requiring phylogenetic reconstruction.

2.5 Discussion

In this study, we highlight the importance of WGS for RSV genotype assignment. We statistically compare the phylogeny derived from different RSV gene datasets with the likelihood score test and evolutionary rate estimation from different gene datasets. Our results indicate WGS should be used for genotype assignment. Our analysis is based on a recombination-free dataset to avoid inferential biases. 12 RSV sequences in our initial dataset showed some evidence of potential recombination. Since genomic recombination in RSV is believed to be extremely rare, it is most likely that these recombinants arose as a result of PCR or sequencing artifacts [27]. Even though we did not observe a significant statistical difference in our analysis, the G gene duplication should be considered as a single biological event and we implement an indel coding method to improve phylogenetic resolution.

There are several expectations for a widely acceptable RSV genotyping system. First, since more than 30 genotypes have been identified, we expect to have a reasonable number of RSV genotypes to simplify the study with different RSV strains. Secondly, genotype assignment should be able to capture the genetic diversity, thereby providing valuable insights into the ongoing evolution of the virus and playing an important role in its mitigation and control [96]. Finally, we expect the novel emergent strains could be classified and easily

Table 2.3: **Accuracy of RSV genotype assignment tool.** a. Annotation column contains the number of sequences in genotype or subclade. b. Classfication tool column contains the number of sequences that are assigned to the correct genotype or subclade using classification tool. c. Accuracies are caluclated by the percentage of sequences that are assigned to the correct genotype or subclade using classification tool.

Genotype	**Subclade**	**Annotation**	**Sequences for training module**	**Classification tool**	**Accuracy(%)**
A.1		6	6	6	100.00%
A.2		151	65	151	100.00%
	A.2.1	16	16	16	100.00%
	A.2.2	41	20	41	100.00%
	A.2.3	9	9	9	100.00%
	A.2.4	85	20	85	100.00%
A.3		6	6	6	100.00%
A.4		15	15	15	100.00%
A.5		516	196	516	100.00%
	A.5.1	20	20	20	100.00%
	A.5.2	13	13	13	100.00%
	A.5.3	105	20	105	100.00%
	A.5.4	39	20	38	97.40%
	A.5.5	11	11	10	90.90%
	A.5.6	8	8	4	50.00%
	A.5.7	27	27	27	100.00%
	A.5.8	61	20	57	93.40%
	A.5.9	83	20	83	100.00%
	A.5.10	17	17	17	100.00%
	A.5.11	132	20	120	90.90%
B.1		11	11	11	100.00%
B.2		42	20	10	23.81%
B.3		16	16	16	100.00%
B.4		4	4	4	100.00%
B.5		424	149	419	98.80%
	B.5.1	7	7	7	100.00%
	B.5.2	4	4	4	100.00%
	B.5.3	53	20	53	100.00%
	B.5.4	9	9	9	100.00%
	B.5.5	43	20	43	97.70%
	B.5.6	17	17	17	100.00%
	B.5.7	15	15	14	93.30%
	B.5.8	118	20	118	100.00%
	B.5.9	17	17	17	100.00%
	B.5.10	141	20	141	100.00%
Total		1202	477	1147	95.42%

Table 2.4: **Accuracy of RSV genotype assignment tool with different test datasets.** Accuracy is evaluated by the number of sequences that are correctly annotated over the number of test sequences.

Test Data	Sequence Count	Minimum sequence length (nt)	Maximal sequence length (nt)	Accuracy
SVM training set self-validation)	1176	14902	15333	95.4%
RSV full length sequences submitted after April, 2019	582	14906	15276	100%
G gene sequences from SVM training set	1176	896	970	91.8%

added into the revised nomenclature system. Competing RSV genotype systems have been proposed (Table 2.5), including an influenza-like system for RSV genotype classification based on the highest intra-genotypic p-distance as the minimum threshold to define a genotype [35, 28, 97], which are highly sensitive to sampling bias. Another recent systematic RSV genotype study attempted to use patristic distance (the shortest distance between two tips) instead of p-distance to propose a new classification system [36]. Both p-distance and patristic distance are sensitive to the sequencing error and the length of sequences. In addition, a cutoff value of either genetic distance or patristic distance is always needed to standardize the molecular classification of RSV strains, which is likely to be problematic due to sampling bias in RSV and delineation criteria may need to be re-evaluated with continued surveillance of RSV strains [28, 36]. We calculate pair-wise node distance to assign genotype in our analysis. Our approach relies on the genetic divergence calculated from the tree tips to the most recent common ancestor for each genotype, which is less sensitive to the individual sequence quality and has advantages to the under-sampled RSV sequences. In addition to the genetic differences between RSV genotypes, previous studies have suggested some RSV genotypes may have an advantage in transmission and circulation [98]. Instead of identifying every lineage or strain, one of the major goals in this manuscript is to identify the group of strains that have

the potential to circulate and to keep tracking them since their emergence. Therefore, we include the circulation time as a criterion to characterize the RSV strains that continue to circulate with a potential to be recognized as an emerging subclade. These are strains that may require elevation to genotype level with continued circulation. We expect new genotypes and subclades will be defined as RSV continues to circulate. The subclades identified are those strains among co-circulating variants that are currently most likely to require monitoring. In addition, some reported genotypes may have an increased risk to cause severe symptoms [99], which is an important characterization of a classified genotype. However, these studies have limits that prevent consistent predictions to make firm conclusions about the potential clinical relevance of the different RSV genotypes. More information about the correlations between RSV strains and disease severity is needed to include these features in a genotyping system.

It is crucial to share our updated genotype assignment so that new sequences can easily be added. We deploy our genotype assignment as well as the genotype assignment from previous published studies with Nextstrain, which allows a comparison and a continually updated visualization [95]. We also provide a tool that enables the automated classification of newly generated RSV sequences. By using the platform provided, RSV sequences can be assigned with genotypes and subgroups based on the similarity of the sequences that are included in our system [93]. This fast and accurate RSV genotyping assignment tool will be valuable for the classification of novel sequences in future phylogenetic or diagnostic settings.

There are several limitations that need be addressed with any molecular systematic revision of RSV genotypes. In particular, RSV genotype assignment using WGS is subject to sampling bias. Only a limited number of sequences are currently available in GenBank, especially among older samples that were sequenced prior to the widespread and routine use of WGS, which may affect our genotype assignment. In addition, most samples sequenced were isolated from the regions in the Americas. With more RSV sequencing effort, we would

Table 2.5: **Comparison of RSV molecular systematic proposals.**

Reference	This study	Peret et al 1998 [19]	Agoti et al 2014 [35]	Goya et al 2020 [28]	Ramaekers et al 2020 [36]
Dataset	RSV whole-genome sequence with isolation year in GenBank up to April 2019	G gene sequence in GenBank and sequences obtained in the laboratory	RSV G gene sequence isolated from 2006 through 2011 available in GenBank and sequences obtained in the laboratory	All RSV G ectodomain sequences available in GenBank up to February 2018	All RSV whole-genome sequences available in GenBank up to January 2019
Genotyping region	CDS region of full-length genome	second hypervariable region of G ectodomain	G gene	G ectodomain	full-length genome
Genotyping criteria	genotype: 70% bootstrap and clade detection time 10 years; subgroup: clade detection time 5 years within genotypes	genotype: 70% bootstrap; subtype: 96% nucleotide similarity within genotype	60% bootstrap and average genetic distance cutoff 1.5%	80% bootstrap and p-distance 0.03 subs/site	70% bootstrap and patristic distanc >0.018 subs/site for RSV-A, patristic distance >0.026 subs/site for RSV-B
Genotype name	A.1-A.5, (subgroup: A.2.1-A2.4, A.5.1-A.5.11); B.1-B.5 (subgroup: B5.1-B5.10)	GA1-GA5 (22 subtypes among 5 genotypes), GB1-GB4 (6 subtypes among 4 genotypes)	GA1-GA7, SAA1; GB1-GB4, SAB1-SAB3, BA	GA1-GA3, GB1-GB5;	A1-A23; B1-B6

expect the geographic distribution of sequence could be captured and effectively used for future updates to this genotype system.

In summary, we propose a revised RSV genotyping assignment that reflects the genetic diversity and circulation pattern of RSV. WGS should be used for future RSV genotyping revisions. In addition, the G gene duplication and other indels should be taken into account for the phylogenetic analysis as a single evolutionary event rather than multiple substitution patterns. Overall, a robust RSV genotype assignment based on WGS will greatly assist those working in clinical identification, epidemiological studies, and vaccine development.

Chapter 3

Diversity and Evolution of Computationally Predicted T Cell Epitopes against Human Respiratory Syncytial Virus[1]

[1]Jiani Chen Swan Tan Vasanthi Avadhanula, Leonard Moise, Pedro A Piedra, Anne S De Groot, Justin Bahl, Diversity and Evolution of Computationally Predicted T Cell Epitopes against Human Respiratory Syncytial Virus.

Submitted to Plos Computational Biology, July 2022.

3.1 ABSTRACT

Human respiratory syncytial virus (RSV) is a major cause of lower respiratory infection. Despite more than 60 years of research, there is no licensed vaccine. While B cell response is a major focus for vaccine design, the T cell epitope profile of RSV is also important for vaccine development. Here, we computationally predicted putative T cell epitopes in the Fusion protein (F) and Glycoprotein (G) of RSV wild circulating strains by predicting Major Histocompatibility Complex (MHC) class I and class II binding affinity. We limited our inferences to conserved epitopes in both F and G proteins that have been experimentally validated. We applied multidimensional scaling (MDS) to construct T cell epitope landscapes to investigate the diversity and evolution of T cell profiles across different RSV strains. We find the RSV strains are clustered into three RSV-A groups and two RSV-B groups on this T epitope landscape. These clusters represent divergent RSV strains with potentially different immunogenic profiles. In addition, our results show a greater proportion of F protein T cell epitope content conservation among recent epidemic strains, whereas the G protein T cell epitope content was decreased. Importantly, our results suggest that RSV-A and RSV-B have different patterns of epitope drift and replacement and that RSV-B vaccines may need more frequent updates. Our study provides a novel framework to study RSV T cell epitope evolution. Understanding the patterns of T cell epitope conservation and change may be valuable for vaccine design and assessment.

3.2 Introduction

Human respiratory syncytial virus (RSV) is a negative-strand RNA virus that is classified in the Orthopneumovirus genus of the family Pneumoviridae. It is a major cause of lower respiratory disease in young infants, immunocompromised individuals, and elderly people, resulting in annual epidemics worldwide [100]. The single-stranded RNA genome of RSV is approximate 15.2 kb and encodes 11 viral proteins [101]. The Fusion (F) and Glycoprotein (G)

proteins are the two major surface proteins [25]. F protein is generally thought to be conserved and therefore it is the focus of most current RSV vaccine designs. Although G protein is highly variable, its contribution to disease pathogenesis and its role in the biology of infection suggest it can also be an effective RSV vaccine antigen [102]. Despite the significant burden of RSV infection worldwide, there is no licensed vaccine. The only approved intervention is passive immuno-prophylaxis with palivizumab, which is achieved by administering the monoclonal antibody (mAb) to a highly restricted group of infants under the age of 24 months and treatment must be repeated monthly during the RSV season due to the relatively short half-life of the antibody [103, 104]. Due to the high cost of monoclonal antibody treatments, this intervention is limited to high-risk infants and is generally unavailable in developing countries. An RSV vaccine is an urgent global healthcare priority, and it is likely that different strategies are needed for the various high-risk groups.

A number of research teams have worked on the development of RSV vaccine since its isolation and characterization in 1956 [105, 106]. However, vaccination with the formalin-inactivated, alum precipitated RSV (FI-RSV) vaccine in RSV-naïve infants and young children, led to the development of vaccine enhanced disease (VED) that hampered vaccine development for decades to follow [107]. Many studies have been conducted to explain this undesirable outcome. It is likely that formalin fixation led to a vaccine that mostly presented the post-fusion conformation of RSV F protein, leading to an excess of non-neutralizing antibodies and immune complex formation [108, 109, 110]. Other studies indicated that an impaired T cell response with Th2 skewing [111, 112], as well as complement deposition in the lungs, contributed to enhanced neutrophil recruitment [110]. Recent developments, including the resolution of the F protein [113] and the development of RSV rodent models [114] have contributed to a number of vaccine candidates with novel designs and formulations currently in clinical trials [60, 48, 25].

While most current RSV vaccination strategies focus on a B-cell-induced neutralization immune response, T cell immunity also plays a major role in the resolution of virus infection

and is essential for RSV vaccine development [60, 48]. Once RSV infection of the lower airways is established, CD8 T cells play an important part in viral clearance and CD4 helper T cells can orchestrate cellular immune responses and stimulate B cells to produce antibodies. However, Th2-biased responses have been associated with animal models of RSV VED, and measurement of Th1 and Th2 responses are considered important to predict the safety of vaccine candidates [110]. Therefore, induction of a balanced cell-mediated immune response through vaccination would promote RSV clearance, but caution must be taken to avoid the potential for immunopathology. Taken together, a closer examination of T cell immunity and the virus sequences that induce T cell responses are needed for RSV vaccine development.

Human respiratory syncytial virus has a complex circulation pattern in the human population. Within two antigenic groups, RSV-A and RSV-B, different genotypes can co-circulate within the same community, while novel RSV genotypes with high genomic diversity may arise and potentially replace the previously dominant genotypes [115]. In recent years, several unique genetic modifications in RSV have been identified, including a 72-nucleotide (nt) duplication (ON genotype) in RSV-A G gene and another with a 60-nt duplication (BA genotype) in RSV-B at a similar region [77]. The observed RSV genetic diversity has raised a question about whether it is necessary for an RSV vaccine to include several different strains to be effective. Most current RSV vaccine developments are based on an RSV A2 laboratory strain, which is a chimeric strain that belongs to subtype A [116]. While these treatments hold promise, there is the possibility of viral strains developing escape mutations. For example, palivizumab-resistant strains have been isolated from both RSV rodent models and human [117, 60]. Several lines of evidence also suggest antigenic variation may play a role in the ability of RSV to escape immune response and established infections [78]. While highly conserved T cell epitopes in RSV vaccine may not provide complete protection against infection when cross-protective antibody responses are lacking, highly conserved T cell epitopes in the vaccine may still reduce the severity of the illness and limit the spread of the virus. However, amino-acid variation at the T cell epitope level and the potential

emergence of novel T cell epitopes of recent RSV circulating strains have been reported [64], and further studies are needed to illustrate the effects of amino acid variations on T cell recognition. Hence, characterizing T cell epitope profiles across different strains is very important to understand RSV evolution and can be important for RSV vaccine development.

In this study, we utilize immunoinformatic approaches that are implemented in the iVAX toolkit [118] to predict T cell epitopes in RSV across different strains with a focus on the two major surface proteins F and G. With the analysis of a comprehensive dataset, we evaluate the lineage-specific T cell epitope profile of RSV. We also create sequence-based T cell epitope landscapes based on epitope content comparison across different strains and further correlate RSV T cell immunity change with virus evolution. The proportion of cross conserved T cell epitope content between vaccine candidate strains that developed earlier and RSV circulating strains with different isolated years and locations were also calculated. These analyses may aid in understanding RSV T cell immunity across different strains and contribute to current vaccine design efforts.

3.3 Materials and Methods

3.3.1 Dataset

RSV GenBank records files were retrieved from NCBI's GenBank nucleotide database using the search term "HRSVA" or "HRSVB" on June 22, 2020. F and G gene nucleotide sequences and metadata including country of isolation and collection date were extracted using customized python scripts. Genotype assignments were made with the program "LABEL", using a customized RSV module [45, 93]. Countries of isolation were grouped into 6 WHO regions: African Region, Region of the Americas, South-East Asia Region, European Region, Eastern Mediterranean Region, and Western Pacific Region [119]. The following inclusion and exclusion criteria were applied: (i) each sequence needed to have a known isolated geographic location and isolated year, (ii) each sequence had to be at least 80% of the complete gene sequence in length, (iii) identical sequences with the same isolate country were removed, and (iv) vaccine

derivative and recombinant sequences were removed. Using these criteria, comprehensive datasets of RSV F and G genes were defined (RSV-A F gene = 1010, RSV-B F gene = 894, RSV-A G gene = 1488, RSV-B G gene = 1120). Nucleotide sequences from each dataset were aligned using MAFFT.v7 [120] and were translated into amino acids using EMBOSS.v6.6.0 [121] for immunoinformatic analyses. In addition, two artificial sequences, CP248 and CP52 (cold passage live RSV strains that were previously evaluated as vaccine candidates, Accession No: U63644, AF0132551 respectively) were downloaded from the NCBI's GenBank nucleotide database [122].

3.3.2 Phylogenetic inference

The nucleotide sequences of RSV major surface proteins were used to reconstruct the maximum-likelihood (ML) phylogeny of RSV using RAxML.v8 with GTR+GAMMA substitution model [85]. The best-scoring ML tree was automatically generated from five runs by RAxML. Time-scaled phylogenies were further reconstructed with the best-scoring ML trees using the program "Timetree" [123]. The phylogenies are visualized in the R package "ggtree" [91].

3.3.3 T cell epitope prediction

RSV major surface protein sequences were scored for binding potential against a globally representative panel of Human Leukocyte Antigen (HLA) class I and class II alleles using the EpiMatrix algorithm. This algorithm as well as the ClustiMer, JanusMatrix, and EpiCC algorithms discussed below are part of the iVAX toolkit developed by EpiVax, which is available for use under a license or through academic collaborations [118].

Evaluation of class I epitopes was made based on predictions for four HLA-A and two HLA-B supertype alleles: A*01:01, A*02:01, A*03:01, A*24:02, B*07:02, B*44:03. Class II epitopes were identified for nine HLA-DR supertype alleles: DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01. These

are HLA allele supertypes (alleles sharing common binding preferences) that cover the genetic diversity of more than 95% of human populations globally [124, 125]. EpiMatrix parsed 9-mer sequence frames (each one overlapping the previous one by eight amino acids) from the antigen sequence and assigned a score for each nine-mer/allele pair on a normalized Z distribution. Nine-mer sequences that had Z-scores of at least 1.64 are considered to be in the top 5% of any randomly generated set of 9-mer sequences and to have a high likelihood of binding to HLA molecules and being presented to T cells. Sequences that score above 2.32 on the Z-scale (top 1%) are extremely likely to bind to a particular HLA allele and to be immunogenic. For this analysis, HLA-class I restricted 9-mer sequences that had top 1% binder scores to at least one HLA class I supertype allele were considered to be putative class I epitopes [118]. To identify putative class II epitopes, we used an algorithm called ClustiMer [118] to screen EpiMatrix scoring results for the nine class II alleles. ClustiMer identifies contiguous regions of 15–30 amino acids that have a high density of MHC class II binding potential. Epitope density within a cluster is reported as an EpiMatrix Cluster Score, where scores of 10 and above are likely to be recognized in the context of multiple class II alleles and to be high-quality class II epitopes.

3.3.4 Identification of cross-conservation between putative RSV epitopes and human peptides

We also applied analysis of human homology to this study. After identifying putative T cell epitopes sequences in RSV major surface proteins, the JanusMatrix algorithm [126] was used to assess the potential cross-conservation of T cell epitopes with epitopes restricted by the same HLA alleles in the human proteome database (Uniprot-sourced human proteins, UniProt [127]). JanusMatrix scans input peptides and takes the 9-mer epitope regions that are identified in EpiMatrix to find the human peptides with a compatible HLA facing-agretope (i.e. the agretopes of both the input peptide and its human counterpart are predicted to bind to the same HLA allele) and the same T cell receptor (TCR) facing epitope to compute as

a JanusMatrix Human Homology Score. As defined in retrospective studies, foreign class I epitopes that score greater than 2 and class II epitopes that score greater than 5 may be less immunogenic due to T cell tolerance [118].

3.3.5 Protein-level T cell immunogenic potential evaluation

RSV reference sequences (RSV-A: NC_038235, RSV-B: NC_001781) were downloaded from the NCBI RefSeq database and were used to evaluate the protein-level immunogenic potential of RSV major surface proteins. The protein-level immunogenic potential as represented by the EpiMatrix-defined T cell epitope density score was computed by summing the top 5% binder scores across HLA alleles and normalizing for a 1000-amino acid protein length. Zero on this scale is set to indicate the average number of top 5% binders that would be observed in 10,000 random protein sequences with natural amino acid frequencies. Proteins scoring above +20 have been observed to have the significant immunogenic potential [128]. Fully human proteins generally score lower than zero on the EpiMatrix immunogenicity scale. To investigate the distribution of T cell immunogenic potential across RSV protein sequence regions, we summed up the binding scores of HLA alleles for each nine-mer frame, to get a frame-specific immunogenic potential score and standardized this score to a relative scale. The relative immunogenic potential across protein structure was represented by a color scale and the visualization of F protein structure was built with PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC). Protein data bank (PDB) files 5UDE [129] and 3RRR [130] were used for the pre-fusion and post-fusion forms.

3.3.6 Subsampling strategy

Considering the heavy computational load that would be required to evaluate all available RSV sequences and to correct the overrepresentation of recently sampled strains, the comparative analysis for T cell epitope content was conducted with datasets in which overrepresented groups were reduced. A maximum of five sequences of each isolation year from different WHO

region groups were subsampled randomly from the original datasets (RSV-A F gene = 402, RSV-B F gene = 319, RSV-A G gene = 390, RSV-B G gene = 359).

3.3.7 T cell epitope content comparison

The Epitope Content Comparison (EpiCC) algorithm, which is implemented in iVAX was used to compare T cell epitope content within each subsampled dataset by evaluating cross-conserved T cell epitopes (9-mer peptides with identical TCR-facing residues and are predicted to binding to the same MHC allele) content between different virus strains [131]. We reasoned that epitopes with identical T cell receptor-facing residues (*TCRf*, position 4, 5, 6, 7, 8 for class I epitopes binding core and 2, 3, 5, 7, 8 for class II epitopes binding core), regardless of differences on their MHC-facing (*MHCf*) amino acids, which are also predicted to bind to the same MHC allele, are more likely to induce cross-reactive memory T cells (These epitopes are called cross-conserved T cell epitopes). To simplify the analysis, the binding of 9-mer epitopes within protein sequences are assumed to be mutually exclusive and uniform, which means the T cell immune response of the antigen protein can be represent by summing up all T cell epitopes within the protein sequence.

We use u to represent 9-mer peptides with similar *MHCf* capable of binding the same MHC alleles but bearing different *TCRf* (non-cross conserved T cell epitopes) in two wild circulating strains (w_1 and w_2). Because the T cell immune response to virus is directly related to its T cell epitope content, the T cell immune distance (D) between two strains can be represented by the sum of binding probabilities of these unique 9-mer peptides for a set of HLA alleles (Equation 3.1). $p(u)_a$ is the predicted binding probability of unique 9-mer peptide u against a single class I or class II allele a, which is a member of A, which represents a set of HLA alleles.

$$D\left(w_1, w_2\right) = \sum_{u \in (w_1, w_2)} \sum_{a \in A} p\left(u\right)_a \tag{3.1}$$

Since the calculation of T cell epitope immune distance relies on the predicted epitope binding affinity, we use another T cell epitope prediction tool to evaluate the T cell epitope immune distance generated by the EpiCC algorithm. We apply the Equation 3.1 to re-calculate T cell epitope immune distance with customized Python scripts (available at https://github.com/JianiC/RSV_Epitope/tree/master/NetMHCpan_reproduce) using MHC binding prediction results that are generated from publicly available T cell epitope prediction tool, netMHCpan EL 4.1 methods in the Immune Epitope Database (IEDB) [132]. Eigenvalues of each sequence that were calculated from the pairwise distance matrix with "RSpectra" package were used to statistically examine the correlation of the epitope distances that are computed from the two methods, and Pearson correlation test was used to test the correlation hypothesis.

The capacity for a vaccine to induce a T cell immune response that could be recalled by a wild circulating strain is related to the cross-conservation of the T cell epitopes between the vaccine strain (v) and the wild circulating strain (w). For each pair of 9-mer peptides i (from strain v) and j (from strain w) that are cross-conserved (i.e. bearing identical residues that face the TCR), the probability to recall cross-reactive T cell memory by those two 9-mer peptides via a single HLA allele a can be represented by the joint estimation of the binding probability of these two 9-mer peptides $(p(i)_a * p(j)_a)$.

Therefore, a T cell epitope similarity score (S) between two sequences can be represented by summing the probability to cross-reactive memory T cells by all paired 9-mer peptides that are cross-conserved between the vaccine strain (v) and wild circulating strain (w) against a set HLA alleles A (Equation 3.2).

$$S(v,w) = \sum_{i \in v, j \in w} \sum_{a \in A} (p(i)_a * p(j)_a) \tag{3.2}$$

We further normalized the T cell epitope similarity score between the vaccine strain and wild

circulating strain by the maximum T cell epitope similarity score for the vaccine strain in comparison with itself (Equation 3.3):

$$P(v,w) = \frac{\sum_{i \in v, j \in w} \sum_{a \in A} (p(i)_a * p(j)_a)}{\sum_{i \in v, j \in v} \sum_{a \in A} (p(i)_a * p(j)_a)} \tag{3.3}$$

3.3.8 Dimension reduction

The equation to calculate T cell epitope immune distance was applied iteratively to the subsampled dataset and therefore the pairwise T cell epitope immune distances are structured into an n x n square-distance matrix. Given that each protein is described by a relative distance to the rest of n-1 proteins, the data must be dimensionally reduced to be graphed. Classic (metric) multidimensional scaling (MDS) can be used to preserve the distances between a set of observations in a way that allows the distances to be represented in a two-dimensional space. MDS was performed as previously described by Gower [133]. The MDS method first constructs an n-dimensional Euclidean space using the distance matrix in which all distances are conserved, and then principal component analysis is performed. MDS [134] was carried out using the cmdscale package in R [133]. *K-means* clustering was performed using the kmeans function in base R. Due to the lack of previous characterizations of RSV T cell immunity clusters, the number of T cell immunity groups was determined using the optimized within-cluster sum of square (wss) with Elbows method [135]. To evaluate whether applying *k-means* clustering to classify RSV strains on two-dimensional space can reflect their T epitope profile, we calculated the stress of MDS using the smacof package in R [136]. We also compare the performance of *k-means* clustering on MDS spaces with different numbers of dimensions (Figure 3.1).

3.3.9 Calculation of genetic hamming distance

Genetic hamming distance, which is defined as the number of bases by which two nucleotide sequences differ, was calculated by comparing the number of different bases between each

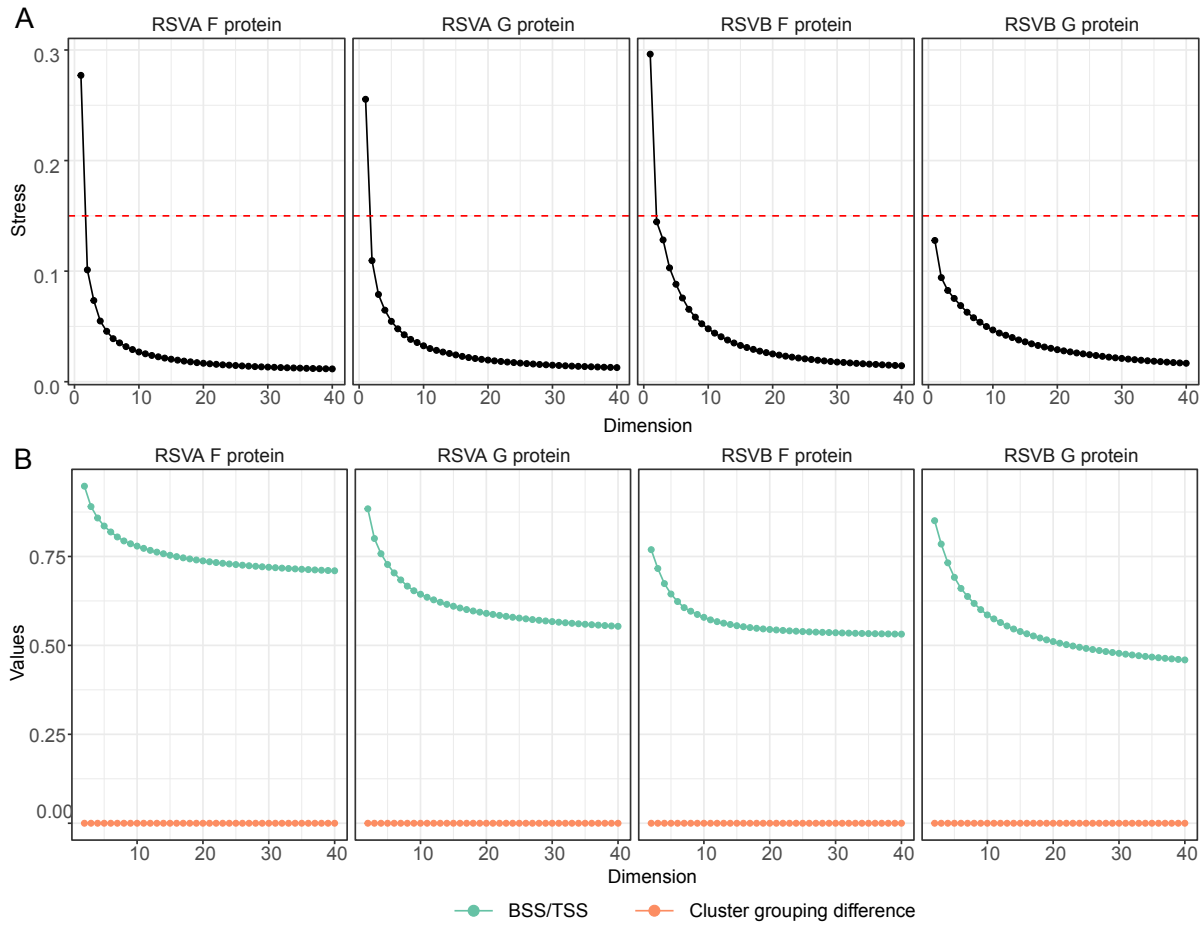

Figure 3.1: **Sensitivity analysis for MDS.** (A) Stress evaluation under the different number of dimensions for RSV distance matrix. Stress less than 0.15 (red dash line) indicates an acceptable precise MDS solution. (B) Performance of *k-means* clustering under the different number of dimensions, the number of clusters is determined at 2-dimensional space. There is no cluster grouping difference at higher dimensional space (orange). The sum square between clusters /sum square of total differences (BSS/ TSS) measures indicates the total variance in the data is explained well under higher dimensional space (green).

sequence in the subsampled datasets. The reconstructed most recent common ancestor (TMRCA) sequences for each dataset (subsampled F and G protein sequences of subtype A and subtype B, respectively) were estimated using the program "Treetime" and were used as root in our analysis [123].

3.4 Results

3.4.1 Distribution of T cell epitopes in RSV surface proteins

We evaluated the T cell immunogenic potential across RSV surface proteins by scanning 9 residue regions to predict the binding probability to MHC class I and class II molecules (Figure 3.2). The epitope density of RSV surface proteins was evaluated using a normalized epitope density score,, which is computed by summing up the predicted peptide-MHC binding score across the protein and normalizing it with the protein length. The score for randomly generated proteins is set to zero and vaccine antigens generally score above 20 on this scale [118] F protein has an epitope density score greater than +20 for both the class I and class II immunogenicity scale analysis, indicating significant immunogenic potential [118]. This contrasts with lower G protein class I and class II epitope density protein scores for both subtypes. The class I epitope density score of G protein was greater than +10 in both subtypes but the class II density was lower than random expectation in the analysis of RSV-B (Figure 3.2A). This result suggests that RSV surface proteins are likely to have the potential to stimulate T cells that are required for protective immunity. We then investigated the distribution of T cell immunogenicity across the proteins and found that there are regions with relatively high T cell immunogenic potential (Figure 3.2B). The distribution of T cell immunogenicity of F protein was mapped onto its protein structure and overlap between protein sequence regions with high T cell immunity potential and the antibody neutralizing targets was observed at antigenic site ∅ and site II.

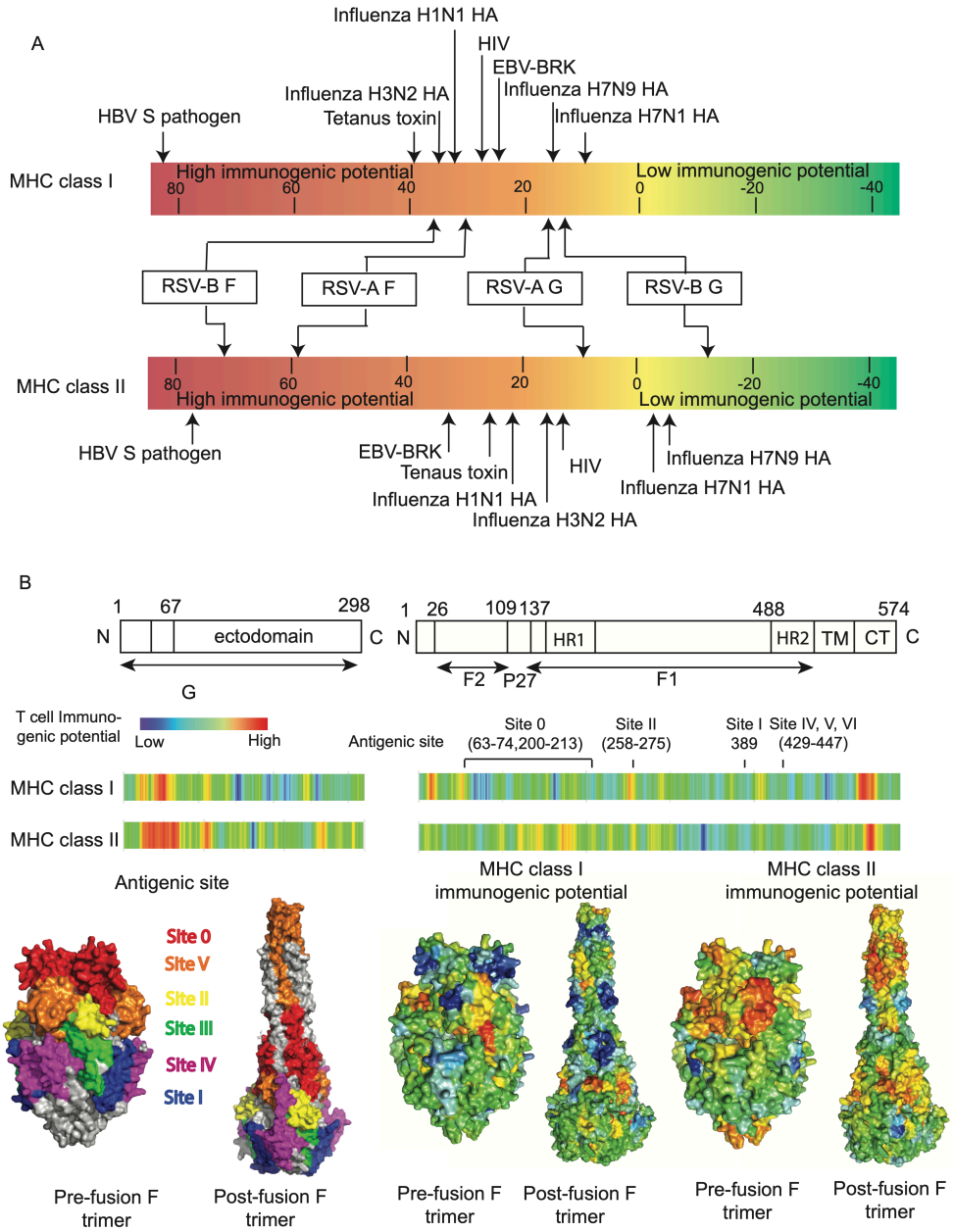

Figure 3.2: **T cell immunogenic potential for RSV surface proteins based on MHC binding prediction.** (A) T cell immunogenic potential of RSV major surface proteins. T cell epitope density scores for RSV major surface proteins and other pathogen proteins are labeled on a scale bar. Low-scoring proteins are known to engender little to no immunogenicity while higher-scoring proteins are known immunogens. Proteins scoring above +20 on this scale are considered to have significant immunogenic potential. (B) Distribution of RSV T cell immunogenic potential across F and G protein in RSV reference strain A2 and RSV F protein main antigenic sites that are determined in previous studies [137]. Prefusion or post-fusion F protein surface was colored by the antigenic sites and relative immunogenetic potential at each location. Analyses are based on the RSV-A reference sequence

3.4.2 Lineage specific T cell epitope profiles

We then extended T cell epitope predictions from RSV representative strains to multiple wild-circulating strains. The distribution and diversity of T cell epitopes across different strains are illustrated in heatmaps with the corresponding time-scaled phylogenies (Figure 3.3 and Figure 3.4). Both F and G proteins contain epitopes that were conserved across all RSV strains in almost 100% of sampled isolates, suggesting that they could serve as high-quality T cell epitope candidates for vaccine design. In contrast, some epitopes were mutated in selected strains, and those epitopes that only occurred in certain clades within the phylogeny could be interpreted as clade-specific "fingerprints".

The G gene duplication events in RSV, which are unique gene signatures, can either shift the position of epitopes (locations are different but the amino acids of epitopes are identical to the G protein isolates without duplication) or cause the emergence of novel epitopes Two novel class I epitopes, (no. 31 and no. 40 in Supplementary Figure 3.4A), were found in RSV-A strains that contain G gene duplication. In addition, an emergent class II epitope (no. 25 in Supplementary Figure 3.4A) was identified in RSV-A sequences that contain G gene duplication, which was a shift from an epitope (no. 24) that has been observed in other strains. From RSV-B strains that contain the G gene duplication event, we also observed multiple lineage-specific class I T cell epitopes, which are caused by a 2-aa deletion (aa157 and aa158) in these strains instead of directly due to the 60-nt duplication event. RSV-B G proteins that have the duplication event contain multiple novel epitopes (no. 22, 23, 26, 28, 30, 37) but do not contain several epitopes (no. 24, 25, 27, 29, 31, 38) that are identified in other strains (Figure 3.4B).

To further determine whether the computationally predicted T cell epitopes with high MHC binding potential are immunogenic, we utilized the JanusMatrix [126] algorithm to identify the T cell epitopes that are likely to be cross-conserved with human peptides and thereby tolerated by the immune system. Based on this analysis, 6.45% of putative class I epitopes and 1.12% of putative class II epitopes of RSV major surface proteins are cross-

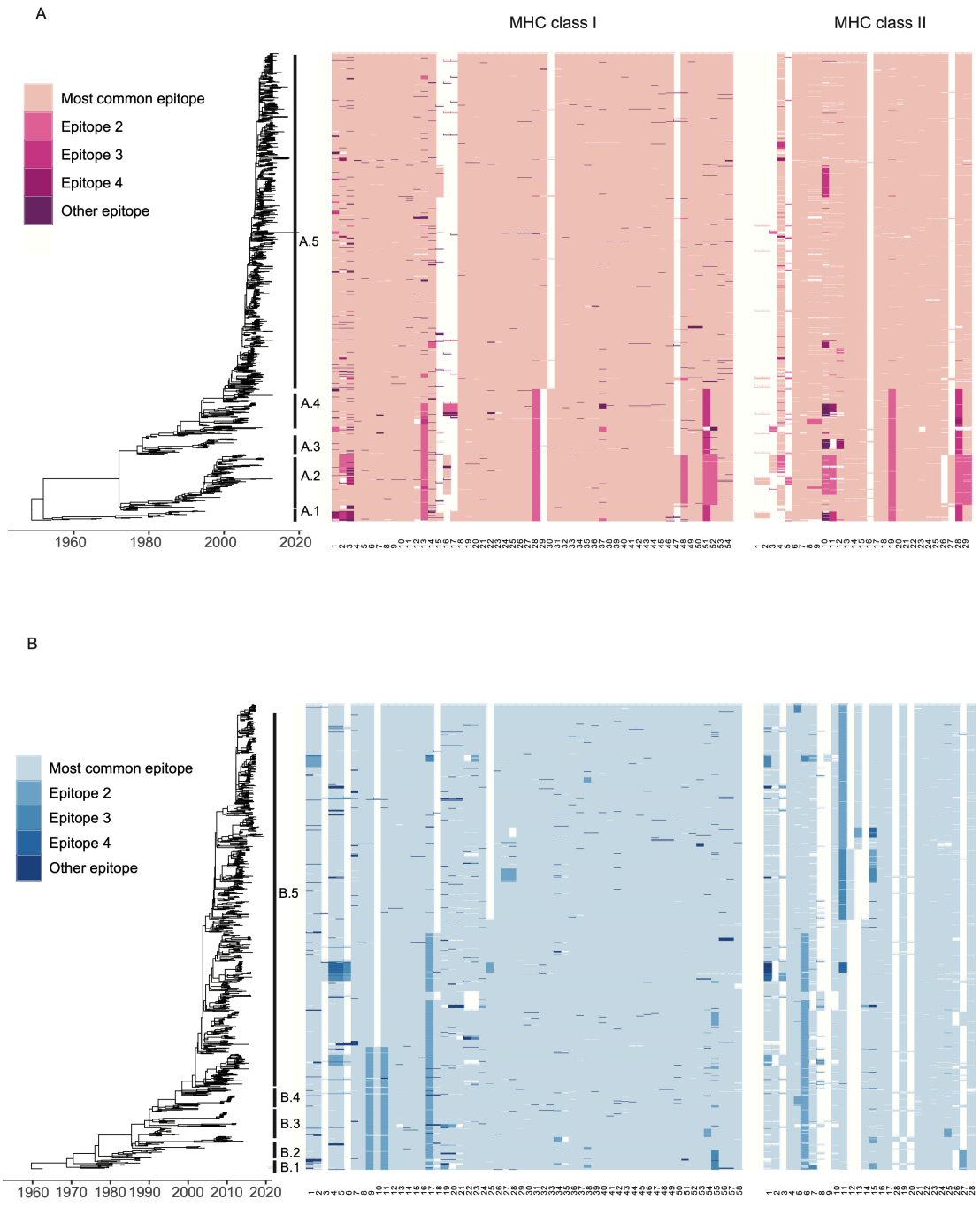

Figure 3.3: **Distribution and diversity of T cell epitopes in RSV F protein.** The tree panel on the left is a time-scaled phylogeny build with RSV-A (A) or RSV-B (B) F gene nucleotide sequences using the ML approach. Determined genotypes are labeled on the right with black bars. Each color column on the right side represents the presence of an MHC class I or class II epitope. Only the epitopes that are present in more than 1% of sampled isolates are displayed. The column color indicates different numbers of epitope sequences at the same location.

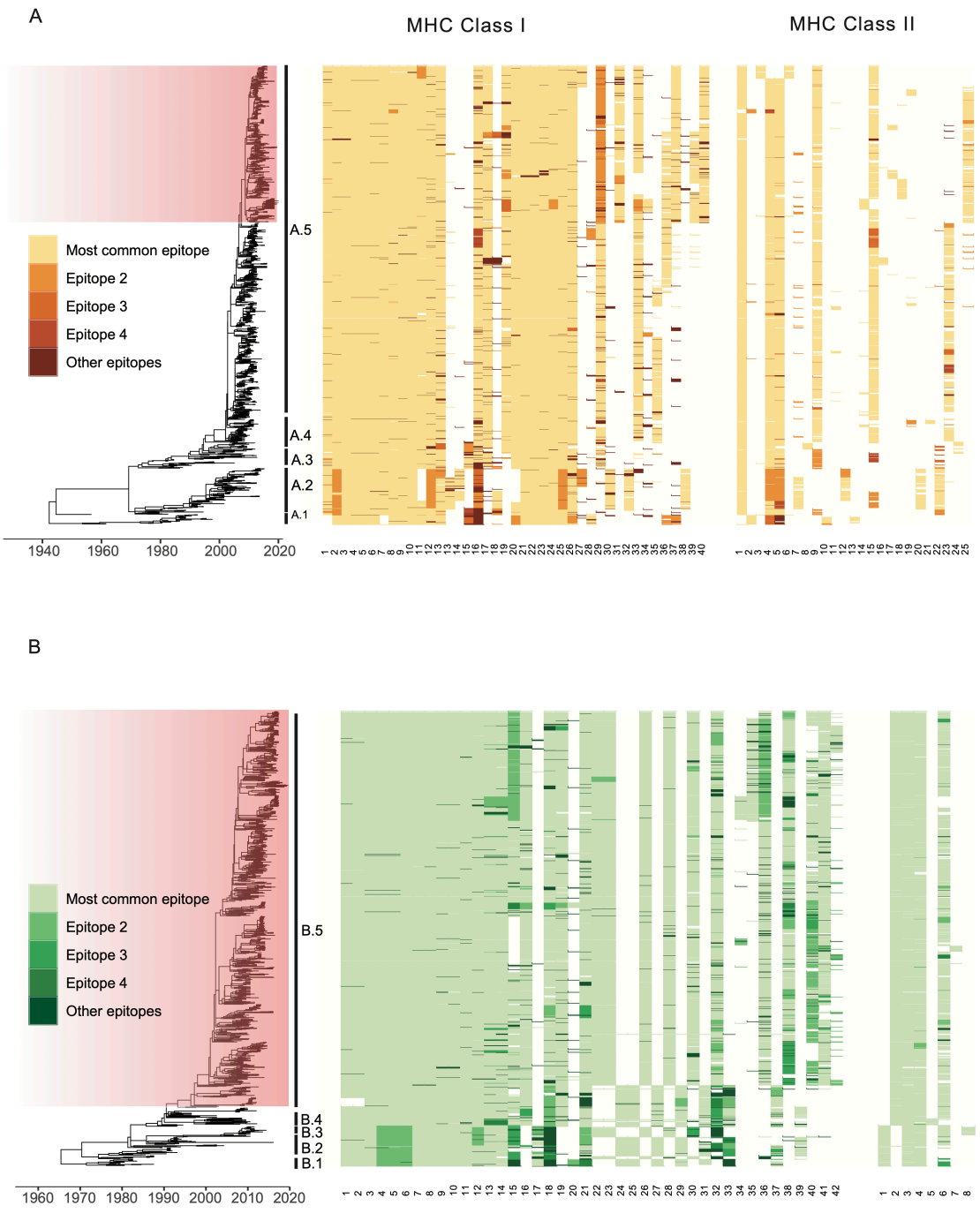

Figure 3.4: **Distribution and diversity of T cell epitopes in RSV G protein.** The tree panel on the left is a time-scaled phylogeny build with RSV-A (A) or RSV-B (B) G gene nucleotide sequences using the ML approach. The clades that contain novel 72-nt or 60-nt duplication at the second hypervariable region of G gene were highlighted in red. Determined genotypes are labeled on the right with black bars. Each color column on the right side represents the presence of an MHC class I or class II epitope. Only the epitopes that are present in more than 1% of sampled isolates were displayed. The column color indicates different numbers of epitope sequences at the same location.

Table 3.1: **Number of computationally predicted conserved RSV T cell epitopes and experimentally identified RSV T cell epitopes.** a. The conserved epitopes are identified with at least 60% presence across all RSV-A or RSV-B sequences that are publicly available. b. The number of experimentally identified epitopes include RSV peptides that are positive in MHC class I/ class II ligand assays from the IEDB database after removing duplicates.

MHC allele	RSV protein	No. of computationally predicted conserved epitopes	No. of experimentally identified epitopes	Computationally predicted conserved epitopes with experimental validation	Experimentally identified epitopes with computational identification
class I	F	77	20	12	19
	G	44	4	3	4
class II	F	31	46	18	30
	G	7	4	2	3

conserved with human proteome-derived epitopes at TCR-facing residues. As these peptides have similar HLA binding preferences that are contained in human proteins (Figure 3.5), they were therefore assumed not to be immunogenic. After excluding the high-JanusMatrix score epitopes identified above, we were able to identify T cell epitopes that were conserved in more than 60% of currently circulating RSV strains. We searched the IEDB epitope database to determine if these epitopes were related to experimentally validated RSV T cell epitopes or HLA ligands (Table 3.1). The conserved RSV T cell epitope sequences that may be important for future vaccine development are shown in Table 3.2 and Table 3.3.

3.4.3 Predicted RSV T cell epitope landscapes

To investigate the evolution of RSV on T cell immunity profiles, we use a multidimensional scaling (MDS) approach to visualize the T cell immunity profile of multiple RSV strains on a landscape. We performed a T cell epitope content pairwise comparison between RSV strains using in silico predicted peptide-HLA allele binding affinity. The pairwise T cell epitope distances were then calculated using the algorithm reported in this study (Equation 3.1). We

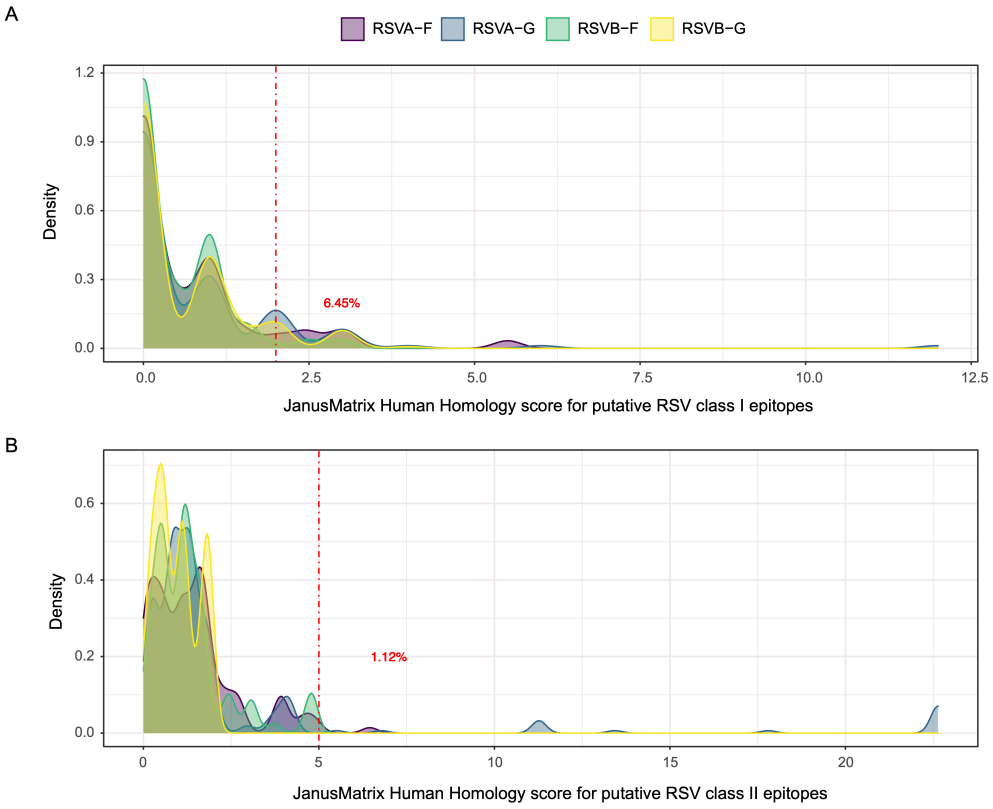

Figure 3.5: **Distribution of JanusMatrix Human Homology score for putative RSV MHC class I and class II epitopes.** The cross-reactive potential of identified putative T cell epitopes and human host was represented with a JanusMatrix Human Homology score. 6.45% identified putative class I epitopes and 1.12% class II epitopes are cross-conserved on the TCR face with human epitopes.

Table 3.2: **Experimentally validated conserved MHC class I epitopes peptides in RSV major surface proteins.** a. This table contains putative MHC class I epitopes that have already been experimentally validated in publications. Only putative class I epitopes that have positive results in MHC class I ligand assays with the same computationally predicted binding HLAs are shown in the table. b. Epitopes sequences that are conserved in both RSV-A and RSV-B are in bold. c. HLAs that have the top 1% binder scores in EpiMatrix for epitope sequence. d. The conservation is evaluated by the presence of epitope peptides across all RSV-A or RSV-B sequences that are publicly available (only epitope sequences with at least 60% conservation are shown in the table). e. Count of human peptides found in the search database. JanusMatrix was used to search human peptides that are predicted to bind to the same allele as the RSV epitope and share TCR-facing contacts with the RSV epitope.

Subgroup	**Protein**	**Epitope address**	**Epitope sequence**	**Binding HLAs**	**Conservation**	**Number of human matches**	**Epitope id in IEDB**
RSV-A & RSV-B	F	45-53	**LSALRTGWY**	A0101	99.55%(A)& 74.24%(B)	1	158982
		140-148	**FLLGVGSAI**	A0201	99.59%(A)& 97.98%(B)	0	156869
		250-258	**YMLTNSELL**	A0201, A2402	99.59%(A) & 99.33%(B)	0	156979
		272-280	**KLMSSNVQI**	A0201	66.64%(A) & 96.08%(B)	3	156902
		273-281	**LMSSNVQIV**	A0201	66.56%(A) & 96.08%(B)	1	156915
		449-457	**TVSVGNTLY**	A0101	99.75%(A) & 99.33%(B)	0	97017
RSV-A	F	10-18	AITTILAAV	A0201	84.69%	3	156844
		111-119	LPRFMNYTL	B0702	91.18%	0	158975
		170-178	ALLSTNKAV	A0201	99.67%	2	156847
		383-391	NIDIFNPKY	A0101	95.86%	0	159045
	G	25-33	FISSCLYKL	A0201	99.26%	0	158759
		61-69	FIASANHKV	A0201	82.08%	0	158751
RSV-B	F	525-533	IMITAIIIV	A0201	89.25%	0	156892
		540-548	SLIAIGLLL	A0201	97.65%	5	156960
	G	25-33	VISSCLYKL	A0201	90.91%	0	158759
		61-69	FIISANHKV	A0201	99.02%	0	158751

Table 3.3: **Experimentally validated conserved MHC class II epitopes peptides in RSV major surface proteins.** a. This table contains putative MHC class II epitopes that share the identical binding groove sequence, which represent the nine-mer frames with the greatest potential to bind class II HLA (epitope sequences with underlines), with the RSV class II epitopes that have already been experimentally validated in publications. Only the putative class II epitopes that have positive results in MHC class II ligand assays with the same computationally predicted binding HLAs are shown in the table. b. Underlined sequences represent the nine-mer frames with the greatest potential to bind class II HLA. Epitope sequences that are in bold indicate sequences are predicted to bind class II HLA and are conserved in both RSV-A and RSV-B. c. Conservation is evaluated by the presence of epitope peptides across all RSV-A or RSV-B sequences that are publicly available (Only epitope sequences with at least 60% conservation are shown in the table). d. Count of human peptides found in the search database. JanusMatrix was used to search human peptides that are predicted to bind to the same allele as the RSV epitope and share TCR-facing contacts with the RSV epitope.

Subtype	Protein	Epitope address	Epitope sequence	Conservation	Number of human matches	Epitope id in IEDB
RSV-A	F	29 - 44	TEEF**YQSTCSAVS**KGY	98.53%	3	956680
		50 - 70	TGW**YTSVITIELSNIK**ENKCN	97.75%	1	153700
		167 - 192	IKSALLSTNKAVVSLSNGVSVLTSKV	93.14%	4	545502
		218 - 234	ETVIEFQQKNNRLLEIT	98.86%	3	1087566
		247 - 268	VSTYMLTNSELLSLINDMPITN	98.98%	8	99471
		288 - 310	IMSIIKEEVLAYVVQLPLYGVID	98.57%	5	99334
		399 - 418	KTDVSSSV**ITSLGAIVS**CYG	99.14%	0	545603
		453 - 470	GNTLYYVNKQEGKSLYVK	98.37%	1	99691
		492 - 510	ISQVNEKI**NQSLAFIR**KSD	80.32%	1	153713
		543 - 560	AVG**LLLYCKARSTPV**TLS	79.26%	6	153641
	G	19 - 43	TLNHLLFISSCLYKLNLKSIAQITL	93.13%	8	1087567
RSV-B	F	29 - 44	TEE**FYQSTCSAVS**RGY	99.78%	3	956680
		50 - 70	TGW**YTSVITIELSNIK**ETKCN	93.95%	1	153700
		192 - 218	VLDLKNYINNQLLPIVNQQSCRISNIE	83.43%	4	153636
		247 - 268	LSTYMLTNSELLSLINDMPITN	98.54%	8	99471
		399 - 418	KTDISSSV**ITSLGAIVS**CYG	98.88%	0	545603
		453 - 470	GNTLYYVNKLEGKNLYVK	98.77%	0	99691
		492 - 510	ISQVNEKI**NQSLAFIR**RSD	97.42%	1	153713
		543 - 560	AIGL**LLYCKAKNTPV**TLS	94.96%	4	153641
	G	51 - 74	STSLIIAAIIFIISANHKVTLTTV	94.66%	8	158751

then applied a multidimensional scaling (MDS) approach using these estimated pair-wise T epitope distances to map RSV strains to a landscape to characterize their T-cell immunity profile. We found both Class I and Class II T cell immunity profiles of F and G proteins of different RSV strains were clustered into groups on this T cell epitope landscapes (Figure 3.6). Combining the Class I and Class II T-cell epitope biding profiles, RSV-A major surface protein isolates can be divided into three clusters and RSV-B major surface protein isolates can be divided into two clusters (Figure 3.8, Figure 3.1). We observe that the G gene sequence isolates that contain 72-nt (RSV-A) or 60-nt (RSV-B) duplications clustered together with other sequences instead of forming isolated groups. To further investigate the T cell epitope diversity, we correlated this clustering pattern with the phylogenetic histories (Figure 3.7). The phylogenetic tree topologies of the RSV-A F gene and G gene are similar. The F gene cluster 1 is paraphyletic, while cluster 2 and 3 are monophyletic. Cluster 1 is the closest to the ancestral sequence and mapping this group onto the phylogeny show that this cluster has a basal relationship with clusters 2 and 3 indicating that the phylogenetic divergence occurred prior to epitope drift. The RSV-B F and G gene genealogies are very different. In particular, the RSV-B F gene topologies is indicative of strong immune selection, similar to observed human influenza A virus or within host HIV phylogenies [138]. In contrast, the RSV-B G gene phylogeny shows the co-circulation of multiple lineages, though this could reflect the sequencing bias of G genes (Figure 3.7B). We then calculated the T-cell epitope immune distance of each strain from a reconstructed ancestral sequence (Figure 3.7C). These distances were then plotted against the year of isolation and colored according to the cluster identified in Figure 3.7A. RSV-A shows that multiple predicted immune phenotypes co-circulate and persist for long periods (>2 decades). Analysis of RSV-B shows a turnover of the predicted immune phenotypes with short periods of co-circulation (<5 years) for F and G protein T cell epitopes. The limited periods of co-circulation is again consistent with phenotype patterns observed for viruses under strong immune selection (e.g H3N2 influenza A virus) [139, 140]. In contrast, genetic distances from the reconstructed ancestral sequence plotted against year

of isolation show patterns typical of gradual genetic drift, except in the G gene where a 72-nt and 60-nt insertion is present (Figure 3.7D). Taken together, these results suggest that genetic and predicted T-cell epitope immune diversity are different and may be an important factor to consider when evaluating RSV vaccine efficacy.

There are multiple methods available to predict T cell epitopes [141], which may result in different reconstructed landscapes if there is a systematic bias in the prediction method. We used the NetMHCpan method [142] to predict T cell epitopes and perform the same landscape reconstruction using MHC class I binding predictions for RSV-A F protein. Our analysis showed a consistent clustered pattern of RSV T epitope profile on the landscape regardless of T cell epitope prediction method (Figure 3.9).

3.4.4 Assessment of vaccine candidate strains with T cell epitope content

T cell epitopes that are similar between vaccine strains and wild strains (cross-conserved T epitopes, which are defined as epitopes that share identical T cell receptor-facing residues and are restricted by the same alleles [131]) may be responsible for the T immune protection of the vaccine. To quantitatively evaluate whether it might be necessary to include multiple RSV strains to prepare an effective vaccine, two live attenuated RSV strains that are previously considered as vaccine candidates, CP248, a recombinant virus that belongs to subtype A, and CP52, which is a recombinant RSV-B strain, were included in our analysis and we evaluated their T cell epitope conservation with different RSV wild-type strains. We calculated the average proportion of cross-conserved T cell epitope content between the selected vaccine strains and wild-circulating strains from different isolation years and WHO regional groups (Figure 3.10, Figure 3.11). Different proportions of cross-conserved T cell epitope content against isolates from two different subtypes, A and B, were observed in both the F and G protein analyses. In the comparison of the vaccine strains and wild strains belonging to the same subtype, the proportion of cross-conserved T cell epitope in RSV F protein is relatively stable in different groups, all are higher than 78% for RSV-A and higher than 85% for RSV-B.

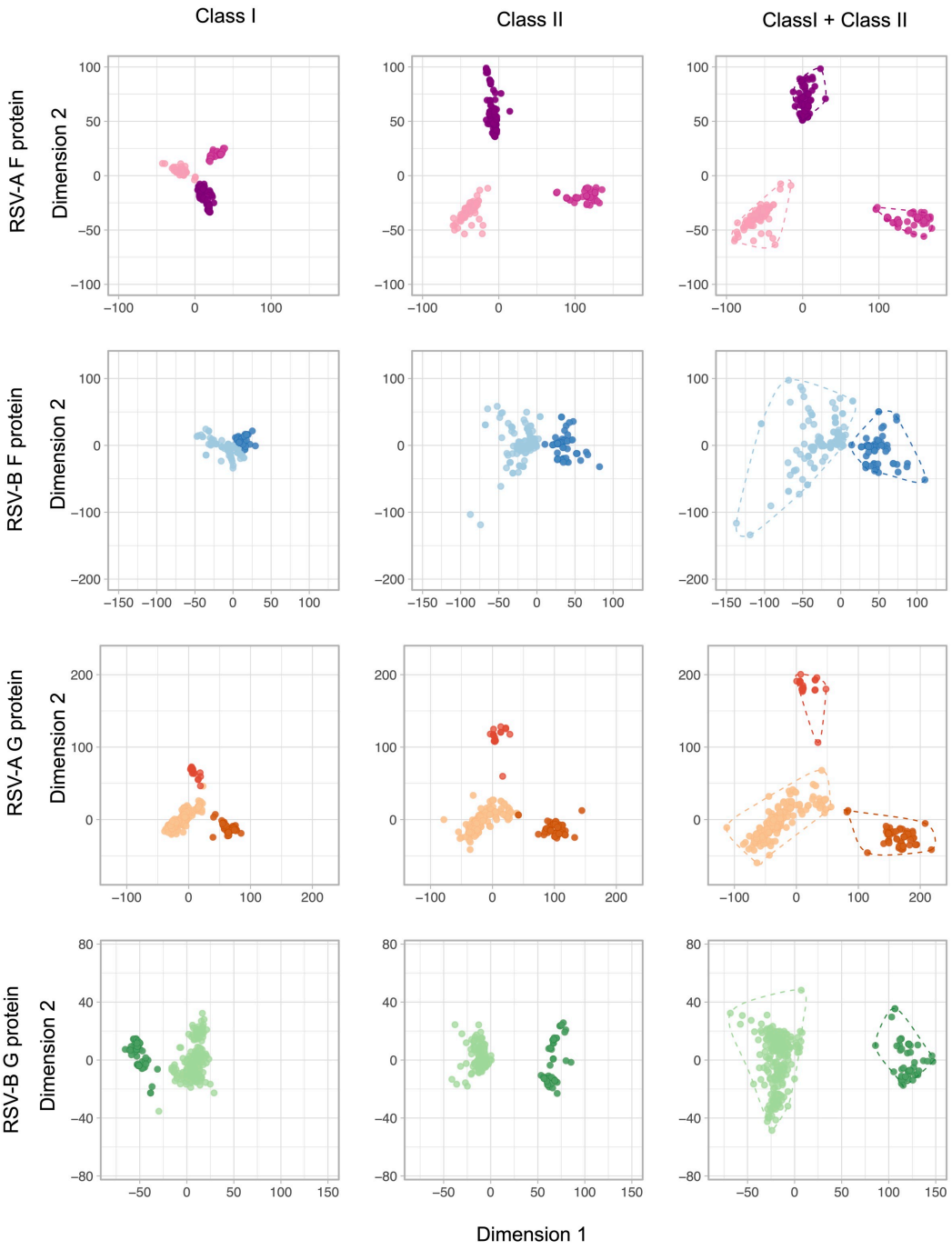

Figure 3.6: **Predicted T cell epitope landscapes of RSV surface proteins.** RSV T cell epitope landscapes were built with sequenced-based MHC class I epitope binding prediction (left), MHC class II epitope binding prediction (middle) or combining class I and class II epitope biding prediction (right). Sequences are colored by the epitope cluster determined by epitope landscapes built with combining Class I and Class II epitope prediction.

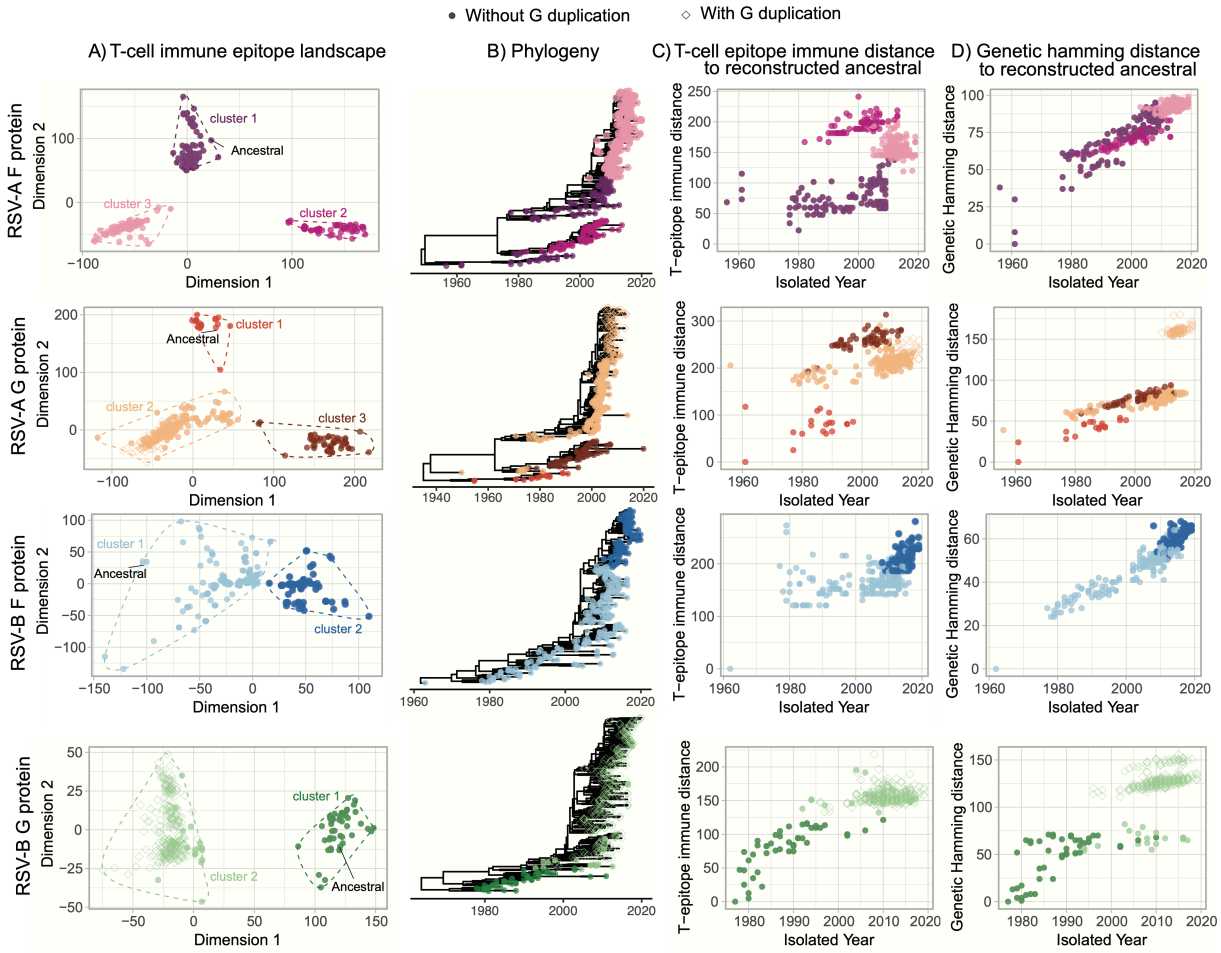

Figure 3.7: **Predicted T cell epitope landscapes and genetic evolution of RSV surface proteins.** Filled circles indicate RSV F protein isolates or G protein isolates without duplication. Diamonds indicate G protein isolates with gene duplication. (A) Epitope landscapes of RSV major surface proteins are built with MHC class I and class II epitope content comparison across different strains. T cell immunity clusters are determined with *k-means* method and are used to color the sequenced isolates in the following panels. (B) The corresponding time-scaled phylogenies are reconstructed with the Maximum Likelihood (ML) approach. (C) T cell epitope immune distance and (D) genetic hamming distance from the estimated TMRCA are plotted against the isolated time of each sequence.

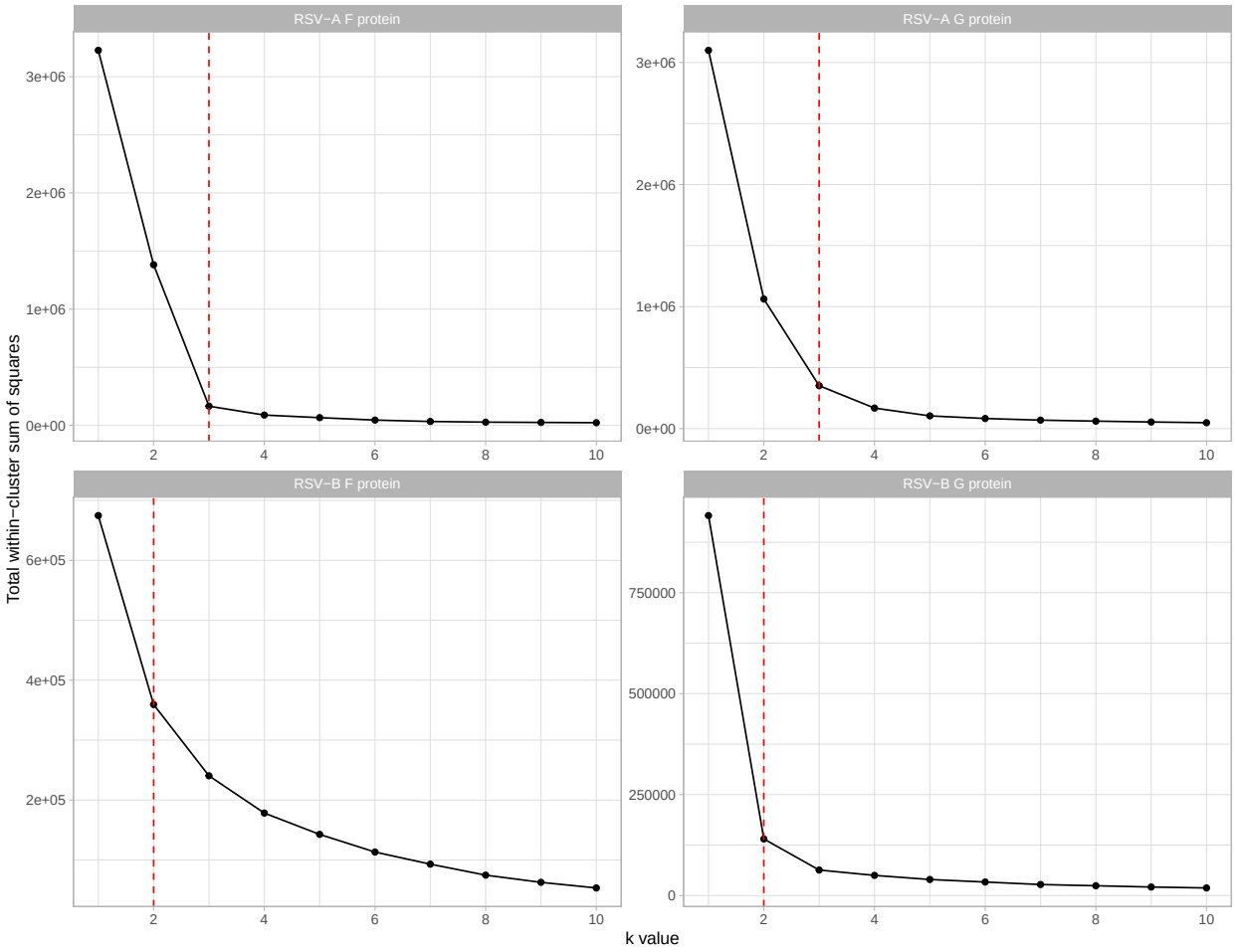

Figure 3.8: **Total within sum of squares (wss) using *k-means* algorithm.** Totals within sum of squares in epitope topographies were calculated after clustering into k (from 1 to 10) groups with *k-means*. The optimal number of clusters is determined to be 3 in the analysis of RSV-A F and G proteins and is determined to be 2 in the analysis of RSV-B F and G proteins using the Elbow method.

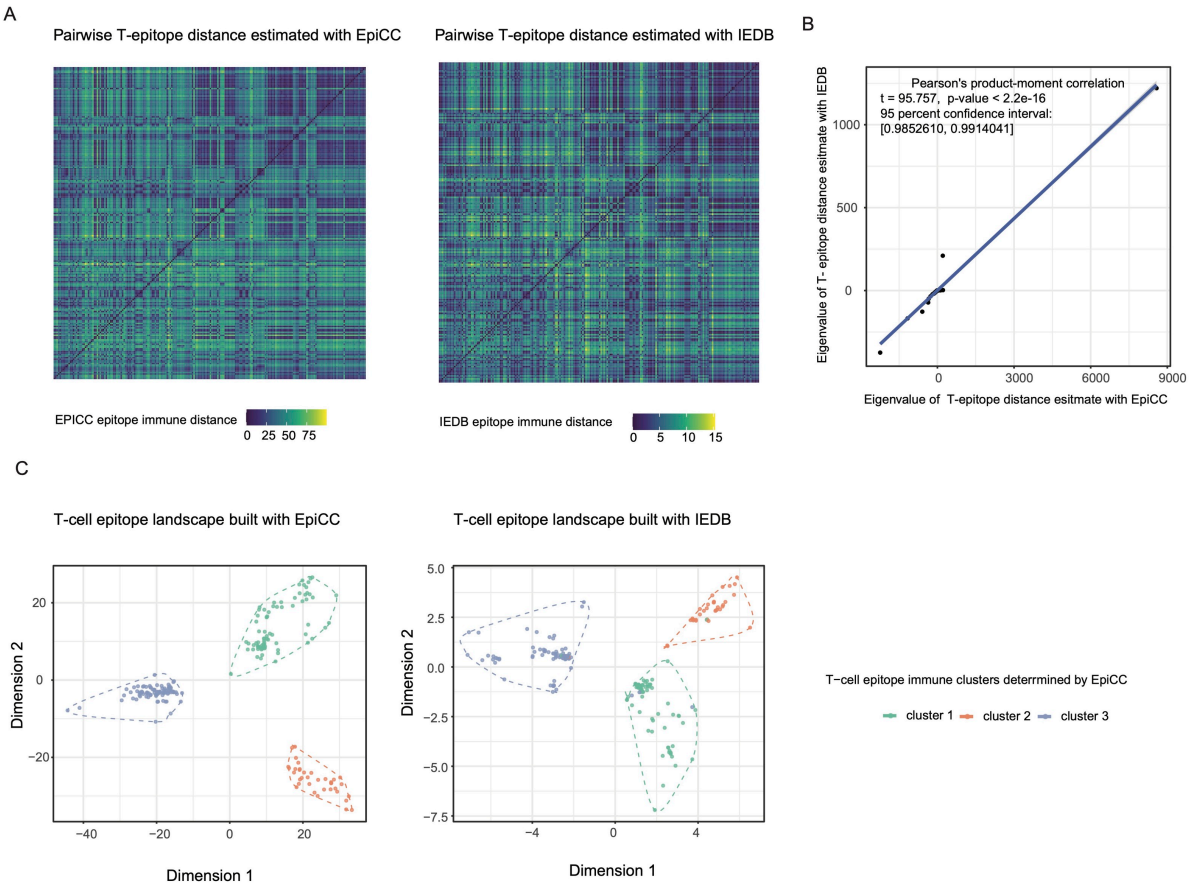

Figure 3.9: **Validation of T cell epitope distance estimation using the IEDB analysis resource.** Validation is performed with MHC class I epitope binding prediction of RSV-A F protein. (A) Heatmaps for pairwise MHC class I epitope distance estimated in iVAX toolkits or calculated with custom python scripts using MHC class I molecule binding prediction that is implemented in IEDB. (B) Eigenvalues for each sequence are calculated from pairwise distance matrices using "RSpectra" package in R. The Pearson correlation test significantly supports a non-zero correlation between T cell epitope distance estimated with EpiCC and T cell epitope distance estimated with IEDB. (C) T cell epitope topographies are built with pairwise epitope distances estimated from EpiCC or IEDB. Both methods resulted in a similar cluster pattern for the CD8 T cell epitope profile of RSV-A F protein.

In contrast, changes in the proportion of cross-conserved T cell epitopes were detected among groups within the same RSV subtype, especially in different temporal groups in the G protein analysis (Figure 3.10). Vaccine strain CP248 appears to have a relatively higher proportion of cross-conserved T cell epitopes within G protein when compared to the RSV-A strains that were isolated before 1991 (>70%) and a relatively lower degree of conservation against recently isolated strains. A similar decrease in T cell epitope conservation with time was identified for vaccine strain CP52 among circulating RSV-B strains.

3.5 Discussion

Although both CD4 and CD8 T cells contribute to protection against RSV-induced disease following primary infection [143], T cell epitopes have received limited attention in the RSV research effort. We demonstrate RSV surface proteins appear to have significant potential to drive T cell immunity using a computational approach, based on their T cell epitope density scores as determined by MHC molecular binding prediction. The relatively high putative T-cell epitope density might make F protein a good target for RSV vaccine. In addition to the analysis of T cell epitope density and distribution in RSV major surface proteins, we also demonstrated lineage-specific variations in T cell epitope content. Even though RSV F protein is believed to be well conserved and G protein is reported to be highly variable, epitope mutations are observed across different lineages within the F protein, and potential conserved T cell epitopes can still be found in the highly variable G protein, suggesting that studying the lineage-specific T cell epitopes in RSV can provide insight into the impact of immune selection on viral diversity and persistence. In contrast to the conserved F protein, RSV G protein is reported to be highly variable, we still observed potential conserved T cell epitopes across different strains, which suggests a great interest of G protein conserved domain as potential vaccine targets on T cell immune protection. While experimental validation is needed, this analysis highlights the importance of understanding population-level epitope

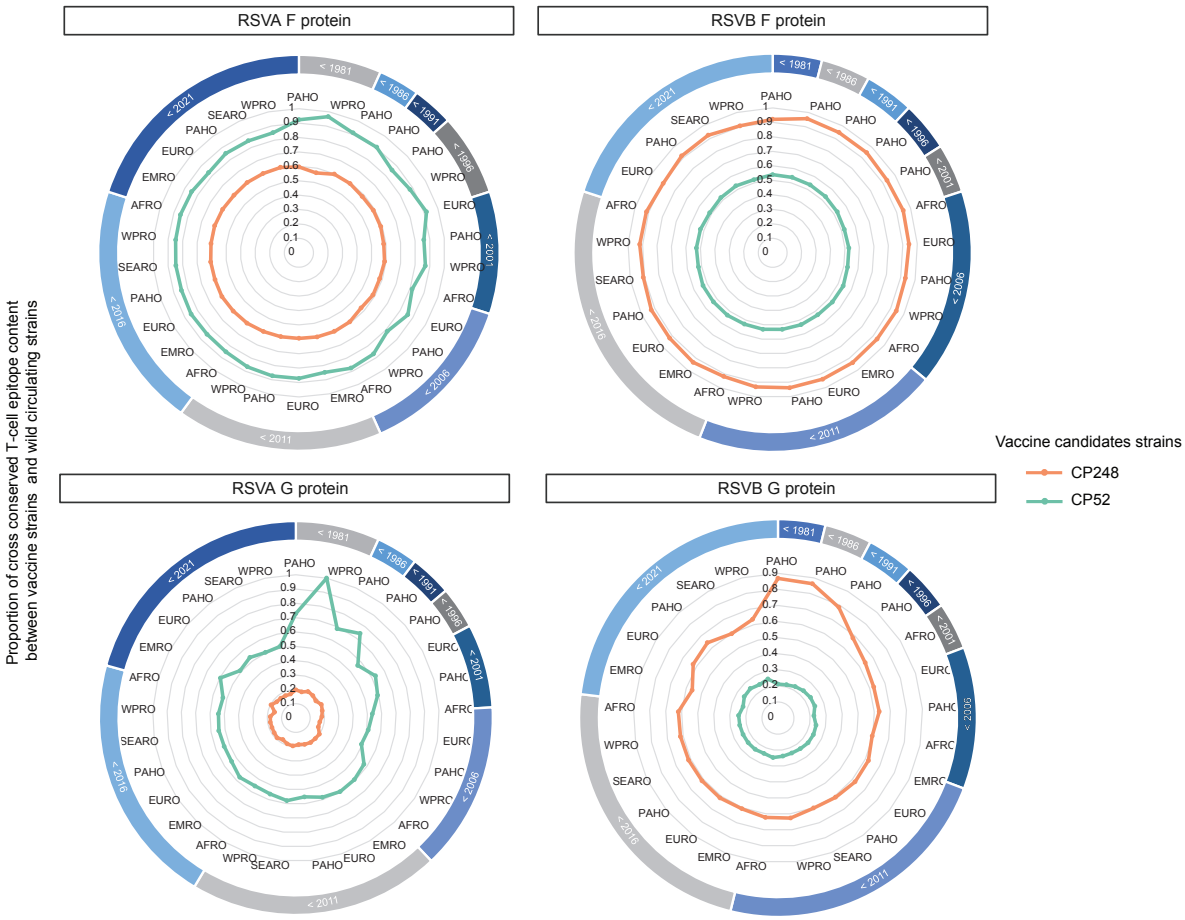

Figure 3.10: **Evaluation of previously used RSV vaccine candidate strains with T cell epitope content of circulating strains.** RSV-A and RSV-B major surface protein sequences are subsampled and then grouped by isolation year and 6 isolated WHO regions. African Region (AFRO), Region of the Americas (PAHO), South-East Asia Region (SEARO), European Region (EURO), Eastern Mediterranean Region (EMRO) and Western Pacific Region (WPRO). The proportion of cross-conserved T cell epitope content between live attenuated strains (CP248 or CP52) and wild circulating strains are displayed as radar plots.

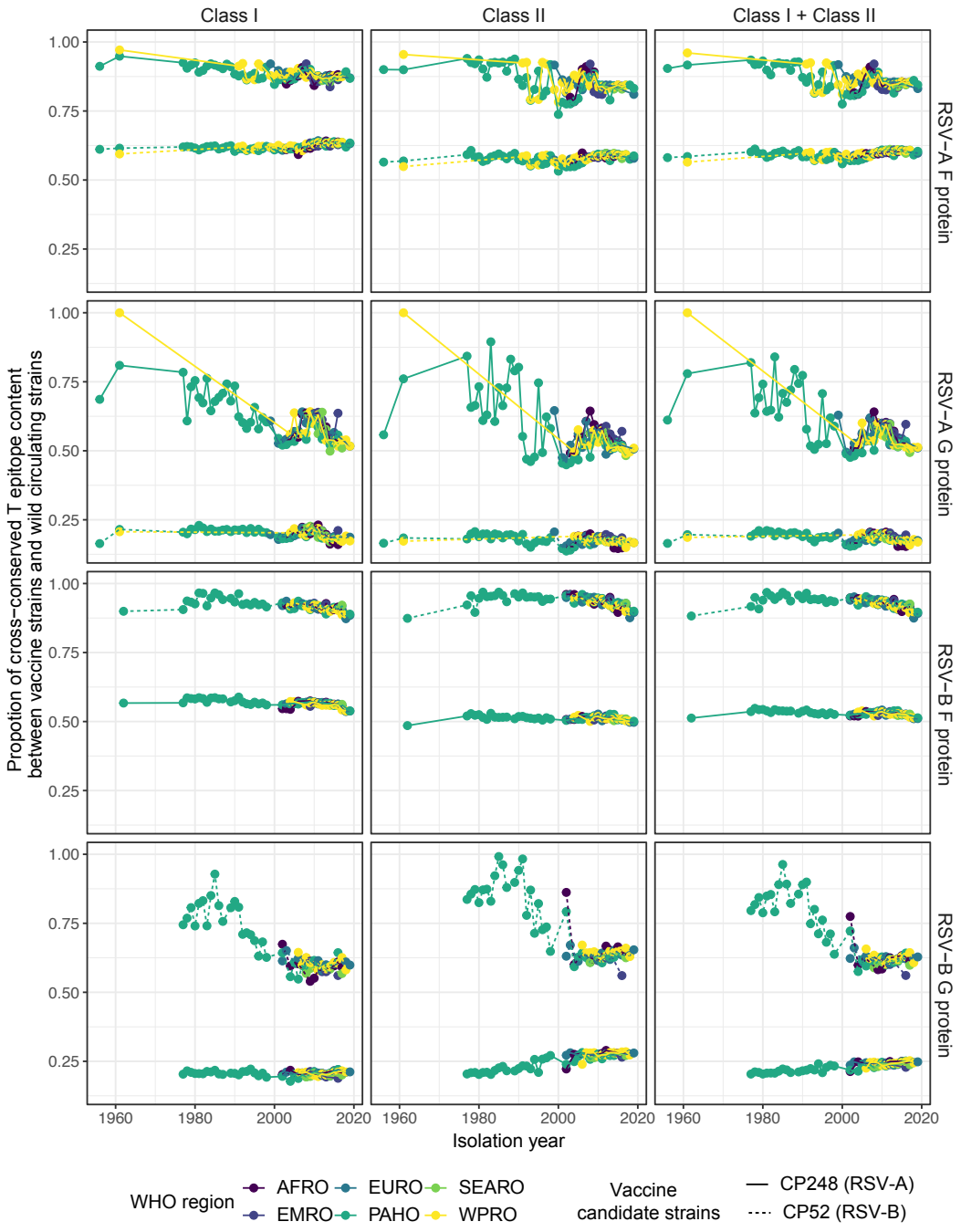

Figure 3.11: **Evaluation of RSV vaccine candidate strains with T cell epitope content in different WHO regions.** RSV-A and RSV-B major surface protein sequences were grouped by isolation year and 6 isolated WHO regions, African Region (AFRO), Region of the Americas (PAHO), South-East Asia Region (SEARO), European Region (EURO), Eastern Mediterranean Region (EMRO) and Western Pacific Region (WPRO). The proportion of cross-conserved T cell epitope content between vaccine strains (CP248 or CP52) and wild circulating strains in different isolation years and different WHO regions were represented.

conservation as it may provide important insight into the development of T cell epitope-driven vaccines against RSV infection.

A major focus of our work is the development of a sequence-based method to map the evolution of T cell immunity across different stains. Following a previously pivotal work that used MDS method to map the evolutionary adaptation of influenza A virus-induced by CD8 T cell using the presence and absence of MHC class I epitopes [144], we constructed RSV T cell immunity landscapes using immune distances that were generated by T cell epitope cross-conservation analyses, which allows for easy visualization and intuitive understanding of the potential for T cell immunity relationships among different strains. When comparing across strains, we found that the T cell epitope content of RSV surface proteins from different strains can be clustered, as has been observed for the antigenic relationship reported in other pathogens [123, 91]. Our results also demonstrate the correspondence between RSV T cell immunity clusters and their corresponding phylogeny, with sequences in the same clade generally belonging to the same T cell immune cluster. Importantly, we also observe different patterns of T-cell epitope evolution of RSV wild strains compared with their genetic evolution. We find the RSV strains with G gene duplications can still cluster with previous RSV isolate on the T cell epitope MDS space. Our results highlight the importance of characterizing T cell epitope changes in RSV.

We identified highly conserved RSV T cell epitopes in this study, some of which have already been experimentally validated and published in the IEDB database. However, we also identified several other conserved T cell epitopes that have not been previously described. These may be valuable for vaccine design, although experimental validation will be needed. Furthermore, the homology of selected RSV epitopes to human peptides suggests that some predicted RSV T cell epitopes might be tolerated by the human immune system, or could induce a harmful cross-reactive immune response against human proteins when administered with an adjuvant [126]. Certain aspects of immunity to RSV were not addressed by this study. For example, neutralizing antibody responses are currently considered to be the most

important correlate of immunity. While neutralizing antibodies would not directly be elicited by a T cell epitope-driven vaccine, helper (CD4) T cell epitopes are required to generate high affinity, high specificity antibodies. We also note that we have limited our focus on the two major RSV surface proteins in our current analysis, but other RSV proteins like N, M, or M2-2 proteins might also contribute to vaccine efficacy [145].

An effective vaccine against variable viruses should contain T cell epitopes that are highly conserved among circulating strains [146]. Vaccine efficacy can be diminished if T cell epitopes in a vaccine strain do not match when new strains of pathogens emerge. In this study, we used an immunoinformatic-based approach to estimate cross-conserved T cell epitope contents between two live attenuated vaccine candidate strains and RSV circulating wild strains. We found that there was a low proportion of cross-conserved T cell epitope content with vaccine strains that belonged to different antigenic groups, which indicates the risk of using a single-subtype strain in RSV vaccines. In addition, we observed a lower proportion of cross-conserved G-protein T cell epitope content between vaccine strains and recent circulating strains in the same antigenic group, which suggests that including T cell epitopes from different strains in the same antigenic group might also be important for RSV vaccine development. Although we did not observe a significant change in cross-conserved T cell epitope content in F protein, we cannot rule out the possibility that variation of F protein in the future could render a single-strain-based vaccine less effective. Our current analysis is based on reduced datasets due to the heavy computational capacity required to perform epitope content comparison. We constructed these representative datasets by randomly subsampling the complete datasets according to geographical regions and isolated years. Our findings may reflect the T cell epitope diversity of publicly available RSV strains, however, additional RSV surveillance efforts may be required to get a full picture of the T cell epitope variability of RSV.

Our current study is limited by lacking experimental validation of T cell epitope prediction. We focus on computationally predicted MHC binding to identify T cell epitopes. Although the

strength of MHC binding is the key parameter that determines a peptide's immunogenicity, but not sufficient for a module to be immunogenic [147]. Other aspects associated with pathogen-induced T cell immune response, such as appropriate antigen-processing [148] and T cell receptor recognition [147] are not considered in this study, which might cause bias in the computational-based T cell epitope landscapes. However, the observed clustered pattern of RSV surface proteins on the T cell epitope landscape in this study reflects the diversity of T cell epitopes within different strains. This finding provides valuable insights into virus evolution in the aspects of T cell immunity and can contribute to the strain selection for vaccine design.

Overall, this study provides a focused analysis of T cell epitopes in RSV major surface proteins using computational tools. We performed a comprehensive T cell epitope prediction for RSV showing the immunological relationship of T cell epitopes in RSV surface proteins. This study demonstrates that T cell epitope evolution may differ from genetic variation and provides a framework for developing an integrated epitope-based RSV vaccine and evaluation methods that could be used to optimize vaccination strategies.

Chapter 4

Characterizing Potential Interaction Between Respiratory Syncytial Virus and Seasonal Influenza in the U.S.[1]

[1]Jiani Chen, Deven V. Gokhale, Liang Liu Pejman Rohani, Justin Bahl, Characterizing Potential Interaction Between Respiratory Syncytial Virus and Seasonal Influenza in the U.S.
Submitted to *PNAS*, November 2022.

4.1 ABSTRACT

RSV and seasonal influenza are two of the most important causes of respiratory infection that consistently peak during winter months in the U.S. Here, we characterized the circulation of these viruses in the U.S. with weekly positive case reports and genetic surveillance and used a mathematical modeling approach to explore their potential interaction at an HHS regional level. Our analyses showed RSV and seasonal influenza co-circulate with various relatively epidemic sizes and seasonal overlaps across seasons and regions. We found RSV might have different evolutionary dynamics compared to seasonal influenza, with local persistence may play a role in underling annual epidemics. Our analysis supports a competitive interaction between RSV and seasonal influenza in most HHS regions and we speculate that cross-immunity after infection might be the major driver of viral competition. Together, our work supports the competition between RSV and seasonal influenza across the U.S. at a population level. Our findings are important for the future development of protective strategies against these respiratory viruses.

4.2 Introduction

Interaction among infectious diseases has been very well documented in various polymicrobial disease systems [149]. In the context of respiratory infections, a well-known example is the association between the influenza virus and the bacterium pneumococcus, driven by enhanced susceptibility to secondary bacterial colonization subsequent to influenza infection [150, 151, 152]. Another important example comes from the 2009 influenza A pandemic (H1N1pdm09), with data from several European countries indicating that the spread of the virus might have been interrupted by the annual autumn rhinovirus epidemic [153]. These pathogen-pathogen interactions may lead to cooperation or competition, and their co-occurrence within the same host population can profoundly alter the pathogenesis and spread of infectious diseases with substantial public health consequences [154].

Respiratory Syncytial Virus (RSV) and seasonal influenza virus (Flu) are two of the most important causes of respiratory infection [11, 155, 68]. Both RSV and seasonal influenza activity consistently peak during winter months in the US. However, the relative disease burden and similarity between the epidemic timing of these viruses vary substantially among regions and seasons [12, 68]. The burden of influenza varies from season to season, depending in part on the dominant virus type or subtypes in circulation [156]. Influenza subtypes A/H3N2 and A/H1N1 (Flu A) have become the leading cause of seasonal influenza illness and death in the U.S. over the last 50 years [157]. Additionally, two distinct lineages of the influenza B (Flu B) virus, Victoria and Yamagata, co-circulate or alternate with Flu A and have received greater attention in recent years [158]. Intensive studies of the molecular evolution of the influenza virus have provided important insights into its seasonal emergence and spread in human populations [156, 159, 158]. Similar to seasonal influenza, two distinct antigenic subgroups of RSV have been identified, RSV-A and RSV-B, which show clear phylogenetic divergence. Genotype ON (RSV-A) and BA (RSV-B) are the two most prevalent RSV genotypes that are circulating in the U.S. in recent years [160, 161, 79]. However, the molecular epidemiology of RSV is largely unknown both on local and global scales [162]. Since RSV and seasonal influenza have similar seasonality and clinical presentation, researchers often try to extrapolate knowledge gleaned from the study of seasonal influenza to RSV to understand viral dynamics and develop protective strategies against these respiratory viruses-induced illnesses together [163]. Therefore, elucidating the exact nature of the interaction between RSV and seasonal influenza can greatly enhance our ability to forecast future epidemics, and substantially inform the development and implementation of control strategies.

Both biological and epidemiological studies suggest potential competitive interaction between RSV and seasonal influenza [73]. The proposed biological mechanism for the competition between these viruses is the activation of the innate “antiviral response” by an infection that can inhibit further or subsequent infection of another virus, resulting in a period of cross-protection during or after infection [162, 164]. Epidemiological data show

that when rates of infection with RSV are high, influenza infections are low; the converse is also true [165]. The interaction between respiratory viruses has the potential to impact virus genetic diversity and epidemics pattern and therefore would lead to the observed patterns in phylogenetics or surveillance data [156, 166]. Previous studies have demonstrated the application of dynamic modeling to test biological hypotheses on pathogen interaction mechanisms and it could be used to infer the interaction between RSV and seasonal influenza [167]. However, very few studies examine potential interaction between RSV and seasonal influenza on a population level and the explicit mechanism of their interaction has not been well-studied.

In this study, we initially carry out a phylogenetic analysis to elucidate differences in the evolutionary trajectories of RSV and seasonal influenza. We then apply a seasonally forced, two-pathogen, mechanistic transmission model to explicitly investigate the nature of ecological interference between RSV and seasonal influenza and quantify its effects in driving transmission dynamics in different regions in the U.S. Our analysis provides evidence of the potential negative interaction of RSV and seasonal influenza at a population level and the cross-immunity after infection might be an important mechanism.

4.3 Materials and Methods

4.3.1 RSV and seasonal influenza surveillance data

PCR reports from HHS regional-level surveillance data for RSV that are collected from 2011 to 2019 were requested from CDC's National Respiratory and Enteric Virus Surveillance System (NREVSS). HHS regional-level type-specific seasonal influenza (Flu A / Flu B) positive test reports were downloaded from FluView Interactive [168], and the viral surveillance reported by Clinical Labs was used after the year 2015. The national-level surveillance data were aggregated from the reports of 10 HHS regions.

4.3.2 Correlation coefficient estimation

To measure the association between RSV and the seasonal influenza epidemic, we computed correlation coefficients with weekly surveillance reports of different pairs of viruses in each HHS region in R v4.1. The weekly positive cases / million individuals of each virus were calculated by the estimated population for each HHS region in given year, which were collected from the U.S. Census Bureau using the "State Population Totals and Components of Change: 2010-2019" data set [169]. We employed the Shapiro-Wilk normality test to test for the normality of the error distribution of the time series being compared. Pearson's correlation coefficient was used to quantify the linear association among the co-circulating viral dynamics and the association was taken to be statistically significant at a p-value < 0.05 in a t-test. The time-lagged correlation coefficient was estimated in R package astsa v1.15 [170].

To examine the correlation of the genetic diversity for RSV and seasonal influenza, approx() interpolating function in R was used to generate the effective population sizes of different virus strains at the same time points from 2011-2019 from their corresponding Bayesian "Skyride" analysis. After examining the normality of the data using Shapiro-Wilk normality test, we used Pearson's correlation coefficient test to measure the correlation of RSV and seasonal influenza

4.3.3 Genetic datasets

HA segments of seasonal influenza (H3, H1, Victoria, and Yamagata) were retrieved from the GISAID database spanning the 2011-2019 epidemic seasons (generally considered as the October of the start year through May of the following year) in the U.S. Due to the lack of sampling for RSV in the US and North America regions, RSV G gene sequences (at least 70% length of complete G gene) that are collected from available northern hemisphere countries during the same period were retrieved from the GISAID database. RSV genotypes were determined with a reference dataset using NCBI BLAST [171, 172] and further confirmed with

phylogenetic analyses, and sequences belonging to the dominant genotypes "ON" (RSV-A) and "BA" (RSV-B) were selected for the following analysis.

Nucleotide sequences that belong to each genotype were aligned using MAFFT.v7 [173] separately and initial maximum likelihood (ML) phylogenetic trees were built using RAxML.v8 [85]. The temporal signal of each dataset was analyzed using TempEst v1.5.3 [86] and temporal outliers were removed for further studies. These ML phylogenies for seasonal influenza in different subtypes were further used to subsample to a maximum of 100 isolates in each season using the Phylogenetic Diversity Analyzer (PDA) [174]. These subsampled datasets were used for the subsequent Bayesian phylogenetic analyses. The final datasets contained 1004 ON1, 769 BA RSV G gene sequences, 867 H3, 871 H1, 796 Victoria, and 789 Yamagata HA gene segments of seasonal influenza.

4.3.4 Bayesian phylogenetic analysis

Evolutionary dynamics of RSV and seasonal influenza were estimated with a Bayesian phylogenetic approach using BEAST v.1.10.4 [175]. A generalized time-reversible nucleotide substitution model with gamma rate heterogeneity was used for the analysis of RSV and the SRD06 codon position model was used for the analysis of seasonal influenza [176, 177]. Uncorrelated lognormal relaxed clock and GMRF Skyride coalescent model [178] were used to estimate the population dynamics of the virus during each season for both pathogens. At least 4 independent MCMC chains of 150-200 million generations were simulated to ensure a sufficient effective sample size (ESS $>$ 200) as diagnosed in Tracer v1.7 [179]. LogCombiner v1.10.4 was used to combine the multiple runs and the Maximum Clade Credibility (MCC) tree was summarized in TreeAnnotator v1.10.4 after the removal of the 10% burn-in. Temporal-sliced descriptive statistics, including TMRCA estimates, diversity measurements, and mean statistics including coalescent rate, diversity, and Tajima's D scores were calculated from BEAST sampled Bayesian phylogenetic trees using the software package PACT v0.9.3 [180].

4.3.5 Two pathogen transmission model

We applied a two-pathogen transmission model to study the potential interaction of RSV and seasonal influenza [70]. The population is compartmentalized into Susceptible (S), Infectious (I), Cross-protected (C), and Recovered (R) for each pathogen i, ($i\epsilon 1, 2$) [181, 70]. Transitions among these compartments were governed by the following processes. Susceptible individuals with no immunity are infected at rates λ_i and enter the corresponding I compartment. Following the infectious period (with a mean duration of $1/\gamma_i$), individuals recover and move to the C compartment, where individuals are fully protected against homologous re-infection and have some cross-immunity against another virus. Individuals then lose this cross-immunity but retain pathogen-specific immunity in moving to the R compartment at rate ρ. Pathogen-specific immunity wanes as individuals become susceptible again at rate ω. The birth-death rate at each compartment is μ.

We hypothesize that the infection with another pathogen might be less likely either during or after the infection of the first pathogen, corresponding to either the I or C compartments, and therefore we define the infection rate of another pathogen with the potential of reduction. In the I compartment, an interaction corresponds to a reduction in the infection rate of the second pathogen at $\psi\lambda_i$, where the inhibition to be co-infected is modeled by scaling the rates of infection using a value between 0 and 1 ($\psi \epsilon [0, 1]$). In the C compartment, we similarly model a reduced infection rate $\chi\lambda_i$ for the second virus, wherein the strength of cross-immunity, χ, takes a value between 0 and 1 ($\chi \epsilon [0, 1]$)). Parameters and estimated values used in the model are shown in Appendix A, Table A.3. In a deterministic setting, the model is described by 16 ordinary differential equations (Appendix A, Figure A.4, equations for each state can be read directly from Appendix A, equation A.2 to equation A.17).

The inhibition of co-infection was modeled by a reduction in the infection rate to the second virus, $\psi\lambda_i(\psi \epsilon [0, 1])$, the equations governing these dynamics are given by:

$$\frac{dX_{SI}}{dt} = \omega_1 X_{RI} + \lambda_2 X_{SS} - \gamma_2 X_{SI} - \psi\lambda_1 X_{SI} - \mu X_{SI} \tag{4.1}$$

$$\frac{dX_{IS}}{dt} = \lambda_1 X_{SS} + \omega_2 X_{IR} - \gamma_1 X_{IS} - \psi\lambda_2 X_{IS} - \mu X_{IS} \tag{4.2}$$

$$\frac{dX_{II}}{dt} = \psi\lambda_1 X_{SI} + \psi\lambda_2 X_{IS} - \gamma_1 X_{II} - \gamma_2 X_{II} - \mu X_{II}. \tag{4.3}$$

The cross-immunity after the infection was modeled by a reduced infection rate to the second virus $\chi\lambda_i(\chi \epsilon [0,1])$ at the C compartment:

$$\frac{dX_{IC}}{dt} = \chi\lambda_1 X_{SC} + \gamma_2 X_{II} - \gamma_1 X_{IC} - \rho_2 X_{IC} - \mu X_{IC} \tag{4.4}$$

$$\frac{dX_{CI}}{dt} = \gamma_1 X_{II} + \chi\lambda_2 X_{CS} - \rho_1 X_{CI} - \gamma_2 X_{CI} - \mu X_{CI} \tag{4.5}$$

The overall population in this model is the sum of the populations in each compartment:

$$\begin{aligned} N = X_{SS} + X_{SI} + X_{SC} + X_{SR} + X_{IS} + X_{II} + X_{IC} + X_{IR} + \\ X_{CS} + X_{CI} + X_{CC} + X_{CR} + X_{RS} + X_{RI} + X_{RC} + X_{RR} \end{aligned} \tag{4.6}$$

For virus i($i \in 1,2$) in this model, seasonality in the transmission rate ($\beta_i(t)$) was incorporated in the transmission model, where R_0^i is the basic reproductive number, b_i is the amplitude of seasonality, and t_0^i is the peak transmission day during the season:

$$\beta_i(t) = \gamma_i R_0^i \left[1 + b_i \cos\left(2\pi \frac{t - t_0^i}{T}\right)\right] \tag{4.7}$$

The force of infection depends on the number of individuals infected with the virus i and the rate at which cases are imported into the population from external sources η_i.

$$\lambda_1(t) = \beta_1(t) \frac{X_{IS} + X_{II} + X_{IC} + X_{IR} + \eta_1}{N} \tag{4.8}$$

$$\lambda_2(t) = \beta_2(t) \frac{X_{SI} + X_{II} + X_{CI} + X_{RI} + \eta_2}{N}. \tag{4.9}$$

The deterministic model was implemented in R using the package "pomp" [182]. For the likelihood-based inference, we generated the observed cumulative number of cases (K_i) according to the following equations, where δ_i is the case report rate of the virus.

$$(dK_1)/dt = \delta_1(\gamma_1 X_I S + \gamma_1 X_I I + \gamma_1 X_I C + \gamma 1 X_I R \tag{4.10}$$

$$(dK_1)/dt = \delta_1(\gamma_1 X_I S + \gamma_1 X_I I + \gamma_1 X_I C + \gamma 1 X_I R \tag{4.11}$$

4.3.6 Hypothesis testing and parameter estimation

The model was fit with RSV and type-specific seasonal incidence surveillance that are collected in 10 HHS regions from 2014 to 2017. Maximum-likelihood estimates (MLEs) for the unknown parameters were found using trajectory matching. Differential evolution algorithm (DEoptim) for global optimization implemented in R package "DEoptim" [183] was used to find the estimates that maximize the MLE. We calculated AIC using MLE estimates under each hypothesis. If a model is more than 2 AIC units lower than another, we then consider it significantly better. RMSE was calculated to quantify how well the model fits with surveillance data. As proof of concept, we performed an out-of-sample simulation to demonstrate the suitability of the MLE model; that is, we generated the forecast for the following season of the MLE model with the estimated parameters and then compare the simulated cases with RSV and seasonal influenza surveillance that are collected in each HHS region.

We performed a simulation study to estimate the uncertainty of the competitive interaction parameters. We created likelihood profiles for the respective parameters by fitting a smooth line through the log of repeated likelihood estimates in which the respective parameter is a fixed value. The 95% CI is taken to $\chi^2_1(0.95)/21.92$ log-likelihood units below the maximum likelihood estimates using the χ^2 distribution.

4.4 Results

4.4.1 Seasonality overlaps between RSV and seasonal influenze

Different relative epidemic size and peaking time of RSV and seasonal influenza can be observed across different HHS regions in the U.S. (Appendix A, Figure A.1, represented regions are shown in Figure 4.1A). We computed the correlation coefficient of these viruses using weekly surveillance reports and found the positive case of RSV and seasonal influenza increased with another (Figure 4.1B, Appendix A, Table A.1), suggesting epidemic seasonality similarities between these viruses in the U.S. In addition, the magnitude of the correlation between RSV and seasonal influenza are various across HHS regions, with a relatively high value in HHS Regions 6, 7, and 10 between RSV and Flu A and in HHS regions 6, 7, and 8 between RSV and Flu B. We further computed the time-lagged correlation of different pairs of viruses (Figure 4.1C) to measure the temporal association between RSV and seasonal influenza epidemics in the U.S. In our analysis, we observed the maximize correlation with zero to little (positive/negative) time lags in different regions, indicating the seasonality overlap between RSV and seasonal influenza are different across regions and might have different interaction patterns.

4.4.2 Population dynamics of RSV and seasonal influenza

We characterized the population dynamics of RSV and seasonal influenza using a phylogenetic approach. The phylogenies of seasonal influenza lineages were estimated using Hemagglutinin (HA) gene segments that were collected in the U.S. during the 2011-2019 seasons. The analyses of RSV were based on G gene sequences from northern hemisphere countries in the same period and we focus on genotypes ON and BA. We calculated the time-sliced statistics (Materials and Methods, TMRCA: Time to the most recent ancestor, and Diversity: branch distance between pairs of tips) from RSV and seasonal influenza phylogenies (Figure 4.2A). A periodic reduction in TMRCA and diversity across time can be observed from

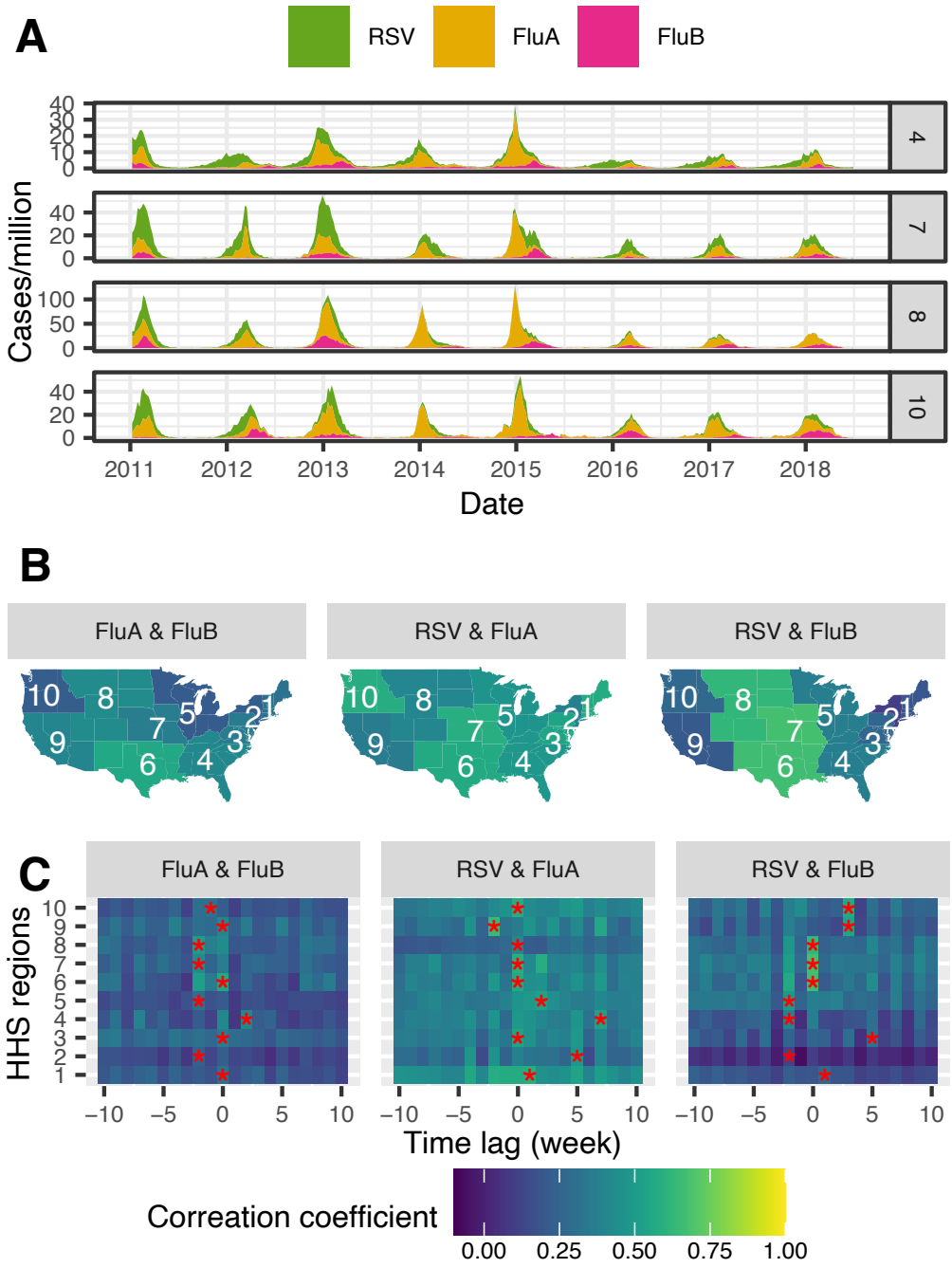

Figure 4.1: **Correlation coefficient of RSV and seasonal influenza epidemics in the U.S.** (A) Weekly RSV and seasonal influenza epidemics in represented HHS regions (HHS Region 4,7,8,10) (B) Correlation coefficient between RSV and seasonal influenza weekly surveillance reports (Appendix A, Table A.1). Relative color in HHS regions indicates the normalized correlation coefficient of weekly positive samples of the different pairs of viruses. (C) Time-lagged correlation coefficient between RSV and seasonal influenza (Appendix A, Table A.2). The time lags (week) that maximize the correlation coefficient in each HHS region are labeled with *. Negative values indicate a leading correlation and positive values indicate a lagging correlation. Lighter-colored indicate a higher correlation coefficient between two viruses and darker-colored indicate a lower correlation coefficient.

seasonal influenza phylogenies (H1, H3, Victoria, and Yamagata). In contrast, TMRCA and diversity scores of RSV phylogenies increased across time, except for the analysis of BA in recent seasons, which might be due to a lack of RSV sampling during recent sampling times. In addition, Bayesian "Skyride" coalescent reconstructions demonstrate strong periodicity of seasonal influenza viruses, which are correlated with their seasonal epidemic patterns, though we observed very low genetic diversity of Flu B lineages in some years. The population dynamics of RSV did not show seasonal fluctuation and both ON and BA genotypes presented peaks from 2011 to 2019 (Figure 4.2B, Figure 4.2C). These different phylogenetic patterns suggest local persistence may play a role in underlying annual epidemics of RSV and in contrast, influenza strains are under constant negative selection.

We computed the correlation coefficient using these estimated time series effective population sizes of RSV and seasonal influenza to explore their potential association at the virus population level (Appendix A, Table A.2). Significant negative correlation coefficients can be observed between RSV and seasonal influenza strains: RSV-A genotype ON with influenza H3 (CI: -0.513, -0.128), H1 (CI: -0.482, -0.087) and Victoria (CI: -0.482, -0.087); RSV-B genotype BA with influenza H1 (CI: -0.506, -0.138) and Yamagata (CI: -0.564, -0.217), which indicate these viruses might have a negative association.

4.4.3 Potential competition between RSV and seasonal influenza

The co-circulation of RSV and seasonal influenza suggests a transmission model that allows the co-infection of these viruses is required. Therefore, we adapted a two-pathogen transmission model to investigate the potential competition between RSV and seasonal influenza (Appendix A, Figure A.4, Appendix A Table A.3) in the U.S. This model accommodates two mechanisms of competition between pathogens. Competition may arise through the inhibition of co-infection, which is modeled by scaling the transmission rates of secondary infections with parameter ψ. In addition, following infection with one pathogen, individuals may gain short-term cross-immunity against another pathogen and can be similarly modeled by scaling the

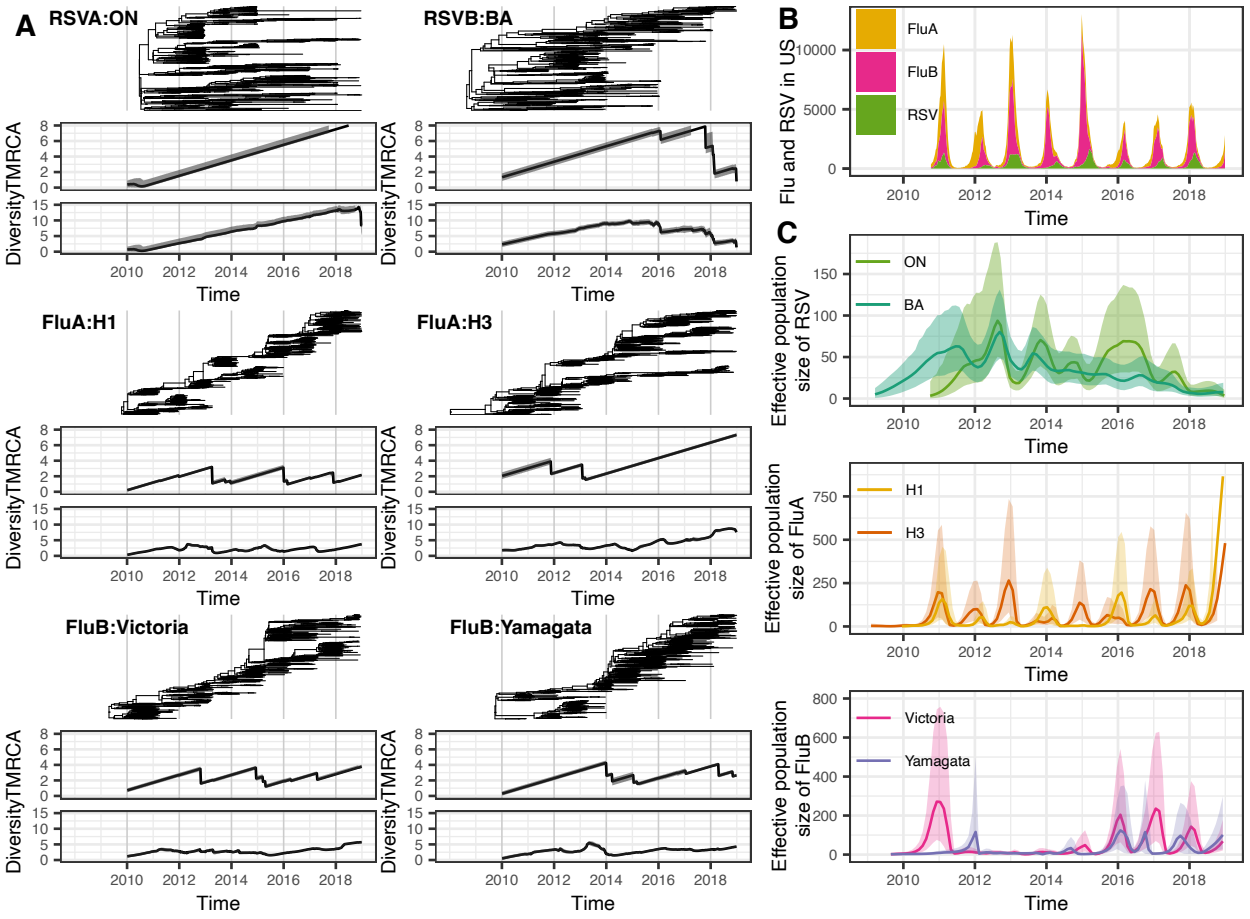

Figure 4.2: **Population dynamics of RSV and seasonal influenza in the U.S.** (A) Phylogenies and corresponding time-slice statistics. The time-slice TMRCA and Diversity are calculated by breaking the phylogeny into multiple temporal sections. TMRCA, time to the most common ancestor of all tips within temporal sections. Diversity is the average time to coalescent for pairs of lineages within temporal sections. TMRCA and Diversity are measured in the unit of years. Solid lines represent mean values and gray outlines represent 95% CI across MCMC replicates. (B) Surveillance curve and (C) reconstructed effective population size of RSV and seasonal influenza using GMRF Bayesian "Skyride" analysis. Solid lines represent mean values and outlines represent 95% CI across MCMC replicates.

force of secondary infection by χ. Changing the value of these two parameters can influence the timing, magnitude, and shape of observed epidemics (Figure 4.3).

We fit this model to RSV and type-specific seasonal influenza (Flu A or Flu B) incidence data that are collected from 10 HHS Regions from 2014 to 2017 and use Akaike Information Criterion (AIC) values to compare four different hypotheses regarding the interaction between RSV and seasonal influenza: (1) no interaction ($\psi = \chi = 1$); (2) inhibition of co-infection $\psi < 1, \chi = 1$); (3) transient cross-immunity after infection ($\psi = 1, \chi < 1$); and (4) both the inhibition of co-infection and transient cross-immunity ($\psi < 1, \chi < 1$). In these analyses, the duration of short-lived cross-protection is fixed at 6 months [164]. We estimated parameters that could optimize the likelihood estimates of the model under each hypothesis in 10 HHS regions (Appendix A, Table A.6 to Table A.15). The virus-specific simulated trajectories under four hypotheses for 10 HHS regions are shown in Appendix A, Figure A.5. We found that adding competitive interaction between RSV and seasonal influenza provides a better explanation of the surveillance data in most regions except for HHS Region 8 (Appendix A, Table A.5), and the epidemic trajectories from surveillance data in different HHS regions can be well fit under the corresponding best hypothesis as examined by Root Mean Square Error (RMSE) (Appendix A, Figure A.6). As described in Table 4.1, different levels of cross-immunity after infection were estimated across HHS regions under the corresponding best hypothesis, and the inhibition of co-infections was also supported in some regions (HHS Regions 1, 3, 4, 5, 6, 7 for RSV and Flu A, HHS Regions 3, 5, 6, 10 for RSV and Flu B). We used HHS Region 1 as an example region to demonstrate our findings: our analysis suggests RSV and Flu A are fully inhibited to be co-infected and there was also a short-term cross-immunity, with a 45% reduction in the infection rate of another pathogen. By fitting the RSV and Flu B surveillance, there was a 39% reduction in the infection of the second virus and the inhibition of co-infection was not supported (Figure 4.4, Table 4.2).

We further estimated the credible interval (CI) of the two competitive parameters under the best-fit hypothesis in 10 HHS regions (Appendix A, Figure A.7). As shown in the analysis

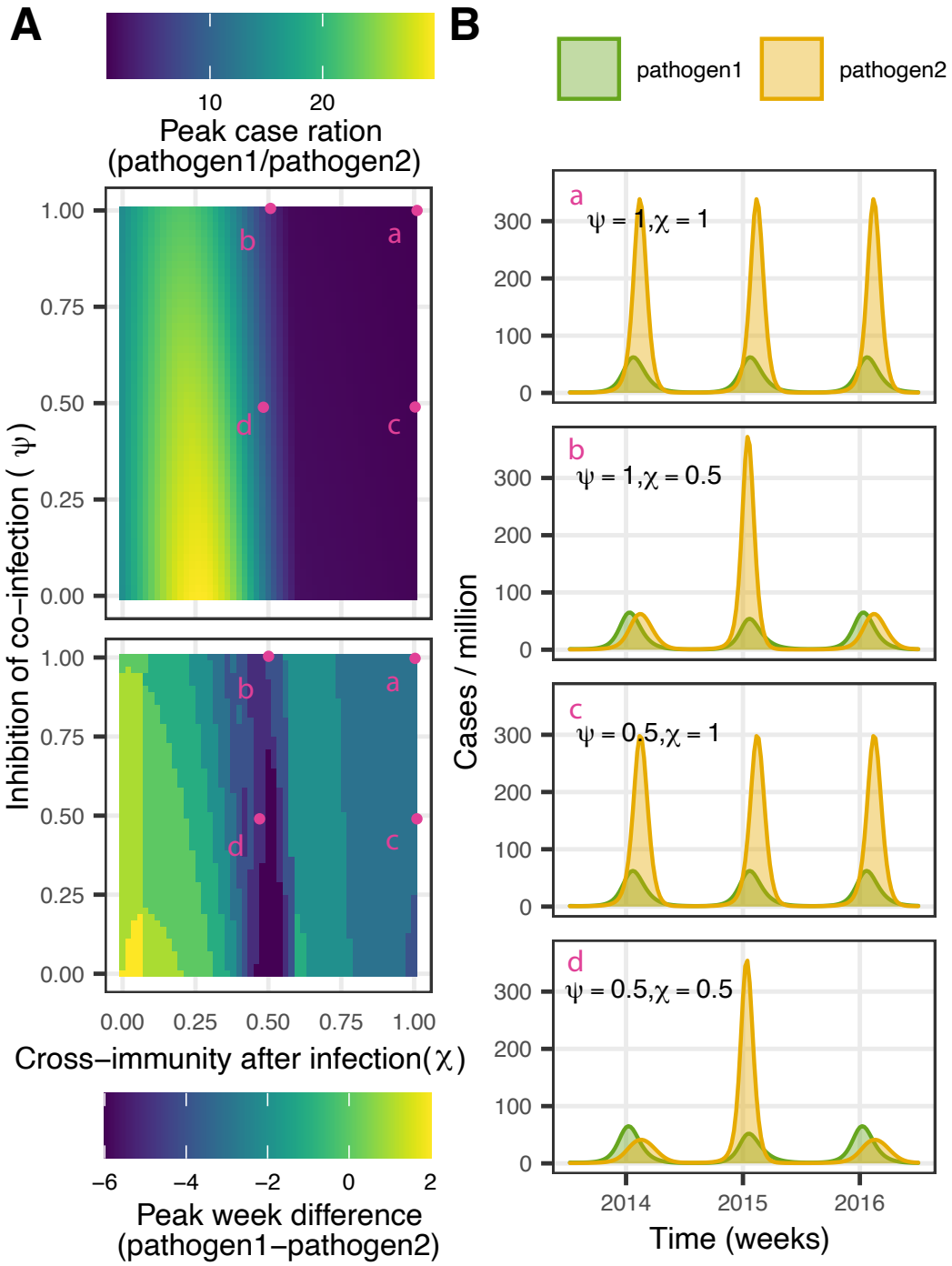

Figure 4.3: **Simulation study to examine the effects of the proportion of inhibition of co-infection and cross-immunity after infection using a two-pathogen transmission model.** (A) Changes in relative peak case ratio and the relative time of peaks based on the values ψ and χ in the model. (B) Simulated trajectories resulting from parameter values of ψ and χ selected at points a, b, c, and d in panel (A).

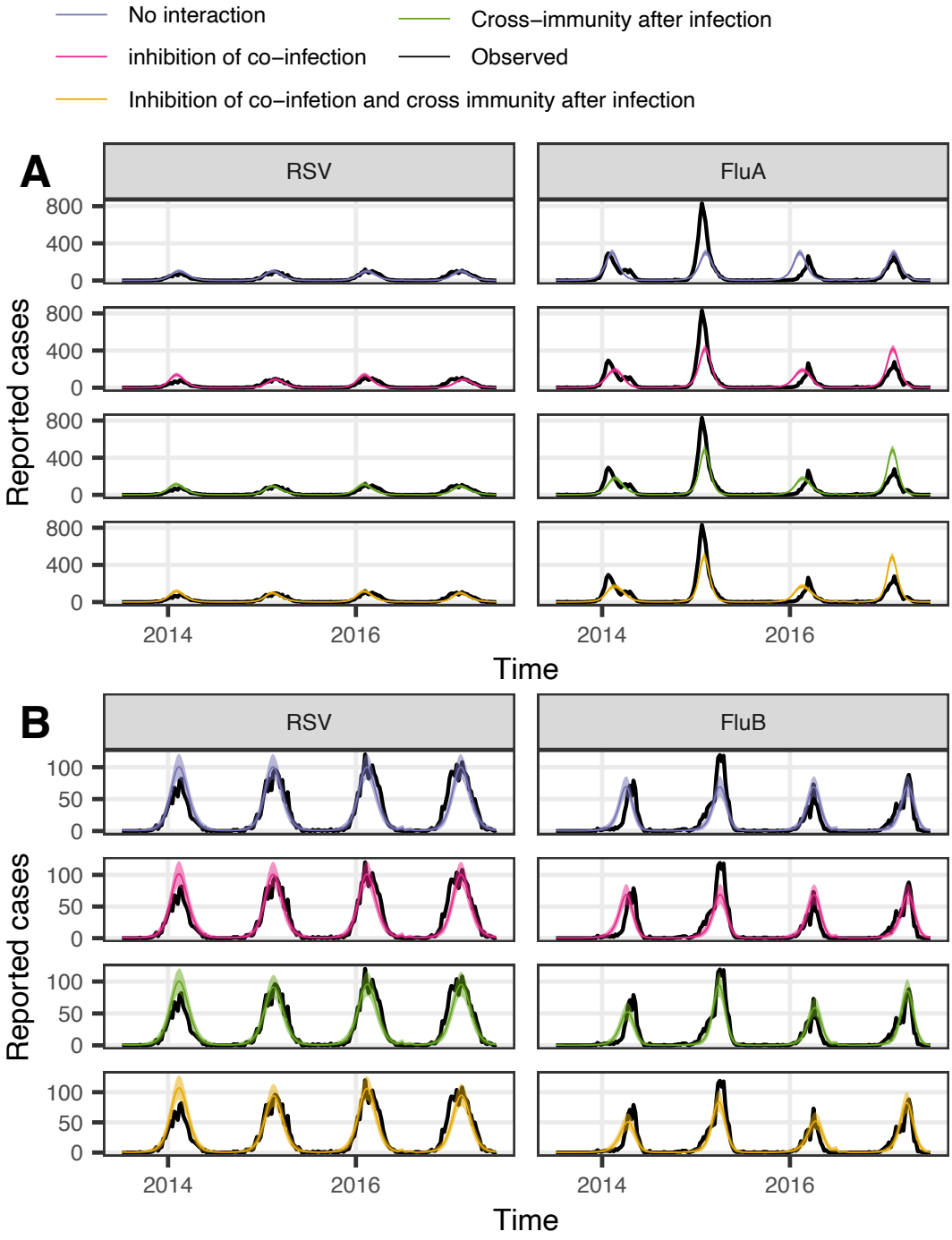

Figure 4.4: **Relative fits of four epidemiological hypotheses for season 2014-2017 in HHS Region 1.** Matched virus-specific simulated trajectories under four hypotheses are shown in different colors (Mean in solid line and 95% CI in shading). The black solid line represented the HHS regional level weekly case report data. (A) The test of the interaction between RSV and Flu A. (B) The test of the interaction between RSV and Flu B.

Table 4.1: **Competition parameter estimates under the best fit hypothesis for seasons 2014-2018 in 10 HHS regions.**

HHS region	RSV-Flu A		RSV-Flu B	
	Inhibition of co-infection (ψ)	Cross-immunity after infection (χ)	Inhibition of co-infection (ψ)	Cross-immunity after infection (χ)
1	0	0.55	1	0.64
2	1	0.69	1	0.72
3	0	0.59	0	0.8
4	0.57	0.37	0	0.87
5	0	0.78	1	0.61
6	0	0.77	1	0.45
7	0	0.52	0.95	0.64
8	1	1	1	1
9	1	0.79	0.2	0.6
10	1	0.5	1	0.2

of HHS Region 1 (Figure 4.5), changing the values of ψ have relatively small effects on the likelihood profile compared with the strength of cross-protection (χ), and the CI of χ can be well identified (RSV and Flu A: 0.522, 0. 567, RSV and Flu B: 0.628, 0.638). Our results indicate the short-term cross-immunity after the infection might be the major mechanism to explain the competition between RSV and seasonal influenza.

4.5 Discussion

In this study, we demonstrate the circulation dynamics of RSV and seasonal influenza in the U.S and examine their potential competition and mechanism at a population level. Our phylogenetic analyses initially suggest there might be a negative association between RSV and seasonal influenza by studying the correlation of their estimated time-series genetic diversity. We further study this potential negative interaction explicitly by applying a two-pathogen transmission model with HHS regional-level surveillance reports. Taken together, our analyses provide statistical support for the negative interaction between RSV and seasonal influenza after infection.

Table 4.2: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 1.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.49	1.46	1.84	1.89
R_0^{Flu}	1.37	1.05	1.31	1.27
b_{RSV}	0.26	0.33	0.13	0.15
b_{Flu}	0.1	0.27	0.17	0.23
t_0^{RSV}	0.91	0.93	0.98	0.99
t_0^{Flu}	0.1	0.03	0.12	0.11
ψ	1	0	1	0
χ	1	1	0.47	0.55
δ_{RSV}	2.49E-04	2.45E-04	2.07E-04	1.97E-04
δ_{Flu}	6.49E-04	1.07E-03	8.94E-04	9.34E-04
loglik	-4946.37	-3966.74	-3673.11	-3463.34
AIC	9908.74	7951.49	7364.22	6946.68
ΔAIC	2962.06	1004.81	417.54	0
RMSE	62.47	52.51	50.36	49.3
RSV-Flu B				
R_0^{RSV}	1.49	1.49	1.34	1.36
R_0^{Flu}	1	1	1	1.09
b_{RSV}	0.26	0.26	0.29	0.3
b_{Flu}	0.28	0.32	0.38	0.32
t_0^{RSV}	0.91	0.92	0.9	0.9
t_0^{Flu}	0.14	0.14	0.16	0.16
ψ	1	0	1	0.02
χ	1	1	0.64	0.61
δ_{RSV}	2.49E-04	2.49E-04	3.33E-04	3.17E-04
δ_{Flu}	2.69E-04	2.69E-04	2.74E-04	2.47E-04
loglik	-1504.85	-1491.16	-1353.99	-1376.56
AIC	3025.7	3000.32	2725.99	2773.13
ΔAIC	299.71	274.33	0	47.14
RMSE	11.14	11.28	11.03	12.06

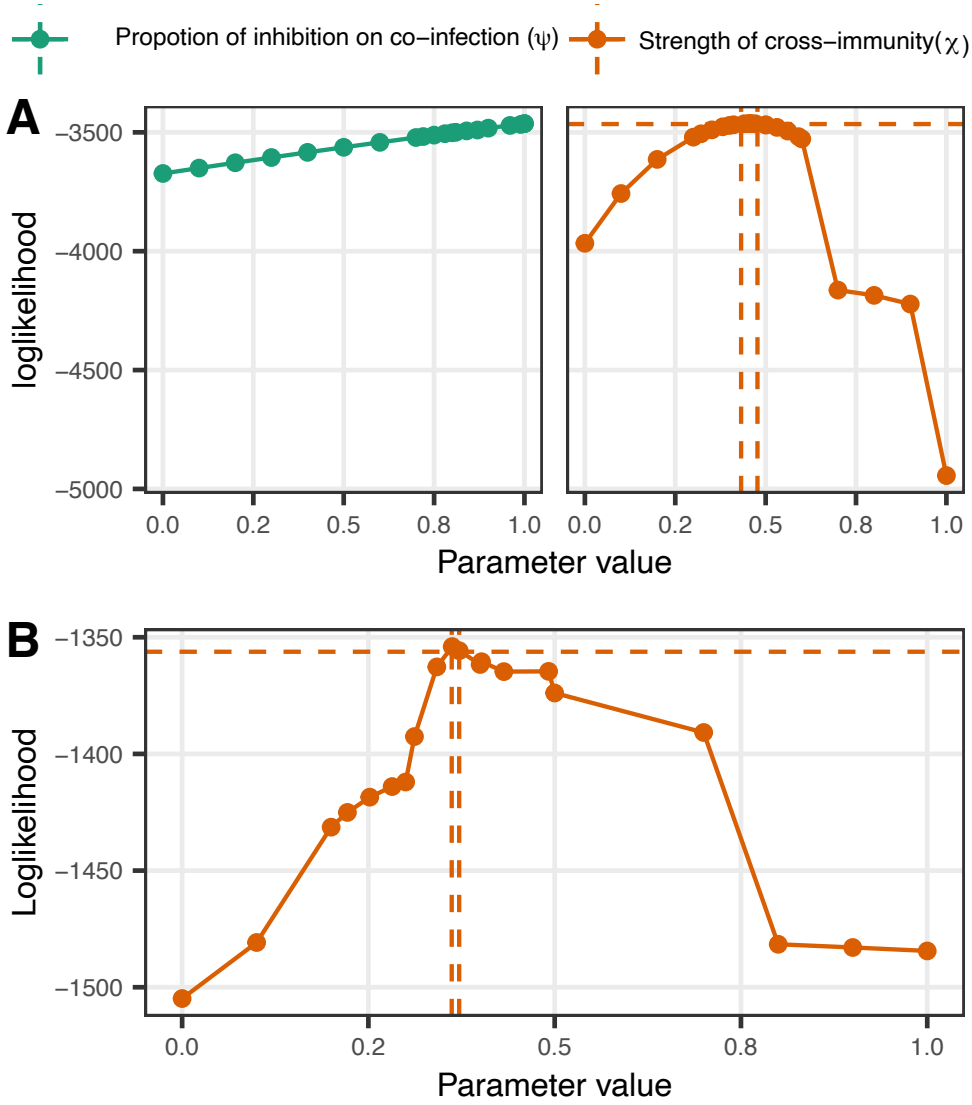

Figure 4.5: **Likelihood profile tests of competition interaction parameters for RSV and seasonal influenza, inferred in HHS Region 1 from 2014 to 2017.** Plotted in each graph is the likelihood profile for the inhibition of co-inhibition (ψ) in green and the cross-immunity after infection (χ) in orange, which are lines connected by repeated likelihood estimates (N =20, shown in colored solid circles). The values within the two dashed lines are within the estimated 95% CI. (A) The test of the interaction between RSV and Flu A. (B) The test of the interaction between RSV and Flu B.

By studying the co-circulation of RSV and seasonal influenza in different regions of the U.S, we confirm the temporal correlation between these viruses. We found there might be little to zero number of weeks of time-lags to maximize the correlation coefficient between RSV and different subtypes of influenza. These differences in relative epidemic peaks of RSV and seasonal influenza across different HHS regions in the U.S suggests they might interact with different severity.

Phylogenetic analysis can be used to estimate the virus's genetic diversity through time and therefore has the potential to reflect the effects of interactions between viruses. Previous phylogenetic studies of seasonal influenza demonstrated a periodicity in the relative genetic diversity in temperate zones where regional epidemics have coinciding epidemic peaks and seasonal patterns [184]. A similar seasonality of RSV was also reported in multiple Northern Hemisphere countries experiencing a temperate climate, with the annual epidemics starting in the autumn, peaking in winter, and ending in spring [185, 186, 187]. Therefore, we expect a similar seasonal periodicity in relative genetic diversity could be observed in the RSV phylogenetic analyses with sequences that are collected from available northern hemisphere countries. However, our phylogenetic analyses of RSV suggest a different pattern of seasonal influenza and indicate the potential of regional persistence, which is consistent with previous phylogenetic studies of RSV in other regions [188, 189, 190]. The antigenic drift of seasonal influenza creates competition among strains, resulting in the emergence of selective forces and previously circulating strains can be prone to local extinction each year [191]. This negative selection including the effect of the vaccine causes rapid influenza population turnover and therefore it only takes a few years for contemporaneous strains to find a common ancestor [192, 180]. In contrast, RSV phylogenies harbor many deep branches in our analysis (Figure 4.2). This reconstructed population dynamics of RSV might indicate latency supported by low levels of circulation during the summer season that is not captured by current RSV genomic surveillance [190], and RSV seasonal epidemics could be driven by other ecological factors like humidity, temperature, and human contact or travel [187, 193]. These differences between RSV

and influenza suggest in phylogenies suggest different protection and vaccine strategies may be needed against these viruses [191, 180, 180]. We observe a negative correlation between RSV and seasonal influenza lineages with time series effective population sizes that are estimated from genetic data, which suggests a potentially negative interaction between these viruses at a population level. However, our analyses are limited by the lack of genetic surveillance effort and more explicit analyses are still needed to study the potential competition between RSV and seasonal influenza.

Our transmission modeling work provides additional evidence of the negative interaction between RSV across regions and seasonal influenza and further demonstrates the potential mechanism of interaction. We find adding inhibition on co-infection to the model has relatively small effects on epidemic trajectories from a simulation study (Figure 4.3), and the inhibition on co-infection is not supported by the surveillance data from some HHS regions. This might be due to the short infection period of RSV and seasonal influenza, and only a small proportion of the population has the potential to be infected by two viruses simultaneously and therefore do not leave a strong dynamical footprint in population-level weekly surveillance data. In our analyses, there are very few individuals to be infected by RSV and seasonal influenza simultaneously, which is consistent with previous surveillance studies that suggest the observed incidence of co-infectivity of RSV and influenza was significantly less than the expected [68]. To further confirm our estimation, the exact number of cases with co-infection is needed. In addition, the severities of competition between these viruses across HHS regions are estimated to be different from our analysis. The population structure (e.g. age) in different regions might cause different levels of cross-pathogen immunity [162]. It is also important to note that the surveillance reporting institutions vary across regions, which may result in different levels of competition that are reflected in the surveillance data. These location-specific features need to be considered when generalizing the findings.

Much of the previous evidence for the interaction between RSV and seasonal influenza is on an individual, biological level [194, 166, 164, 195]. Previous evidence of the interaction of RSV

and seasonal influenza on population level implies that prevention of one could inadvertently lead to an increase in the burden of the other [167, 196]. Recent modeling work suggests a competition between RSV and influenza with infection reducing heterologous acquisition by 41% for 10 days among children after infection in Nha Trang, Vietnam [197]. We provide additional evidence for the competition between RSV and influenza at the population level in a more diverse region in the U.S. and the level of competition could be differentiated by location. While our findings are consistent with previous experimental studies and modeling work suggests a competitive interaction between RSV and influenza [73, 164, 162, 198], it is important to emphasize social behavioral changes after infection (sequestration during convalescence) with a virus may also play a role [199].

To simplify the analyses with the transmission model, we assume RSV-influenza interaction is bidirectional and the strength and duration of interaction that influenza exhibits on RSV are the same and vice versa. We do not model RSV and two subtypes of seasonal influenza together because adding another virus would significantly increase the complexity of the model and we believe there is a low chance for individuals to be infected with RSV and two subtypes of seasonal influenza within a short time. To reduce the number of parameters that need to be estimated, we fix the cross-protection period with a relatively long time of 6 months [183] and we assume the cross-immunity could protect the individual to be less likely infected by another pathogen at the same season. Our analysis is also limited by current RSV and seasonal influenza surveillance availability. We are lacking RSV subtype information in the current surveillance support. In addition, the reporting institutions of RSV and seasonal influenza surveillance vary between years, and spikes in detections may reflect increased testing within a given surveillance year. Hence, additional high-quality surveillance data in the future might be able to draw a more reliable conclusion.

Overall, we characterize the seasonal overlaps and evolutionary dynamics of RSV and seasonal influenza and focus on their potential interaction. Our work highlights the use of mathematical mechanistic models to test the interaction hypothesis of RSV and seasonal

influenza. Although more effort is needed, our results suggest a cross-immunity-induced negative interaction of RSV and influenza at a population level and can be identified in surveillance data. This study is helpful to have a better understanding of the different dynamics and the potential interaction between RSV and seasonal influenza, which are critical for predicting the effects of alteration of their ecological balance and designing vaccines against these viruses.

Chapter 5

Conclusions

5.1 Summary

The preceding chapters provide a demonstration of the adaption and development of computational frameworks to study the evolution and circulation of RSV. While RSV causes a significant clinical burden every year, we are currently lacking computational and statistical frameworks to track its evolution and circulation at the population level [172, 19]. The major objective of this dissertation is to examine the evolutionary and epidemiology dynamics of RSV at the population level. The analyses presented in this dissertation explore the RSV genotyping system, the evolution of RSV immunological profiles, and the co-circulation of RSV with other respiratory pathogens.

In Chapter 2, we propose a novel nomenclature system based on RSV whole-genome phylogeny to aid in classifying circulating RSV strains to a genotype [45]. While RSV genetic diversity has been widely reported and several research groups have been working on RSV genotyping, we are still lacking a standard RSV genotyping system capable of reflecting the viral genetic diversity and circulation patterns [28, 36, 31, 27, 39]. Complicated and excluded RSV genotype designations often cause confusion in RSV research and limit current surveillance efforts [79, 200]. Our new genotyping system extends previous genotyping research [19, 21] by three means. First, most RSV genotyping studies focus on the monophyletic clades from complete or partial G gene-derived phylogenies. We use whole-genome phylogenies and period of detection as the basis for RSV genotyping classification, which enables us to capture a more complete evolutionary history of RSV [200, 28]. Secondly, most RSV genotyping studies use genetic p-distance, which is calculated from the sequence's genetic similarities

and is used to set a minimum threshold to define an RSV genotype. However, sequence-based p-distance is highly sensitive to sequence errors and has to be updated with newly isolated sequences within the same phylogenetic clade [28, 36]. In our work, we calculate pair-wise node distances (ie., genetic distance between estimated common ancestors for a group of sequences) from RSV phylogenies and carry out a simulation study to justify the threshold to define the genotype groups. To make this genotyping system more accessible to the scientific community, we additionally provide a tool that enables the automated genotype classification of newly generated whole genome sequences as well as G gene sequences. This novel RSV nomenclature system can provide a better understanding of RSV evolution and strengthen scientific communication and RSV public health control measures.

The sequenced-based T cell landscape approach developed in Chapter 3 allows for an examination of the potential link between the genetic diversity of RSV. As described in Chapter 2, RSV genetic evolution has the potential to cause escape mutations to evade host T cell immunity. Compared to other respiratory viruses, such as influenza, we lack experimental studies of RSV T cell epitopes. Therefore, we apply an *in-silico* T cell epitope prediction approach to develop an algorithm to evaluate the T cell epitope differences between different strains of RSV. The Multidimensional Scaling (MDS) approach [133] is adapted to generate a visualization of T cell epitope patterns of the RSV population. Our study provides an important methodology to evaluate immune profiles predicted from genomic sequences of circulating viral strains. With an extensive examination of the RSV T epitope profiles at the population level, we find that patterns of T cell epitope evolution may differ from corresponding genetic evolution. The diversity of predicted T cell epitope phenotypes could be an important factor to be considered in RSV vaccine development.

Chapter 4 investigates the potential interaction between RSV and seasonal influenza. The interaction between respiratory pathogens has been widely reported [201, 202, 203]. However, understanding the possible influence of pathogen-to-pathogen interaction remains a key question for broader public health. Our work focuses on RSV and seasonal influenza, which

are both important directly transmitted viral respiratory diseases that are circulating during winter months in the U.S. Our phylogenetic comparison between these two pathogens suggests local persistence might play an important role in the circulation of RSV [190, 189, 44]. In contrast, the ongoing evolution (i.e., antigenic drift) of seasonal influenza leads to its annual epidemics [192, 191]. We initially examine the potential interaction between RSV and seasonal influenza with the virus's effective population size estimated from phylogenetic analysis. A potential negative association between these viruses was found. We then apply a previously described two-pathogen transmission model to investigate their interaction mechanistically using HHS regional-level weekly surveillance reports for these viruses. Our statistical modeling analysis provides statistical support for the negative association between these viruses and suggests the magnitude of competition might be different across different regions in the U.S. While the potential competition between RSV and seasonal influenza has been reported, we study the effects of pathogen interaction on virus population and observed epidemic trajectories. Our findings are important for the future development of protective strategies against these respiratory viruses.

While the focus of this dissertation is RSV, the computational and statistical approaches developed here can be widely applied to the surveillance work of other emerging viruses. The nomenclature system developed in Chapter 2 provides a new framework to develop genotyping systems for viruses with known sampling bias. In Chapter 3, we develop a sequenced-based T cell epitope landscape approach to study the T cell epitope evolution of RSV by integrating an *in-silico* epitope prediction approach, which also has possibilities to be applied to other pathogens with high genetic diversity. In Chapter 4, we provide statistical evidence of the competition between RSV and seasonal influenza in the U.S, which illustrates the potential of pathogen-to-pathogen interaction to impact the infectious disease dynamics and persistence of a much wider range of infections.

5.2 Challenges

5.2.1 Sampling bias

Sampling bias is a potential problem in any epidemiologic study, especially when analyses are based on publicly available data for pathogens with limited surveillance efforts. Because of the lack of surveillance efforts for RSV, sampling bias is a special challenge for the study of RSV at a population level. For example, in the RSV genotype classification work described in Chapter 2, the RSV whole genome sequences, including the cutoffs used to assign genotypes, may need further updating in the future with the prospect of increasing population turnover and sequence sampling [36]. Because of the ever-present threat of sampling bias and the lack of standard guidelines for its handling [204], we used node distance instead of traditional genetic distance to assign genotypes to attempt to limit the impact of sampling discrepancies within the sequence data. The node distance is determined by the sum of the branch length that is estimated from the reconstructed most recent common ancestors of different groups of sequences [205]. Compared with traditional genetic distance approaches, node distance incorporates the phylogenetic model and therefore is less sensitive to the potential sampling bias [204, 206].

The number of pathogen-derived genome sequences, including the genome sequence of RSV, deposited into repositories such as NCBI and GISAID is increasing rapidly. This wealth of data provides a great opportunity and challenge for biologists. Large datasets are usually difficult to visualize and use in downstream analyses [207]. To overcome these limitations, methods that reduce the size of datasets based on either sequence similarity or phylogenetic diversity have been developed [94, 208, 209]. In the RSV T cell epitope landscape development described in Chapter 3, the calculation of pair-wise T-cell immunity distances from a great number of all available RSV sequences requires heavy computational resources and will also create challenges for visualization to include the T cell epitope information from a great number of virus strains into a single landscape. We applied a phylogenetic

diversity subsampling approach with the software phylogenetic diversity analyzer (PDA) [94] to generate a subsample dataset that could also correct the overrepresentation of recently sampled strains in our analysis. Previous work demonstrated that it is more appropriate to assess biodiversity based on phylogenetic trees than on the concept of species richness. Phylogenetic diversity (PD) is a popular measure of the amount of evolutionary history encompassed by the species under consideration [210]. Therefore, PDA based subsampling approach has the advantage to reserve the genetic diversity of the original dataset and we apply it in our analyses with RSV genetic sequences.

5.2.2 *In silico* T cell epitope prediction

In our T cell epitope prediction work in Chapter 3, we identified highly conserved T cell epitopes from RSV surface proteins. Computational prediction of T cell epitope candidates has been used in many epitope identification and vaccine discovery studies in recent years [211, 212]. Indeed, computational-based pre-screening peptides on the basis of predicted MHC binding affinity can save valuable time and resources needed for the epitope identification process [211, 212]. While there have been continuous improvements in the algorithmic performance of MHC binding predictions, there are several considerations that may threaten or limit the success of *in silico* T cell epitope predictions [213]. For example, *in silico* T cell epitope prediction relies on the HLA supertypes that share epitope binding similarities [214]. However, there is extensive binding promiscuity that goes well beyond the specific allele. Its assigned supertype and for which commonly used minimal cutoffs of binding affinities would not be able to identify reactive peptides [215, 213]. Another critical limitation that is not properly addressed by most epitope prediction tools is the need to identify appropriate antigen-processing sites that could give rise to the predicted epitopes that are presented [213]. Therefore, we further searched the IEDB epitope database in Chapter 3 to determine if these epitopes were related to experimentally validated RSV T cell epitopes or HLA ligands [216]. It is important to note that the T cell epitopes have not been previously described but

are predicted with high MHC affinity in our analyses and may play a role in RSV-induced immune response. Therefore, our RSV T cell epitope landscapes are constructed with *in-silico* T cell epitope prediction results. However, to further characterize the T cell epitope profile of RSV at a population level, an unbiased experimental validation for RSV-specific T cell responses with classical and alternative T-cell effector functions is still needed.

5.3 FUTURE WORK

Findings and strategies from RSV evolutionary and epidemiology studies from this dissertation carry important insights into vaccination schemes and prevention measures for RSV as well as other circulating pathogens. How to apply the findings from this dissertation to RSV vaccine development and optimizing vaccination strategy and how to come up with effective infectious disease molecular surveillance based on these new approaches will be discussed in this section.

5.3.1 RSV VACCINE DESIGN

Despite the significant burden of RSV infection worldwide, there is no licensed vaccine [13]. Traditional approaches have failed to produce stable and protective vaccines for hypervariable and rapidly evolving viral pathogens, including RSV [217, 218]. Fortunately, the growth of databases containing genome sequences sampled throughout global epidemics, increased computational power, and improved theoretical algorithms allow complex data sources to be integrated into a unified framework providing the opportunity for a more holistic understanding of the pathogen and host features. This makes computational approaches valuable in providing novel insights into vaccine selection and design.

A standardized nomenclature for virus sequence upload would facilitate curation and evolutionary analyses on databases such as GenBank (https://www.ncbi.nlm.nih.gov/genbank/) [219] or GISAID (www.gisaid.org) [220]. A consistent approach to genotyping would facilitate the study of genotype/phenotype relationships and also simplify the study

of the emergence, growth, replacement, and extinction of different virus strains in a global context. Two recent publications have moved towards a more consistent approach to viral genotyping but work remains to establish a system that is accepted and used by the RSV research community as a whole [28, 36]. In this dissertation, we propose another extendable nomenclature system for RSV in Chapter 2. Such work will be key in predicting and tracking whether genomics variability is likely to impact the effectiveness of vaccines and therapeutics [172].

Secondly, computational approaches to identify candidates for influenza vaccine design have been used with a variety of novel vaccine production strategies in development, including epitope-based design [217, 212]. While most current RSV vaccination strategies focus on a B-cell-induced neutralization immune response, T cell immunity also plays a major role in the resolution of virus infection and is essential for RSV vaccine development [60, 48]. In chapter 3, we examine the T cell epitope profile of current circulation RSV strains at a population level. We identified the T cell epitopes that are more conserved and less likely to induce human autoimmune responses in the RSV surface proteins using computation T cell epitope prediction tool from iVAX toolkit [118]. Though current RSV vaccine development is not based on computational-based approaches, we systematically evaluate the cross-immunity between RSV strains with previous candidate strains. Our work provides insights into the potential different efficacy of vaccines for different subtypes and genotypes of RSV viruses.

5.3.2 Development of molecular surveillance tool for emerging viruses

An important aspect of public health study is to monitor pathogen epidemiology and evolution, which required a quick phylogenetic inference from novel isolated genomic sequences incorporated with spatial/temporal content and other phenotype profiles. The most common mechanisms of genetic and accumulated antigenic changes in RNA viruses are high point mutation rates (usually cause “antigenic drift”), immune selection and migration [221]. While several systematic nomenclature systems and virus classification tools have been proposed

for well-characterized viruses [96, 222], significant efforts are still required to monitor and update these virus classification systems due to constant evolutionary pressure. Therefore, the novel and extendable RSV genotyping classification pipeline described in Chapter 2 and the sequence-based T cell epitope landscapes approach that was developed in Chapter 3 have the potential to be extended as molecular surveillance tools for multiple circulating pathogens.

The classification of viruses into genotypes is the first step in understanding the molecular epidemiology of emerging viruses and can inform the subsequent implementation of control measures. Not only the evolutionary and epidemiological patterns of previous and current circulating viruses can be better characterized with a strong virus nomenclature system, but also novel genotypes can be identified. In addition, it is important to develop automated pipelines to facilitate the rapid classification of the novel isolated virus with the defined nomenclature system. Tools for lineage assignment, such as the machine learning approaches have been used for lineage assignment for specific viruses using gene nucleotide sequences. For example, in Chapter 2, we provide an RSV-specific module to perform genotyping assignments with the existing lineage assignment platform "LABEL" [93].

It is important to provide insight into the infection risks of pathogens. Pathogen surveillance needs to assess the level of immune protection the population has against infections through immune surveillance [223]. The T cell epitope landscape approach that we developed in Chapter 3 can provide information on the T-cell immunity of RSV circulating strains before experimental validation. From a T cell landscape, we can relate pathogen genetic information to its immunological profile and identify mutations that have the possibility to become new genotypes. This approach has the potential to apply to the continuous monitoring of other circulating viruses with significant genetic variation.

In addition, surveillance relies on data or insight sharing among public health. The Nextstrain platform [95] is designed to integrate a database of pathogen genomes into an analysis pipeline of phylodynamic models and provides a flexibly interactive visualization [95]. In Chapter 2, we deployed the novel and extendable RSV nomenclature system that we

developed via the Nextstrain platform. Both the genetically derived genotype information and predicted T cell profile can be incorporated with this Nextstrain platform to share key information with the molecular surveillance community.

5.3.3 New platform to study co-circulation of multiple pathogens

There is a growing interest in understanding the nature and consequences of interactions among infectious agents [70]. Previous studies have demonstrated that pathogen-pathogen interactions can lead to cooperation or competition, and their co-occurrence within the same host population can profoundly alter the pathogenesis and spread of infectious diseases with substantial public health consequences [149]. In Chapter 4, we use genetic surveillance and weekly case reports to examine the potential interaction between RSV and seasonal influenza at a population level and its potential imprints on virus populations and observed epidemics. The approaches we used to study the interaction between RSV and influenza can be applied to other pathogens as well. For example, similar approaches could be used to investigate the disruption of seasonal transmission patterns of RSV and seasonal influenza under the ongoing COVID-19 pandemic [201, 224].

Previous studies have demonstrated that noisy time series incident data can contain sufficient information to allow correct inference of interactions in multi-pathogen systems [150]. In addition, phylodynamic models have the potential to provide insights into epidemiological processes that are difficult to assess through traditional surveillance means using genetic information [225]. In this dissertation, we examined the potential interaction between RSV and seasonal influenza via genetic data estimated effective population size. Previous studies have suggested phylodynamic approaches have the possibility of more directly revealing these otherwise hidden transmission patterns [226, 225]. For example, the phylogenetic framework described in the "PhyDyn" package can be used to estimate important epidemic features from genetic data, while genetic surveillance is required to make a consistent conclusion under this framework [227].

5.4 Conclusions

I applied multiple computational approaches in this dissertation. I characterized the genetic diversity of RSV with a novel nomenclature system in Chapter 2 using a phylogenetic approach. An automated RSV genotyping classification tool was also provided to assign new RSV isolates with genotypes based on the genetic similarity of the sequences that are included in our system without time-consuming phylogeny reconstruction and manual annotation. In Chapter 3 of this dissertation, I attempted to relate the high genetic variation of RSV to its potential T cell immune-profile diversity with computationally predicted T cell epitopes. In Chapter 4, mathematical modeling approaches were used to study the potential interaction between RSV and seasonal influenza. The results that are presented in each chapter enhance our understanding of RSV evolutionary and outbreak dynamics. Overall, this dissertation highlights the importance of applying computational and statistical strategies to help us have a better understanding of the evolution, transmission, and epidemiology of RSV and other circulating pathogens.

Appendix A

Chapter 4 Supplementary materials[1]

A.1 RSV and seasonal influenza surveillance data correlation test

We provide the weekly RSV and seasonal influenza in all 10 HHS regions in Figure A.1. In Figure 4.1B, we demonstrate the correlation coefficient of the weekly epidemic size of different pairs of pathogens across the U.S. In Table A.1, we provide the complete results of the correlation coefficient estimates including the 95% confidence interval (CI) in 10 HHS regions.

A.2 Additional results for population dynamics of RSV and seasonal influenza in the U.S.

We use genetic data to study the population dynamics of RSV and seasonal influenza in the main text. Here we provide additional results in Table A.2, Figure A.2, Figure A.3

Table A.2 is the correlation coefficients of the time series of effective population sizes estimated from genomic data for RSV and seasonal influenza (Figure 4.2A). This analysis is used to study the potential interaction of RSV and seasonal influenza by genetic diversity.

In Figure A.2, we estimate the basic reproduction number (R_0) of RSV and seasonal influenza using two approaches, either surveillance reports or Bayesian Skyride coalescent estimated effective population size dynamics. To calculate the R_0 using surveillance reports, we use a serial interval fitted to a gamma distribution of 3.2 days with a standard deviation

[1]Supplementary materials for
Jiani Chen, Deven V. Gokhale, Liang Liu Pejman Rohani, Justin Bahl, Characterizing Potential Interaction Between Respiratory Syncytial Virus and Seasonal Influenza in the U.S.
Submitted to *PNAS*, November 2022.

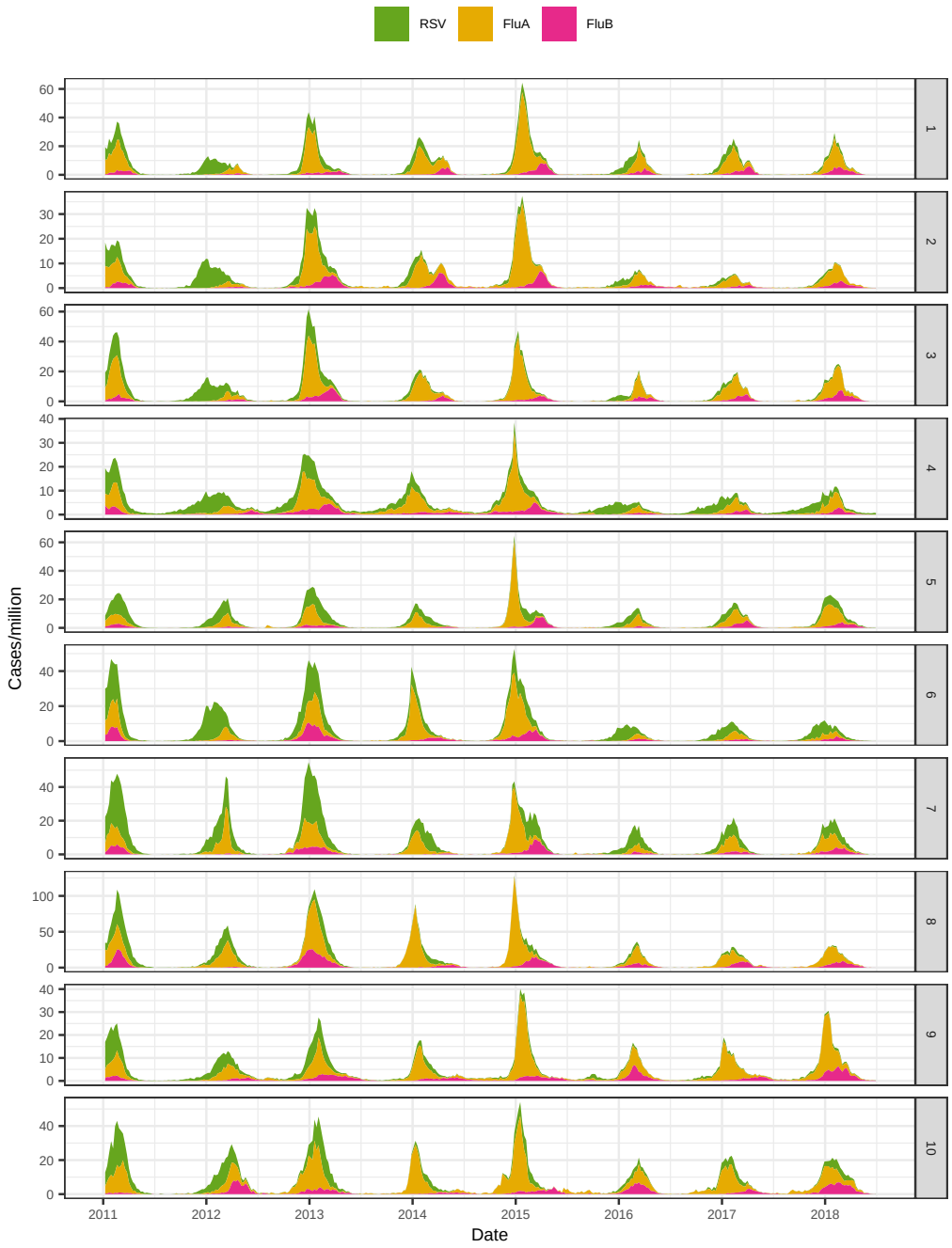

Figure A.1: **Weekly HHS regional-level positive cases of RSV and seasonal influenza.**

Table A.1: **Epidemic correlation between different pairs of viruses inferred from surveillance data.**

Pathogens	HHS region	Correlation coefficient (mean)	Correlation coefficient (95% CI low)	Correlation coefficient (95% CI high)
Flu A - Flu B	1	0.31	0.22	0.4
	2	0.23	0.13	0.31
	3	0.36	0.27	0.44
	4	0.41	0.33	0.49
	5	0.21	0.12	0.3
	6	0.56	0.49	0.62
	7	0.36	0.28	0.44
	8	0.43	0.35	0.5
	9	0.4	0.32	0.48
	10	0.23	0.14	0.31
RSV - Flu A	1	0.57	0.51	0.63
	2	0.38	0.3	0.46
	3	0.54	0.47	0.61
	4	0.48	0.41	0.55
	5	0.47	0.39	0.54
	6	0.57	0.5	0.63
	7	0.59	0.53	0.65
	8	0.4	0.32	0.48
	9	0.36	0.27	0.43
	10	0.59	0.52	0.65
RSV - Flu B	1	0.2	0.1	0.28
	2	0.1	0.01	0.19
	3	0.26	0.17	0.35
	4	0.37	0.28	0.45
	5	0.36	0.28	0.44
	6	0.66	0.61	0.71
	7	0.67	0.62	0.72
	8	0.62	0.56	0.68
	9	0.21	0.12	0.29
	10	0.27	0.18	0.35

of 0.35 for the generation time of RSV [228] and a gamma distribution of 2.6 days with a standard deviation of 1 day was used as the generation time for the R_0 estimation of Flu A and Flu B [229]. The "exponential growth" method that is implemented in R package "R_0" was used for the calculation. The Bayesian Skyride population dynamic based R_0 was calculated from [230], which allows the estimation of R_0 in genotype levels.

$$\Delta = \gamma(R_0 - 1) \tag{A.1}$$

where Δ is the exponential growth rate of the virus population and $1/\gamma$ is the mean duration of the infectious period, in this case, assumed to be 9 days (SD =3) for RSV and 3 (SD = 1.5) for Flu A /Flu B [231, 73]. To estimate Δ, we use linear regression to calculate the slope of the best-fit line of the exponential growth period for each season, which was chosen by maximizing the R-squared statistic of the linear regression. We only estimate R0 for the season with an observed peak and at least 3 points need to be included in the estimate of the exponential growth rate. As shown in Figure A.2, different genotypes of the same pathogens might have similar or different values of R_0 in each season, which suggests a complex epidemic dynamic at the genotype level for both RSV and seasonal influenza virus, and the virus interaction might play a role in this complex virus population dynamics.

In Figure 4.2C, we provide the time-sliced phylogenic statistics of RSV and seasonal influenza in different genotypes. Here, we additionally provide the mean statistics of RSV and seasonal influenza phylogenies in Figure A.3. Similar to the results as shown in sliced phylogenetic analysis, the mean statistic also suggests RSV lineages have a relatively slower coalescent rate and higher diversity in the analysis of compared with seasonal influenza. Tajima's D values of both RSV (ON and BA) and seasonal influenza virus (H1, H3, Victoria, and Yamagata) were negative, suggesting a rapid population growth in both RSV and seasonal influenza virus populations.

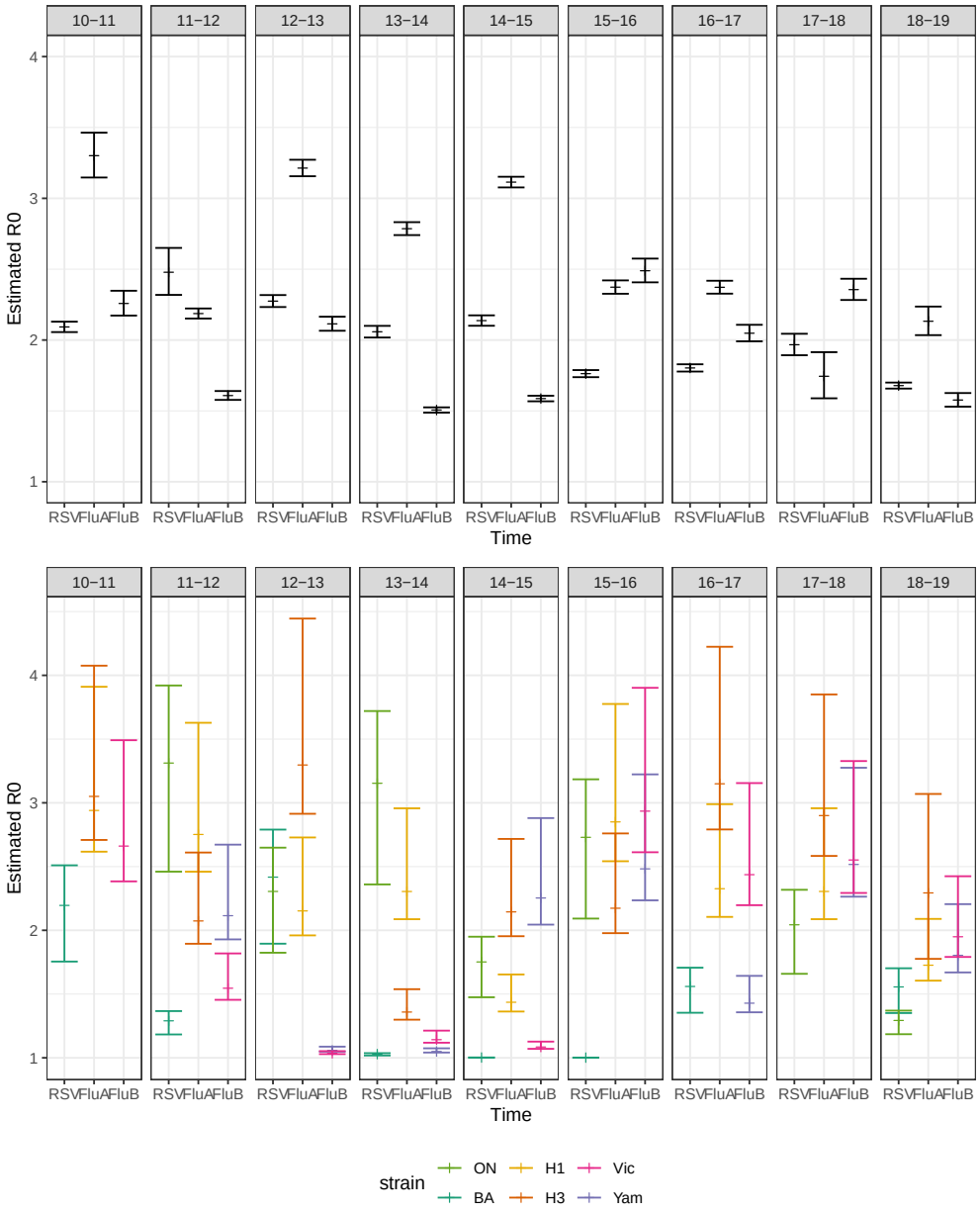

Figure A.2: **R0 estimation of RSV and seasonal influenza in the U.S. from 2010-2019.** Standard deviations are shown as vertical bars. (A) R_0 estimation from surveillance curve (B) R_0 estimation from GMRF Bayesian Skyride analysis.

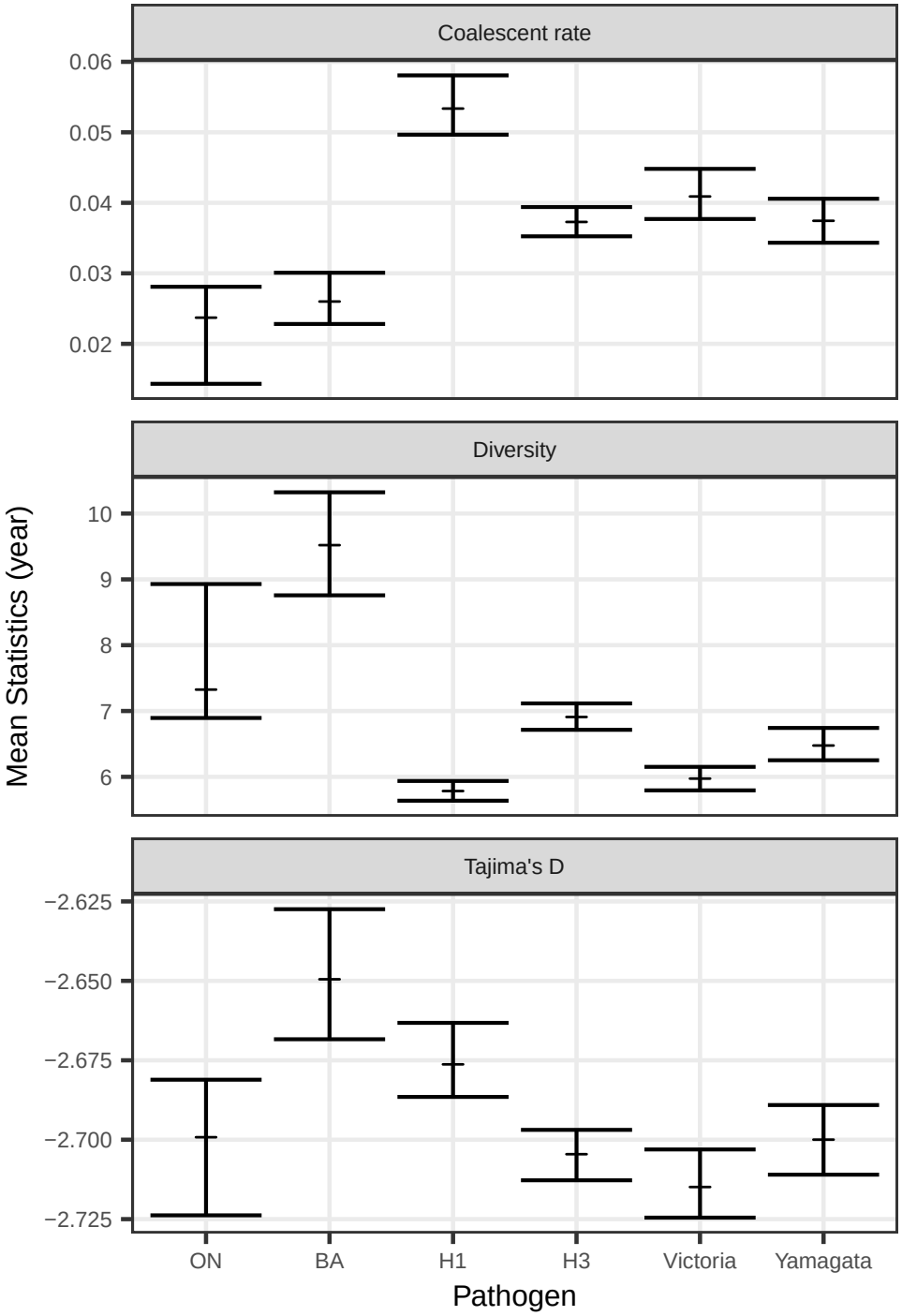

Figure A.3: **Mean phylogenetics statistics of RSV and seasonal influenza in different lineages.**

Table A.2: **Epidemic correlation between different pairs of viruses inferred from genetic data.**

Virus pairs	Virus 1	Virus 2	Correlation coefficient (mean)	Correlation coefficient (95% CI low)	Correlation coefficient (95% CI high)	P value
Flu - Flu	H3	H1	0.62	0.48	0.74	0
		Victoria	0.44	0.25	0.59	0
		Yamagata	0.34	0.14	0.51	0
	H1	Victoria	0.27	0.06	0.45	0.01
		Yamagata	0.55	0.38	0.68	0
	Victoria	Yamagata	0.19	-0.02	0.38	0.08
RSV - RSV	ON	BA	0.41	0.21	0.57	0
RSV - Flu	ON	H3	-0.33	-0.51	-0.13	0
		H1	-0.3	-0.48	-0.09	0.01
		Victoria	-0.3	-0.48	-0.09	0.01
		Yamagata	0.05	-0.17	0.26	0.66
	BA	H3	-0.14	-0.34	0.07	0.18
		H1	-0.34	-0.51	-0.14	0
		Victoria	-0.14	-0.33	0.07	0.2
		Yamagata	-0.41	-0.56	-0.22	0

A.3 Two-pathogen transmission model with competitive interaction

Here we proved details of the two-pathogen transmission model that is used to explore the potential competitive interaction between RSV and seasonal influenza [70]. Figure 4.4 illustrates the model structure, and the deterministic skeleton of this model can be described by the 16 equations as shown in A.2 to A.17. The parameters that are used to infer the interaction between RSV and seasonal influenza are described in Table A.3.

Equations to describe the deterministic skeleton of the two-pathogen transmission model with the potential for competition:

$$\frac{dX_{SS}}{dt} = \mu N + \omega_2 X_{RS} + \omega_1 X_{SR} - \lambda_1 X_{SS} - \lambda_2 X_{SS} - \mu X_{SS} \tag{A.2}$$

$$\frac{dX_{IS}}{dt} = \lambda_1 X_{SS} + \omega_2 X_{IR} - \gamma_1 X_{IS} - \psi\lambda_2 X_{IS} - \mu X_{IS} \tag{A.3}$$

$$\frac{dX_{CS}}{dt} = \gamma_1 X_{IS} + \omega_2 X_{CR} - \rho_1 X_{CS} - \chi\lambda_2 X_{CS} - \mu X_{CS} \tag{A.4}$$

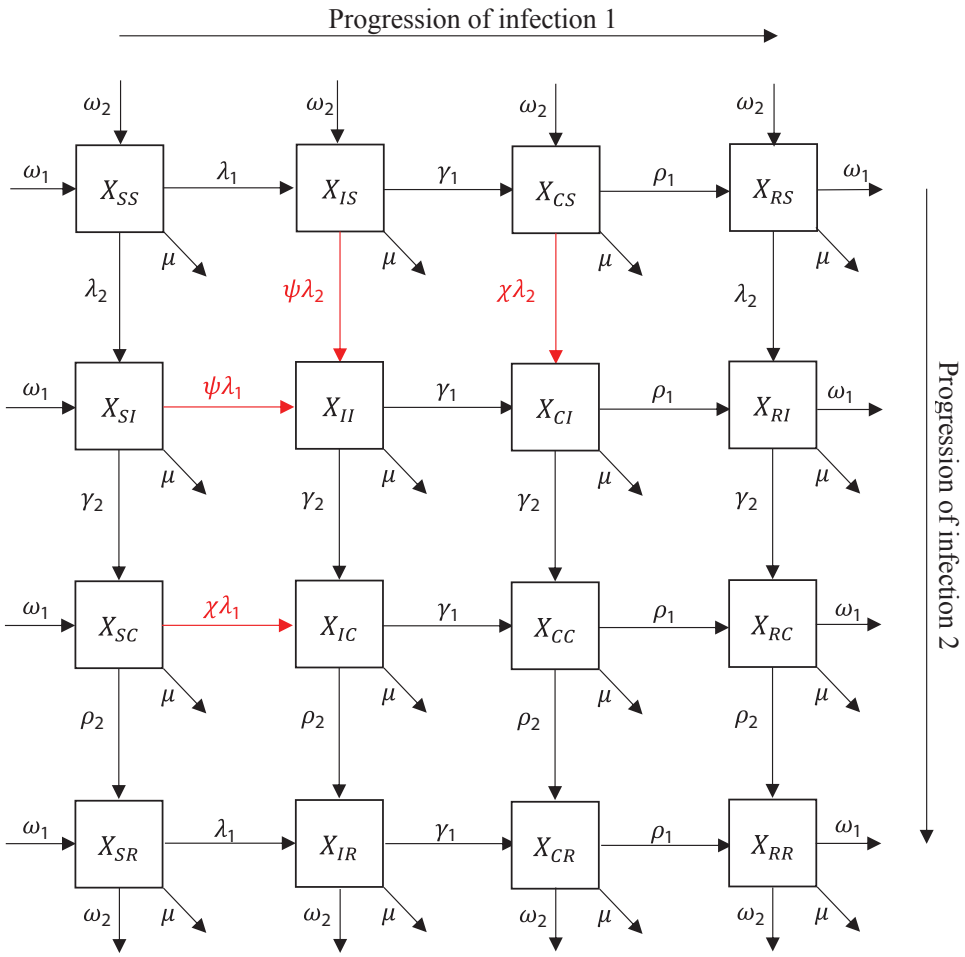

Figure A.4: **Schematics of a two-pathogen model with competitive interaction.** Each box represents a possible host state, with individual X_{i_i} indicating the states of the two pathogens, Susceptible (S), Infectious (I), Cross-protected (C), and Recovered (R). The horizontal arrows follow the progression of individuals' infection of the first pathogen, and the vertical arrows follow the progression of the second pathogen. λ indicates the infectious rate of the pathogen, γ indicates the recovery rate of the pathogen, ρ, the rate to lose cross-protected immunity, and ω, the rate to lost pathogen-specific immunity. The transitions denoted by the red arrow are affected by pathogen completion. ψ and χ modulate the force of infection of another pathogen by the individual in each of the I and C classes. The diagonal arrow represents disease-independent births and deaths. The mechanism of the models is detailed in the methodology and the parameters used are detailed in Table A.3.

$$\frac{dX_{RS}}{dt} = \rho_1 X_{CS} + \omega_2 X_{RR} - \lambda_2 X_{RS} - \omega_1 X_{RS} - \mu X_{RS} \tag{A.5}$$

$$\frac{dX_{SI}}{dt} = \omega_1 X_{RI} + \lambda_2 X_{SS} - \gamma_2 X_{SI} - \psi\lambda_1 X_{SI} - \mu X_{SI} \tag{A.6}$$

$$\frac{dX_{II}}{dt} = \psi\lambda_1 X_{SI} + \psi\lambda_2 X_{IS} - \gamma_1 X_{II} - \gamma_2 X_{II} - \mu X_{II} \tag{A.7}$$

$$\frac{dX_{CI}}{dt} = \gamma_1 X_{II} + \chi\lambda_2 X_{CS} - \rho_1 X_{CI} - \gamma_2 X_{CI} - \mu X_{CI} \tag{A.8}$$

$$\frac{dX_{RI}}{dt} = \rho_1 X_{CI} + \lambda_2 X_{Rs} + \gamma_1 X_{II} - \gamma_2 X_{RI} - \mu X_{RI} \tag{A.9}$$

$$\frac{dX_{SC}}{dt} = \omega_1 X_{RC} + \gamma_2 X_{SI} - \rho_2 X_{SC} - \chi\lambda_1 X_{SC} - \mu X_{SC} \tag{A.10}$$

$$\frac{dX_{IC}}{dt} = \chi\lambda_1 X_{SC} + \gamma_2 X_{II} - \gamma_1 X_{IC} - \rho_2 X_{IC} - \mu X_{IC} \tag{A.11}$$

$$\frac{dX_{CC}}{dt} = \gamma_1 X_{IC} + \gamma_2 X_{CI} - \rho_2 X_{CC} - \rho_1 X_{CC} - \mu X_{CC} \tag{A.12}$$

$$\frac{dX_{RC}}{dt} = \rho_1 X_{CC} + \gamma_2 X_{RI} - \omega_1 X_{RC} - \rho_2 X_{RC} - \mu X_{RC} \tag{A.13}$$

$$\frac{dX_{SR}}{dt} = \rho_2 X_{SC} + \omega_1 X_{RR} - \lambda_1 X_{SR} - \omega X_{SR} - \mu X_{SR} \tag{A.14}$$

$$\frac{dX_{IR}}{dt} = \lambda_1 X_{SR} + \rho_2 X_{IC} - \gamma_1 X_{IR} - \omega X_{IR} - \mu X_{IR} \tag{A.15}$$

$$\frac{dX_{CR}}{dt} = \gamma_1 X_{IR} + \rho_2 X_{CC} - \omega_2 X_{CR} - \rho_1 X_{CR} - \mu X_{CR} \tag{A.16}$$

$$\frac{dX_{RR}}{dt} = \rho_1 X_{CR} + \rho_2 X_{RC} - \omega_1 X_{RR} - \omega_2 X_{RR} - \mu X_{RR} \tag{A.17}$$

A.4 Complete results of RSV and seasonal competition evaluation in 10 HHS regions

We use the result of HHS1 as a representative to illustrate the potential competition interaction of RSV and seasonal influenza in the main text. Here we provide the complete results of 10 HHS regions in the U.S. The relative model fits among four epidemiological hypotheses (Table A.4) for the season 2014-2017 in 10 HHS regions are shown in Figure A.5. The models with different hypotheses were compared with AIC value and relative model fits in 10 HHS regions that are further quantified with RMSE as shown in Table A.5. In Figure A.6, we illustrate the model fitting in 10 HHS regions using the best hypothesis (lowest AIC value) with the

Table A.3: **Parameters that are used in two-pathogen transmission model.**

Symbol	Value	Definition
N		Total population in each HHS region
R_0^{RSV}	[0.5,10]	Reproductive number of RSV
R_0^{Flu}	[0.5,10]	Reproductive number of Flu A/ Flu B
$1/\gamma_{Flu}$	9	Infectious period of RSV
$1/\gamma_{Flu}$	3	Infectious period of Flu A/ Flu B
$1/\rho_{Flu}$	180/365.25	Cross protection period after RSV infection
$1/\rho_{Flu}$	180/365.25	Cross protection period after Flu infection
$1/\omega_{RSV}$	1y	Duration of RSV immunity
$1/\omega_{Flu}$	1y	Duration of Flu immunity
b_{RSV}	[0,1]	The amplitude of seasonality for RSV
b_{Flu}	[0,1]	The amplitude of seasonality for Flu
t_0^{RSV}	[1/365.25,1]	Timing of seasonal peak for RSV
t_0^{Flu}	[1/365.25,1]	Timing of seasonal peak for Flu
δ_{RSV}	(0,1]	Case report rate for RSV
δ_{Flu}	(0,1]	Case report rate for Flu
ψ	[0,1]	Parameter to modulate the infectious rate by inhibition of co-infection
χ	[0,1]	Parameter to modulate the infectious rate by short-term cross-immunity
μ	1/80	Birth and death rate

simulated trajectory from 2014 to 2017, and we further predict the epidemic curve in the following 2018 season using these estimated parameters. We detailed all parameter estimates and the relative model fits in 10 HHS regions from Table A.6 to Table A.15, respectively.

A.5 Likelihood profiles for competitive interaction parameters in 10 HHS regions

We show the likely Likelihood profiles for the proportion of inhibition on co-inhibition (ψ) and strength of cross-immunity (χ) in HHS region1 in Figure 4.5 as an example to illustrate the nature of competitive interaction parameters under the best hypothesis. Here we show the complete results of the likelihood profile of competitive interaction parameters in 10 HHS regions in Figure A.7. We only estimate the likelihood profile of estimated parameters is not zero under the best-supported hypothesis. (ψ in HHS regions 1, 3, 4, 5, 6, for the analysis of

Table A.4: **Epidemiological hypotheses formulation.**

Hypothesis name	Parameters
No-interaction	$\psi = 0; \chi = 0;\ 1/\rho_{RSV} = 1/\rho_{Flu} = 180/360$
Inhibition of co-infection	$\psi\epsilon(0,1]; \chi = 0;\ 1/\rho_{RSV} = 1/\rho_{Flu} = 180/360$
Cross-immunity after infection	$\psi = 0; \chi\epsilon(0,1];\ 1/\rho_{RSV} = 1/\rho_{Flu} = 180/360$
Inhibition of co-infection and cross-immunity after infection	$\psi\epsilon(0,1]; \chi\epsilon(0,1];\ 1/\rho_{RSV} = 1/\rho_{Flu} = 180/360$

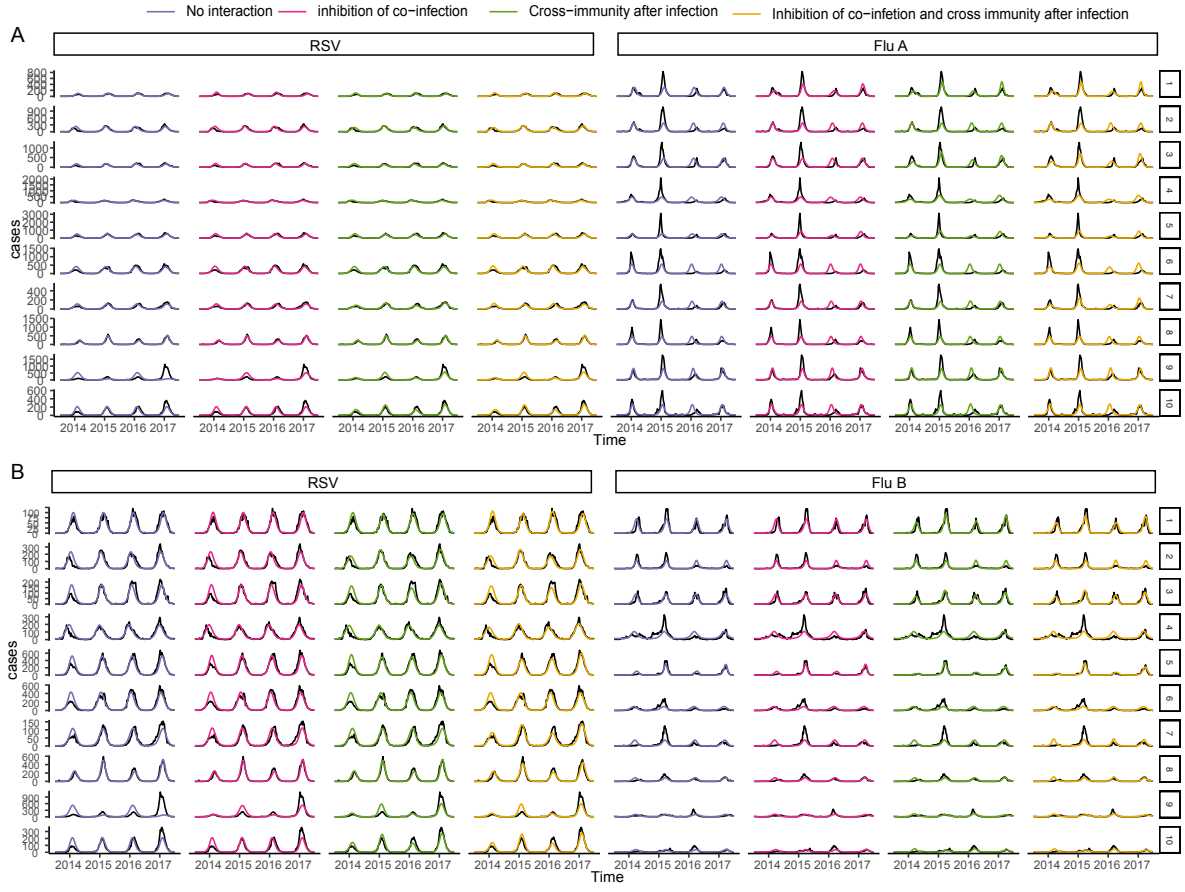

Figure A.5: **Relative fits between four epidemiological hypotheses for season 2014-2017 in 10 HHS regions.** Matched virus-specific simulated trajectories under four hypotheses are shown in different colors. The black solid line represented the HHS regional-level weekly case report data. (A) The test of the interaction between RSV and Flu A. (B) The test of the interaction between RSV and Flu B.

Table A.5: ΔAIC **and goodness of fit among four epidemiological hypotheses for season 2014-2017 in 10 HHS regions.**

HHS region	Parameter	RSV-Flu A				RSV-Flu B			
		No -interaction	Inhibition of co -infection	Cross -immunity after infection	Inhibition of co -infection and cross -immunity after infection	No -interaction	Inhibition of co -infection	Cross -immunity after infection	Inhibition of co -infection and cross -immunity after
1	AIC	9908.74	7951.49	7364.22	6946.68	3025.7	3000.32	2725.99	2773.13
	ΔAIC	2962.06	1004.81	417.54	0	299.71	274.33	0	47.14
	RMSE	62.47	52.51	50.36	49.3	11.14	11.28	11.03	12.06
2	AIC	16479.77	16345.37	16340.83	16342.83	16479.77	16345.37	16340.83	16342.83
	ΔAIC	138.94	4.54	0	2	138.94	4.54	0	2
	RMSE	91.62	91.38	91.28	91.28	91.62	91.38	91.28	91.28
3	AIC	17078.88	16623.49	14230.38	13351.05	4915.46	4887.51	4616.48	4604.05
	ΔAIC	3727.83	3272.44	879.33	0	311.41	283.46	12.43	0
	RMSE	97.2	96.2	85.9	85.37	28.88	28.9	29.12	28.09
4	AIC	25385.76	25261.83	23197.95	23190.48	9258.08	9217.32	9347.18	9195.17
	ΔAIC	2195.27	2071.35	7.47	0	62.91	22.15	152.01	0
	RMSE	134.76	134.29	126.89	126.79	28.06	28.69	28.82	27.36
5	AIC	35364.97	30144.32	29693.37	29072.77	8915.13	8853.34	6966.77	7363.08
	ΔAIC	6292.2	1071.55	620.6	0	1948.36	1886.57	0	396.31
	RMSE	203.48	191.87	191.2	189.74	71.44	73.65	68.27	74.36
6	AIC	28194.92	28101.06	28196.92	27758.76	9947.21	9942.01	8455.16	9303.85
	ΔAIC	436.16	342.3	438.16	0	1492.05	1486.85	0	848.69
	RMSE	143.43	143.41	143.43	142.85	63.08	63.11	62.5	63.86
7	AIC	9010.22	8931.1	8705.66	8214.94	4764.54	4866.03	4762.08	4532.76
	ΔAIC	795.27	716.16	490.72	0	231.78	333.27	229.32	0
	RMSE	42.27	42.22	41.73	39.6	18.65	19.19	18.55	14.86
8	AIC	16750.04	16752.05	16752.05	16754.06	5368.68	5370.69	5370.69	6629.86
	ΔAIC	0	2	2	4.01	0	2	2	1261.17
	RMSE	99.56	99.56	99.56	99.56	30.39	30.39	30.39	37.8
9	AIC	23156.86	23234.74	21196.38	21212.31	16106.29	15807.21	12999.64	12993.28
	ΔAIC	1960.47	2038.35	0	15.92	3113.01	2813.93	6.37	0
	RMSE	197.4	150.03	146.67	146.81	222.64	124.08	113.82	113.82
10	AIC	10755.37	10722.05	10109.81	10113.13	5658.38	5658.44	4641.26	4647.48
	ΔAIC	645.56	612.24	0	3.32	1017.12	1017.18	0	6.21
	RMSE	50.95	50.81	48.38	48.41	40.23	40.24	30.84	30.95

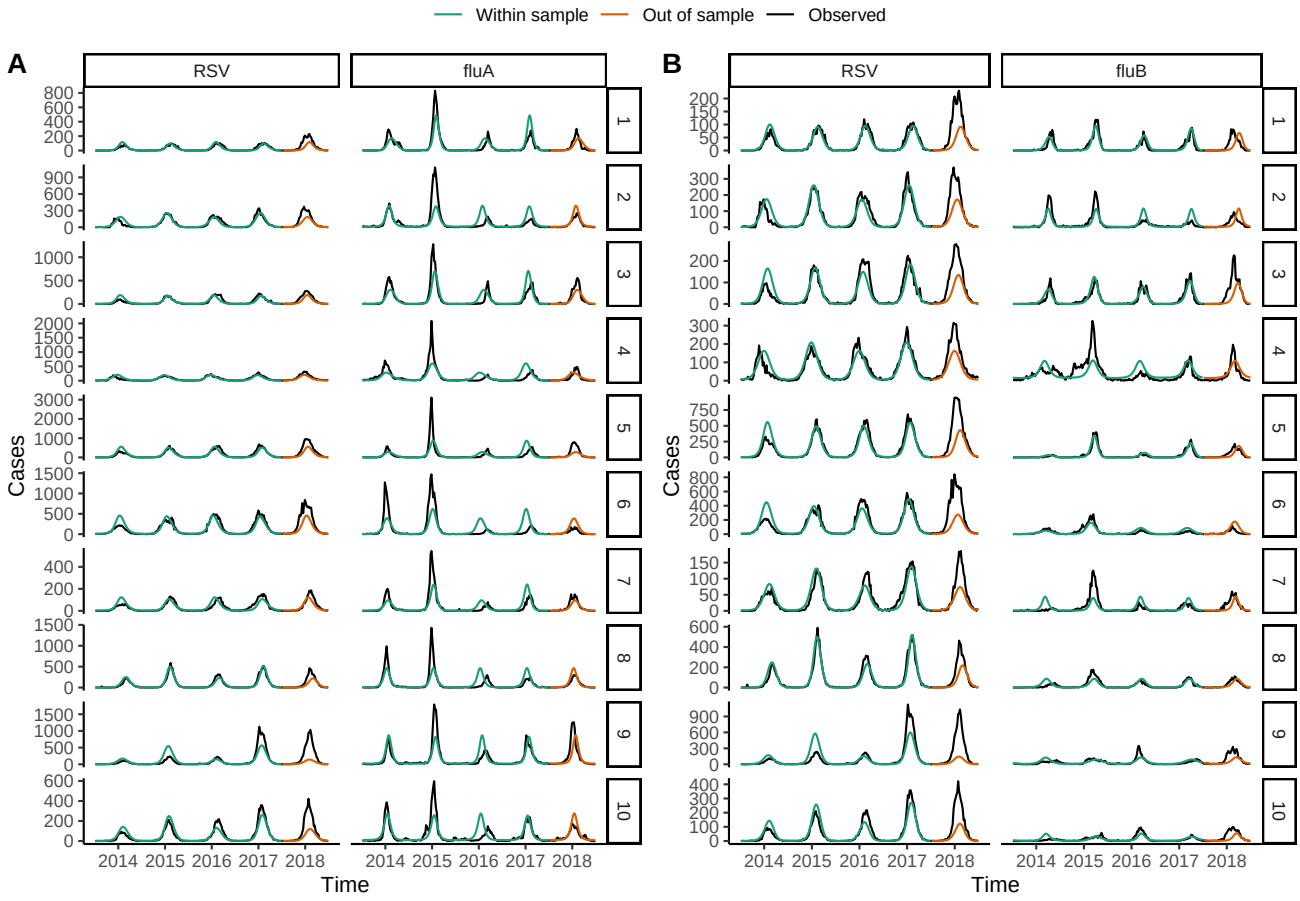

Figure A.6: **Relative fits using the best fit hypothesis for seasons 2014-2018 in 10 HHS regions.** Plots from 2014-2017 show matched pathogen-specific simulated trajectories (green) and plots from 2018 show the prediction of the 2018 season (orange). The black solid line represented the HHS regional-level weekly case report data. (A) The test of the interaction between RSV and Flu A. (B) The test of the interaction between RSV and Flu B.

Table A.6: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 1.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.49	1.46	1.84	1.89
R_0^{Flu}	1.37	1.05	1.31	1.27
b_{RSV}	0.26	0.33	0.13	0.15
b_{Flu}	0.1	0.27	0.17	0.23
t_0^{RSV}	0.91	0.93	0.98	0.99
t_0^{Flu}	0.1	0.03	0.12	0.11
ψ	1	0	1	0
χ	1	1	0.47	0.55
δ_{RSV}	2.49E-04	2.45E-04	2.07E-04	1.97E-04
δ_{Flu}	6.49E-04	1.07E-03	8.94E-04	9.34E-04
loglik	-4946.37	-3966.74	-3673.11	-3463.34
AIC	9908.74	7951.49	7364.22	6946.68
ΔAIC	2962.06	1004.81	417.54	0
RMSE	62.47	52.51	50.36	49.3
RSV-Flu B				
R_0^{RSV}	1.49	1.49	1.34	1.36
R_0^{Flu}	1	1	1	1.09
b_{RSV}	0.26	0.26	0.29	0.3
b_{Flu}	0.28	0.32	0.38	0.32
t_0^{RSV}	0.91	0.92	0.9	0.9
t_0^{Flu}	0.14	0.14	0.16	0.16
ψ	1	0	1	0.02
χ	1	1	0.64	0.61
δ_{RSV}	2.49E-04	2.49E-04	3.33E-04	3.17E-04
δ_{Flu}	2.69E-04	2.69E-04	2.74E-04	2.47E-04
loglik	-1504.85	-1491.16	-1353.99	-1376.56
AIC	3025.7	3000.32	2725.99	2773.13
ΔAIC	299.71	274.33	0	47.14
RMSE	11.14	11.28	11.03	12.06

Table A.7: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 2.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.54	1.56	1.19	1.19
R_0^{Flu}	1.6	1.68	1.64	1.64
b_{RSV}	0.21	0.21	0.28	0.28
b_{Flu}	0.08	0.1	0.09	0.09
t_0^{RSV}	0.82	0.83	0.81	0.81
t_0^{Flu}	0.25	0.23	0.25	0.25
ψ	1	0	1	1
χ	1	1	0.69	0.69
δ_{RSV}	2.78E-04	2.73E-04	7.46E-04	7.46E-04
δ_{Flu}	3.29E-04	3.12E-04	3.30E-04	3.30E-04
loglik	-8231.89	-8163.69	-8161.42	-8161.42
AIC	16479.77	16345.37	16340.83	16342.83
ΔAIC	138.94	4.54	0	2
RMSE	91.62	91.38	91.28	91.28
RSV-Flu B				
R_0^{RSV}	1.54	1.54	1.19	1.19
R_0^{Flu}	1.68	1.68	1.74	1.74
b_{RSV}	0.21	0.21	0.31	0.31
b_{Flu}	0.1	0.1	0.12	0.12
t_0^{RSV}	0.82	0.82	0.8	0.8
t_0^{Flu}	0.44	0.43	0.43	0.43
ψ	2.78E-04	2.78E-04	7.43E-04	7.43E-04
χ	8.40E-05	8.42E-05	8.24E-05	8.24E-05
δ_{RSV}	1	0.51	1	1
δ_{Flu}	1	1	0.72	0.72
loglik	-3371.65	-3370.86	-3217.91	-3217.91
AIC	6759.3	6759.71	6453.82	6455.82
ΔAIC	305.48	305.89	0	2
RMSE	36.17	36.21	30.82	30.82

Table A.8: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 3.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.54	1.98	1.28	1.87
R_0^{Flu}	1.23	1.27	1.15	1.26
b_{RSV}	0.21	0.11	0.25	0.11
b_{Flu}	0.1	0.13	0.19	0.19
t_0^{RSV}	0.85	0.95	0.81	0.92
t_0^{Flu}	0.97	0.01	0	0.04
ψ	1	0	1	0
χ	1	1	0.37	0.59
δ_{RSV}	2.15E-04	1.58E-04	3.96E-04	1.76E-04
δ_{Flu}	7.15E-04	6.65E-04	1.00E-03	8.74E-04
loglik	-8531.44	-8302.75	-7106.19	-6665.53
AIC	17078.88	16623.49	14230.38	13351.05
ΔAIC	3727.83	3272.44	879.33	0
RMSE	97.2	96.2	85.9	85.37
RSV-Flu B				
R_0^{RSV}	1.54	1.54	1.19	1.17
R_0^{Flu}	1	1	1.08	1.08
b_{RSV}	0.21	0.21	0.31	0.32
b_{Flu}	0.21	0.23	0.2	0.21
t_0^{RSV}	0.85	0.85	0.83	0.83
t_0^{Flu}	0.09	0.08	0.1	0.1
ψ	1	0	1	0
χ	1	1	0.75	0.8
δ_{RSV}	2.15E-04	2.14E-04	4.54E-04	4.71E-04
δ_{Flu}	2.95E-04	3.03E-04	2.21E-04	2.20E-04
loglik	-2449.73	-2434.76	-2299.24	-2292.02
AIC	4915.46	4887.51	4616.48	4604.05
ΔAIC	311.41	283.46	12.43	0
RMSE	28.88	28.9	28.12	28.09

Table A.9: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 4.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.45	1.46	1.83	1.82
R_0^{Flu}	1.42	1.48	1.26	1.25
b_{RSV}	0.16	0.16	0.07	0.08
b_{Flu}	0.03	0.04	0.08	0.08
t_0^{RSV}	0.76	0.76	0.83	0.83
t_0^{Flu}	0.13	0.12	0.02	1
ψ	1	0	1	0.57
χ	1	1	0.32	0.37
δ_{RSV}	1.60E-04	1.59E-04	1.17E-04	1.17E-04
δ_{Flu}	3.39E-04	3.16E-04	1.09E-03	1.10E-03
loglik	-12684.88	-12621.91	-11589.98	-11585.24
AIC	25385.76	25261.83	23197.95	23190.48
ΔAIC	2195.27	2071.35	7.47	0
RMSE	134.76	134.29	126.89	126.79
RSV-Flu B				
R_0^{RSV}	1.45	1.63	1.13	1.17
R_0^{Flu}	1.6	1.68	1.6	1.79
b_{RSV}	0.16	0.12	0.24	0.23
b_{Flu}	0.05	0.05	0.05	0.08
t_0^{RSV}	0.76	0.78	0.74	0.74
t_0^{Flu}	0.38	0.36	0.38	0.38
ψ	1	0	1	0
χ	1	1	1	0.87
δ_{RSV}	1.60E-04	1.30E-04	4.15E-04	3.94E-04
δ_{Flu}	8.30E-05	7.87E-05	8.30E-05	7.23E-05
loglik	-4621.04	-4599.66	-4664.59	-4587.58
AIC	9258.08	9217.32	9347.18	9195.17
ΔAIC	62.91	22.15	152.01	0
RMSE	28.06	28.69	28.82	27.36

Table A.10: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 5.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.56	1.58	1.8	1.79
R_0^{Flu}	1.31	1.09	1.19	1.18
b_{RSV}	0.21	0.21	0.14	0.14
b_{Flu}	0.05	0.16	0.09	0.11
t_0^{RSV}	0.86	0.87	0.91	0.9
t_0^{Flu}	0.94	0.92	0.94	0.96
ψ	1	0	1	0
χ	1	1	0.7	0.78
δ_{RSV}	3.72E-04	3.63E-04	3.16E-04	3.14E-04
δ_{Flu}	4.74E-04	8.45E-04	7.93E-04	8.14E-04
loglik	-17674.49	-15063.16	-14837.68	-14526.39
AIC	35364.97	30144.32	29693.37	29072.77
ΔAIC	6292.2	1071.55	620.6	0
RMSE	203.48	191.87	191.2	189.74
RSV-Flu B				
R_0^{RSV}	1.56	1.69	1.52	1.22
R_0^{Flu}	1	1	1.07	1.04
b_{RSV}	0.21	0.17	0.22	0.31
b_{Flu}	0.3	0.33	0.26	0.24
t_0^{RSV}	0.86	0.88	0.86	0.84
t_0^{Flu}	0.11	0.1	0.13	0.11
ψ	1	0	1	0
χ	1	1	0.61	0.69
δ_{RSV}	3.72E-04	3.30E-04	3.98E-04	7.08E-04
δ_{Flu}	2.00E-04	2.05E-04	2.84E-04	2.79E-04
loglik	-4449.57	-4417.67	-3474.39	-3671.54
AIC	8915.13	8853.34	6966.77	7363.08
ΔAIC	1948.36	1886.57	0	396.31
RMSE	71.44	73.65	68.27	74.36

Table A.11: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 6.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.06	1.53	1.06	1.75
R_0^{Flu}	1.44	1.44	1.44	1.18
b_{RSV}	0.33	0.2	0.33	0.14
b_{Flu}	0.06	0.08	0.06	0.14
t_0^{RSV}	0.81	0.83	0.81	0.87
t_0^{Flu}	0.05	0.05	0.05	0.93
ψ	1	0	1	0
χ	1	1	1	0.77
δ_{RSV}	1.73E-03	4.33E-04	1.73E-03	3.64E-04
δ_{Flu}	4.14E-04	4.17E-04	4.14E-04	8.08E-04
loglik	-14089.46	-14041.53	-14089.46	-13869.38
AIC	28194.92	28101.06	28196.92	27758.76
ΔAIC	436.16	342.3	438.16	0
RMSE	143.43	143.41	143.43	142.85
RSV-Flu B				
R_0^{RSV}	1.52	1.52	1.28	1.51
R_0^{Flu}	1.01	1	1.14	1.11
b_{RSV}	0.19	0.19	0.26	0.19
b_{Flu}	0.12	0.14	0.11	0.15
t_0^{RSV}	0.82	0.82	0.81	0.82
t_0^{Flu}	1	0.01	0.06	0.05
ψ	1	0.15	1	1
χ	1	1	0.45	0.75
δ_{RSV}	4.36E-04	4.36E-04	7.03E-04	4.55E-04
δ_{Flu}	4.85E-04	5.54E-04	3.23E-04	2.76E-04
loglik	-4965.61	-4962.01	-4218.58	-4641.92
AIC	9947.21	9942.01	8455.16	9303.85
ΔAIC	1492.05	1486.85	0	848.69
RMSE	63.08	63.11	62.5	62.86

Table A.12: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 7.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.46	1.47	1.22	1.88
R_0^{Flu}	1.41	1.4	1.47	1.23
b_{RSV}	0.2	0.21	0.26	0.11
b_{Flu}	0.09	0.11	0.09	0.17
t_0^{RSV}	0.88	0.88	0.89	0.98
t_0^{Flu}	0.07	0.07	0.1	0.03
ψ	1	0	1	0
χ	1	1	0.74	0.52
δ_{RSV}	3.57E-04	3.55E-04	6.69E-04	2.67E-04
δ_{Flu}	3.71E-04	3.82E-04	3.66E-04	7.04E-04
loglik	-4497.11	-4456.55	-4343.83	-4097.47
AIC	9010.22	8931.1	8705.66	8214.94
ΔAIC	795.27	716.16	490.72	0
RMSE	42.27	42.22	41.73	39.6
RSV-Flu B				
R_0^{RSV}	1	1.46	1	1.24
R_0^{Flu}	1	1	1	1.63
b_{RSV}	0.35	0.2	0.35	0.24
b_{Flu}	0.22	0.22	0.24	0.12
t_0^{RSV}	0.88	0.88	0.89	0.86
t_0^{Flu}	0.08	0.08	0.09	0.32
ψ	1	1	1	0.95
χ	1	1	0.17	0.64
δ_{RSV}	1.51E-03	3.57E-04	2.18E-03	6.92E-04
δ_{Flu}	2.18E-04	2.18E-04	2.37E-04	7.89E-05
loglik	-2374.27	-2424.02	-2372.04	-2256.38
AIC	4764.54	4866.03	4762.08	4532.76
ΔAIC	231.78	333.27	229.32	0
RMSE	18.65	19.19	18.55	14.86

Table A.13: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 8.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.31	1.31	1.31	1.31
R_0^{Flu}	1.49	1.48	1.48	1.48
b_{RSV}	0.42	0.42	0.42	0.42
b_{Flu}	0.1	0.1	0.1	0.1
t_0^{RSV}	0.94	0.94	0.94	0.94
t_0^{Flu}	0.11	0.11	0.11	0.11
ψ	1	1	1	1
χ	1	1	1	1
δ_{RSV}	1.12E-03	1.12E-03	1.12E-03	1.12E-03
δ_{Flu}	1.07E-03	1.07E-03	1.07E-03	1.07E-03
loglik	-8367.02	-8367.02	-8367.02	-8367.03
AIC	16750.04	16752.05	16752.05	16754.06
ΔAIC	0	2	2	4.01
RMSE	99.56	99.56	99.56	99.56
RSV-Flu B				
R_0^{RSV}	1.31	1.31	1.31	1.26
R_0^{Flu}	1.21	1.21	1.21	1.67
b_{RSV}	0.42	0.42	0.42	0.33
b_{Flu}	0.1	0.1	0.1	0.13
t_0^{RSV}	0.94	0.94	0.94	0.92
t_0^{Flu}	0.14	0.14	0.14	0.34
ψ	1	1	1	0.99
χ	1	1	1	0.46
δ_{RSV}	1.12E-03	1.12E-03	1.11E-03	1.85E-03
δ_{Flu}	4.10E-04	4.12E-04	4.11E-04	2.53E-04
loglik	-2676.34	-2676.34	-2676.34	-3304.93
AIC	5368.68	5370.69	5370.69	6629.86
ΔAIC	0	2	2	1261.17
RMSE	30.39	30.39	30.39	37.8

Table A.14: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 9.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.31	1.31	1.31	1.31
R_0^{Flu}	1.49	1.48	1.48	1.48
b_{RSV}	0.42	0.42	0.42	0.42
b_{Flu}	0.1	0.1	0.1	0.1
t_0^{RSV}	0.94	0.94	0.94	0.94
t_0^{Flu}	0.11	0.11	0.11	0.11
ψ	1	1	1	1
χ	1	1	1	1
δ_{RSV}	1.12E-03	1.12E-03	1.12E-03	1.12E-03
δ_{Flu}	1.07E-03	1.07E-03	1.07E-03	1.07E-03
loglik	-8367.02	-8367.02	-8367.02	-8367.03
AIC	16750.04	16752.05	16752.05	16754.06
ΔAIC	0	2	2	4.01
RMSE	99.56	99.56	99.56	99.56
RSV-Flu B				
R_0^{RSV}	1.31	1.31	1.31	1.26
R_0^{Flu}	1.21	1.21	1.21	1.67
b_{RSV}	0.42	0.42	0.42	0.33
b_{Flu}	0.1	0.1	0.1	0.13
t_0^{RSV}	0.94	0.94	0.94	0.92
t_0^{Flu}	0.14	0.14	0.14	0.34
ψ	1	1	1	0.99
χ	1	1	1	0.46
δ_{RSV}	1.12E-03	1.12E-03	1.11E-03	1.85E-03
δ_{Flu}	4.10E-04	4.12E-04	4.11E-04	2.53E-04
loglik	-2676.34	-2676.34	-2676.34	-3304.93
AIC	5368.68	5370.69	5370.69	6629.86
ΔAIC	0	2	2	1261.17
RMSE	30.39	30.39	30.39	37.8

Table A.15: **Parameters estimates and epidemiological hypotheses evaluation for season 2014-2017 in HHS regions 10.**

Parameter	No interaction	Inhibition of co-infection	Cross-immunity after infection	Inhibition of co-infection and cross-immunity after infection
RSV-Flu A				
R_0^{RSV}	1.48	1.48	1.28	1.27
R_0^{Flu}	1.64	1.61	1.68	1.68
b_{RSV}	0.29	0.3	0.34	0.34
b_{Flu}	0.1	0.11	0.11	0.11
t_0^{RSV}	0.89	0.89	0.91	0.91
t_0^{Flu}	0.21	0.19	0.23	0.23
ψ	1	0.22	1	1
χ	1	1	0.5	0.5
δ_{RSV}	5.18E-04	5.19E-04	9.61E-04	9.71E-04
δ_{Flu}	5.09E+01	5.08E+01	4.84E+01	4.84E+01
loglik	-5369.69	-5352.02	-5045.91	-5046.57
AIC	10755.37	10722.05	10109.81	10113.13
ΔAIC	645.56	612.24	0	3.32
RMSE	50.95	50.81	48.38	48.41
RSV-Flu B				
R_0^{RSV}	1.48	1.48	1.28	1.27
R_0^{Flu}	1.14	1.2	1.59	1.63
b_{RSV}	0.29	0.29	0.35	0.35
b_{Flu}	0.08	0.09	0.11	0.11
t_0^{RSV}	0.89	0.89	0.87	0.88
t_0^{Flu}	0.12	0.12	0.31	0.33
ψ	1	0.29	1	0.71
χ	1	1	0.2	0.27
δ_{RSV}	5.18E-04	5.17E-04	1.03E-03	1.05E-03
δ_{Flu}	2.45E-04	2.11E-04	1.33E-04	1.25E-04
loglik	-2821.19	-2820.22	-2311.63	-2313.74
AIC	5658.38	5658.44	4641.26	4647.48
ΔAIC	1017.12	1017.18	0	6.21
RMSE	40.23	40.24	30.84	30.95

RSV and Flu A, in HHS regions 3, 4, 7, 9 in the analyses of RSV and Flu B, χ in HHS regions 1, 2, 3, 4, 5, 6, 7, 9, 10 for the analysis of RSV and both subtypes of seasonal influenza)

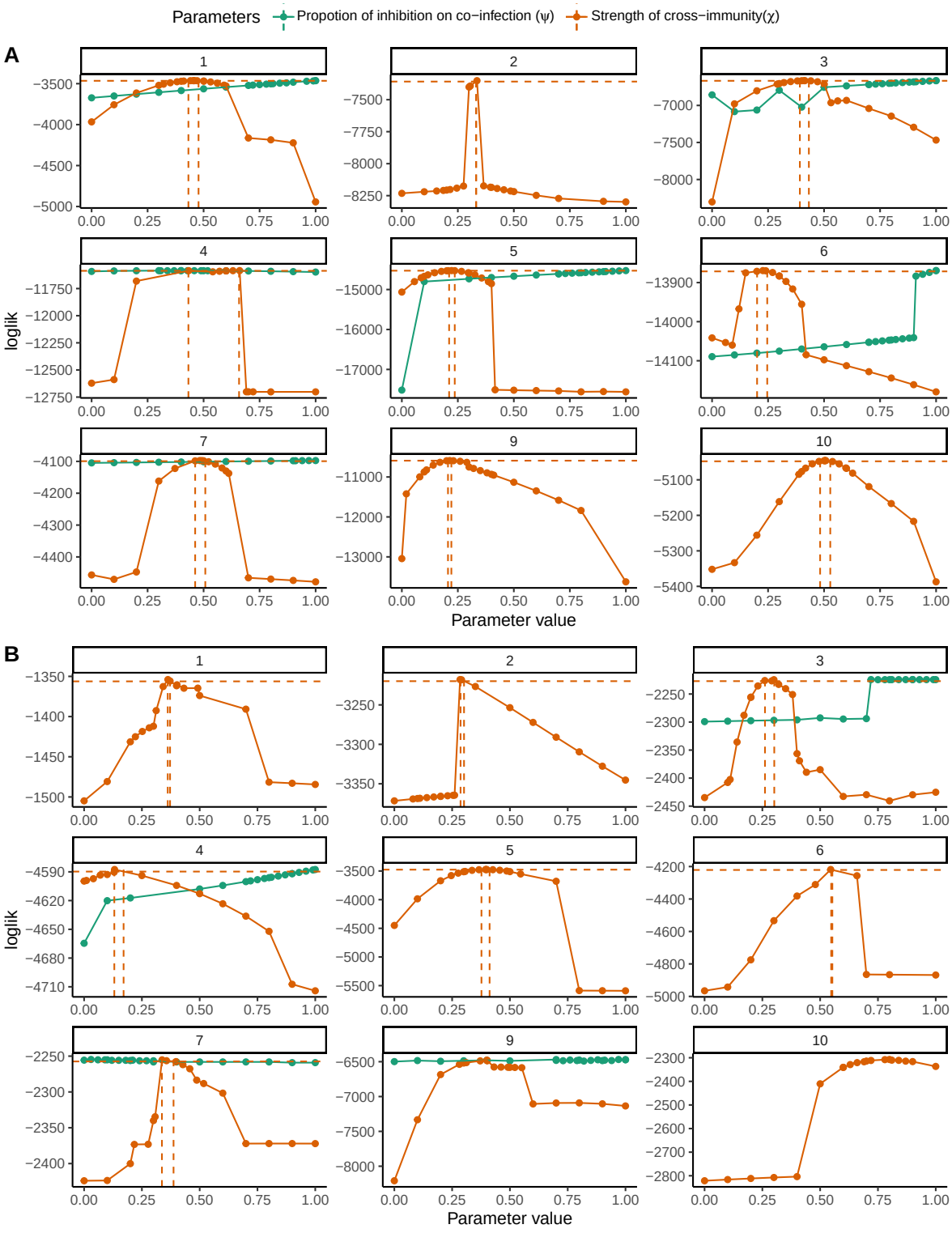

Figure A.7: **Likelihood profile tests of competition interaction parameters for RSV and seasonal influenza, inferred in 10 HHS regions from 2014 to 2017.** Plotted in each graph is the likelihood profile for the inhibition of co-inhibition (ψ) in green and the cross-immunity after infection (χ) in orange, which are lines connected by repeated likelihood estimates (N =20, shown in colored solid circles). The values within the two dashed lines are within the estimated 95% CI. (A) The test of the interaction between RSV and Flu A. (B) The test of the interaction between RSV and Flu B.

Bibliography

[1] J A Morris, R E Blount, and R E Savage. Recovery of cytopathogenic agent from chimpanzees with goryza. *Proceedings of the Society for Experimental Biology and Medicine*, 92(3):544–549, 7 1956. doi: 10.3181/00379727-92-22538.

[2] Katherine L O'Brien, Henry C Baggett, W Abdullah Brooks, Daniel R Feikin, Laura L Hammitt, Melissa M Higdon, Stephen R C Howie, Maria Deloria Knoll, Karen L Kotloff, Orin S Levine, Shabir A Madhi, David R Murdoch, Christine Prosperi, J Anthony G Scott, Qiyuan Shi, Donald M Thea, Zhenke Wu, Scott L Zeger, V Peter Adrian, Pasakorn Akarasewi, Trevor P Anderson, Martin Antonio, Juliet O Awori, Vicky L Baillie, Charatdao Bunthi, James Chipeta, Mohammod Jobayer Chisti, Jane Crawley, Andrea N DeLuca, Amanda J Driscoll, Bernard E Ebruke, Hubert P Endtz, Nicholas Fancourt, Wei Fu, Doli Goswami, Michelle J Groome, Meredith Haddix, Lokman Hossain, Yasmin Jahan, E Wangeci Kagucia, Alice Kamau, Ruth A Karron, Sidi Kazungu, Nana Kourouma, Locadiah Kuwanda, Geoffrey Kwenda, Mengying Li, Eunice M Machuka, Grant Mackenzie, Nasreen Mahomed, Susan A Maloney, Jessica L McLellan, Joanne L Mitchell, David P Moore, Susan C Morpeth, Azwifarwi Mudau, Lawrence Mwananyanda, James Mwansa, Micah Silaba Ominde, Uma Onwuchekwa, Daniel E Park, Julia Rhodes, Pongpun Sawatwong, Phil Seidenberg, Arifin Shamsul, Eric A F Simões, Seydou Sissoko, Somwe Wa Somwe, Samba O Sow, Mamadou Sylla, Boubou Tamboura, Milagritos D Tapia, Somsak Thamthitiwat, Aliou Toure, Nora L Watson, Khalequ Zaman, and Syed M A Zaman. Causes of severe pneumonia requiring hospital admission in children without hiv infection from africa and asia: the perch multi-country case-control study. *The Lancet*, 394(10200):757–779, 2019.

[3] Lorenz Schubert, Johanna Steininger, Felix Lötsch, Anna Nele Herdina, Monika Redlberger-Fritz, Selma Tobudic, Michael Kundi, Robert Strassl, and Christoph Steininger. Surveillance of respiratory syncytial virus infections in adults, austria, 2017 to 2019. *Scientific Reports*, 11(1):8939, 2021.

[4] Paul S McNamara and Rosalind L Smyth. The pathogenesis of respiratory syncytial virus disease in childhood. *British Medical Bulletin*, 61(1):13–28, 3 2002.

[5] Caroline Breese Hall, Geoffrey A Weinberg, Marika K Iwane, Aaron K Blumkin, Kathryn M Edwards, Mary A Staat, Peggy Auinger, Marie R Griffin, Katherine A Poehling, Dean Erdman, Carlos G Grijalva, Yuwei Zhu, and Peter Szilagyi. The burden of respiratory syncytial virus infection in young children. *New England Journal of Medicine*, 360(6):588–598, 2 2009. doi: 10.1056/NEJMoa0804877.

[6] Brian Rha, Aaron T Curns, Joana Y Lively, Angela P Campbell, Janet A Englund, Julie A Boom, Parvin H Azimi, Geoffrey A Weinberg, Mary A Staat, Rangaraj Selvarangan, Natasha B Halasa, Monica M McNeal, Eileen J Klein, Christopher J Harrison, V John Williams, Peter G Szilagyi, Monica N Singer, Leila C Sahni, Daniella Figueroa-Downing, Darius McDaniel, Mila M Prill, Brett L Whitaker, Laura S Stewart, Jennifer E Schuster, Barbara A Pahud, Gina Weddle, Vasanthi Avadhanula, Flor M Munoz, Pedro A Piedra, Daniel C Payne, Gayle Langley, and Susan I Gerber. Respiratory syncytial virus-associated hospitalizations among young children: 2015-2016. *Pediatrics*, 146(1), 7 2020.

[7] Rafael Lozano, Mohsen Naghavi, Kyle Foreman, Stephen Lim, Kenji Shibuya, Victor Aboyans, Jerry Abraham, Timothy Adair, Rakesh Aggarwal, Stephanie Y Ahn, Miriam Alvarado, H Ross Anderson, Laurie M Anderson, Kathryn G Andrews, Charles Atkinson, Larry M Baddour, Suzanne Barker-Collo, David H Bartels, Michelle L Bell, Emelia J Benjamin, Derrick Bennett, Kavi Bhalla, Boris Bikbov, Aref Bin Abdulhak, Gretchen Birbeck, Fiona Blyth, Ian Bolliger, Soufiane Boufous, Chiara Bucello, Michael Burch,

Peter Burney, Jonathan Carapetis, Honglei Chen, David Chou, Sumeet S Chugh, Luc E Coffeng, Steven D Colan, Samantha Colquhoun, K Ellicott Colson, John Condon, Myles D Connor, Leslie T Cooper, Matthew Corriere, Monica Cortinovis, Karen Courville de Vaccaro, William Couser, Benjamin C Cowie, Michael H Criqui, Marita Cross, Kaustubh C Dabhadkar, Nabila Dahodwala, Diego De Leo, Louisa Degenhardt, Allyne Delossantos, Julie Denenberg, Don C Des Jarlais, Samath D Dharmaratne, E Ray Dorsey, Tim Driscoll, Herbert Duber, Beth Ebel, Patricia J Erwin, Patricia Espindola, Majid Ezzati, Valery Feigin, Abraham D Flaxman, Mohammad H Forouzanfar, Francis Gerry R Fowkes, Richard Franklin, Marlene Fransen, Michael K Freeman, Sherine E Gabriel, Emmanuela Gakidou, Flavio Gaspari, Richard F Gillum, Diego Gonzalez-Medina, Yara A Halasa, Diana Haring, James E Harrison, Rasmus Havmoeller, Roderick J Hay, Bruno Hoen, Peter J Hotez, Damian Hoy, Kathryn H Jacobsen, Spencer L James, Rashmi Jasrasaria, Sudha Jayaraman, Nicole Johns, Ganesan Karthikeyan, Nicholas Kassebaum, Andre Keren, Jon-Paul Khoo, Lisa Marie Knowlton, Olive Kobusingye, Adofo Koranteng, Rita Krishnamurthi, Michael Lipnick, Steven E Lipshultz, Summer Lockett Ohno, Jacqueline Mabweijano, Michael F MacIntyre, Leslie Mallinger, Lyn March, Guy B Marks, Robin Marks, Akira Matsumori, Richard Matzopoulos, Bongani M Mayosi, John H McAnulty, Mary M McDermott, John McGrath, George A Mensah, Tony R Merriman, Catherine Michaud, Matthew Miller, Ted R Miller, Charles Mock, Ana Olga Mocumbi, Ali A Mokdad, Andrew Moran, Kim Mulholland, M Nathan Nair, Luigi Naldi, K M Venkat Narayan, Kiumarss Nasseri, Paul Norman, Martin O'Donnell, Saad B Omer, Katrina Ortblad, Richard Osborne, Doruk Ozgediz, Bishnu Pahari, Jeyaraj Durai Pandian, Andrea Panozo Rivero, Rogelio Perez Padilla, Fernando Perez-Ruiz, Norberto Perico, David Phillips, Kelsey Pierce, C Arden 3rd Pope, Esteban Porrini, Farshad Pourmalek, Murugesan Raju, Dharani Ranganathan, Jürgen T Rehm, David B Rein, Guiseppe Remuzzi, Frederick P Rivara, Thomas Roberts, Felipe Rodriguez De León, Lisa C Rosenfeld, Lesley Rushton,

Ralph L Sacco, Joshua A Salomon, Uchechukwu Sampson, Ella Sanman, David C Schwebel, Maria Segui-Gomez, Donald S Shepard, David Singh, Jessica Singleton, Karen Sliwa, Emma Smith, Andrew Steer, Jennifer A Taylor, Bernadette Thomas, Imad M Tleyjeh, Jeffrey A Towbin, Thomas Truelsen, Eduardo A Undurraga, N Venketasubramanian, Lakshmi Vijayakumar, Theo Vos, Gregory R Wagner, Mengru Wang, Wenzhi Wang, Kerrianne Watt, Martin A Weinstock, Robert Weintraub, James D Wilkinson, Anthony D Woolf, Sarah Wulf, Pon-Hsiu Yeh, Paul Yip, Azadeh Zabetian, Zhi-Jie Zheng, Alan D Lopez, Christopher J L Murray, Mohammad A AlMazroa, and Ziad A Memish. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet (London, England)*, 380(9859):2095–2128, 12 2012.

[8] Ann R Falsey, Patricia A Hennessey, Maria A Formica, Christopher Cox, and Edward E Walsh. Respiratory syncytial virus infection in elderly and high-risk adults. *New England Journal of Medicine*, 352(17):1749–1759, 4 2005. doi: 10.1056/NEJMoa043951.

[9] Ting Shi, Angeline Denouel, Anna K Tietjen, Iain Campbell, Emily Moran, Xue Li, Harry Campbell, Clarisse Demont, Bryan O Nyawanda, Helen Y Chu, Sonia K Stoszek, Anand Krishnan, Peter Openshaw, Ann R Falsey, and Harish Nair. Global disease burden estimates of respiratory syncytial virus–associated acute respiratory infection in older adults in 2015: A systematic review and meta-analysis. *The Journal of Infectious Diseases*, 222(Supplement7):S577–S583, 10 2020.

[10] Ting Shi, David A. McAllister, Katherine L. O'Brien, Eric A.F. Simoes, Shabir A. Madhi, Bradford D. Gessner, Fernando P. Polack, Evelyn Balsells, Sozinho Acacio, Claudia Aguayo, Issifou Alassani, Asad Ali, Martin Antonio, Shally Awasthi, Juliet O. Awori, Eduardo Azziz-Baumgartner, Henry C. Baggett, Vicky L. Baillie, Angel Balmaseda, Alfredo Barahona, Sudha Basnet, Quique Bassat, Wilma Basualdo, Godfrey Bigogo, Louis Bont, Robert F. Breiman, W. Abdullah Brooks, Shobha Broor, Nigel Bruce, Dana

Bruden, Philippe Buchy, Stuart Campbell, Phyllis Carosone-Link, Mandeep Chadha, James Chipeta, Monidarin Chou, Wilfrido Clara, Cheryl Cohen, Elizabeth de Cuellar, Duc Anh Dang, Budragchaagiin Dash-yandag, Maria Deloria-Knoll, Mukesh Dherani, Tekchheng Eap, Bernard E. Ebruke, Marcela Echavarria, Carla Cecília de Freitas Lázaro Emediato, Rodrigo A. Fasce, Daniel R. Feikin, Luzhao Feng, Angela Gentile, Aubree Gordon, Doli Goswami, Sophie Goyet, Michelle Groome, Natasha Halasa, Siddhivinayak Hirve, Nusrat Homaira, Stephen R.C. Howie, Jorge Jara, Imane Jroundi, Cissy B. Kartasasmita, Najwa Khuri-Bulos, Karen L. Kotloff, Anand Krishnan, Romina Libster, Olga Lopez, Marilla G. Lucero, Florencia Lucion, Socorro P. Lupisan, Debora N. Marcone, John P. McCracken, Mario Mejia, Jennifer C. Moisi, Joel M. Montgomery, David P. Moore, Cinta Moraleda, Jocelyn Moyes, Patrick Munywoki, Kuswandewi Mutyara, Mark P. Nicol, D. James Nokes, Pagbajabyn Nymadawa, Maria Tereza da Costa Oliveira, Histoshi Oshitani, Nitin Pandey, Gláucia Paranhos-Baccalà, Lia N. Phillips, Valentina Sanchez Picot, Mustafizur Rahman, Mala Rakoto-Andrianarivelo, Zeba A. Rasmussen, Barbara A. Rath, Annick Robinson, Candice Romero, Graciela Russomando, Vahid Salimi, Pongpun Sawatwong, Nienke Scheltema, Brunhilde Schweiger, J. Anthony G. Scott, Phil Seidenberg, Kunling Shen, Rosalyn Singleton, Viviana Sotomayor, Tor A. Strand, Agustinus Sutanto, Mariam Sylla, Milagritos D. Tapia, Somsak Thamthitiwat, Elizabeth D. Thomas, Rafal Tokarz, Claudia Turner, Marietjie Venter, Sunthareeya Waicharoen, Jianwei Wang, Wanitda Watthanaworawit, Lay Myint Yoshida, Hongjie Yu, Heather J. Zar, Harry Campbell, and Harish Nair. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *The Lancet*, 390(10098):946–958, 9 2017.

[11] M Tin Tin Htar, M S Yerramalla, J C Moïsi, and D L Swerdlow. The burden of respiratory syncytial virus in adults: a systematic review and meta-analysis. *Epidemiology and infection*, 148:e48–e48, 2 2020.

[12] Kimberly Bloom-Feshbach, Wladimir J Alonso, Vivek Charu, James Tamerius, Lone Simonsen, Mark A Miller, and Cécile Viboud. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (rsv): a global comparative review. *PloS one*, 8(2):e54445, 2013.

[13] Andrea T. Borchers, Christopher Chang, M. Eric Gershwin, and Laurel J. Gershwin. Respiratory syncytial virus - a comprehensive review. *Clinical Reviews in Allergy and Immunology*, 45(3):331–379, 12 2013.

[14] You Li, Rachel M Reeves, Xin Wang, Quique Bassat, W Abdullah Brooks, Cheryl Cohen, David P Moore, Marta Nunes, Barbara Rath, Harry Campbell, Harish Nair, Sozinho Acacio, Wladimir J Alonso, Martin Antonio, Guadalupe Ayora Talavera, Darmaa Badarch, Vicky L Baillie, Gisela Barrera-Badillo, Godfrey Bigogo, Shobha Broor, Dana Bruden, Philippe Buchy, Peter Byass, James Chipeta, Wilfrido Clara, Duc-Anh Dang, Carla Cecília de Freitas Lázaro Emediato, Menno de Jong, José Alberto Díaz-Quiñonez, Lien Anh Ha Do, Rodrigo A Fasce, Luzhao Feng, Mark J Ferson, Angela Gentile, Bradford D Gessner, Doli Goswami, Sophie Goyet, Carlos G Grijalva, Natasha Halasa, Orienka Hellferscee, Danielle Hessong, Nusrat Homaira, Jorge Jara, Kathleen Kahn, Najwa Khuri-Bulos, Karen L Kotloff, Claudio F Lanata, Olga Lopez, Maria Renee Lopez Bolaños, Marilla G Lucero, Florencia Lucion, Socorro P Lupisan, Shabir A Madhi, Omphile Mekgoe, Cinta Moraleda, Jocelyn Moyes, Kim Mulholland, Patrick K Munywoki, Fathima Naby, Thanh Hung Nguyen, Mark P Nicol, D James Nokes, Daniel E Noyola, Daisuke Onozuka, Nandhini Palani, Yong Poovorawan, Mustafizur Rahman, Kaat Ramaekers, Candice Romero, Elizabeth P Schlaudecker, Brunhilde Schweiger, Phil Seidenberg, Eric A F Simoes, Rosalyn Singleton, Sujatha Sistla, Katharine Sturm-Ramirez, Nungruthai Suntronwong, Agustinus Sutanto, Milagritos D Tapia, Somsak Thamthitiwat, Ilada Thongpan, Gayani Tillekeratne, Yeny O Tinoco, Florette K Treurnicht, Claudia Turner, Paul Turner, Rogier van Doorn, Marc Van Ranst, Benoit Visseaux, Sunthareeya Waicharoen, Jianwei Wang, Lay-Myint Yoshida, and Heather J

Zar. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. *The Lancet Global Health*, 7(8):e1031–e1045, 2019.

[15] Erica Billig Rose, Alexandra Wheatley, Gayle Langley, Susan Gerber, and Amber Haynes. Respiratory syncytial virus seasonality - united states, 2014-2017. *MMWR. Morbidity and mortality weekly report*, 67(2):71–76, 1 2018.

[16] Shobha Broor, Harry Campbell, Siddhivinayak Hirve, Siri Hague, Sandra Jackson, Ann Moen, Harish Nair, Rakhee Palekar, Soatiana Rajatonirina, Peter G Smith, Marietjie Venter, Niteen Wairagkar, Maria Zambon, Thedi Ziegler, and Wenqing Zhang. Leveraging the global influenza surveillance and response system for global respiratory syncytial virus surveillance-opportunities and challenges. *Influenza and other respiratory viruses*, 14(6):622–629, 11 2020.

[17] Rodica Gilca, Gaston De Serres, Mireille Tremblay, Marie-Louise Vachon, Eric Leblanc, Michel G Bergeron, Pierre Dery, and Guy Boivin. Distribution and clinical impact of human respiratory syncytial virus genotypes in hospitalized children over 2 winter seasons. *The Journal of infectious diseases*, 193(1):54–58, 1 2006.

[18] M S Imaz, M D Sequeira, C Videla, I Veronessi, R Cociglio, E Zerbini, and G Carballal. Clinical and epidemiologic characteristics of respiratory syncytial virus subgroups a and b infections in santa fe, argentina. *Journal of medical virology*, 61(1):76–80, 5 2000.

[19] T C Peret, J A Golub, L J Anderson, C B Hall, and K C Schnabel. Circulation patterns of genetically distinct group a and b strains of human respiratory syncytial virus in a community. *Journal of General Virology*, 79(9):2221–2229, 9 1998.

[20] Vicente Mas, Harish Nair, Harry Campbell, Jose A Melero, and Thomas C Williams. Antigenic and sequence variability of the human respiratory syncytial virus f glycoprotein

compared to related viruses in a comprehensive dataset. *Vaccine*, 36(45):6660–6673, 10 2018.

[21] Seth A Schobel, Karla M Stucker, Martin L Moore, Larry J Anderson, Emma K Larkin, Jyoti Shankar, Jayati Bera, Vinita Puri, Meghan H Shilts, Christian Rosas-Salazar, Rebecca A Halpin, Nadia Fedorova, Susmita Shrivastava, Timothy B Stockwell, R Stokes Peebles, V Tina Hartert, and Suman R Das. Respiratory syncytial virus whole-genome sequencing identifies convergent evolution of sequence duplication in the c-terminus of the g gene. *Scientific reports*, 6:26311, 2016.

[22] Jinhua Song, Huiling Wang, Jing Shi, Aili Cui, Yanzhi Huang, Liwei Sun, Xingyu Xiang, Chaofeng Ma, Pengbo Yu, Zifeng Yang, Qi Li, Teresa I Ng, Yan Zhang, Rongbo Zhang, and Wenbo Xu. Emergence of ba9 genotype of human respiratory syncytial virus subgroup b in china from 2006 to 2014. *Scientific Reports*, 7(1):16765, 2017.

[23] Guanglin Cui, Yuan Qian, Runan Zhu, Jie Deng, Linqing Zhao, Yu Sun, and Fang Wang. Emerging human respiratory syncytial virus genotype on1 found in infants with pneumonia in beijing, china. *Emerging microbes infections*, 2(4):e22–e22, 4 2013.

[24] Julia Tabatabai, Christiane Prifert, Johannes Pfeil, Jürgen Grulich-Henn, and Paul Schnitzler. Novel respiratory syncytial virus (rsv) genotype on1 predominates in germany during winter season 2012–13. *PLOS ONE*, 9(10):e109191, 10 2014.

[25] Jason S McLellan, William C Ray, and Mark E Peeples. Structure and function of respiratory syncytial virus surface glycoproteins. *Current topics in microbiology and immunology*, 372:83–104, 2013.

[26] Anne M Hause, David M Henke, Vasanthi Avadhanula, Chad A Shaw, Lorena I Tapia, and Pedro A Piedra. Sequence variability of the respiratory syncytial virus (rsv) fusion gene among contemporary and historical genotypes of rsv/a and rsv/b. *PloS one*, 12(4):e0175792–e0175792, 4 2017.

[27] Lydia Tan, Frank E J Coenjaerts, Lieselot Houspie, Marco C Viveen, Grada M van Bleek, Emmanuel J H J Wiertz, Darren P Martin, and Philippe Lemey. The comparative genomics of human respiratory syncytial virus subgroups a and b: genetic variability and molecular evolutionary dynamics. *Journal of virology*, 87(14):8213–26, 7 2013.

[28] Stephanie Goya, Mónica Galiano, Inne Nauwelaers, Alfonsina Trento, Peter J. Openshaw, Alicia S. Mistchenko, Maria Zambon, and Mariana Viegas. Toward unified molecular surveillance of rsv: A proposal for genotype definition. *Influenza and Other Respiratory Viruses*, 14(3):274–285, 5 2020.

[29] Teresa C. T. Peret, Caroline B. Hall, Gregory W. Hammond, Pedro A. Piedra, Gregory A. Storch, Wayne M. Sullender, Cecilia Tsou, and Larry J. Anderson. Circulation patterns of group a and b human respiratory syncytial virus genotypes in 5 communities in north america. *The Journal of Infectious Diseases*, 181(6):1891–1896, 6 2000.

[30] Marietjie Venter, Shabir A Madhi, Caroline T Tiemessen, and Barry D Schoub. Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in south africa: identification of new subgroup a and b genotypes. *The Journal of general virology*, 82(Pt 9):2117–2124, 9 2001.

[31] A Blanc, A Delfraro, S Frabasile, and J Arbiza. Genotypes of respiratory syncytial virus group b identified in uruguay. *Archives of Virology*, 150(3):603–609, 2005.

[32] Shobugawa Yugo, Saito Reiko, Sano Yasuko, Zaraket Hassan, Suzuki Yasushi, Kumaki Akihiko, Dapat Isolde, Oguma Taeko, Yamaguchi Masahiro, and Suzuki Hiroshi. Emerging genotypes of human respiratory syncytial virus subgroup a among patients in japan. *Journal of Clinical Microbiology*, 47(8):2475–2482, 8 2009. doi: 10.1128/JCM.00115-09.

[33] Francesca Di Giallonardo, Jen Kok, Marian Fernandez, Ian Carter, Jemma L. Geoghegan, Dominic E. Dwyer, Edward C. Holmes, and John Sebastian Eden. Evolution

of human respiratory syncytial virus (rsv) over multiple seasons in new south wales, australia. *Viruses*, 10(9), 9 2018.

[34] Laura Gimferrer, Magda Campins, María Gema Codina, María del Carmen Martín, Francisco Fuentes, Juliana Esperalba, Andreu Bruguera, Luz María Vilca, Lluís Armadans, Tomàs Pumarola, and Andrés Antón. Molecular epidemiology and molecular characterization of respiratory syncytial viruses at a tertiary care university hospital in catalonia (spain) during the 2013–2014 season. *Journal of Clinical Virology*, 66:27–32, 2015.

[35] Charles N Agoti, Lillian M Mayieka, James R Otieno, Jamal A Ahmed, Barry S Fields, Lilian W Waiboci, Raymond Nyoka, Rachel B Eidex, Nina Marano, Wagacha Burton, Joel M Montgomery, Robert F Breiman, and D James Nokes. Examining strain diversity and phylogeography in relation to an unusual epidemic pattern of respiratory syncytial virus (rsv) in a long-term refugee camp in kenya. *BMC infectious diseases*, 14:178, 4 2014.

[36] Kaat Ramaekers, Annabel Rector, Lize Cuypers, Philippe Lemey, Els Keyaerts, and Marc Van Ranst. Towards a unified classification for human respiratory syncytial virus genotypes. *Virus Evolution*, 6(2):veaa052, 7 2020.

[37] C PANAYIOTOU, J RICHTER, M KOLIOU, N KALOGIROU, E GEORGIOU, and C CHRISTODOULOU. Epidemiology of respiratory syncytial virus in children in cyprus during three consecutive winter seasons (2010–2013): age distribution, seasonality and association between prevalent genotypes and disease severity. *Epidemiology and Infection*, 142(11):2406–2411, 2014.

[38] Imène Fodha, Astrid Vabret, Leila Ghedira, Hassen Seboui, Slaheddine Chouchane, John Dewar, Neji Gueddiche, Abdelhalim Trabelsi, Noureddine Boujaafar, and François Freymuth. Respiratory syncytial virus infections in hospitalized infants: association between viral load, virus subgroup, and disease severity. *Journal of medical virology*, 79(12):1951–1958, 12 2007.

[39] Susanna Esposito, Antonio Piralla, Alberto Zampiero, Sonia Bianchini, Giada Di Pietro, Alessia Scala, Raffaella Pinzani, Emilio Fossali, Fausto Baldanti, and Nicola Principi. Characteristics and their clinical relevance of respiratory syncytial virus types and genotypes circulating in northern italy in five consecutive winter seasons. *PLOS ONE*, 10(6):e0129369, 6 2015.

[40] Wanwei Li, Yanlan Wang, Bo Yu, Qiqi Tan, Jijian Zhou, Jingjing Hu, Yuanbin Wu, Bo Wang, and Hongjian Li. Disease severity of respiratory syncytial virus (rsv) infection correlate to a novel set of five amino acid substitutions in the rsv attachment glycoprotein (g) in china. *Virus Research*, 281:197937, 2020.

[41] Akinobu Hibino, Reiko Saito, Kiyosu Taniguchi, Hassan Zaraket, Yugo Shobugawa, Tamano Matsui, Hiroshi Suzuki, and for the Japanese HRSV Collaborative Study Group. Molecular epidemiology of human respiratory syncytial virus among children in japan during three seasons and hospitalization risk of genotype on1. *PLOS ONE*, 13(1):e0192085, 1 2018.

[42] Ivy K Kombe, Charles N Agoti, Patrick K Munywoki, Marc Baguelin, D James Nokes, and Graham F Medley. Integrating epidemiological and genetic data with different sampling intensities into a dynamic model of respiratory syncytial virus transmission. *Scientific Reports*, 11(1):1463, 2021.

[43] Charles N Agoti, My V T Phan, Patrick K Munywoki, George Githinji, Graham F Medley, Patricia A Cane, Paul Kellam, Matthew Cotten, and D James Nokes. Genomic analysis of respiratory syncytial virus infections in households and utility in inferring who infects the infant. *Scientific Reports*, 9(1):10076, 2019.

[44] Lirong Zou, Lina Yi, Jie Wu, Yingchao Song, Guofeng Huang, Xin Zhang, Lijun Liang, Hanzhong Ni, Oliver G. Pybus, Changwen Ke, and Jing Lu. Evolution and transmission of respiratory syncytial group a (rsv-a) viruses in guangdong, china 2008–2015. *Frontiers in Microbiology*, 07(AUG):1263, 8 2016.

[45] Jiani Chen, Xueting Qiu, Samuel Shepard, Do-Kyun Kim, James Hixson, Pedro Piedra, Vasanthi Avadhanula, and Justin Bahl. Novel and extendable genotyping system for human respiratory syncytial virus based on whole-genome sequence analysis. *Authorea Preprints*, 6 2021.

[46] R CHANOCK and L FINBERG. Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent (cca). ii. epidemiologic aspects of infection in infants and young children. *American journal of hygiene*, 66(3):291–300, 11 1957.

[47] Lydia J Atherton, Patricia A Jorquera, Abhijeet A Bakre, and Ralph A Tripp. Determining immune and mirna biomarkers related to respiratory syncytial virus (rsv) vaccine types. *Frontiers in Immunology*, 10, 2019.

[48] Natalie I Mazur, Deborah Higgins, Marta C Nunes, José A Melero, Annefleur C Langedijk, Nicole Horsley, Ursula J Buchholz, Peter J Openshaw, Jason S McLellan, Janet A Englund, Asuncion Mejias, Ruth A Karron, Eric Af Simões, Ivana Knezevic, Octavio Ramilo, Pedro A Piedra, Helen Y Chu, Ann R Falsey, Harish Nair, Leyla Kragten-Tabatabaie, Anne Greenough, Eugenio Baraldi, Nikolaos G Papadopoulos, Johan Vekemans, Fernando P Polack, Mair Powell, Ashish Satav, Edward E Walsh, Renato T Stein, Barney S Graham, and Louis J Bont. The respiratory syncytial virus vaccine landscape: lessons from the graveyard and promising candidates. *The Lancet. Infectious diseases*, 18(10):e295–e311, 10 2018.

[49] Sarah C Gilbert. Clinical development of modified vaccinia virus ankara vaccines. *Vaccine*, 31(39):4241–4246, 9 2013.

[50] Jerald Sadoff, Els De Paepe, Wouter Haazen, Edmund Omoruyi, Arangassery R. Bastian, Christy Comeaux, Esther Heijnen, Cynthia Strout, Hanneke Schuitemaker, and Benoit Callendret. Safety and immunogenicity of the ad26.rsv.pref investigational vaccine coadministered with an influenza vaccine in older adults. *Journal of Infectious Diseases*, 223(4):699–708, 2 2021.

[51] Paola Cicconi, Claire Jones, Esha Sarkar, Laura Silva-Reyes, Paul Klenerman, Catherine de Lara, Claire Hutchings, Philippe Moris, Michel Janssens, Laurence A Fissette, Marta Picciolato, Amanda Leach, Antonio Gonzalez-Lopez, Ilse Dieussaert, and Matthew D Snape. First-in-human randomized study to assess the safety and immunogenicity of an investigational respiratory syncytial virus (rsv) vaccine based on chimpanzee-adenovirus-155 viral vector-expressing rsv fusion, nucleocapsid, and antitermination viral protein. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 70(10):2073–2081, 5 2020.

[52] Gsk reports its rsv vaccine is efficacious in older adults, 0. [Online; accessed 2022-06-13].

[53] Dpx-rsv :: Imv inc. (imv), 0. [Online; accessed 2022-06-13].

[54] Wo2013049342a1 - recombinant nanoparticle rsv f vaccine for respiratory syncytial virus - google patents, 0. [Online; accessed 2022-06-13].

[55] Natalija Van Braeckel-Budimir, Bert Jan Haijema, and Kees Leenhouts. Bacterium-like particles for efficient immune stimulation of existing vaccines and new subunit vaccines in mucosal applications. *Frontiers in Immunology*, 4, 2013.

[56] Gabriel J Robbie, Ryan Criste, William F Dall'acqua, Kathryn Jensen, Nita K Patel, Genevieve A Losonsky, and M Pamela Griffin. A novel investigational fc-modified humanized monoclonal antibody, motavizumab-yte, has an extended half-life in healthy adults. *Antimicrobial agents and chemotherapy*, 57(12):6147–6153, 12 2013.

[57] Ruth A Karron, Cindy Luongo, Bhagvanji Thumar, Karen M Loehr, Janet A Englund, Peter L Collins, and Ursula J Buchholz. A gene deletion that up-regulates viral gene expression yields an attenuated rsv vaccine with improved antibody responses in children. *Science translational medicine*, 7(312):312ra175, 11 2015.

[58] Pablo F Céspedes, Emma Rey-Jurado, Janyra A Espinoza, Claudia A Rivera, Gisela Canedo-Marroquín, Susan M Bueno, and Alexis M Kalergis. A single, low dose of a

cgmp recombinant bcg vaccine elicits protective t cell immunity against the human respiratory syncytial virus infection and prevents lung pathology in mice. *Vaccine*, 35(5):757–766, 2 2017.

[59] Susan M Bueno, Pablo A González, Kelly M Cautivo, Jorge E Mora, Eduardo D Leiva, Hugo E Tobar, Glenn J Fennelly, Eliseo A Eugenin, William R Jr Jacobs, Claudia A Riedel, and Alexis M Kalergis. Protective t cell immunity against respiratory syncytial virus is efficiently induced by recombinant bcg. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52):20822–20827, 12 2008.

[60] Brittani N Blunck, Wanderson Rezende, and Pedro A Piedra. Profile of respiratory syncytial virus prefusogenic fusion protein nanoparticle vaccine. *Expert review of vaccines*, 20(4):351–364, 4 2021.

[61] William F Dall'Acqua, Peter A Kiener, and Herren Wu. Properties of human igg1s engineered for enhanced binding to the neonatal fc receptor (fcrn). *The Journal of biological chemistry*, 281(33):23514–23524, 8 2006.

[62] Annefleur C Langedijk, Robert Jan Lebbink, Christiana Naaktgeboren, Anouk Evers, Marco C Viveen, Anne Greenough, Terho Heikkinen, Renato T Stein, Peter Richmond, Federico Martinón-Torres, Marta Nunes, Mitsuaki Hosoya, Christian Keller, Monika Bauck, Robert Cohen, Jesse Papenburg, Jeffrey Pernica, Marije P Hennus, Hong Jin, David E Tabor, Andrev Tovchigrechko, Alexey Ruzin, Michael E Abram, Deidre Wilkins, Joanne G Wildenbeest, Leyla Kragten-Tabatabaie, Frank E J Coenjaerts, Mark T Esser, and Louis J Bont. Global molecular diversity of rsv – the "inform rsv" study. *BMC Infectious Diseases*, 20(1):450, 2020.

[63] Marietjie Venter, Michael Rock, Adrian J Puren, Caroline T Tiemessen, and James E Crowe Jr. Respiratory syncytial virus nucleoprotein-specific cytotoxic t-cell epitopes in a south african population of diverse hla types are conserved in circulating field strains. *Journal of virology*, 77(13):7319–7329, 7 2003.

[64] Xiangpeng Chen, Baoping Xu, Jiayun Guo, Changchong Li, Shuhua An, Yunlian Zhou, Aihuan Chen, Li Deng, Zhou Fu, Yun Zhu, Chunyan Liu, Lili Xu, Wei Wang, Kunling Shen, and Zhengde Xie. Genetic variations in the fusion protein of respiratory syncytial virus isolated from children hospitalized with community-acquired pneumonia in china. *Scientific Reports*, 8(1):4491, 2018.

[65] Ribas-Aparicio Rosa María. The impact of bioinformatics on vaccine design and development. page Ch. 7. IntechOpen, Rijeka, 2017.

[66] Onyeka S Chukwudozie, Vincent C Duru, Charlotte C Ndiribe, Abdullahi T Aborode, Victor O Oyebanji, and Benjamin O Emikpe. The relevance of bioinformatics applications in the discovery of vaccine candidates and potential drugs for covid-19 treatment. *Bioinformatics and Biology Insights*, 15:11779322211002168, 1 2021. doi: 10.1177/11779322211002168.

[67] Ilada Thongpan, Nungruthai Suntronwong, Preeyaporn Vichaiwattana, Nasamon Wanlapakorn, Sompong Vongpunsawad, and Yong Poovorawan. Respiratory syncytial virus, human metapneumovirus, and influenza virus infection in bangkok, 2016-2017. *PeerJ*, 7:e6748, 2019.

[68] Sarah D Meskill, Paula A Revell, Lakshmi Chandramohan, and Andrea T Cruz. Prevalence of co-infection between respiratory syncytial virus and influenza in children. *The American Journal of Emergency Medicine*, 35(3):495–498, 2017.

[69] O H Price, S G Sullivan, C Sutterby, J Druce, and K S Carville. Using routine testing data to understand circulation patterns of influenza a, respiratory syncytial virus and other respiratory viruses in victoria, australia. *Epidemiology and infection*, 147:e221–e221, 1 2019.

[70] Sourya Shrestha, Aaron A King, and Pejman Rohani. Statistical inference for multipathogen systems. *PLOS Computational Biology*, 7(8):e1002135, 8 2011.

[71] Shrestha Sourya, Foxman Betsy, Weinberger Daniel M., Steiner Claudia, Viboud Cécile, and Rohani Pejman. Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Science Translational Medicine*, 5(191):191ra84–191ra84, 6 2013. doi: 10.1126/scitranslmed.3005982.

[72] Lello Joanne. Coinfection: Doing the math. *Science Translational Medicine*, 5(191):191fs24–191fs24, 6 2013. doi: 10.1126/scitranslmed.3006565.

[73] Naomi R Waterlow, Stefan Flasche, Amanda Minter, and Rosalind M Eggo. Competition between rsv and influenza: Limits of modelling inference from surveillance data. *Epidemics*, 35:100460, 2021.

[74] L. J. Anderson, J. C. Hierholzer, C. Tsou, R. M. Hendry, B. F. Fernie, Y. Stone, and K. McIntosh. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *Journal of Infectious Diseases*, 151(4):626–633, 4 1985.

[75] Uzma Bashir Aamir, Muhammad Salman, Nadia Nisar, Nazish Badar, Mohammad Masroor Alam, Jamil Ansari, and Syed Sohail Zahoor Zaidi. Molecular characterization of circulating respiratory syncytial virus genotypes in pakistani children, 2010–2013. *Journal of Infection and Public Health*, 13(3):438–445, 3 2020.

[76] Venkata R. Duvvuri, Andrea Granados, Paul Rosenfeld, Justin Bahl, Alireza Eshaghi, and Jonathan B. Gubbay. Genetic diversity and evolutionary insights of respiratory syncytial virus a on1 genotype: Global and local transmission dynamics. *Scientific Reports*, 5, 9 2015.

[77] Anwar Ahmed, Shakir H. Haider, Shama Parveen, Mohammed Arshad, Hytham A. Alsenaidy, Alawi Omar Baaboud, Khalid Fahad Mobaireek, Muslim Mohammed AlSaadi, Abdulrahman M. Alsenaidy, and Wayne Sullender. Co-circulation of 72bp duplication group a and 60bp duplication group b respiratory syncytial virus (rsv) strains in riyadh, saudi arabia during 2014. *PLoS ONE*, 11(11):166145, 11 2016.

[78] Wayne M. Sullender. Respiratory syncytial virus genetic and antigenic diversity. *Clinical Microbiology Reviews*, 13(1):1–15, 2000.

[79] Juan Carlos Muñoz-Escalante, Andreu Comas-García, Sofía Bernal-Silva, Carla Daniela Robles-Espinoza, Guillermo Gómez-Leal, and Daniel E. Noyola. Respiratory syncytial virus a genotype classification based on systematic intergenotypic and intragenotypic sequence analysis. *Scientific Reports*, 9(1):1–12, 12 2019.

[80] Alfonsina Trento, Leyda Ábrego, Rosa Rodriguez-Fernandez, Maria Isabel González-Sánchez, Felipe González-Martínez, Adriana Delfraro, Juan M. Pascale, Juan Arbiza, and José A. Melero. Conservation of g-protein epitopes in respiratory syncytial virus (group a) despite broad genetic diversity: Is antibody selection involved in virus evolution? *Journal of Virology*, 89(15):7776–7785, 8 2015.

[81] Who | definition of regional groupings. *WHO*, 2017.

[82] Kazutaka Katoh, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. Technical report, 0.

[83] Mathieu Fourment and Edward C. Holmes. Seqotron: A user-friendly sequence editor for mac os x. *BMC Research Notes*, 9(1):106, 2 2016.

[84] Marc A Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. Logcombiner version 1.10.4. *Virus Evolution*, 4(1), 3 2018.

[85] Alexandros Stamatakis. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *BIOINFORMATICS APPLICATIONS*, 30(9):1312–1313, 2014.

[86] Exploring the temporal structure of heterochronous sequences using tempest (formerly path-o-gen), 0. [Online; accessed 2020-08-07].

[87] Lam-Tung Nguyen, Heiko A Schmidt, Arndt Von Haeseler, and Bui Quang Minh. Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. 0.

[88] Pavel Sagulenko, Vadim Puller, and Richard A Neher. Treetime: Maximum-likelihood phylodynamic analysis. 0.

[89] Silvia Castiglione, Gianmarco Tesone, Martina Piccolo, Marina Melchionna, Alessandro Mondanaro, Carmela Serio, Mirko Di Febbraro, and Pasquale Raia. A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution*, 9(4):974–983, 4 2018. https://doi.org/10.1111/2041-210X.12954.

[90] Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. Ape: Analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290, 1 2004.

[91] Guangchuang Yu, David K. Smith, Huachen Zhu, Yi Guan, and Tommy Tsan Yuk Lam. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36, 1 2017.

[92] Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. Mega x: Molecular evolutionary genetics analysis across computing platforms. 0.

[93] Samuel S. Shepard, C. Todd Davis, Justin Bahl, Pierre Rivailler, Ian A. York, and Ruben O. Donis. Label: Fast and accurate lineage assignment with assessment of h5n1 and h9n2 influenza a hemagglutinins. *PLoS ONE*, 9(1):e86921, 1 2014.

[94] Olga Chernomor, Bui Quang Minh, Félix Forest, Steffen Klaere, Travis Ingram, Monika Henzinger, and Arndt von Haeseler. Split diversity in constrained conservation prioritization using integer linear programming. *Methods in Ecology and Evolution*, 6(1):83–91, 1 2015.

[95] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. 0.

[96] Andrew Rambaut, Edward C. Holmes, Verity Hill, Áine O'Toole, JT McCrone, Chris Ruis, du Louis Plessis, and Oliver G. Pybus. A dynamic nomenclature proposal for sars-cov-2 to assist genomic epidemiology. *bioRxiv*, page 2020.04.17.046086, 4 2020.

[97] Giovanni Cattoli, Isabella Monne, Alice Fusaro, Tony M. Joannis, Lami H. Lombin, Mona M. Aly, Abdel S. Arafa, Katherine M. Sturm-Ramirez, Emmanuel Couacy-Hymann, Joseph A. Awuni, Komla B. Batawui, Kodzo A. Awoume, Gilbert L. Aplogan, Adama Sow, Andrè C. Ngangnou, Imam M. El Nasri Hamza, Djibo Gamatié, Gwenaelle Dauphin, Joseph M. Domenech, and Ilaria Capua. Highly pathogenic avian influenza virus subtype h5n1 in africa: A comprehensive phylogenetic analysis and molecular characterization of isolates. *PLoS ONE*, 4(3):1–9, 2009.

[98] Everlyn Kamau, James R Otieno, Clement S Lewa, Anthony Mwema, Nickson Murunga, D James Nokes, and Charles N Agoti. Evolution of respiratory syncytial virus genotype ba in kilifi, kenya, 15 years on. *Scientific Reports*, 10(1):21176, 2020.

[99] James R. Otieno, Everlyn M. Kamau, Charles N. Agoti, Clement Lewa, Grieven Otieno, Ann Bett, Mwanajuma Ngama, Patricia A. Cane, and D. James Nokes. Spread and evolution of respiratory syncytial virus a genotype on1, coastal kenya, 2010-2015. *Emerging Infectious Diseases*, 23(2):264–271, 2 2017.

[100] James E Crowe Jr and V John Williams. Paramyxoviruses: Respiratory syncytial virus and human metapneumovirus. *Viral Infections of Humans: Epidemiology and Control*, pages 601–627, 2 2014.

[101] Wan-Ji Lee, You-jin Kim, Dae-Won Kim, Han Saem Lee, Ho Yeon Lee, and Kisoon Kim. Complete genome sequence of human respiratory syncytial virus genotype a with

a 72-nucleotide duplication in the attachment protein g gene. *Journal of Virology*, 86(24):13810 LP – 13811, 12 2012.

[102] Jeehyun Lee, Laura Klenow, Elizabeth M Coyle, Hana Golding, and Surender Khurana. Protective antigenic sites in respiratory syncytial virus g attachment protein outside the central conserved and cysteine noose domains. *PLoS pathogens*, 14(8):e1007262–e1007262, 8 2018.

[103] Updated guidance for palivizumab prophylaxis among infants and young children at increased risk of hospitalization for respiratory syncytial virus infection. *Pediatrics*, 134(2):415–420, 8 2014.

[104] 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(3 Pt 1):531–537, 9 1998.

[105] Carlotta Biagi, Arianna Dondi, Sara Scarpini, Alessandro Rocca, Silvia Vandini, Giulia Poletti, and Marcello Lanari. Current state and challenges in developing respiratory syncytial virus vaccines. *Vaccines*, 8(4):672, 11 2020.

[106] Tino F Schwarz, Casey Johnson, Christine Grigat, Dan Apter, Peter Csonka, Niklas Lindblad, Thi Lien-Anh Nguyen, Feng F Gao, Hui Qian, Antonella N Tullio, Ilse Dieussaert, Marta Picciolato, and Ouzama Henry. Three dose levels of a maternal respiratory syncytial virus vaccine candidate are well tolerated and immunogenic in a randomized trial in non-pregnant women. *The Journal of infectious diseases*, 6 2021.

[107] H W Kim, J G Canchola, C D Brandt, G Pyles, R M Chanock, K Jensen, and R H Parrott. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *American journal of epidemiology*, 89(4):422–434, 4 1969.

[108] B R Murphy, V A Sotnikov, L A Lawrence, S M Banks, and G A Prince. Enhanced pulmonary histopathology is observed in cotton rats immunized with formalin-inactivated

respiratory syncytial virus (rsv) or purified f glycoprotein and challenged with rsv 3-6 months after immunization. *Vaccine*, 8(5):497–502, 10 1990.

[109] April M Killikelly, Masaru Kanekiyo, and Barney S Graham. Pre-fusion f is absent on the surface of formalin-inactivated respiratory syncytial virus. *Scientific reports*, 6:34108, 9 2016.

[110] Fernando P Polack, Michael N Teng, Peter L Collins, Gregory A Prince, Marcus Exner, Heinz Regele, Dario D Lirman, Richard Rabold, Scott J Hoffman, Christopher L Karp, Steven R Kleeberger, Marsha Wills-Karp, and Ruth A Karron. A role for immune complexes in enhanced respiratory syncytial virus disease. *The Journal of experimental medicine*, 196(6):859–865, 9 2002.

[111] M Connors, N A Giese, A B Kulkarni, C Y Firestone, H C 3rd Morse, and B R Murphy. Enhanced pulmonary histopathology induced by respiratory syncytial virus (rsv) challenge of formalin-inactivated rsv-immunized balb/c mice is abrogated by depletion of interleukin-4 (il-4) and il-10. *Journal of virology*, 68(8):5321–5325, 8 1994.

[112] M E Waris, C Tsou, D D Erdman, S R Zaki, and L J Anderson. 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. *Journal of virology*, 70(5):2852–2860, 5 1996.

[113] Morgan S A Gilman, Polina Furmanova-Hollenstein, Gabriel Pascual, Angélique B. van 't Wout, Johannes P M Langedijk, and Jason S McLellan. Transient opening of trimeric prefusion rsv f proteins. *Nature Communications*, 10(1):2105, 2019.

[114] Geraldine Taylor. Animal models of respiratory syncytial virus infection. *Vaccine*, 35(3):469–480, 1 2017.

[115] Xuan Liang, Dong-Hai Liu, De Chen, Li Guo, Hui Yang, Yong-Sheng Shi, Yong-Jun Wang, Wei-Kai Wang, Zhi-Ping Xie, Han-Chun Gao, Zhao-Jun Duan, and Rong-Fang

Zhang. Gradual replacement of all previously circulating respiratory syncytial virus a strain with the novel on1 genotype in lanzhou from 2010 to 2017. *Medicine*, 98(19), 2019.

[116] Daiyin Tian, Michael B Battles, Syed M Moin, Man Chen, Kayvon Modjarrad, Azad Kumar, Masaru Kanekiyo, Kevin W Graepel, Noor M Taher, Anne L Hotard, Martin L Moore, Min Zhao, Zi-Zheng Zheng, Ning-Shao Xia, Jason S McLellan, and Barney S Graham. Structural basis of respiratory syncytial virus subtype-dependent neutralization by an antibody targeting the fusion glycoprotein. *Nature Communications*, 8(1):1877, 2017.

[117] Koichi Hashimoto and Mitsuaki Hosoya. Neutralizing epitopes of rsv and palivizumab resistance in japan. *Fukushima journal of medical science*, 63(3):127–134, 12 2017.

[118] Anne S De Groot, Leonard Moise, Frances Terry, Andres H Gutierrez, Pooja Hindocha, Guilhem Richard, Daniel Fredric Hoft, Ted M Ross, Amy R Noe, Yoshimasa Takahashi, Vinayaka Kotraiah, Sarah E Silk, Carolyn M Nielsen, Angela M Minassian, Rebecca Ashfield, Matt Ardito, Simon J Draper, and William D Martin. Better epitope discovery, precision immune engineering, and accelerated vaccine design using immunoinformatics tools. *Frontiers in Immunology*, 11:442, 2020.

[119] Who | world health statistics 2011. *WHO*, 2011.

[120] John Rozewicki, Songling Li, Karlou Mar Amada, Daron M Standley, and Kazutaka Katoh. Mafft-dash: integrated protein sequence and structural alignment. *Nucleic Acids Research*, 47(W1):W5–W10, 7 2019.

[121] Peter Rice, Ian Longden, and Alan Bleasby. Emboss: The european molecular biology open software suite. *Trends in Genetics*, 16(6):276–277, 6 2000.

[122] S S Whitehead, M G Hill, C Y Firestone, M St. Claire, W R Elkins, B R Murphy, and P L Collins. Replacement of the f and g proteins of respiratory syncytial virus (rsv)

subgroup a with those of subgroup b generates chimeric live attenuated rsv subgroup b vaccine candidates. *Journal of Virology*, 73(12):9773 LP – 9780, 12 1999.

[123] Pavel Sagulenko, Vadim Puller, and Richard A Neher. Treetime: Maximum-likelihood phylodynamic analysis. *Virus Evolution*, 4(1), 1 2018.

[124] A Sette and J Sidney. Nine major hla class i supertypes account for the vast preponderance of hla-a and -b polymorphism. *Immunogenetics*, 50(3-4):201–212, 11 1999.

[125] S Southwood, J Sidney, A Kondo, M F del Guercio, E Appella, S Hoffman, R T Kubo, R W Chesnut, H M Grey, and A Sette. Several common hla-dr types share largely overlapping peptide binding repertoires. *Journal of immunology (Baltimore, Md. : 1950)*, 160(7):3363–3373, 4 1998.

[126] Lu He, Anne S De Groot, Andres H Gutierrez, William D Martin, Lenny Moise, and Chris Bailey-Kellogg. Integrated assessment of predicted mhc binding and cross-conservation with self reveals patterns of viral camouflage. *BMC bioinformatics*, 15 Suppl 4(Suppl 4):S1, 2014.

[127] The UniProt Consortium . Uniprot: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489, 1 2021.

[128] Leonard Moise, Andres Gutierrez, Farzana Kibria, Rebecca Martin, Ryan Tassone, Rui Liu, Frances Terry, Bill Martin, and Anne S De Groot. ivax: An integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines. *Human vaccines immunotherapeutics*, 11(9):2312–2321, 2015.

[129] Qing Zhu, Jason S McLellan, Nicole L Kallewaard, Nancy D Ulbrandt, Susan Palaszynski, Jing Zhang, Brian Moldt, Anis Khan, Catherine Svabek, Josephine M McAuliffe, Daniel Wrapp, Nita K Patel, Kimberly E Cook, Bettina W M Richter, Patricia C Ryan, Andy Q Yuan, and JoAnn A Suzich. A highly potent extended half-life

antibody as a potential rsv vaccine surrogate for all infants. *Science Translational Medicine*, 9(388):eaaj1928, 5 2017.

[130] Jason S McLellan, Yongping Yang, Barney S Graham, and Peter D Kwong. Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. *Journal of virology*, 85(15):7788–7796, 8 2011.

[131] Andres H Gutiérrez, Vicki J Rapp-Gabrielson, Frances E Terry, Crystal L Loving, Leonard Moise, William D Martin, and Anne S De Groot. T-cell epitope content comparison (epicc) of swine h1 influenza a virus hemagglutinin. *Influenza and other respiratory viruses*, 11(6):531–542, 11 2017.

[132] Edita Karosiene, Claus Lundegaard, Ole Lund, and Morten Nielsen. Netmhccons: a consensus method for the major histocompatibility complex class i predictions. *Immunogenetics*, 64(3):177–186, 2012.

[133] J C GOWER. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338, 12 1966.

[134] J Sánchez. Mardia, k. v., j. t. kent, j. m. bibby: Multivariate analysis. academic press, london-new york-toronto-sydney-san francisco 1979. xv, 518 pp., 61.00. *Biometrical Journal*, 24(5):502, 1 1982. https://doi.org/10.1002/bimj.4710240520.

[135] Malika Charrad, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. Nbclust: An r package for determining the relevant number of clusters in a data set. *Journal of Statistical Software; Vol 1, Issue 6 (2014)*, 2014.

[136] Jan de Leeuw and Patrick Mair. Multidimensional scaling using majorization: Smacof in r. *Journal of Statistical Software*, 31(3 SE - Articles):1–30, 8 2009.

[137] Morgan S A Gilman, Carlos A Castellanos, Man Chen, Joan O Ngwuta, Eileen Goodwin, Syed M Moin, Vicente Mas, José A Melero, Peter F Wright, Barney S Graham, Jason S

McLellan, and Laura M Walker. Rapid profiling of rsv antibody repertoires from the memory b cells of naturally infected adult donors. *Science Immunology*, 1(6):eaaj1879, 12 2016. doi: 10.1126/sciimmunol.aaj1879.

[138] B. T. Grenfell. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327–332, 1 2004.

[139] Trevor Bedford, Marc A Suchard, Philippe Lemey, Gytis Dudas, Victoria Gregory, Alan J Hay, John W McCauley, Colin A Russell, Derek J Smith, and Andrew Rambaut. Integrating influenza antigenic dynamics with molecular evolution. *eLife*, 3, 2 2014.

[140] Derek J. Smith, Alan S. Lapedes, Jan C. De Jong, Theo M. Bestebroer, Guus F. Rimmelzwaan, Albert D.M.E. Osterhaus, and Ron A.M. Fouchier. Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682):371–376, 7 2004.

[141] Bette Korber, Montiago LaBute, and Karina Yusim. Immunoinformatics comes of age. *PLoS computational biology*, 2(6):e71–e71, 6 2006.

[142] Birkir Reynisson, Bruno Alvarez, Sinu Paul, Bjoern Peters, and Morten Nielsen. Netmhcpan-4.1 and netmhciipan-4.0: improved predictions of mhc antigen presentation by concurrent motif deconvolution and integration of ms mhc eluted ligand data. *Nucleic Acids Research*, 48(W1):W449–W454, 7 2020.

[143] Clark D Russell, Stefan A Unger, Marc Walton, and Jürgen Schwarze. The human immune response to respiratory syncytial virus infection. *Clinical microbiology reviews*, 30(2):481–502, 4 2017.

[144] Rutger G Woolthuis, Christiaan H van Dorp, Can Keşmir, Rob J de Boer, and Michiel van Boven. Long-term adaptation of the influenza a virus by escaping cytotoxic t-cell recognition. *Scientific Reports*, 6(1):33334, 2016.

[145] Jie Liu, Tracy J Ruckwardt, Man Chen, Teresa R Johnson, and Barney S Graham. Characterization of respiratory syncytial virus m- and m2-specific cd4 t cells in a murine model. *Journal of virology*, 83(10):4934–4941, 5 2009.

[146] Cécile Viboud, Katelyn Gostic, Martha I Nelson, Graeme E Price, Amanda Perofsky, Kaiyuan Sun, Nídia Sequeira Trovão, Benjamin J Cowling, Suzanne L Epstein, and David J Spiro. Beyond clinical trials: Evolutionary and epidemiological considerations for development of a universal influenza vaccine. *PLOS Pathogens*, 16(9):e1008583, 9 2020.

[147] Anna-Lisa Schaap-Johansen, Milena Vujović, Annie Borch, Sine Reker Hadrup, and Paolo Marcatili. T cell epitope prediction and its application to immunotherapy. *Frontiers in Immunology*, 12, 2021.

[148] Ramgopal R Mettu, Tysheena Charles, and Samuel J Landry. Cd4+ t-cell epitope prediction using antigen processing constraints. *Journal of immunological methods*, 432:72–81, 5 2016.

[149] Brian M Peters, Mary Ann Jabra-Rizk, Graeme A O'May, J William Costerton, and Mark E Shirtliff. Polymicrobial interactions: impact on pathogenesis and human disease. *Clinical microbiology reviews*, 25(1):193–213, 1 2012.

[150] Sourya Shrestha, Betsy Foxman, Daniel M Weinberger, Claudia Steiner, Cécile Viboud, and Pejman Rohani. Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Science translational medicine*, 5(191):191ra84, 6 2013.

[151] Marc Lipsitch, Osman Abdullahi, Alexander D'Amour, Wen Xie, Daniel M Weinberger, Eric Tchetgen Tchetgen, and J Anthony G Scott. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in kenya with a markov transition model. *Epidemiology*, 23(4), 2012.

[152] Lulla Opatowski, Emmanuelle Varon, Claire Dupont, Laura Temime, Sylvie van der Werf, Laurent Gutmann, Pierre-Yves Boëlle, Laurence Watier, and Didier Guillemot. Assessing pneumococcal meningitis association with viral respiratory infections and antibiotics: insights from statistical and mathematical models. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764):20130519, 8 2013. doi: 10.1098/rspb.2013.0519.

[153] Anchi Wu, Valia T Mihaylova, Marie L Landry, and Ellen F Foxman. Interference between rhinovirus and influenza a virus: a clinical data analysis and experimental infection study. *The Lancet Microbe*, 1(6):e254–e262, 2020.

[154] Francesco Pinotti, Fakhteh Ghanbarnejad, Philipp Hövel, and Chiara Poletto. Interplay between competitive and cooperative interactions in a three-player pathogen system. *Royal Society Open Science*, 7(1):190305, 3 2022. doi: 10.1098/rsos.190305.

[155] João Nunes Caldeira Marinho Matos, Sofia Rodrigues Sousa, Maria Braz, Yvette Martins, and Fernando Barata. Characterization of influenza a, b and rsv infections in a pulmonology ward during the 17/18 flu-season. *European Respiratory Journal*, 54(suppl 63):PA4566, 9 2019.

[156] Aubree Gordon and Arthur Reingold. The burden of influenza: a complex problem. *Current epidemiology reports*, 5(1):1–9, 2018.

[157] Mohsen Moghadami. A narrative review of influenza: A seasonal and pandemic disease. *Iranian journal of medical sciences*, 42(1):2–13, 1 2017.

[158] Ramandeep K Virk, Jayanthi Jayakumar, Ian H Mendenhall, Mahesh Moorthy, Pauline Lam, Martin Linster, Julia Lim, Cui Lin, Lynette L E Oon, Hong Kai Lee, Evelyn S C Koay, Dhanasekaran Vijaykrishna, Gavin J D Smith, and Yvonne C F Su. Divergent evolutionary trajectories of influenza b viruses underlie their contemporaneous epidemic activity. *Proceedings of the National Academy of Sciences*, 117(1):619 LP – 628, 1 2020.

[159] Martha I Nelson, Laurel Edelman, David J Spiro, Alex R Boyne, Jayati Bera, Rebecca Halpin, Elodie Ghedin, Mark A Miller, Lone Simonsen, Cecile Viboud, and Edward C Holmes. Molecular epidemiology of a/h3n2 and a/h1n1 influenza virus during a single epidemic season in the united states. *PLOS Pathogens*, 4(8):e1000133, 8 2008.

[160] Bin Lu, Hui Liu, David E Tabor, Andrey Tovchigrechko, Yanping Qi, Alexey Ruzin, Mark T Esser, and Hong Jin. Emergence of new antigenic epitopes in the glycoproteins of human respiratory syncytial virus collected from a us surveillance study, 2015–17. *Scientific Reports*, 9(1):3898, 2019.

[161] Juan C Muñoz-Escalante, Andreu Comas-García, Sofía Bernal-Silva, and Daniel E Noyola. Respiratory syncytial virus b sequence analysis reveals a novel early genotype. *Scientific Reports*, 11(1):3452, 2021.

[162] Stephanie Ascough, Suzanna Paterson, and Christopher Chiu. Induction and subversion of human protective immunity: Contrasting influenza and respiratory syncytial virus. *Frontiers in Immunology*, 9(MAR):323, 3 2018.

[163] Gilberto Gonzàlez-Parra, Filip De Ridder, Dymphy Huntjens, Dirk Roymans, Gabriela Ispas, and Hana M Dobrovolny. A comparison of rsv and influenza in vitro kinetic parameters reveals differences in infecting time. *PLOS ONE*, 13(2):e0192645, 2 2018.

[164] G Walzl, S Tafuro, P Moss, P J Openshaw, and T Hussell. Influenza virus lung infection protects from respiratory syncytial virus-induced immunopathology. *The Journal of experimental medicine*, 192(9):1317–1326, 11 2000.

[165] Liselotte van Asten, Paul Bijkerk, Ewout Fanoy, Annemarijn van Ginkel, Anita Suijkerbuijk, Wim van der Hoek, Adam Meijer, and Harry Vennema. Early occurrence of influenza a epidemics coincided with changes in occurrence of other respiratory virus infections. *Influenza and other respiratory viruses*, 10(1):14–26, 1 2016.

[166] Stacey M Hartwig, Ann M Miller, and Steven M Varga. Respiratory syncytial virus provides protection against a subsequent influenza a virus infection. *The Journal of Immunology*, 208(3):720 LP – 731, 2 2022.

[167] Lulla Opatowski, Marc Baguelin, and Rosalind M Eggo. Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: A key role for mathematical modelling. *PLOS Pathogens*, 14(2):e1006770, 2 2018.

[168] Deborah H Charbonneau and LaTeesa N James. Fluview and flunet: Tools for influenza activity and surveillance. *Medical Reference Services Quarterly*, 38(4):358–368, 10 2019. doi: 10.1080/02763869.2019.1657734.

[169] State population totals: 2010-2019, 0. [Online; accessed 2022-07-04].

[170] Time series analysis and its applications: With r examples - tsa4, 0. [Online; accessed 2022-07-04].

[171] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 10 1990.

[172] Sonnie Kim, Thomas C Williams, Cecile Viboud, Harry Campbell, Jiani Chen, and David J Spiro. Rsv genomic diversity and the development of a globally effective rsv intervention. *Vaccine*, 39(21):2811–2820, 2021.

[173] Kazutaka Katoh and Daron M Standley. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780, 4 2013.

[174] Bui Quang Minh, Steffen Klaere, and Arndt von Haeseler. Taxon selection under split diversity. *Systematic biology*, 58(6):586–594, 12 2009.

[175] Marc A Suchard, Philippe Lemey, Guy Baele, Daniel L Ayres, Alexei J Drummond, and Andrew Rambaut. Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *Virus Evolution*, 4(1), 1 2018.

[176] Simon Tavaré et al. Some probabilistic and statistical problems in the analysis of dna sequences. *Lectures on mathematics in the life sciences*, 17(2):57–86, 1986.

[177] Beth Shapiro, Andrew Rambaut, and Alexei J. Drummond. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, 23(1):7–9, 1 2006.

[178] Vladimir N Minin, Erik W Bloomquist, and Marc A Suchard. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471, 7 2008.

[179] Andrew Rambaut, Alexei J Drummond, Dong Xie, Guy Baele, and Marc A Suchard. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Systematic Biology*, 67(5):901–904, 9 2018.

[180] Trevor Bedford, Sarah Cobey, and Mercedes Pascual. Strength and tempo of selection revealed in viral gene genealogies. *BMC evolutionary biology*, 11:220, 7 2011.

[181] Carol Y Lin. Modeling infectious diseases in humans and animals by keeling, m. j. and rohani, p. *Biometrics*, 64(3):993, 9 2008. https://doi.org/10.1111/j.1541-0420.2008.01082.

[182] Aaron A King, Dao Nguyen, and Edward L Ionides. Statistical inference for partially observed markov processes via the r package pomp. *Journal of Statistical Software*, 69(12 SE - Articles):1–43, 3 2016.

[183] Sonja J Olsen, Amber K Winn, Alicia P Budd, Mila M Prill, John Steel, Claire M Midgley, Krista Kniss, Erin Burns, Thomas Rowe, Angela Foust, Gabriela Jasso,

Angiezel Merced-Morales, C Todd Davis, Yunho Jang, Joyce Jones, Peter Daly, Larisa Gubareva, John Barnes, Rebecca Kondor, Wendy Sessions, Catherine Smith, David E Wentworth, Shikha Garg, Fiona P Havers, Alicia M Fry, Aron J Hall, Lynnette Brammer, and Benjamin J Silk. Changes in influenza and other respiratory virus activity during the covid-19 pandemic - united states, 2020-2021. *MMWR. Morbidity and mortality weekly report*, 70(29):1013–1019, 7 2021.

[184] Bahl Justin, Nelson Martha I., Chan Kwok H., Chen Rubing, Vijaykrishna Dhanasekaran, Halpin Rebecca A., Stockwell Timothy B., Lin Xudong, Wentworth David E., Ghedin Elodie, Guan Yi, Peiris J S Malik, Riley Steven, Rambaut Andrew, Holmes Edward C., and Smith Gavin J D. Temporally structured metapopulation dynamics and persistence of influenza a h3n2 virus in humans. *Proceedings of the National Academy of Sciences*, 108(48):19359–19364, 11 2011. doi: 10.1073/pnas.1109314108.

[185] Eva Grilc, Katarina Prosenc Trilar, Jaro Lajovic, and Maja Sočan. Determining the seasonality of respiratory syncytial virus in slovenia. *Influenza and Other Respiratory Viruses*, 15(1):56–63, 1 2021. https://doi.org/10.1111/irv.12779.

[186] You Li, Xin Wang, Eeva K Broberg, Harry Campbell, and Harish Nair. Seasonality of respiratory syncytial virus and its association with meteorological factors in 13 european countries, week 40 2010 to week 39 2019. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 27(16), 4 2022.

[187] Lisa Staadegaard, Saverio Caini, Sonam Wangchuk, Binay Thapa, Walquiria Aparecida Ferreira de Almeida, Felipe Cotrim de Carvalho, Rodrigo A Fasce, Patricia Bustos, Jan Kyncl, Ludmila Novakova, Alfredo Bruno Caicedo, Domenica Joseth de Mora Coloma, Adam Meijer, Mariëtte Hooiveld, Q Sue Huang, Tim Wood, Raquel Guiomar, Ana Paula Rodrigues, Vernon Jian Ming Lee, Li Wei Ang, Cheryl Cohen, Jocelyn Moyes, Amparo Larrauri, Concepción Delgado-Sanz, Clarisse Demont, Mathieu

Bangert, Michel Dückers, Jojanneke van Summeren, and John Paget. Defining the seasonality of respiratory syncytial virus around the world: National and subnational surveillance data from 12 countries. *Influenza and other respiratory viruses*, 15(6):732–741, 11 2021.

[188] Mark Robertson, John-Sebastian Eden, Avram Levy, Ian Carter, Rachel L Tulloch, Elena J Cutmore, Bethany A Horsburgh, Chisha T Sikazwe, Dominic E Dwyer, David W Smith, and Jen Kok. The spatial-temporal dynamics of respiratory syncytial virus infections across the east–west coasts of australia during 2016–17. *Virus Evolution*, 7(2):veab068, 12 2021.

[189] Hsin Chi, Kuang-Liang Hsiao, Li-Chuan Weng, Chang-Pan Liu, and Hsin-Fu Liu. Persistence and continuous evolution of the human respiratory syncytial virus in northern taiwan for two decades. *Scientific Reports*, 9(1):4704, 2019.

[190] Jurgen Schwarze, Diarmund R O'Donnell, Angela Rohwedder, and Peter J M Openshaw. Latency and persistence of respiratory syncytial virus despite t cell immunity. *American Journal of Respiratory and Critical Care Medicine*, 169(7):801–805, 4 2004. doi: 10.1164/rccm.200308-1203OC.

[191] Daniel Zinder, Trevor Bedford, Sunetra Gupta, and Mercedes Pascual. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLOS Pathogens*, 9(1):e1003104, 1 2013.

[192] Trevor Bedford, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, Alexander Klimov, Nicola S Lewis, Xiyan Li, John W McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu, Eugene Skepner, Derek J Smith, Marc A Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin A Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, 7 2015.

[193] Terezinha M Paiva, Maria A Ishida, Margarete A Benega, Clóvis R A Constantino, Daniela B B Silva, Kátia C O Santos, Maria I Oliveira, Helena A Barbosa, Telma R M P Carvalhanas, Cynthia Schuck-Paim, and Wladimir J Alonso. Shift in the timing of respiratory syncytial virus circulation in a subtropical megalopolis: implications for immunoprophylaxis. *Journal of medical virology*, 84(11):1825–1830, 11 2012.

[194] Lubna Pinky and Hana M Dobrovolny. Coinfections of the respiratory tract: Viral competition for resources. *PLOS ONE*, 11(5):e0155589, 5 2016.

[195] Karen L Laurie, Teagan A Guarnaccia, Louise A Carolan, Ada W C Yan, Malet Aban, Stephen Petrie, Pengxing Cao, Jane M Heffernan, Jodie McVernon, Jennifer Mosse, Anne Kelso, James M McCaw, and Ian G Barr. Interval between infections and viral hierarchy are determinants of viral interference following influenza virus infection in a ferret model. *The Journal of Infectious Diseases*, 212(11):1701–1710, 12 2015.

[196] Yaron Drori, Jasmine Jacob-Hirsch, Rakefet Pando, Aharona Glatman-Freedman, Nehemya Friedman, Ella Mendelson, and Michal Mandelboim. Influenza a virus inhibits rsv infection via a two-wave expression of ifit proteins. *Viruses*, 12(10), 10 2020.

[197] Naomi R Waterlow, Michiko Toizumi, Edwin van Leeuwen, Hien-Anh Thi Nguyen, Lay Myint-Yoshida, Rosalind M Eggo, and Stefan Flasche. Evidence for influenza and rsv interaction from 10 years of enhanced surveillance in nha trang, vietnam, a modelling study. *PLOS Computational Biology*, 18(6):e1010234, 6 2022.

[198] Yaowu Yang, Zhong Wang, Lili Ren, Wei Wang, Guy Vernet, Gláucia Paranhos-Baccalà, Qi Jin, and Jianwei Wang. Influenza a/h1n1 2009 pandemic and respiratory virus infections, beijing, 2009–2010. *PLOS ONE*, 7(9):e45807, 9 2012.

[199] P Rohani, D J Earn, B Finkenstädt, and B T Grenfell. Population dynamic interference among childhood diseases. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1410):2033–2041, 11 1998.

[200] Mansi C Pandya, Sean M Callahan, Kyryll G Savchenko, and Christopher C Stobart. A contemporary view of respiratory syncytial virus (rsv) biology and strain-specific differences. *Pathogens (Basel, Switzerland)*, 8(2), 5 2019.

[201] Rachel E Baker, Sang Woo Park, Wenchang Yang, Gabriel A Vecchi, C Jessica E Metcalf, and Bryan T Grenfell. The impact of covid-19 nonpharmaceutical interventions on the future dynamics of endemic infections. *Proceedings of the National Academy of Sciences*, 117(48):30547–30553, 12 2020. doi: 10.1073/pnas.2013182117.

[202] Matthew R Hendricks, Lauren P Lashua, Douglas K Fischer, Becca A Flitter, Katherine M Eichinger, Joan E Durbin, Saumendra N Sarkar, Carolyn B Coyne, Kerry M Empey, and Jennifer M Bomberger. Respiratory syncytial virus infection enhances pseudomonas aeruginosa biofilm growth through dysregulation of nutritional immunity. *Proceedings of the National Academy of Sciences*, 113(6):1642–1647, 2 2016. doi: 10.1073/pnas.1516979113.

[203] Nickbakhsh Sema, Mair Colette, Matthews Louise, Reeve Richard, Johnson Paul C D., Thorburn Fiona, von Wissmann Beatrix, Reynolds Arlene, McMenamin James, Gunson Rory N., and Murcia Pablo R. Virus–virus interactions impact the population dynamics of influenza and the common cold. *Proceedings of the National Academy of Sciences*, 116(52):27142–27150, 12 2019. doi: 10.1073/pnas.1911083116.

[204] Daniel Magee and Matthew Scotch. The effects of random taxa sampling schemes in bayesian virus phylogeography. *Infection, Genetics and Evolution*, 64:225–230, 2018.

[205] Mathieu Fourment and Mark J Gibbs. Patristic: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC evolutionary biology*, 6:1, 1 2006.

[206] Sandrine Pavoine, Sébastien Ollier, Dominique Pontier, and Daniel Chessel. Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theoretical population biology*, 73(1):79–91, 2 2008.

[207] Fabrizio Menardo, Chloé Loiseau, Daniela Brites, Mireia Coscolla, Sebastian M Gygli, Liliana K Rutaihwa, Andrej Trauner, Christian Beisel, Sonia Borrell, and Sebastien Gagneux. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics*, 19(1):164, 2018.

[208] W Li, L Jaroszewski, and A Godzik. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics (Oxford, England)*, 17(3):282–283, 3 2001.

[209] Kresimir Sikic and Oliviero Carugo. Protein sequence redundancy reduction: comparison of various method. *Bioinformation*, 5(6):234–239, 11 2010.

[210] Daniel P Faith. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1):1–10, 1992.

[211] Leonard Moise, Julie A McMurry, Soren Buus, Sharon Frey, William D Martin, and Anne S De Groot. In silico-accelerated identification of conserved and immunogenic variola/vaccinia t-cell epitopes. *Vaccine*, 27(46):6471–6479, 10 2009.

[212] Sinu Paul, John Sidney, Alessandro Sette, and Bjoern Peters. Tepitool: A pipeline for computational prediction of t cell epitope candidates. *Current protocols in immunology*, 114:18.19.1–18.19.24, 8 2016.

[213] Sandra Silva-Arrieta, Philip J R Goulder, and Christian Brander. In silico veritas? potential limitations for sars-cov-2 vaccine development based on t-cell epitope prediction. *PLOS Pathogens*, 16(6):e1008607, 6 2020.

[214] John Sidney, Bjoern Peters, Nicole Frahm, Christian Brander, and Alessandro Sette. Hla class i supertypes: a revised and updated classification. *BMC immunology*, 9:1, 1 2008.

[215] Nicole Frahm, Karina Yusim, ToddJ. Suscovich, Sharon Adams, John Sidney, Peter Hraber, HannahS. Hewitt, CaitlynH. Linde, DanielG. Kavanagh, Tonia Woodberry, LeahM. Henry, Kellie Faircloth, Jennifer Listgarten, Carl Kadie, Nebojsa Jojic, Kaori Sango, NancyV. Brown, Eunice Pae, M.Tauheed Zaman, Florian Bihl, Ashok Khatri, Mina John, Simon Mallal, FrancescoM. Marincola, BruceD. Walker, Alessandro Sette, David Heckerman, BetteT. Korber, and Christian Brander. Extensive hla class i allele promiscuity among viral ctl epitopes. *European Journal of Immunology*, 37(9):2419–2433, 9 2007. https://doi.org/10.1002/eji.200737365.

[216] Iedb.org: Free epitope database and prediction resource, 0. [Online; accessed 2022-09-09].

[217] Linling He and Jiang Zhu. Computational tools for epitope vaccine design and evaluation. *Current opinion in virology*, 11:103–112, 4 2015.

[218] Julia L Hurwitz. Respiratory syncytial virus vaccine development. *Expert review of vaccines*, 10(10):1415–1433, 10 2011.

[219] J Rodney Brister, Danso Ako-Adjei, Yiming Bao, and Olga Blinkova. Ncbi viral genomes resource. *Nucleic acids research*, 43(Database issue):D571–7, 1 2015.

[220] Yuelong Shu and John McCauley. Gisaid: Global initiative on sharing all influenza data - from vision to reality. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 22(13), 3 2017.

[221] Zeinab Abdelrahman, Mengyuan Li, and Xiaosheng Wang. Comparative review of sars-cov-2, sars-cov, mers-cov, and influenza a respiratory viruses. *Frontiers in Immunology*, 11, 2020.

[222] Gavin J D Smith, Ruben O Donis, World Health Organization/World Organisation for Animal Health/Food Group, and Agriculture Organization (WHO/OIE/FAO) H5 Evolution Working. Nomenclature updates resulting from the evolution of avian influenza a(h5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013–2014. *Influenza and Other Respiratory Viruses*, 9(5):271–276, 9 2015. https://doi.org/10.1111/irv.12324.

[223] Gerco den Hartog, Rob van Binnendijk, Anne-Marie Buisman, Guy A M Berbers, and Fiona R M van der Klis. Immune surveillance for vaccine-preventable diseases. *Expert Review of Vaccines*, 19(4):327–339, 4 2020. doi: 10.1080/14760584.2020.1745071.

[224] Helen E Groves, Pierre-Philippe Piché-Renaud, Adriana Peci, Daniel S Farrar, Steven Buckrell, Christina Bancej, Claire Sevenhuysen, Aaron Campigotto, Jonathan B Gubbay, and Shaun K Morris. The impact of the covid-19 pandemic on influenza, respiratory syncytial virus, and other seasonal respiratory virus circulation in canada: A population-based study. *The Lancet Regional Health - Americas*, 1:100015, 2021.

[225] Erik M. Volz, Katia Koelle, and Trevor Bedford. Viral phylodynamics. *PLoS Computational Biology*, 9(3), 2013.

[226] Erik M. Volz. Complex population dynamics and the coalescent under neutrality. *Genetics*, 190(1):187–201, 2012.

[227] Erik M Volz and Igor Siveroni. Bayesian phylodynamic inference with complex models. *PLoS computational biology*, 14(11):e1006546, 2018.

[228] Hirono Otomaru, Taro Kamigaki, Raita Tamaki, Michiko Okamoto, Portia Parian Alday, Alvin Gue Tan, Joanna Ina Manalo, Edelwisa Segubre-Mercado, Marianette Tawat Inobaya, Veronica Tallo, Socorro Lupisan, and Hitoshi Oshitani. Transmission of respiratory syncytial virus among children under 5 years in households of rural communities, the philippines. *Open Forum Infectious Diseases*, 6(3):ofz045, 3 2019.

[229] Fabrice Carrat, Elisabeta Vergu, Neil M Ferguson, Magali Lemaitre, Simon Cauchemez, Steve Leach, and Alain-Jacques Valleron. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American journal of epidemiology*, 167(7):775–785, 4 2008.

[230] Rebecca K Borchering, Christian E Gunning, V Deven Gokhale, K Bodie Weedop, Arash Saeidpour, Tobias S Brett, and Pejman Rohani. Anomalous influenza seasonality in the united states and the emergence of novel influenza b viruses. *Proceedings of the National Academy of Sciences*, 118(5):e2012327118, 2 2021.

[231] Julie A Spencer, Deborah P Shutt, Sarah K Moser, Hannah Clegg, Helen J Wearing, Harshini Mukundan, and Carrie A Manore. Distinguishing viruses responsible for influenza-like illness. *medRxiv*, page 2020.02.04.20020404, 1 2021.