MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

Ung Thi Hong Trang

MOLECULAR CHARACTERISTICS OF RESPIRATORY SYNCYTIAL VIRUS CAUSING SEVERE ACUTE RESPIRATORY INFECTION IN PEDIATRIC PATIENTS IN THE NATIONAL CHILDREN'S HOSPITAL AND QUANG NINH GENERAL HOSPITAL

SUMMARY OF DISSERTATION ON BIOLOGY

Major: Microbiology

Code: 9 42 01 07

Ha Noi – 2025

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy of Science and Technology

Supervisors:

1. Supervisor 1: Dr. Hoang Vu Mai Phuong - National Institute of Hygiene and Epidemiology (NIHE)
2. Supervisor 2: Prof. Dr. Nguyen Huy Hoang - Institute of Biology, Vietnam Academy of Science and Technology (VAST)

Referee 1: Prof. Dr. Nguyen Duy Bac

Referee 2: Assoc. Prof. Dr. Vu Chi Dung

Referee 3: Assoc. Prof. Dr. Do Thi Phuc

The dissertation is examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology At: 14h00 date: 30th month : June year: 2025

The dissertation can be found at:

1. Graduate University of Science and Technology Library
2. National Library of Vietnam

INTRODUCTION

Respiratory syncytial virus (RSV), which was discovered more than 60 years ago, is a common respiratory virus in winter and causes disease in all ages. It is estimated that up to 95% of children are infected with RSV before they are 2 years old. This is also the age group with the highest mortality rate due to RSV in the community all over the world. The prevalences of RSV epidemics differ in countries and climates. In temperate countries, the epidemic season is mainly winter and continuous until spring. In tropical and subtropical countries, the epidemic occurs for whole year.

There are not many researches on RSV performed in Vietnam. These studies were all conducted before 2015. In particular, genetic characteristics were analyzed on samples collected before 2012. The results obtained in this study are the first results to determine the circulation and molecular biology characteristics of respiratory syncytial virus at the National Children's Hospital and Quang Ninh General Hospital in particular and in Vietnam in general since 2015. The results on the circulation, seasonality of epidemics as well as the genomic characteristics, and the relationship of strains in this thesis with strains in the world contribute to supporting the research and development of RSV vaccines.

Objectives

- Determine the circulation of respiratory syncytial virus in pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital.
- Determine the molecular biology characteristics of respiratory syncytial virus at the National Children's Hospital and Quang Ninh General Hospital.

Main research contents of the thesis

- Collecting respiratory specimens from pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital from 2017 to 2020.

- Extracting viral RNA genetic material, real-time RT-PCR to diagnose and subtype RSV. Analyzing the results of RSV circulation.

- Isolate RSV on infected cell lines, and check the replication of virus strains.

- Sequencing the entire genome of RSV strains and some positive respiratory specimens. Determining the genome sequence, location, and size of genes on the genome. Determining mutations in genes.

- Constructing a phylogenetic tree of gene G. Calculating the genetic distance and evolutionary rate of each RSV subgroup

Chapter 1. LITERATURE REVIEW

1.1. General characteristics of respiratory syncytial virus

1.1.1. Molecular structure and function of respiratory syncytial virus

Respiratory syncytial virus (RSV) belongs to the Mononegavirales order, a member of the Pneumovirinae family, and the Orthopneumovirus genus. RSV is an enveloped, spherical virus whose genetic material is a single-stranded, linear, non-segmented RNA. The molecular mass is 5x106 KDa, and the average size is 120-300 nm (Figure 1.2).

The RSV genome has 10 genes in the following order: 3' end NS1-NS2-N-P-M-SH-G-F-M2-L 5' end. Each gene encodes a corresponding mRNA. Each mRNA encodes a single main protein except M2. In addition to the coding regions, the RSV genome also includes a short non-coding region at both ends (leader region LE and trailer region TR) and intergenic regions that separate genes. Four proteins are components of the nucleocapsid and polymerase complex: nucleoprotein (N), phosphoprotein (P), transcription factor (M2-1), and RNA-dependent RNA polymerase (L).

Three transmembrane glycoproteins: membrane-bound glycoprotein (G), membrane-integrating glycoprotein (F), and small hydrophobic ion channel protein (SH), which together with the M protein form the viral envelope. There are also non-structural proteins (NS1, NS2) that play a role in reducing the host's immune response to RSV and the regulatory protein M2-2.

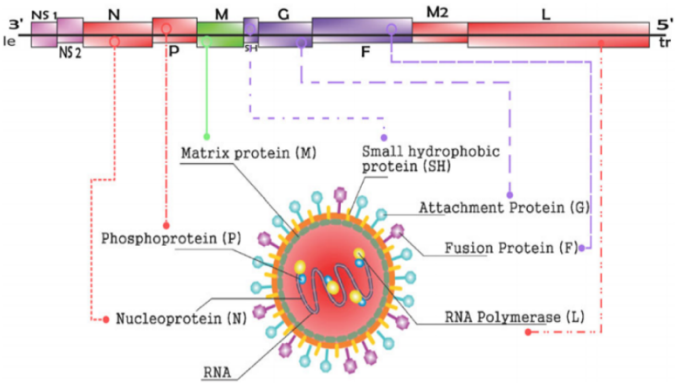

Figure 1.1. Respiratory syncytial virus (RSV) genome (top) and virion structure (bottom).

The two glycoproteins, F and G, are important surface antigens for viral infection and pathogenicity. They are immunological antigens and are the main targets of neutralization. Glycoprotein G functions to bind to the host cell membrane. The F protein plays a role in the initial entry of the virus into the host cell by fusing the viral membrane with the host cell and promoting viral spread by merging infected cells with neighboring healthy cells, thus creating the syncytial characteristic of RSV.

1.1.2. Replication

Several receptors are thought to be involved in the entry of RSV into host cells, including annexin II, CX3C-chemokine receptor 1 (CX3CR1), epidermal growth factor receptor (EGF), calcium-dependent lectins, Toll-

like receptor 4 (TLR4), intercellular adhesion molecule 1 (ICAM-1), nucleolin, and heparan sulfate proteoglycans (HSPGs).

RSV RNA replication is carried out by the polymerase facing the first nucleotide at the 3' end of the genome. The polymerase ignores the GS, GE signals and produces a single-stranded positive-sense copy of the full-length genome (antigenome). A portion of the antigenome is supplemented with 1-3 nucleotides at the 3' end, which are not copied into the RNA genome. The number of antigenomes generated is at least 10 times the number of initial genomes, because the promoter of the antigenome is more efficient than that of the genome and because the synthesis from the genome is divided between RNA replication and transcription.

1.1.3. Genetic diversity of respiratory syncytial virus

RSV is classified into two subgroups, RSV-A and RSV-B, based on antigenic variation and mainly on the G gene sequence. To date, RSV-A has been divided into 12 genotypes (GA1-GA7, SAA1, NA1-2, ON1-2) and RSV-B has been divided into 32 genotypes (GB1-5, BA1-14, SAB1-4, URU1-2, NZB 1-2, BA-CCA, BA-CCB, BA-C, CBB and CB1). From multi-year studies, RSV-A appears to be more dominant than RSV-B worldwide, with RSV-A exhibiting a higher degree of genetic variation in the G protein than RSV-B.

The RSV-A ON1 genotype, identified in Canada in 2010, has a 72-nucleotide duplication in the second hypervariable region (HVR) of the G gene, resulting in seven additional O-glycosylation sites in the protein. Similarly, the BA genotype, first identified in Buenos Aires in 1999 result of a 60-nucleotide duplication in the HVR region of the G gene. Since then, the ON1 and BA genotypes have rapidly spread to become dominant globally, continuing to mutate and adapt to each outbreak in different geographic areas.

1.2. Epidemiological characteristics of respiratory syncytial virus

In general, RSV hospitalization rates peak when annual temperatures are lowest and rainfall is highest. However, when tropical locations around the world are compared within a season, epidemics are not closely related or influenced by temperature, humidity, and rainfall. Furthermore, within a given year, the timing of RSV outbreaks varies considerably between tropical regions. This suggests that although RSV epidemics occur in tropical regions, it is difficult to predict the RSV outbreak season in these regions.

The greatest risk factor for RSV infection is age. In addition, the following factors also affect RSV infection: overcrowded living environment, exposure to smoke (cooking and tobacco), children with comorbidities, congenital diseases, mothers with asthma, economic status lower than the average per capita of the living area, chronic diseases, premature birth, having older siblings living together or delayed vaccination of infants.

1.3. Laboratory diagnostic techniques

Several studies have used nucleic acid amplification methods to detect RSV-A and RSV-B from clinical specimens in the laboratory. There were PCR-based technologies or other methods such as RT-LAMP. Among them, PCR-based diagnostic methods such as RT-PCR, real-time RT-PCR, or multiplex RT-PCR are used in many laboratories.

RSV immunoassays which mainly performed for research and treatment support purposes are rarely used directly detect the acute phase of the disease. Commonly used methods also include direct fluorescent antibody (DFA), indirect fluorescent antibody (IFA), ELISA, etc.

A number of cell lines have been used in RSV isolation such as Hep-2, SIAT, A549, RD, MDCK, RhMK cells, and other cell lines such as BEAS-2B. However, the isolation method is currently only used for research due to

its low sensitivity, time-consuming and labor-intensive.

Genetic sequencing, especially next-generation sequencing, is one of the most modern methods currently used in RSV diagnosis. The biggest problem with this method is that the RNA genome sequence of RSV has segments that are sequenced much more than the rest. In addition, the viral load in clinical samples must be large enough to perform this method.

1.4. Research situation

1.4.1. Research situation in the world

One of the most studied issues on RSV today is the burden of disease caused by this virus. Research on RSV in children has been conducted relatively a lot, especially in children under 2 years old. More than 80% of children with lower respiratory tract infections are related to RSV.

The World Health Organization, the European Center for Disease Control, and the United States have conducted RSV surveillance in their influenza surveillance programs. This has helped the research on RSV in developed countries in general to be continuous, systematic and to assess the disease status relatively fully. However, in developing countries, RSV surveillance and research are only carried out on a small scale, sporadically, and lack a continuous surveillance system.

Molecular and evolutionary characteristics of RSV to serve vaccine research also is an area of interest to many scientists. Research on the molecular biology of RSV focuses mainly on the G gene (molecular epidemiology, evolution) or the F gene (vaccine development, monoclonal antibody research). However, the HVR region of the G gene is the most studied.

1.4.2. Research situation in Vietnam

There are not many studies on RSV in Vietnam. These mainly conducted in the period 2009-2012 and concentrated in the Central and

Southern regions. The studies mainly provide information on RSV infection rates, seasonal circulation and co-infection/interaction with other respiratory agents. To date, there have been very few studies on the molecular biological characteristics of RSV. Only a few studies have conducted gene sequencing (HVR gene G region) and were performed before 2015. Since then, no molecular studies of RSV have been performed.

Chapter 2: MATERIALS AND METHODS

2.1. Study subjects

The study subjects were respiratory clinical specimens (throat and nasopharyngeal secretions) from pediatric patients diagnosed with severe acute respiratory infection according to the World Health Organization case definition of Severe Acute Respiratory Infection (SARI). Sample selection and epidemiological information collection followed the SARI Surveillance Guidelines issued under Decision No. 1271/QD-BYT dated April 3, 2017.

2.2. Time and location

From January 2017 to December 2020, respiratory specimens were collected from the National Children's Hospital and Quang Ninh General Hospital. Then, these specimens were transported to the Laboratory for RSV diagnosis and other analyses.

2.3. Ethics

The study was conducted based on the plan and ethics of Decision No. 4608/QD-BYT dated October 11th 2017 of the Ministry of Health on promulgating the "Plan for key surveillance of influenza syndrome, severe acute respiratory infections and hand, foot and mouth disease in Vietnam in the period 2017-2020".

2.4. Instruments and equipment

The equipment: refrigerated centrifuge, class II biological safety cabinet, ABI 7500 fast realtime PCR machine, MiniAmp PCR thermal

cycler, JEM 1010 transmission electron microscope, Illumina Miseq gene sequencer... Experimental equipment used includes: transport medium tubes, single and multi-channel micropipettes, eppendorf tubes, 8-tube PCR strips, filter tips of various types...

Chemicals: QIAamp Viral RNA Mini Kit, SuperScript III Platinum One-Step qRT-PCR realtime reagent, Nextera XT DNA Library Prepration Kit …

2.5. Methods

2.5.1. Collection, transportation

The study collected 1563 clinical specimens using throat swabs and nasopharyngeal swabs. The specimens were transported to the laboratory weekly and stored at -70^{0}C.

Sample selection criteria: Respiratory specimens collected properly have a case investigation form, a test request form, and a consent form from the child's parents or guardians. Samples that do not meet these standards and are spilled during transportation will be excluded from the study.

2.5.2. RNA extraction, real-time RT-PCR

Clinical specimens were extracted using the QIAampViral RNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions.

Determination of the presence and subgrouping of viruses by realtime RT-PCR using the AgPath-ID One-Step RT-PCR kit (Life Technology, USA) and Invitrogen SuperScript III Platinum One-Step qRT-PCR Kit (Invitrogen, USA). Specific primers and probes for detecting RSV follow the procedure of the US Centers for Disease Control.

2.5.3. Isolation and identification

Samples with positive results for RSV with the Ct value <25 will be taken for clinical examination to isolate the virus on the Hep-2 cell line. After 7 days, the samples will be tested for virus replication and subcultured 2 more

times. The presence of the virus will be tested by realtime RT-PCR and negative staining for examination under a transmission electron microscope.

2.5.4. Next-generation sequencing

Follow the protocol of the Illumina preparation library.

2.5.5. Analysis next-generation sequencing

From the results of the sequencing machine, the data will be analyzed using CLC Genomics Workbench 11 software to create a consensus sequence for each sample. Then, the start and end positions of each gene, intergenic regions, positions, sizes of proteins, and amino acids that are different from the reference strain will be determined to create annotations. Use the RSV Server and GISAID systems to analyze and find the positions of amino acids that have changed.

2.5.6. Phylogenetic tree

Construct a phylogenetic tree based on the G gene region of the two RSV subgroups. Select reference sequences as RSV-A and RSV-B sequences in the world and Vietnam in recent years. Use BEAST v2.7.0 software and Bayesian Markov Chain Monte Carlo (MCMC) method to construct a phylogenetic tree over time and calculate the evolutionary dynamics of each RSV subgroup.

2.5.7. Data analysis

Patient information and test results were entered into the Filemaker Pro sample management software. Stata version 14.0 software was used for analyzing epidemiological parameters, Microsoft Excel was used to analyze epidemiological characteristics and draw charts.

Whole genome sequences of 56 RSV-A and 50 RSV-B isolates and clinical samples with accessions EPI_ISL_16282709 to EPI_ISL_16282764 (RSV-A), EPI_ISL_16289246 to EPI_ISL_16289295 (RSV-B) in the GISAID database.

Chương 3. KẾT QUẢ

1.1. Determination of the circulation of respiratory syncytial virus in pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital

3.1.1. *Characteristics of collected respiratory specimens*

A total of 1,563 specimens were collected from pediatric patients with severe acute respiratory infections. Of , 967 specimens were collected at the National Children's Hospital and 596 specimens were collected at Quang Ninh General Hospital. During the 4 years of the study, the number of collected specimens decreased year by year. The gender distribution of the specimens in the study was 1.5 times higher in males than females (929 males compared to 634 females).

3.1.2. Circulation of respiratory syncytial virus

The overall positive rate of the entire study was 28.02%. The positive rate at the National Children's Hospital was 33.3% and at Quang Ninh General Hospital was 19.46% ($p<0.05$).

The RSV positive rate was statistically different between the years of sampling and by gender. The group under 1 year old not only had a large number of samples collected but also had a higher RSV-positive rate in the 4 years of study than the other groups (43.84%, $p<0.05$).

The epidemic caused by respiratory syncytial virus is cyclical. The number of RSV positive cases was detected almost all year round but the positive rate increased from June to November. RSV peaked annually in August or September (Figure 3.5). In 2017 and 2018, the two groups RSV-A and RSV-B circulated together and had almost the same positive rate. Notably, in 2020, the RSV-A/RSV-B ratio was 2.14 times. In contrast, 2019 was the only year in which the RSV-B ratio was higher than RSV-A at 60.3% compared to 39.7%.

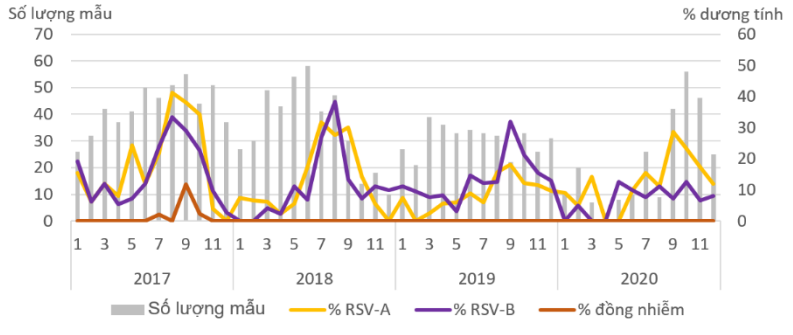

Figure 3.1. Circulation of RSV subtypes over time.

3.2. Molecular biological characteristics of respiratory syncytial virus in pediatric patients with severe acute respiratory tract infection at the National Children's Hospital and Quang Ninh General Hospital

3.2.1. Results of respiratory syncytial virus isolation

A total of 25 RSV strains were successfully amplified (Figure 3.7). The overall isolation rate of the entire study was 22.32% (25/112). Of the 25 isolated strains, 18 strains (72%) were RSV-A and 7 strains (28%) were RSV-B, with 22 isolates (88%) collected from the National Children's Hospital and 3 isolates (12%) collected from the Quang Ninh General Hospital

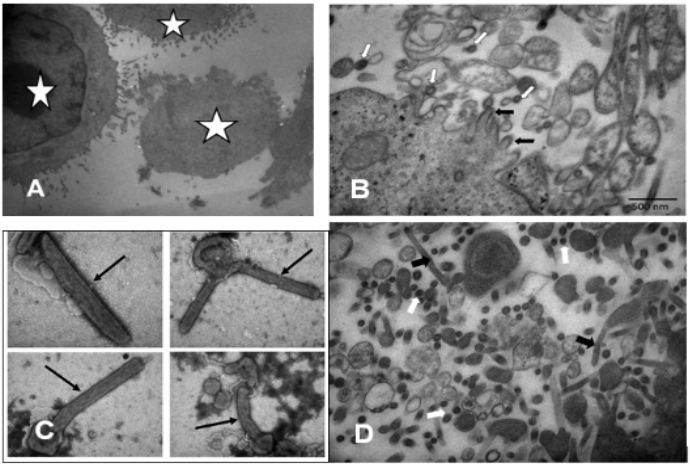

Figure 3.7. Image of respiratory syncytial virus (RSV) in the supernatant of virus isolation on Hep-2 cell line taken under transmission electron microscopy (TEM) with magnification from 1,000-20,000 times. (A). Image of virus replication in the cell, fusion of adjacent cell membranes to form syncytial forms (white stars). (B). Image of RSV budding out of Hep-2 cell membrane and forming elongated structures (black arrows). Virus particles completely released from Hep-2 cell membranes have typical spherical structures in the supernatant of the virus isolation medium (white arrows). (C). Different shapes of RSV after budding from Hep-2 cell membranes when observed under transmission electron microscopy. (D). RSV virus particles are completely cleaved (white arrows) from the initial budding structures (black arrows) and become the majority in the virus culture medium.

3.2.2. Next-generation sequencing

A total of 56 RSV-A and 50 RSV-B samples (strains and clinical specimens) were subjected to whole-genome sequencing using next-generation sequencing techniques. Of the 56 RSV-A samples, 18 were

isolated strains and 38 were clinical specimens. RSV-B included 7 isolated strains and 43 were clinical specimens that underwent NGS gene sequencing.

The whole-genome sequencing data of RSV samples were obtained after 4 runs on the Illumina Miseq system. After analysis, each sample had a large enough number of reads for analysis (over 500,000 reads/sample). The average coverage depth of reads per nucleotide ranged from 2000X-3450X for RSV-A and from 2330X-4010X for RSV-B. The coverage rate >1000X of the samples was above 95%. All samples covered the entire length of the reference sequence and there were no gaps in the genome.

3.2.3. Molecular biological characteristics of the obtained sequences

T he total genome length from the first nucleotide position of the NS1 gene to the last nucleotide of the L gene of RSV-A was 14,885-14,889 nucleotides and of RSV-B was 14,958-14,967 nucleotides.

Despite the different genome sizes within the same subgroup, the sequences had the same length for each gene (except for the G gene). In the RSV-A group, the G gene has two lengths of 894 and 897 nucleotides, corresponding to 298 amino acids and 299 amino acids, respectively. In the RSV-B group, the G gene has two lengths of 933 and 954 nucleotides, corresponding to 311 amino acids and 318 amino acids, respectively. The SH, M2-1, and L genes of RSV-B are also 3 nucleotides longer than those of RSV-A. The remaining genes in both subgroups have the same length.

Table 3.9 shows the results of comparing the study sequences with the corresponding reference strains. The sequences in the study have 100% amino acid coverage of the genes of the reference sequence (except for the G and F genes).

Table 3.9. Amino acid mutations and corresponding positions in the RSV genome

Gene coding protein	Identification (%)	Amino acid (%)	Number of mutation	List of common amino acid mutations
RSV-A Reference: EPI_ISL_412866-hRSV/A/England/397/2017				
NS1	99,3-100	100	0-1	
NS2	99,2-100	100	0-1	
N	99,7	100	1	V352A
P	97,4-98,8	100	2-5	L55P
M	99,6-100	100	0-1	
SH	98,5-100	100	0-1	
G	94,9-97,6	92,5	31-37	Q261del, K262del,E263del, T264del, I265del, H266del, S267del, T268del, T269del, S270del, E271del, G272del, Y273del, P274del, S275del, P276del, S277del, Q278del, V279del, Y280del, T281del, T282del, S283del, D284del
F	99,7-100	100	0-2	
M2-1	99,0-99,5	100	1-2	S176P
M2-2	96,6-98,9	100	1-3	Y24C, S44N
L	99,6-99,7	99,8-100	6-11	P171L, R256K, L1438Q, N1723S, G1731D
RSV-B Reference: EPI_ISL_1653999-hRSV/B/Australia/VIC-RCH056/2019				
NS1	99,3-100	100	0-1	
NS2	99,2100	100	0-1	
N	99,2-99,7	100	1-3	V97I
P	99,6-100	100	0-2	
M	100	100	0	
SH	98,5-100	100	0-1	D64N
G	95,8-98,8	99,7-100	1-13	A74V
F	99,5-100	100	0-3	
M2-1	99,5-100	100	0-1	
M2-2	97,8-99,8	100	1-2	I2T
L	99,6-99,9	99,6-100	3-8	T1987I

3.2.4. Phylogenetic tree of G gene and amino acid differences of RSV-A

Among 56 RSV-A isolates/clinical specimens collected in Northern Vietnam during 2017-2020, there were 53 sequences belonging to the RSV-A ON1 genotype and 3 sequences belonging to the RSV-A non-ON1 genotype (Figure 3.12). Of these, 53 sequences belonging to the RSV-A ON1 genotype were divided into 4 lineages (lineage 1-4).

Lineage 1 consists of 29 sequences, of which 22 sequences share 3 characteristic mutations T200P, P215L, and N255D. Lineage 2 consists of 5 sequences, separated from other sequences on the phylogenetic tree by the mutation Y304H. The nucleotide change that changes the stop codon to Glycine (G) (stop322G) and then the stop codon appears makes the samples in lineage 2 not only have 1 more amino acid at the C terminal of the G gene compared to the comparison sequence but also 1 amino acid longer than the remaining 51 sequences of RSV-A.

Lineage 3 consists of 9 sequences and has typical mutations to distinguish this lineage: T113I, V131D, N178G, H25Q, and H290L. Lineage 4 consists of 10 sequences with characteristic amino acid changes: K134I, E286K.

Lineage 5 consists of 3 sequences belonging to the non-ON1 genotype and this lineage is separated from the ON1, NA1, NA2 or GA2 genotypes of RSV-A by 8 mutations D6N, K11R, R15K, L24I, F25V, T42A, I45V, A63I.

Regardless of the RSV-A ON1 or RSV-A non-ON1 genotype, all RSV-A samples collected in Northern Vietnam in 2017-2020 did not have the 72 nucleotide repeat (corresponding to 24 amino acids) in the HVR region of the G gene.

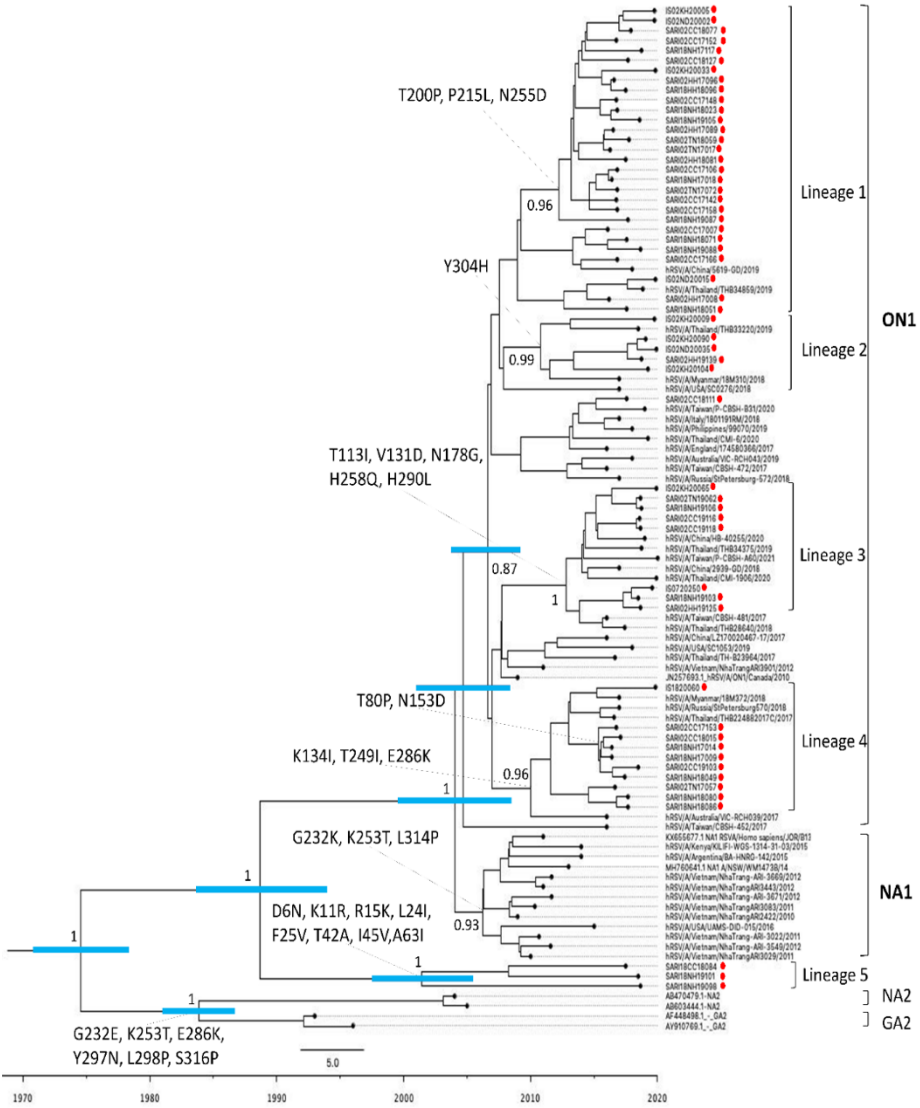

Figure 3.12. Phylogenetic tree of RSV-A subgroup G gene of samples collected at the National Children's Hospital and Quang Ninh General Hospital in the period 2017-2020.

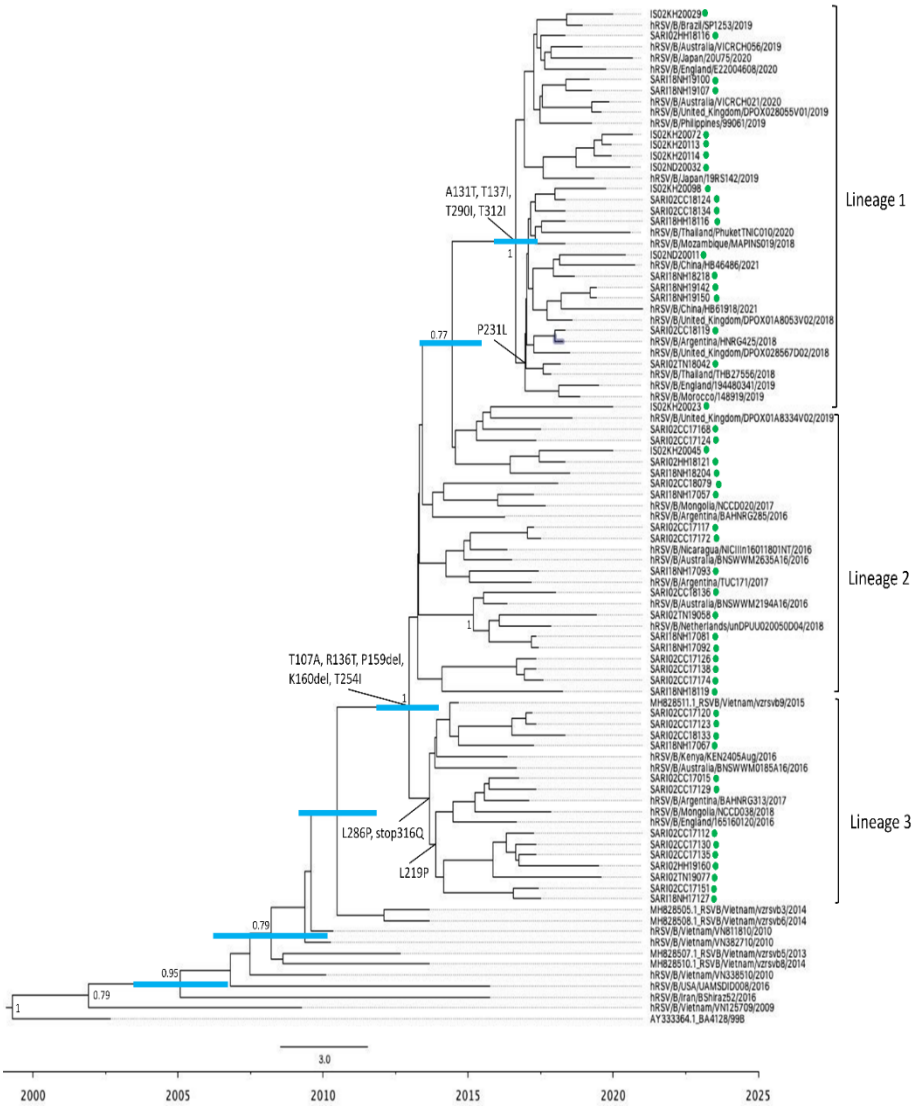

Figure 3.13. Phylogenetic tree of RSV-B subgroup G gene BA9 genotype of samples collected at the National Children's Hospital and Quang Ninh General Hospital in the period 2017-2020.

3.2.5. Phylogenetic tree of G gene and amino acid differences of RSV-B

All 50 RSV-B clinical samples/strains sequenced by NGS in the study collected in Northern Vietnam in the period 2017-2020 belonged to the RSV-B BA9 genotype (Figure 3.13). These sequences were divided into 3 lineages: lineage 1-3.

Lineage 1 consists of 18 sequences and has 4 characteristic amino acid changes: A131T, T137I, T290I, T312I. Lineage 2 consists of 19 sequences collected mainly in 2017-2018 and has no amino acid changes characteristic of this lineage. Samples belonging to lineage 1 and lineage 2 have a length of 312 amino acids, 3 amino acids shorter than the standard strain BA sequence.

Lineage 3 consists of 13 sequences and has two amino acid mutations specific to this lineage, L286P and Stop316Q. Samples belonging to this lineage have a length of 319 amino acids instead of 315 amino acids as the standard strain BA sequence due to the mutation from stop codon to Glutamine (stop316Q).

3.2.6. Evolutionary dynamics

The genetic distance between the 4 lineages from lineage 1 to lineage 4 ranges from 0.0185-0.0315. Meanwhile, lineage 5 has a much larger genetic difference than the remaining 4 lineages with p-distance = 0.0761-0.0876. At the node of lineage 5 on the phylogenetic tree, the posterior probability value of 1 proves the difference between this lineage and other genotypes.

The overall evolutionary rate of RSV-A G-gene in Northern Vietnam during 2017-2020 was estimated to be 1.32×10-3 substitutions/site/year [95% HPD: 8.81×10-4-1.74×10-3]. The time-scaled MCC phylogenetic tree of RSV-A subgroup G-gene MCC demonstrated that the time of the most recent common ancestor (tMRCA) was estimated to be around 2007 (95% HPD: 2004-2009) for the ON1 genotype. For the non-ON1 genotype (lineage

5), the time of the most recent common ancestor was estimated to be around 2002 (95% HPD: 1998-2006). The evolutionary difference between the samples in the study and the reference samples in the world belonging to the same genotype BA9 is not high, p-distance = 0.0236. According to the calculation of the molecular evolution rate over time of the RSV-B subgroup G gene, the sequences of Vietnam in the period 2017-2020 are 8.29x10-4 substitutions/position/year [95% HPD: 1.96x10-4 – 1.47x10-3]. Based on the estimate on the phylogenetic tree, the time of appearance of the most recent common ancestor of the RSV-B sequences of Vietnam genotype BA9 in the study is around 2013 (95%HPD: 2012-2014).

Chapter 4: DISCUSSION

4.1. Circulation of respiratory syncytial virus in pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital

The RSV positivity rate in this study was similar to studies in the Philippines (28.1%), Myanmar (24.5%), and China (23.7%). However, this rate was higher than reported in Thailand (13.2%), Colombia (12.02%), the United States (7.7%), and lower than Laos (41%), Brazil (56%), and Europe (46%). This shows that there is no similarity in the RSV infection rate in countries at the same time. Although having the same geographical location and climatic conditions, the RSV positivity rate is localized, and different countries have different RSV infection rates.

In general, the RSV positivity rate depends mainly on age. Detailed analysis of RSV infection rates by age groups in the 4 years of study, the group under 1-year-old not only had a large number of collected samples but also had a much higher RSV positivity rate than the other groups.

4.2. Molecular biological characteristics of respiratory syncytial virus in pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital

4.2.1. Isolation

The use of Hep-2 cells for isolation is consistent with the research practice due to its high isolation efficiency. However, it is difficult to give a general success rate of RSV isolation for all studies. If all samples positive for real-time RT-PCR with RSV are isolated without regard to high or low viral load and a cell line other than Hep-2 is selected, the success rate of isolation will be greatly reduced. Another issue is that the isolation success rate between the two RSV subtypes is significantly different. The number of clinical samples of the two subtypes isolated is almost the same, but the number of strains obtained is very different. This difference in isolation success rate is statistically significant ($p<0.05$), showing that RSV-A isolation is more efficient than RSV-B.

4.2.2. Next-generation sequencing

Currently, the trend of using clinical samples to determine genome sequences is growing and gradually replacing the isolation method. Mutations/changes in nucleotides in the viral genome can occur during the isolation process, so direct gene sequencing from clinical specimens helps to analyze the genome more accurately.

In the scope of RSV vaccine development, isolated strains are not the first choice of researchers. Based on the above facts, along with the low isolation rate, the disproportion between RSV strains according to subgroups, locations, and ages of research, the choice of combining sequencing of both isolated virus strains and clinical specimens is appropriate for molecular biological studies.

For the first time, the RSV virus gene sequence in Vietnam collected after 2015 was sequenced and published internationally. Although the start and end positions are not the same, the length of the NS1, NS2, N, P, M, F and M2-2 genes of the two subgroups are the same. The SH, M2-1 and L genes of the RSV-B subgroup are 3 nucleotides longer than those of RSV-A. This is a normal trend because RSV-A and RSV-B belong to different genotypes.

The clear difference between the two subgroups is seen in the G gene. RSV-A has a G gene of 894 or 897 nucleotides while RSV-B has a G gene of 933 or 954 nucleotides. The reason for this difference is that RSV-A in the study belongs to the ON1 and non-ON1 genotypes, both of which do not have the 72 nucleotide repeat in the HVR region. In contrast, RSV-B belongs to the BA9 genotype, which has 60 nucleotide repeats in the HVR region.

4.2.3. Molecular characteristics

Lineage 5 is the only lineage that does not belong to the ON1 genotype but to the non-ON1 genotype and has 8 amino acid differences compared to the standard strain sequence of the ON1 genotype. However, these changes are N-terminal and are all located in the viral transmembrane region and not in the extracellular ectodomain. Therefore, they do not affect the infectious function of the virus. However, if only analyzing the ectodomain or HVR region of the G gene as the majority of studies are doing, the 3 sequences of this lineage will be similar to other sequences of Vietnam with only significant differences in the 72 nucleotide repeats. Therefore, these 3 sequences will be classified on the phylogenetic tree as genotype ON1. This shows that the difference when considering the genetic characteristics of RSV on a part of the G gene and the entire length of the G gene is very different.

The genetic distance between lineage 5 and the remaining lineages/groups is large enough to identify it as a new genotype based on the average genetic distance value and the posterior probability value. This demonstrates a relatively large and significant genetic difference between the 3 Vietnamese samples of lineage 5 compared to the remaining genotypes on the phylogenetic tree.

Unlike RSV-A, RSV-B sequences are split into fewer groups, and the number of amino acid changes specific to each lineage is also less. This suggests that the genetic stability of RSV-B is relatively better than that of RSV-A.

The evolutionary rates of RSV-A and RSV-B are often similar or not much different in countries. However, RSV sequences collected in Northern Vietnam have the evolutionary rate of the G gene of RSV-A 1.6 times faster than that of RSV-B. Most studies have shown that the evolutionary rate of RSV-B is faster in Vietnam and is equivalent to that of RSV-A in the same study. Thus, the molecular evolutionary rate of the G gene of RSV-B genotype BA9 in Vietnam is generally slower than that of surrounding countries.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Results of Objective 1:

1. There is a data set on the RSV-positive rate in pediatric patients with severe acute respiratory infections at the National Children's Hospital and Quang Ninh General Hospital in the period of 2017-2020, which is 28.02%. The group under 1 year old has a higher risk of RSV infection than other age groups. Males tend to be more infected with RSV than females during the study period.

2. The circulation of two subgroups RSV-A and RSV-B at the National Children's Hospital and Quang Ninh General Hospital over time has been determined. RSV is detected almost all year round, the epidemic season is usually from June to November, and the peak epidemic is in August - September every year. 2020 is the only year where only RSV-A peaked while RSV-B circulated sporadically throughout the year.

Results of Objective 2:

3. 25 strains were isolated (18 RSV-A strains and 7 RSV-B strains) at two research sites. The entire genomes of 25 strains and 81 respiratory specimens belonging to two subgroups (56 RSV-A sequences and 50 RSV-B sequences) were sequenced.

4. The entire RSV genome sequences of the isolated strains and some representative samples were studied and determined, thereby finding out the molecular biological characteristics of the RSV genome in the study. The RSV-A sequences belong to the ON1 genotype and the non-ON1 genotype. The RSV-B sequences belong to the BA9 genotype. All RSV-A sequences in the study did not have 72 nucleotide repeats in the second hypervariable region of the G gene.

5. The molecular evolution rate of the G gene of RSV at the National Children's Hospital and Quang Ninh General Hospital in the period of 2017-2020 was assessed as 1.32×10−3 substitutions/position/year (RSV-A) and 8.29×10−4 substitutions/position/year (RSV-B). The evolutionary rate of RSV-A is similar to that of many countries. The evolutionary rate of RSV-B is slower than that of other countries.

Novelty of the research

- The circulation characteristics, seasonality, and circulation subgroups of RSV strains at the National Children's Hospital and Quang Ninh General Hospital in the period of 2017-2020 were determined.

- The entire RSV genome sequence of some representative samples was studied and determined, thereby identifying the molecular biological characteristics of the genome in the studied strains. The molecular evolution rate of the G gene of RSV at the National Children's Hospital and Quang Ninh General Hospital in the period 2017-2020 was estimated.

Recommendations

It is necessary to continue analyzing the genetic and evolutionary changes in the F and SH genes of the strains and respiratory specimens that have been sequenced in this study. The results will serve the goal of research and development of RSV vaccines and monoclonal antibodies in the future.

LIST OF THE PUBLICATIONS RELATED TO THE DISSERTATION

1. **Ung Thi Hong Trang**, Hoang Vu Mai Phuong, Nguyen Huy Hoang, Nguyen Vu Son, Le Thi Thanh, Nguyen Le Khanh Hang, Vuong Duc Cuong, Tran Thi Thu Huong, Nguyen Thi Hien, Nguyen Phuong Anh, Le Quynh Mai, "Circulation of human respiratory syncytial virus and new ON1 genotype in northern Viet Nam, 2017–2020", *Western Pacific Surveillance and Response Journal*, vol. 14, no. 4, pp. 1–9, 2023.
2. **Ung Thi Hong Trang,** Hoang Vu Mai Phuong, Nguyen Huy Hoang, Le Thi Thanh, Nguyen Vu Son, Nguyen Co Thach, Vuong Duc Cuong, Tran Thi Thu Huong, Pham Thi Hien, Nguyen Phuong Anh, Nguyen Thi Sang, Nguyen Le Khanh Hang, Le Quynh Mai, "Molecular characteristics of respiratory syncytial virus (RSV) isolated in the North Vietnam, 2017-2018", *Academia Journal of Biology*, vol. 44, no. 4, pp. 101–100, 2022.
3. **Ung Thi Hong Trang,** Hoang Vu Mai Phuong, Nguyen Huy Hoang, "Molecular epidemiology of respiratory syncytial virus", *Vietnam Journal of Preventive Medicine*, vol 32, no. 2, pp. 20–31, 2022.
4. Nguyen Thi Sang, Nguyen Vu Son, **Ung Thi Hong Trang**, Vuong Đuc Cuong, Pham Thi Hien, Tran Thi Thu Huong, Nguyen Phuong Anh, Le Thi Thanh, Nguyen Co Thach, Hoang Vu Mai Phuong, Le Thi Quynh Mai, Nguyen Le Khanh Hang, "Respiratory viruses causing severe acute respiratory infection (SARI) in hospitalized patients in Quang Ninh General Hospital, 2018-2019", *Vietnam Journal of Preventive Medicine*, vol 32, no. 2, pp. 129-138, 2022.
5. **Thi Hong Trang Ung**, Vu Mai Phuong Hoang, Huy Hoang Nguyen, Le Khanh Hang Nguyen, Phuong Anh Nguyen, Quynh Mai Le, "Molecular epidemiology of human Respiratory Syncytial virus in Viet Nam National Children's Hospital", *International Conference on Health Science and Technology (ICHST-2023),* Viet Nam, pp. 1-5, 2024.