Manuscript number: (if known)

Running title: Molecular evolution of RSV-A NA1 / ON1 in Vietnam

Keywords:

ARI, RSV, ON1, Molecular evolution, Vietnam, Paediatric infectious diseases

Title:

Molecular evolution of Respiratory Syncytial Virus subgroup A genotype NA1 and ON1 attachment glycoprotein (*G*) gene in central Vietnam

Authors list:

Keisuke Yoshihara [a,b], Minh Nhat Le [a,c], Koo Nagasawa [d], Hiroyuki Tsukagoshi [e], Hien Anh Nguyen [c], Michiko Toizumi [a], Hiroyuki Moriuchi [f,g], Masahiro Hashizume [a], Koya Ariyoshi [h], Duc Anh Dang [c], Hirokazu Kimura [d] and Lay-Myint Yoshida [a,*]

Affiliations:

a. Department of Paediatric Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, Nagasaki, 852-8523, Japan;
b. Leading Program, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, 852-8523, Japan;
c. National Institute of Hygiene and Epidemiology, Hanoi, Vietnam;
d. Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Musashimurayama, Tokyo, 208-0011, Japan;
e. Gunma Prefectural Institute of Public Health and Environmental Sciences, Maebashi, Gunma, 371-0052, Japan;
f. Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, 852-8523, Japan;
g. Department of Paediatrics, Nagasaki University Hospital, Nagasaki, 852-8102, Japan;
h. Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University, Nagasaki, 852-8523, Japan.

***Corresponding author:**

Lay-Myint Yoshida

Department of Paediatric Infectious Diseases

Institute of Tropical Medicine, Nagasaki University

1-12-4 Sakamoto, Nagasaki-city, Nagasaki, 852-8523, Japan

Phone: +81(0)95-819-7284

Fax: +81(0)95-819-7843

E-mail: lmyoshi@nagasaki-u.ac.jp

Abbreviations:

MCMC, Markov Chain Monte Carlo; AICM, Akaike's Information Criterion through MCMC; ESS, effective sample size; MCC, Maximum Clade Credibility; HPD, highest probability density; tMRCA, time to the most recent common ancestor; BSP, Bayesian Skyline Plot; EPS, effective population size; ML, Maximum Likelihood; SLAC, Conservative Single Likelihood Ancestor Counting; FEL, Fixed Effects Likelihood; IFEL, Internal Fixed Effects Likelihood; MEME, Mixed Effects Model for Episodic Diversifying Selection

Financial disclosure:

The authors have no financial relationships relevant to this article to disclose.

Funding:

This study was supported by the Japan Initiative for Global Research Network on Infectious Diseases (10008012), Japan Agency for Medical Research and Development (AMED) and Grants-in-Aid for Scientific Research (Japan Society for the Promotion of Science).

Competing interests:

The authors have declared no competing interests.

Abstract

We performed molecular evolutionary analyses of the *G* gene C-terminal 3rd hypervariable region of RSV-A genotypes NA1 and ON1 strains from the paediatric acute respiratory infection patients in central Vietnam during the 2010-2012 study period. Time-scaled phylogenetic analyses were performed using Bayesian Markov Chain Monte Carlo (MCMC) method, and pairwise distances (*p*-distances) were calculated. Bayesian Skyline Plot (BSP) was constructed to analyze the time-trend relative genetic diversity of central Vietnam RSV-A strains. We also estimated the N-glycosylation sites within *G* gene hypervariable region. Amino acid substitutions under positive and negative selection pressure were examined using Conservative Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Internal Fixed Effects Likelihood (IFEL) and Mixed Effects Model for Episodic Diversifying Selection (MEME) models. The majority of central Vietnam ON1 strains detected in 2012 were classified into lineage 1 with few positively selected substitutions. As for the Vietnamese NA1 strains, four lineages were circulating during the study period with a few positive selection sites. Shifting patterns of the predominantly circulating NA1 lineage were observed in each year during the investigation period. Median *p*-distance of central Vietnam NA1 strains was wider (*p*-distance = 0.028) than that of ON1 (*p*-distance = 0.012). The molecular evolutionary rate of central Vietnam ON1 strains was estimated to be 2.55×10^{-2} (substitutions/site/year) and was faster than NA1 (7.12×10^{-3} (substitutions/site/year)). Interestingly, the evolutionary rates of both genotypes ON1 and NA1 strains from central Vietnam were faster than the global strains respectively. Furthermore, the shifts of N-glycosylation pattern within the *G* gene 3rd hypervariable region of Vietnamese NA1 strains were observed in each year. BSP analysis indicated the rapid growth of RSV-A effective population size in early 2012. These results suggested that the molecular evolution of RSV-A *G* gene detected in central Vietnam was fast with unique evolutionary dynamics.

1. Introduction

Human respiratory syncytial virus (RSV) belongs to the genus *Pneumovirus*, the family of *Paramyxoviridae* and is one of the leading causes of acute respiratory infections (ARIs) including bronchitis, bronchiolitis and pneumonia in humans (Collins and Karron, 2013). The global death risk of RSV infection may be greater than that of seasonal influenza (Weiss and McMichael, 2004). The under 5-childhood mortality due to RSV infection has a huge socioeconomic burden in developing countries (Anderson et al., 1990; Nair et al., 2010). In addition, RSV reinfections among elderly people may result in severe ARIs such as bronchiolitis and pneumonia (Lee et al., 2013).

RSV genome encodes 10 genes that translate into 11 proteins (Collins and Karron, 2013). Of them, attachment glycoprotein (*G*) and fusion (*F*) proteins are the two major surface antigens, which play pivotal roles during infection to the host respiratory epithelial cells, although their virological functions are distinct (Collins and Karron, 2013). Moreover, the *G* gene has C-terminal 3rd hypervariable region with high genetic variability and contains epitopes that induce neutralizing antibody response (Kim et al., 2014; Melero et al., 1997; Palomo et al., 1991). Substitutions of amino acid composition within the *G* gene hypervariable region may be responsible for lifetime recurrence of RSV infection and respiratory illnesses in human (Hall et al., 1991; Tan et al., 2013).

RSV is classified into two major subgroups, RSV-A and B, based on genetic variability and antigenic characterization of the *G* gene (Anderson et al., 1985; Mufson et al., 1985). Each subgroup is further subdivided into numerous genotypes: 12 genotypes (GA1-7, SAA1, NA1-2 and ON1-2) for RSV-A and 20 genotypes (GB1-4, SAB1-4, BA1-10 and URU1-2) for RSV-B (Eshaghi et al., 2012; Hirano et al., 2014; Shobugawa et al., 2009; Trento et al., 2006). Of all the RSV genotypes, RSV-A ON1-2 and NA1, while RSV-B BA9 and 10 are the dominant types in various regions worldwide (Duvvuri et al., 2015; Hirano et al., 2014; Nagasawa et al., 2015). RSV-A genotype ON1 emerged in Canada in November 2010 (Eshaghi et al., 2012) and had rapidly spread and replaced the previously dominant NA1 in some countries (Agoti et al., 2014; Auksornkitti et al., 2013; Cui, 2013; Khor et al.,

2013; Kim et al., 2014; Panayiotou et al., 2014; Pierangeli et al., 2014; Prifert et al., 2013; Tsukagoshi et al., 2013; Valley-Omar et al., 2013). Furthermore, it was suggested that ON1 with a 72-nucleotide tandem repeat insertion within *G* gene C-terminal 3rd hypervariable region was derived from the ancestral genotype NA1 (Eshaghi et al., 2012; Hirano et al., 2014). However, detailed molecular epidemiological information on ON1 has not been clearly understood up to date in many regions including Southeast Asia.

In Vietnam, the first RSV-A ON1 related paediatric ARI hospitalization case was detected in March 2012. The emergence of genotype ON1 in our study site was associated with the dramatic increase in paediatric ARI hospitalization incidences and clinical severity of paediatric respiratory illnesses compared to previously dominant genotype NA1 (Yoshihara et al., 2016). Therefore, in the current study, we further analyzed the molecular evolutionary and antigenic characteristics of RSV-A genotypes NA1 and ON1 circulating in central Vietnam to gain a better understanding of a molecular epidemiological aspect of RSV in Vietnam.

2. Material and methods

2.1. RSV-A strains used in the current study

An ongoing paediatric ARI surveillance at Khanh Hoa province, Nha Trang, central Vietnam was utilized in the present study (Yoshida et al., 2010). After obtaining informed consent from the guardians, all the children with ARI symptoms hospitalized to the paediatric ward of Khanh Hoa General Hospital (KHGH) were enrolled in the study. Respiratory virus screening including RSV and nucleotide sequencing of the RSV *G* gene C-terminal 3rd hypervariable region for the RSV-A confirmed paediatric ARI samples were performed as previously described (Yoshida et al., 2010; Yoshihara et al., 2016). The approximate lengths of the analyzed region within *G* gene 3rd hypervariable region were 336bp for genotype ON1 and 264bp for NA1 respectively. For phylogenetic and molecular evolutionary analyses in the current study, we included a total of 236 RSV-A global reference strains from GenBank (including 93 ON1 and 125 NA1 strains) used in the

119 previous study (Hirano et al., 2014). The nucleotide sequences of RSV-A partial *G* gene from central
120 Vietnam used in the current study have been submitted to GenBank under accession numbers:
121 KX946220 – KX946477.
122 **2.2. Phylogenetic and molecular evolutionary analyses with Bayesian Markov Chain Monte**
123 **Carlo (MCMC) and Maximum Likelihood (ML) methods**
124 Nucleotide sequences of RSV-A *G* gene 3rd hypervariable region were aligned and edited using
125 ClustalW within MEGA ver.6.0.6 (Tamura et al., 2013). KAKUSAN4
126 (http://www.fifthdimension.jp/products/kakusan/) was utilized for the selection of best-fit nucleotide
127 substitution model (Tanabe, 2011). Phylogenetic and molecular evolutionary analyses were
128 performed with Bayesian Markov Chain Monte Carlo (MCMC) method using BEAST ver.1.8.0
129 (Drummond and Rambaut, 2007; Nagasawa et al., 2015; Tsukagoshi et al., 2013). In the current
130 study, four clock models (Strict clock, Uncorrelated lognormal relaxed clock, Uncorrelated
131 exponential clock and Random local clock) and four demographic models (Constant size,
132 Exponential growth, Logistic growth and Expansion growth) were compared to select the best-fit
133 model for each sequence dataset based on the value of Akaike's Information Criterion through
134 MCMC (AICM) (Suchard et al., 2001) using Tracer ver.1.6 (http://tree.bio.ed.ac.uk/software/tracer/).
135 The model with the lowest AICM value was selected to be the best-fit model in each sequence
136 dataset and used for analysis (Kimura et al., 2015; Nagasawa et al., 2015) (Supplementary Table 1).
137 The detailed condition for each Bayesian MCMC analysis was summarized in Supplementary Table
138 2. The MCMC chains were run for 200,000,000 steps for all the analyses to achieve convergence
139 with sampling every 2,000 steps. The convergence was assessed using Tracer ver.1.6, and the
140 parameters with effective sample sizes (ESS) of 200 or greater after 10% burn-in were accepted
141 (Kushibuchi et al., 2013). The time-scaled Maximum Clade Credibility (MCC) trees were generated
142 by TreeAnnotator ver.1.8.0 after removing the first 10% of trees as burn-in. The time-scaled MCC
143 trees were viewed and edited with FigTree ver.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Furthermore, the molecular evolution rates were estimated using BEAST ver.1.8.0 under the models summarized in Supplementary Table 2 (Drummond and Rambaut, 2007).

Also, the phylogenetic trees of RSV-A genotypes ON1 and NA1 were generated using Maximum Likelihood (ML) method under HKY85-gamma nucleotide substitution model with 1,000 bootstrap replications using MEGA ver.6.0.6 to estimate the evolutionary distances.

2.3. Estimation of the pairwise distance (*p*-distance) frequency distributions

In order to investigate the genetic variability of central Vietnam RSV-A NA1 and ON1 strains, the frequency distributions of pairwise distance (*p*-distance) were estimated using MEGA ver.6.0.6 as previously described (Tamura et al., 2013; Tsukagoshi et al., 2013). Strains with 100% nucleotide sequence identity were excluded from the analyses.

2.4. Bayesian Skyline Plot (BSP) analysis

To assess the time course trend of effective population size (EPS) of overall RSV-A strains circulating in central Vietnam during the investigation period, Bayesian Skyline Plot (BSP) was constructed using BEAST ver.1.8.0 as previously described (Drummond and Rambaut, 2007; Kimura et al., 2015; Nagasawa et al., 2015). KAKUSAN4 was used for the selection of best-fit nucleotide substitution model. The best-fit clock model was selected using Tracer ver.1.6 based on the AICM value comparison among four clock models (Supplementary Table 3). The MCMC chains were run for 200,000,000 steps with sampling every 2,000 steps under the uncorrelated exponential relaxed clock model and HKY85-gamma substitution model (Supplementary Table 2).

2.5. Estimation of potential N-glycosylated sites

Potential N-glycosylation sites within the *G* gene 3rd hypervariable region were analyzed using NetNGlyc ver1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). Amino acid sequences containing Asn-Xaa-Ser/Thr stretch where Xaa were any amino acids except Proline were considered to be potential N-glycosylation sites.

2.6. Positive and negative selection pressure analyses

Amino acid substitutions within the *G* gene 3rd hypervariable region under positive and negative selection pressure for central Vietnam RSV-A NA1 and ON1 strains were estimated by calculating synonymous (*d*S) and non-synonymous (*d*N) substitution rates at every codon with Datamonkey (http://www.datamonkey.org/) (Pond and Frost, 2005). Four selective pressure models, Conservative Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Internal Fixed Effects Likelihood (IFEL) and Mixed Effects Model for Episodic Diversifying Selection (MEME) models were performed for accurate estimation (Kobayashi et al., 2015). In the selection pressure analyses, *p*-values less than 0.05 were considered to be statistically significant (Kimura et al., 2015; Kobayashi et al., 2015).

2.7. Statistical analyses

Mann-Whitney U test was performed to compare the median values of *p*-distance between central Vietnam RSV-A ON1 and NA1 strains. For the proportion comparisons between two independent groups, two-tailed Fisher's exact tests were performed. Furthermore, mean molecular evolutionary rates between two independent groups were compared using Welch's test. All the statistical analyses were performed using STATA ver.12.1 (StataCorp LP, College Station, TX, USA), and *p*-values less than 0.05 were considered to be statistically significant.

3. Results

3.1. Central Vietnam RSV-A strains in the current study

During the three years investigation period (January 2010 - December 2012), we detected 123 RSV-A genotype ON1 and 138 NA1 strains as previously reported (Yoshihara et al., 2016). As shown in Table 1, RSV-A genotype NA1 was the most prevalent type in both 2010 and 2011 season. The first ON1 related paediatric ARI hospitalization case was detected in March 2012 in our study site and rapidly replaced the previously dominant NA1. Along with the dramatic genotype

192 replacement by genotype ON1, we observed an increased incidence of paediatric ARI hospitalization
193 with lower respiratory tract infection (LRTI).

194 **3.2. Phylogenetic analyses of RSV-A using Maximum Likelihood (ML) and Bayesian Markov**
195 **Chain Monte Carlo (MCMC) method**

196 To understand the genetic relationship and evolutionary distances of central Vietnam RSV-A
197 strains compared to globally circulating strains, we constructed phylogenetic trees of RSV-A
198 genotypes ON1 and NA1 respectively using Maximum Likelihood (ML) method (Fig. 1 (a) and (b)).
199 It was previously reported that both RSV-A genotypes NA1 and ON1 diverged into four genetically
200 distinct lineages (Duvvuri et al., 2015; Hirano et al., 2014). With respect to the genotype ON1 strains
201 circulating in central Vietnam, lineage 1 was the most prevalent type (n=120, 97.6%) (Table 2). As
202 for central Vietnam NA1 strains, we observed the co-circulation of multiple NA1 lineages in each
203 season during the investigation period (Table 2). However, one NA1 lineage predominated over the
204 other lineages in each year; lineage 2 predominated in 2010 (n=56, 84.9%), lineage 4 in 2011 (n=19,
205 55.9%) and lineage 1 in 2012 (n=26, 68.4%) respectively. The shifting patterns of circulating NA1
206 lineages were statistically significant during the three years investigation period.

207 Next, phylogenetic trees with Bayesian Markov Chain Monte Carlo (MCMC) method were
208 constructed in order to estimate the time-scaled evolution of RSV *G* gene hypervariable region for
209 Vietnamese NA1, ON1 and global strains, (Fig. 2). The time-scaled Maximum Clade Credibility
210 (MCC) tree illustrated that time to the most recent common ancestor (tMRCA) were estimated to be
211 around 1998 (95% highest probability density (HPD): 1996 - 2000) for NA1 and 2009 (95%HPD:
212 2007 - 2011) for ON1 genotype respectively.

213 Regarding the RSV-A genotype ON1, majority of central Vietnam ON1 strains formed a unique
214 genetic cluster within lineage 1 that possessed genetic homogeneity with Yamaguchi/ST164/2013
215 (AB808777), 1251-066AN (KC858199) and HD12101 (KJ710387) strains, all of which possessed
216 E262K amino acid substitution (Fig. 3). The time-scaled MCC tree presented that this particular

genetic cluster of ON1 lineage 1 with E262K diverged around August 2011 (95%HPD: March 2011 - January 2012).

3.3. Molecular evolution rate comparison of central Vietnam RSV-A NA1 and ON1 with global strains

The molecular evolution rate of overall central Vietnam RSV-A strains was estimated to be 8.67 $\times 10^{-3}$ (substitutions/site/year) (95%HPD: 6.27×10^{-3} - 1.16×10^{-2}), which was faster than that of global RSV-A strains reported in the previous study (5.36×10^{-3} (substitutions/site/year) (95%HPD: 4.42 - 6.39×10^{-3})) (*p*-value < 0.001) (Hirano et al., 2014) (Table 3). Furthermore, the molecular evolution rate of central Vietnam genotype ON1 strains was estimated to be 2.55×10^{-2} (substitutions/site/year) (95%HPD: 3.98×10^{-3} - 5.04×10^{-2}) and was faster than that of global ON1 strains (6.03×10^{-3} (substitutions/site/year) (95%HPD: 3.43×10^{-3} - 9.10×10^{-3}) (*p*-value < 0.001)). Similar tendency was observed in the evolutionary rate comparison between central Vietnam NA1 and global NA1 strains (*p*-value < 0.001).

3.4. Pairwise distance (*p*-distance) frequency distributions of central Vietnam RSV-A NA1 and ON1 strains

A total of 89 (42 ON1 and 47 NA1) central Vietnam RSV-A genotype confirmed strains were analyzed for pairwise distance (*p*-distance) frequency distribution. The median *p*-distances were 0.039 (IQR: 0.023 - 0.053) for overall central Vietnam RSV-A strains, 0.012 (IQR: 0.006 - 0.028) for ON1 and 0.028 (IQR: 0.019 - 0.036) for NA1 strains respectively (Supplementary Fig. 1). The difference in the median *p*-distances between central Vietnam RSV-A ON1 and NA1 strains was significant (*p*-value < 0.001), which indicated the greater genetic variability in genotype NA1 strains compared to ON1.

3.5. Phylodynamics of central Vietnam RSV-A strains with Bayesian Skyline Plot (BSP)

We assessed the time-course trend of the relative genetic diversity of *G* gene 3rd hypervariable region based on the effective population size (EPS) estimated for overall RSV-A strains circulating

in central Vietnam during the three-years investigation period. The Bayesian Skyline Plot (BSP) presented that EPS remained relatively steady during first two years of the investigation period (January 2010 - December 2011), in which NA1 had been circulating as the only RSV-A genotype (Fig. 4). However, starting from January until March 2012, the rapid increase in the size of EPS was observed, which was followed by a short and relatively stable stage of EPS until the end of the investigation period. In fact, the timing of rapid EPS growth detected in the current study was consistent with the first appearance of ON1 related ARI hospitalization case in central Vietnam in March 2012 (Yoshihara et al., 2016).

3.6. Estimation of N-glycosylation sites in central Vietnam RSV-A NA1 and ON1 strains

Regarding the N-glycosylation sites within *G* gene 3rd hypervariable region of central Vietnam NA1 strains, the patterns of N-glycosylation site shifted every year during the investigation period (Table 4). The majority of NA1 strains from 2010 possessed two conserved N-glycosylation sites at 251-254 (NTTG) and 273-276 (NLSP). In addition, N-glycosylation at 318-321(NTTK / NTTE) became prevalent in 2011 due to I320T substitution. The overall N-glycosylation pattern of central Vietnam NA1 strains between 2010 and 2011 seasons was significantly different (p-value < 0.001). The NA1 strains from 2012 had three common N-glycosylated sites at 237-240 (NTTK), 251-254 (NTTG) and 318-321 (NTTK / NTTE), the pattern of which was significantly different from that of 2011 (p-value < 0.001). In 2012, N-glycosylation at 237-240 (NTTK) became prevalent due to D237N substitution whereas 273-276 (NLSP) became less common due to N273Y. Importantly, each genetically distinct NA1 lineage (1 - 4) detected in the current study possessed distinct N-glycosylation pattern (Supplementary Table 4).

With respect to the N-glycosylation pattern of central Vietnam ON1 strains, two conserved N-glycosylation sites at 237-240 (NTTK / NTSK) and 318-321 (NTTK) were detected (Table 4). In comparison with the N-glycosylation pattern of overall central Vietnam NA1 strains (January 2010 - December 2012), N-glycosylation at 251-254 (NTTG) and 273-276 (NLSP) were less prevalent due

to common T253K and N273Y substitutions among ON1 strains. Overall, N-glycosylation pattern of central Vietnam NA1 and ON1 strains were significantly different (p-value < 0.001) (Table 4).

3.7. Selection pressure analyses among central Vietnam RSV-A NA1 and ON1 strains

We estimated the amino acid substitution sites under positive selection pressure using four selection models: SLAC, FEL, IFEL and MEME. Several positively selected amino acid substitutions were detected among the central Vietnam NA1 strains (Table 5). In 2010, substitutions at L215P, N237D, S250F, K262E, P274L and T320I were under positive selection pressure, of which K262E and P274L were agreed by all four selection models. There was no positive selection site among NA1 strains in 2011. NA1 strains from 2012 season possessed D237N, F250S, N273Y and I320T positively selected substitutions.

Regarding the central Vietnam genotype ON1 strains, amino acid substitutions at D237N, E262K and P314L were under positive selection pressure (Table 5). Among them, substitution at E262K was agreed by all four selection models.

In addition to the positively selected amino acid substitution sites, we also estimated the amino acid substitutions under negative selection pressure using three selection models: SLAC, FEL and IFEL models. There was a total of six amino acid substitutions, I236, T245, T246, T249, E263 and S317, under negative selection pressure among the overall central Vietnam RSV-A strains (Supplementary Table 5).

4. Discussion

In the present study, we performed molecular evolutionary analyses of the *G* gene C-terminal 3rd hypervariable region in RSV-A genotypes NA1 and ON1 strains detected from the paediatric ARI surveillance in Khanh Hoa General Hospital, Nha Trang, Vietnam during 2010 - 2012 season. In the previous study, we described that NA1 was the dominant genotype among RSV-A during January 2010 - December 2011 in our study site; however, ON1 suddenly emerged in March 2012 and rapidly spread in the studied population (Yoshihara et al., 2016). In the current study, we found that

majority of Vietnamese ON1 strains with a few positively selected amino acid substitutions were classified into lineage 1 with E262K, while co-circulation of multiple NA1 lineages was observed during the investigation period. With respect to central Vietnam NA1 strains, the pattern of circulating lineages drastically shifted in each season. The genetic variation of Vietnamese NA1 strains was greater than that of ON1. Interestingly, the evolutionary rate of the *G* gene 3rd hypervariable region among central Vietnam ON1 strains was faster than that of NA1. Regarding the N-glycosylation pattern within the *G* gene hypervariable region, significant shifts were observed among Vietnamese NA1 strains in each year. Importantly, the overall N-glycosylation pattern of Vietnamese NA1 and ON1 strains were significantly different. Moreover, the Bayesian Skyline Plot (BSP) presented the rapid growth of effective population size (EPS) in early 2012. Therefore, these results suggested that RSV-A genotypes ON1 and NA1 in our study site not be the only prevalent types but also were rapidly evolving and possessed unique antigenic characteristics.

The genotype ON1 was firstly detected in Canada in November 2010 (Eshaghi et al., 2012) and derived from the ancestral genotype NA1 (Eshaghi et al., 2012; Hirano et al., 2014). ON1 has 72-nucleotide tandem repeat insertion (corresponds to 24-amino acid) within the *G* gene C-terminal 3rd hypervariable region. Since the initial discovery of genotype ON1, it has rapidly spread to various countries worldwide (Agoti et al., 2014; Auksornkitti et al., 2013; Avadhanula et al., 2015; Balmaks et al., 2014; Choudhary et al., 2013; Cui, 2013; Eshaghi et al., 2012; Gimferrer et al., 2015; Hirano et al., 2014; Khor et al., 2013; Kim et al., 2014; Lee et al., 2012; Malasao et al., 2015; Panayiotou et al., 2014; Pierangeli et al., 2014; Prifert et al., 2013; Ren et al., 2014; Tabatabai et al., 2014; Tsukagoshi et al., 2013; Valley-Omar et al., 2013). The time-scaled Maximum Clade Credibility (MCC) tree suggested that ON1 diverged from the ancestral NA1 around 2009, which diverged from NA2 around 1999 (Fig. 2). Previous reports also suggested that ON1 emerged around the same period (Agoti et al., 2014; Hirano et al., 2014; Kim et al., 2014).

316 Recent studies described both genotypes NA1 and ON1 had diverged into four genetically
317 distinct lineages and the new ON1 variant, namely ON2 (Duvvuri et al., 2015; Hirano et al., 2014).
318 In the current study, circulation of multiple NA1 lineages with one dominant lineage was detected in
319 each year during the investigation period (Table 2). Among the central Vietnam ON1 strains, lineage
320 1 was the predominant type and possessed genetic similarity with Yamaguchi/ST164/2013
321 (AB808777), 1251-066AN (KC858199) and HD12101 (KJ710387), all of which possessed E262K
322 substitution (Fig. 3). In fact, E262K was located within one of the previously described B-cell
323 epitope regions of ON1 *G* gene 3rd hypervariable region at 251-265 (NTKGNPEHTSQEETL) in
324 GN425/11 (JX627336) and GG818/12 (AB860239) (Kim et al., 2014), which may link with host
325 immune evasion mechanism. ON1 strains from other neighboring countries including Malaysia
326 (2011-2012), Philippine (2012-2013) and Thailand (2010-2012) also presented lineage 1 as the
327 prevalent ON1 type (Duvvuri et al., 2015). These results suggested that ON1 have suddenly emerged
328 in various Asian countries including Vietnam around 2012. To our best knowledge, this study is the
329 first report regarding the molecular evolutionary analysis of RSV-A genotypes ON1 and NA1 strains
330 in Vietnam.
331 The evolutionary rate of the central Vietnam RSV-A genotypes ON1 and NA1 strains were
332 estimated to be 2.55×10^{-2} (substitutions/site/year) and 7.12×10^{-3} (substitutions/site/year),
333 respectively (Table 3). The previous report by Hirano et al. presented that the molecular evolutionary
334 rate of ON1 and NA1 detected in Japan and other countries were estimated to be 6.03×10^{-3}
335 (substitutions/site/year) for ON1 and 4.61×10^{-3} (substitutions/site/year) for NA1 (Hirano et al.,
336 2014). Our current results suggested that each Vietnamese RSV-A ON1 and NA1 strains were
337 evolving with faster rates than other global strains. Although we could not elucidate the exact reason
338 for the difference in evolutionary rates, factors such as the difference in investigation period
339 (Kushibuchi et al., 2013), the herd immunity against RSV in the community, pre- and co-circulating
340 viruses and genotypes in the studied area may affect the viral evolutionary rate. In addition, the

341 difference in RSV seasonality in tropical and temperate climate region may also play an important
342 role in the viral evolution. Further surveillance with a larger population would be necessary to gain a
343 better understating of the molecular evolutionary characteristics of RSV globally.

344 To assess the phylodynamics of the RSV *G* gene hypervariable region of both Vietnamese ON1
345 and NA1 strains, we analyzed the time-trend effective population size (EPS) during the investigation
346 period using Bayesian Skyline Plot (BSP) (Fig. 4). The result of BSP suggested that size in EPS of
347 RSV-A remained relatively constant during January 2010 - December 2011, which was followed by
348 a slight dip in EPS in late 2011. One possible explanation for a slight dip in EPS may be viral
349 interference with other circulating viruses such as influenza viruses and previously circulating RSV
350 genotypes. Indeed, the dip in EPS that we detected around November 2011 - January 2012 period
351 coincided with seasonal influenza season in central Vietnam. However, it is difficult to confirm this
352 factor without detailed influenza and other respiratory viruses' seroprevalence and virus monitoring
353 data during the current investigation period. Furthermore, the rapid EPS growth was detected,
354 starting from January until March 2012. Interestingly, the timing of rapid EPS growth was consistent
355 with the detection of the first ON1 related ARI hospitalization case in our study site. Also, the
356 previous studies from Korea and Philippine demonstrated similar findings (Kim et al., 2014; Malasao
357 et al., 2015). In our study site, we observed dramatic RSV-A genotype replacement of previously
358 dominant NA1 (in 2010 and 2011) by genotype ON1 in 2012 season, which associated with
359 increased LRTI associated paediatric ARI hospitalizations and clinical severity of paediatric
360 respiratory illnesses (Yoshihara et al., 2016). This could be the primary reason why we observed the
361 exponential growth in the size of EPS as the previous literature described that relative genetic
362 diversity of the RSV-A *G* gene hypervariable regions were significantly correlated with the
363 emergence and prevalence of the ON1 genotype (Kim et al., 2014).

364 Furthermore, we assessed the potential N-glycosylated sites within the C-terminal 3rd
365 hypervariable region of *G* gene. Our results presented that the pattern of N-glycosylation among

366 Vietnamese NA1 strains shifted every year (Table 4), which was consistent with the shifting of
367 dominantly circulating NA1 lineage in each year (Supplementary Table 4). Alteration of the NA1
368 N-glycosylation pattern may impart various biological advantages such as viral antigenicity,
369 virulence, host immune evasion, etc. (Vigerust and Shepherd, 2007). Therefore, shifting of the
370 prevalent NA1 lineage with lineage-specific N-glycosylation pattern in each season might be due to
371 the viral evolutionary mechanism for host immune evasion. More importantly, the overall pattern and
372 the number of the *G* gene hypervariable region N-glycosylated site were significantly different
373 between central Vietnam NA1 and ON1 strains (p-value < 0.001) (Table 4). It was previously
374 described that addition of N-glycosylation within haemagglutinin (HA) glycoprotein globular head
375 region of the 1918 H1N1 influenza strain was linked with attenuation of viral virulence (Sun et al.,
376 2013). Therefore, it is necessary to further investigate the biological role of N-glycosylation pattern
377 within *G* gene hypervariable region on RSV-A genotype-specific viral antigenicity and virulence.
378 Selection pressure on RSV may be associated with the molecular evolutionary mechanism of
379 the C-terminal 3rd hypervariable regions of *G* gene (Melero et al., 1997). In the current study, we
380 found several amino acid substitutions under positive selection pressure (Table 5). With regard to the
381 Vietnamese NA1 strains, the positive selection at K262E, P274L and I320T were agreed by all four
382 selection models (SLAC, FEL, IFEL and MEME), which were within or genetically proximity to the
383 potential *G* gene N-glycosylation sites. Furthermore, the positive selections at 237, 250 and 321
384 underwent “flip-flop” substitutions, which may be associated with the viral evolutionary mechanism
385 for host immune evasion as previously described (Botosso et al., 2009; Palomo et al., 2000).
386 Moreover, among several positively selected substitutions in Vietnamese ON1 strains, E262K was
387 located within one of the previously described B-cell epitope regions (Kim et al., 2014). Therefore,
388 these results suggested that positive selection pressure was likely to be associated with the alteration
389 of viral antigenicity and may play a major role in host immune evasion mechanism. In addition to the
390 positive selection pressure, several amino acid substitutions under negative selection pressure were

391 detected within the *G* gene 3rd hypervariable region using SLAC, FEL and IFEL models. These may
392 be associated with the viral mechanism for preventing the deterioration of functional viral protein
393 (Domingo, 2006).

394 **5. Conclusions**

395 In our study site, genotype NA1 was the dominant type in 2010 and 2011, and then ON1
396 emerged and immediately became predominant type in 2012. All four NA1 lineages were observed
397 with the shifting patterns of the dominant lineage in each season while lineage 1 was the major type
398 among ON1 strains during the investigation period. The genetic variability of Vietnamese NA1
399 strains was greater than ON1; however, the faster molecular evolutionary rate was found in ON1
400 strains. Furthermore, significant shifts of the N-glycosylation pattern within *G* gene hypervariable
401 region of NA1 strains were observed every year. In comparison to NA1, fewer and distinct pattern of
402 N-glycosylated sites were found in the Vietnamese ON1 strains. Therefore, our finding in the current
403 study indicated that RSV-A ON1 and NA1 strains circulating in central Vietnam area were rapidly
404 evolving and possessed unique antigenic characteristics.
405

6. Acknowledgement

We would like to express our sincere gratitude to the medical doctors, nurses and laboratory technicians at Khanh Hoa General Hospital and staffs in Khanh Hoa Health Service for their kind support on ARI epidemiological data and clinical sample collection. We are also thankful to Dr. Tsuyoshi Hamada (Nagasaki Advanced Computing Center, Nagasaki University) for assisting the phylogenetic and molecular evolutionary analyses using super-computer “DEGIMA”. Lastly, we would like to express our special appreciation to the staffs and professors in leading program at Nagasaki University, Graduate School of Biomedical Sciences for their kind support during the research period. This study was conducted (in part) at the Joint Usage / Research Center on Tropical Disease, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan.

7. References

Agoti, C.N., Otieno, J.R., Gitahi, C.W., Cane, P.A., Nokes, D.J., 2014. Rapid spread and diversification of respiratory syncytial virus genotype ON1, Kenya. Emerg Infect Dis 20, 950-959.

Anderson, L.J., Hierholzer, J.C., Tsou, C., Hendry, R.M., Fernie, B.F., Stone, Y., McIntosh, K., 1985. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. J Infect Dis 151, 626-633.

Anderson, L.J., Parker, R.A., Strikas, R.L., 1990. Association between respiratory syncytial virus outbreaks and lower respiratory tract deaths of infants and young children. J Infect Dis 161, 640-646.

Auksornkitti, V., Kamprasert, N., Thongkomplew, S., Suwannakarn, K., Theamboonlers, A., Samransamruajkij, R., Poovorawan, Y., 2013. Molecular characterization of human respiratory syncytial virus, 2010-2011: identification of genotype ON1 and a new subgroup B genotype in Thailand. Arch Virol 159, 499-507.

Avadhanula, V., Chemaly, R.F., Shah, D.P., Ghantoji, S.S., Azzi, J.M., Aideyan, L.O., Mei, M., Piedra, P.A., 2015. Infection with novel respiratory syncytial virus genotype Ontario (ON1) in adult hematopoietic cell transplant recipients, Texas, 2011-2013. J Infect Dis 211, 582-589.

Balmaks, R., Ribakova, I., Gardovska, D., Kazaks, A., 2014. Molecular epidemiology of human respiratory syncytial virus over three consecutive seasons in Latvia. J Med Virol 86, 1971-1982.

Botosso, V.F., Zanotto, P.M., Ueda, M., Arruda, E., Gilio, A.E., Vieira, S.E., Stewien, K.E., Peret, T.C., Jamal, L.F., Pardini, M.I., Pinho, J.R., Massad, E., Sant'anna, O.A., Holmes, E.C., Durigon, E.L., Consortium, V., 2009. Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. PLoS Pathog 5, e1000254.

Choudhary, M.L., Wadhwa, B.S., Jadhav, S.M., Chadha, M.S., 2013. Complete Genome Sequences of Two Human Respiratory Syncytial Virus Genotype A Strains from India, RSV-A/NIV1114046/11 and RSV-A/NIV1114073/11. Genome Announc 1, e00165.

Collins, P.L., Karron, R.A., 2013. Respiratory Syncytial Virus and Metapneumovirus, in: Knipe, D.M., Howley, P.M., Cohen, J.I., Griffin, D.E., Lamb, R.A., Martin, M.A., Racaniello, V.R., Roizman, B. (Ed.), Fields Virology, Sixth ed. Lippincott Williams & Wilkins, Philadelphia pp. 1086-1123.

Cui, G., 2013. Emerging human respiratory syncytial virus genotype ON1 found in infants with pneumonia in Beijing, China. Emerging Microbes and Infections 2, e22.

Domingo, E., 2006. Virus evolution, in: Knipe, D.M., Howley, P.M., Griffin, D.E., Martin, M.A., Lamb, R.A., Roizman, B., Straus, S.E. (Ed.), Fields Virology, Fifth ed. Linppincott Williams & Wilkins, Philadelphia, pp. 389-421.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7, 214.

Duvvuri, V.R., Granados, A., Rosenfeld, P., Bahl, J., Eshaghi, A., Gubbay, J.B., 2015. Genetic diversity and evolutionary insights of respiratory syncytial virus A ON1 genotype: global and local transmission dynamics. Sci Rep 5, 14268.

Eshaghi, A., Duvvuri, V.R., Lai, R., Nadarajah, J.T., Li, A., Patel, S.N., Low, D.E., Gubbay, J.B., 2012. Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: a novel genotype with a 72 nucleotide G gene duplication. PLoS One 7, e32807.

Gimferrer, L., Campins, M., Codina, M.G., Martin Mdel, C., Fuentes, F., Esperalba, J., Bruguera, A., Vilca, L.M., Armadans, L., Pumarola, T., Anton, A., 2015. Molecular epidemiology and molecular characterization of respiratory syncytial viruses at a tertiary care university hospital in Catalonia (Spain) during the 2013-2014 season. J Clin Virol 66, 27-32.

Hall, C.B., Walsh, E.E., Long, C.E., Schnabel, K.C., 1991. Immunity to and frequency of reinfection with respiratory syncytial virus. J Infect Dis 163, 693-698.

Hirano, E., Kobayashi, M., Tsukagoshi, H., Yoshida, L.M., Kuroda, M., Noda, M., Ishioka, T., Kozawa, K., Ishii, H., Yoshida, A., Oishi, K., Ryo, A., Kimura, H., 2014. Molecular evolution of human respiratory syncytial virus attachment glycoprotein (G) gene of new genotype ON1 and ancestor NA1. Infect Genet Evol 28, 183-191.

Khor, C.S., Sam, I.C., Hooi, P.S., Chan, Y.F., 2013. Displacement of predominant respiratory syncytial virus genotypes in Malaysia between 1989 and 2011. Infect Genet Evol 14, 357-360.

Kim, Y.J., Kim, D.W., Lee, W.J., Yun, M.R., Lee, H.Y., Lee, H.S., Jung, H.D., Kim, K., 2014. Rapid replacement of human respiratory syncytial virus A with the ON1 genotype having 72 nucleotide duplication in G gene. Infect Genet Evol 26, 103-112.

Kimura, H., Saitoh, M., Kobayashi, M., Ishii, H., Saraya, T., Kurai, D., Tsukagoshi, H., Shirabe, K., Nishina, A., Kozawa, K., Kuroda, M., Takeuchi, F., Sekizuka, T., Minakami, H., Ryo, A., Takeda, M., 2015. Molecular evolution of haemagglutinin (H) gene in measles virus. Sci Rep 5, 11648.

Kobayashi, M., Yoshizumi, S., Kogawa, S., Takahashi, T., Ueki, Y., Shinohara, M., Mizukoshi, F., Tsukagoshi, H., Sasaki, Y., Suzuki, R., Shimizu, H., Iwakiri, A., Okabe, N., Shirabe, K., Shinomiya, H., Kozawa, K., Kusunoki, H., Ryo, A., Kuroda, M., Katayama, K., Kimura, H., 2015. Molecular Evolution of the Capsid Gene in Norovirus Genogroup I. Sci Rep 5, 13806.

Kushibuchi, I., Kobayashi, M., Kusaka, T., Tsukagoshi, H., Ryo, A., Yoshida, A., Ishii, H., Saraya, T., Kurai, D., Yamamoto, N., Kanou, K., Saitoh, M., Noda, M., Kuroda, M., Morita, Y., Kozawa, K., Oishi, K., Tashiro, M., Kimura, H., 2013. Molecular evolution of attachment glycoprotein (G) gene in human respiratory syncytial virus detected in Japan 2008-2011. Infect Genet Evol 18, 168-173.

Lee, N., Lui, G.C., Wong, K.T., Li, T.C., Tse, E.C., Chan, J.Y., Yu, J., Wong, S.S., Choi, K.W., Wong, R.Y., Ngai, K.L., Hui, D.S., Chan, P.K., 2013. High morbidity and mortality in adults hospitalized for respiratory syncytial virus infections. Clin Infect Dis 57, 1069-1077.

Lee, W.J., Kim, Y.J., Kim, D.W., Lee, H.S., Lee, H.Y., Kim, K., 2012. Complete genome sequence of human respiratory syncytial virus genotype A with a 72-nucleotide duplication in the attachment protein G gene. J Virol 86, 13810-13811.

Malasao, R., Okamoto, M., Chaimongkol, N., Imamura, T., Tohma, K., Dapat, I., Dapat, C., Suzuki, A., Saito, M., Saito, M., Tamaki, R., Pedrera-Rico, G.A., Aniceto, R., Quicho, R.F., Segubre-Mercado, E., Lupisan, S., Oshitani, H., 2015. Molecular Characterization of Human Respiratory Syncytial Virus in the Philippines, 2012-2013. PLoS One 10, e0142192.

Melero, J.A., Garcia-Barreno, B., Martinez, I., Pringle, C.R., Cane, P.A., 1997. Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. J Gen Virol 78 (Pt 10), 2411-2418.

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

Nagasawa, K., Hirano, E., Kobayashi, M., Ryo, A., Oishi, K., Obuchi, M., Ishiwada, N., Noda, M., Kuroda, M., Shimojo, N., Kimura, H., 2015. Molecular evolution of the hypervariable region of the attachment glycoprotein gene in human respiratory syncytial virus subgroup B genotypes BA9 and BA10. Infect Genet Evol 36, 217-223.

Nair, H., Nokes, D.J., Gessner, B.D., Dherani, M., Madhi, S.A., Singleton, R.J., O'Brien, K.L., Roca, A., Wright, P.F., Bruce, N., Chandran, A., Theodoratou, E., Sutanto, A., Sedyaningsih, E.R., Ngama, M., Munywoki, P.K., Kartasasmita, C., Simoes, E.A., Rudan, I., Weber, M.W., Campbell, H., 2010. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet 375, 1545-1555.

Palomo, C., Cane, P.A., Melero, J.A., 2000. Evaluation of the antibody specificities of human convalescent-phase sera against the attachment (G) protein of human respiratory syncytial virus: influence of strain variation and carbohydrate side chains. J Med Virol 60, 468-474.

Palomo, C., Garcia-Barreno, B., Penas, C., Melero, J.A., 1991. The G protein of human respiratory syncytial virus: significance of carbohydrate side-chains and the C-terminal end to its antigenicity. J Gen Virol 72 (Pt 3), 669-675.

Panayiotou, C., Richter, J., Koliou, M., Kalogirou, N., Georgiou, E., Christodoulou, C., 2014. 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. Epidemiol Infect 142, 2406-2411.

Pierangeli, A., Trotta, D., Scagnolari, C., Ferreri, M.L., Nicolai, A., Midulla, F., Marinelli, K., Antonelli, G., Bagnarelli, P., 2014. Rapid spread of the novel respiratory syncytial virus A ON1 genotype, central Italy, 2011 to 2013. Euro Surveill 19, 1-10.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531-2533.

Prifert, C., Streng, A., Krempl, C.D., Liese, J., Weissbrich, B., 2013. Novel respiratory syncytial virus a genotype, Germany, 2011-2012. Emerg Infect Dis 19, 1029-1030.

Ren, L., Xia, Q., Xiao, Q., Zhou, L., Zang, N., Long, X., Xie, X., Deng, Y., Wang, L., Fu, Z., Tian, D., Zhao, Y., Zhao, X., Li, T., Huang, A., Liu, E., 2014. The genetic variability of glycoproteins among respiratory syncytial virus subtype A in China between 2009 and 2013. Infect Genet Evol 27, 339-347.

Shobugawa, Y., Saito, R., Sano, Y., Zaraket, H., Suzuki, Y., Kumaki, A., Dapat, I., Oguma, T., Yamaguchi, M., Suzuki, H., 2009. Emerging genotypes of human respiratory syncytial virus subgroup A among patients in Japan. J Clin Microbiol 47, 2475-2482.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18, 1001-1013.

Sun, X., Jayaraman, A., Maniprasad, P., Raman, R., Houser, K.V., Pappas, C., Zeng, H., Sasisekharan, R., Katz, J.M., Tumpey, T.M., 2013. N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. J Virol 87, 8756-8766.

Tabatabai, J., Prifert, C., Pfeil, J., Grulich-Henn, J., Schnitzler, P., 2014. Novel respiratory syncytial virus (RSV) genotype ON1 predominates in Germany during winter season 2012-13. PLoS One 9, e109191.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-2729.

Tan, L., Coenjaerts, F.E., Houspie, L., Viveen, M.C., van Bleek, G.M., Wiertz, E.J., Martin, D.P., Lemey, P., 2013. The comparative genomics of human respiratory syncytial virus subgroups A and B: genetic variability and molecular evolutionary dynamics. J Virol 87, 8213-8226.

Tanabe, A.S., 2011. Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. Mol Ecol Resour 11, 914-921.

Trento, A., Viegas, M., Galiano, M., Videla, C., Carballal, G., Mistchenko, A.S., Melero, J.A., 2006. Natural history of human respiratory syncytial virus inferred from phylogenetic analysis of the attachment (G) glycoprotein with a 60-nucleotide duplication. J Virol 80, 975-984.

Tsukagoshi, H., Yokoi, H., Kobayashi, M., Kushibuchi, I., Okamoto-Nakagawa, R., Yoshida, A., Morita, Y., Noda, M., Yamamoto, N., Sugai, K., Oishi, K., Kozawa, K., Kuroda, M., Shirabe, K., Kimura, H., 2013. Genetic analysis of attachment glycoprotein (G) gene in new genotype ON1 of human respiratory syncytial virus detected in Japan. Microbiol Immunol 57, 655-659.

Valley-Omar, Z., Muloiwa, R., Hu, N.C., Eley, B., Hsiao, N.Y., 2013. Novel respiratory syncytial virus subtype ON1 among children, Cape Town, South Africa, 2012. Emerg Infect Dis 19, 668-670.

Vigerust, D.J., Shepherd, V.L., 2007. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol 15, 211-218.

Weiss, R.A., McMichael, A.J., 2004. Social and environmental risk factors in the emergence of infectious diseases. Nat Med 10, S70-76.

Yoshida, L.M., Suzuki, M., Yamamoto, T., Nguyen, H.A., Nguyen, C.D., Nguyen, A.T., Oishi, K., Vu, T.D., Le, T.H., Le, M.Q., Yanai, H., Kilgore, P.E., Dang, D.A., Ariyoshi, K., 2010. Viral pathogens associated with acute respiratory infections in central vietnamese children. Pediatr Infect Dis J 29, 75-77.

Yoshihara, K., Le, M.N., Okamoto, M., Wadagni, A.C., Nguyen, H.A., Toizumi, M., Pham, E., Suzuki, M., Nguyen, A.T., Oshitani, H., Ariyoshi, K., Moriuchi, H., Hashizume, M., Dang, D.A., Yoshida, L.M., 2016. Association of RSV-A ON1 genotype with Increased Pediatric Acute Lower Respiratory Tract Infection in Vietnam. Sci Rep 6, 27856.

Highlights:

- Shifting patterns of RSV-A NA1 linages were observed in 2010 - 2012 seasons.
- RSV-A ON1 lineage 1 with E262K appeared as the major genetic cluster in 2012.
- Vietnamese RSV-A ON1 and NA1 were evolving with the faster substitution rates.
- N-glycosylation patterns within *G* gene of Vietnamese NA1 shifted every year.
- N-glycosylation pattern of Vietnamese NA1 and ON1 strains were distinct.

Figures

Figure legend

Figure 1 (a) and (b). Phylogenetic trees of RSV-A ON1 / 2 (a) and NA1 (b) *G* gene hypervariable region with Maximum Likelihood (ML) method

RSV-A strains from central Vietnam are indicated as the RED-colored line. Bootstrap values higher than 70 were considered to be statistically significant and shown at each branch node. Scale bar indicates nucleotide substitutions per site.

Figure 2. Time-scaled phylogenetic tree of RSV-A *G* gene hypervariable region with Bayesian Markov Chain Monte Carlo (MCMC)

RSV-A strains from central Vietnams are indicated as the BLACK-filled objects. Uncertainties for the year of divergences are indicated as 95% highest probability density (95%HPD) in the horizontal light BLUE boxes at each branch point. Scale bar represents the unit of time (in year).

Figure 3. Time-scaled phylogenetic tree of RSV-A ON1 / 2 *G* gene hypervariable region with Bayesian Markov Chain Monte Carlo (MCMC)

RSV-A strains from central Vietnam are indicated as "●" with the 4-digit strain specific ID numbers. Uncertainties for the year of divergences are indicated as 95% highest probability density (95%HPD) in the horizontal light BLUE boxes at each branch point. Scale bar represents the unit of time (in year).

Figure 4. Phylodynamics of central Vietnam RSV-A strains *G* gene hypervariable region with Bayesian Skyline Plot (BSP)

The Y-axis represents the effective population size, and the X-axis represents the generation time (in year). The BLACK-solid line indicates the median effective population size, and the two light purple lines present the range for 95% highest probability density (HPD).

Figure 1 (a) and (b). Phylogenetic trees of RSV-A ON1 / 2 (a) and NA1 (b) *G* gene hypervariable region with Maximum Likelihood (ML) method

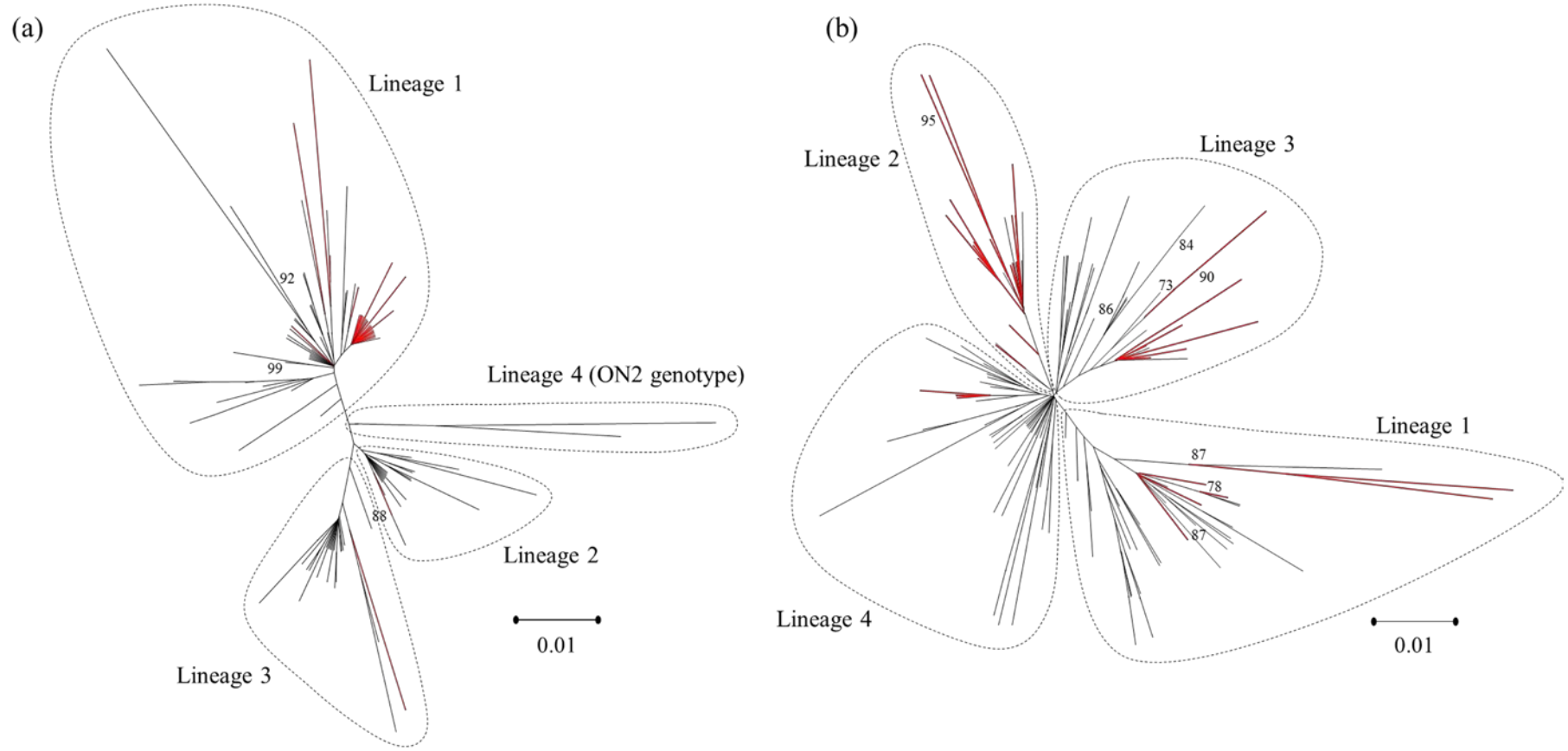


Figure 2. Time-scaled phylogenetic tree of RSV-A *G* gene hypervariable region with Bayesian Markov Chain Monte Carlo (MCMC)

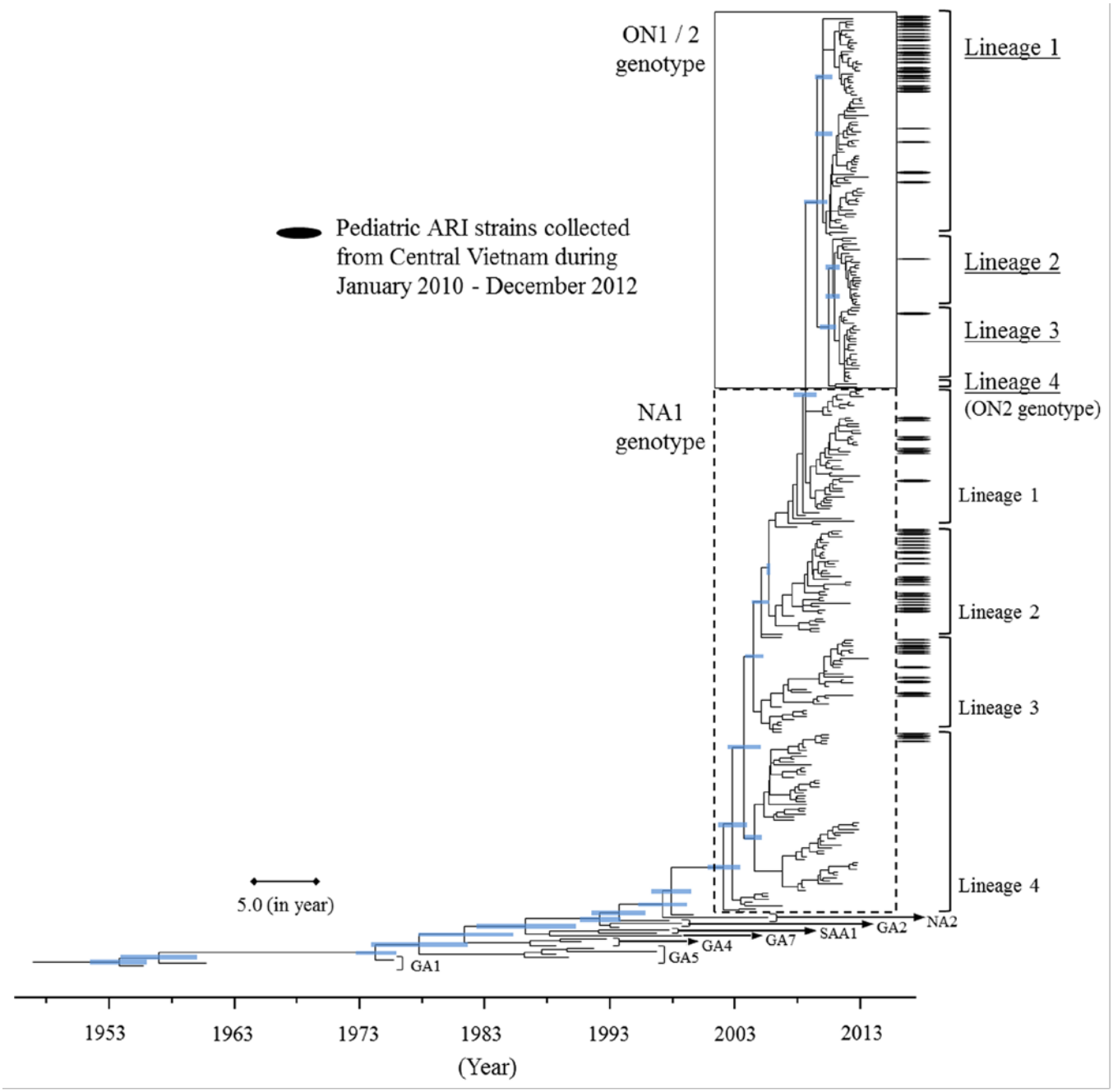


Figure 3. Time-scaled phylogenetic tree of RSV-A ON1 / 2 *G* gene hypervariable region with Bayesian Markov Chain Monte Carlo (MCMC)

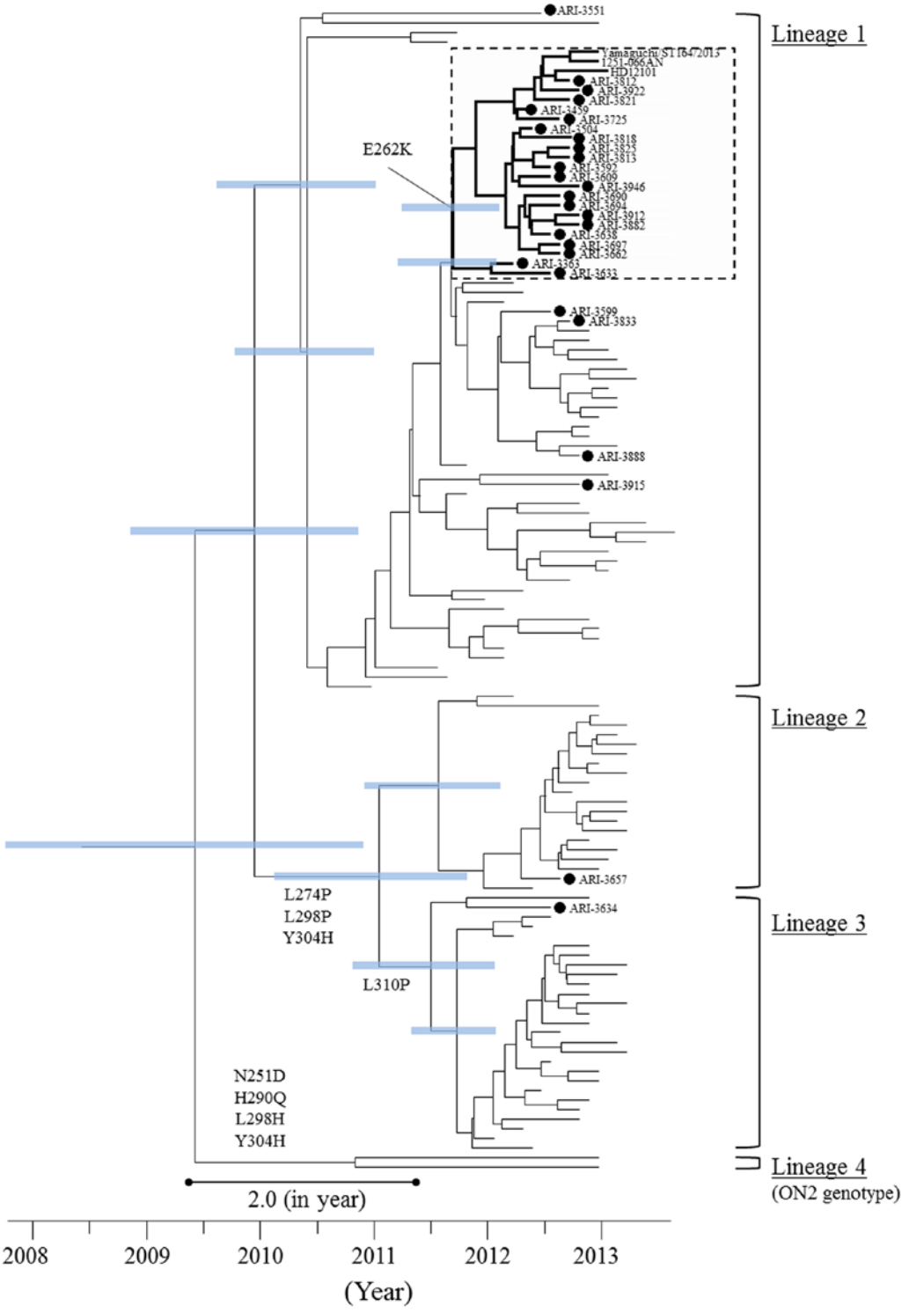
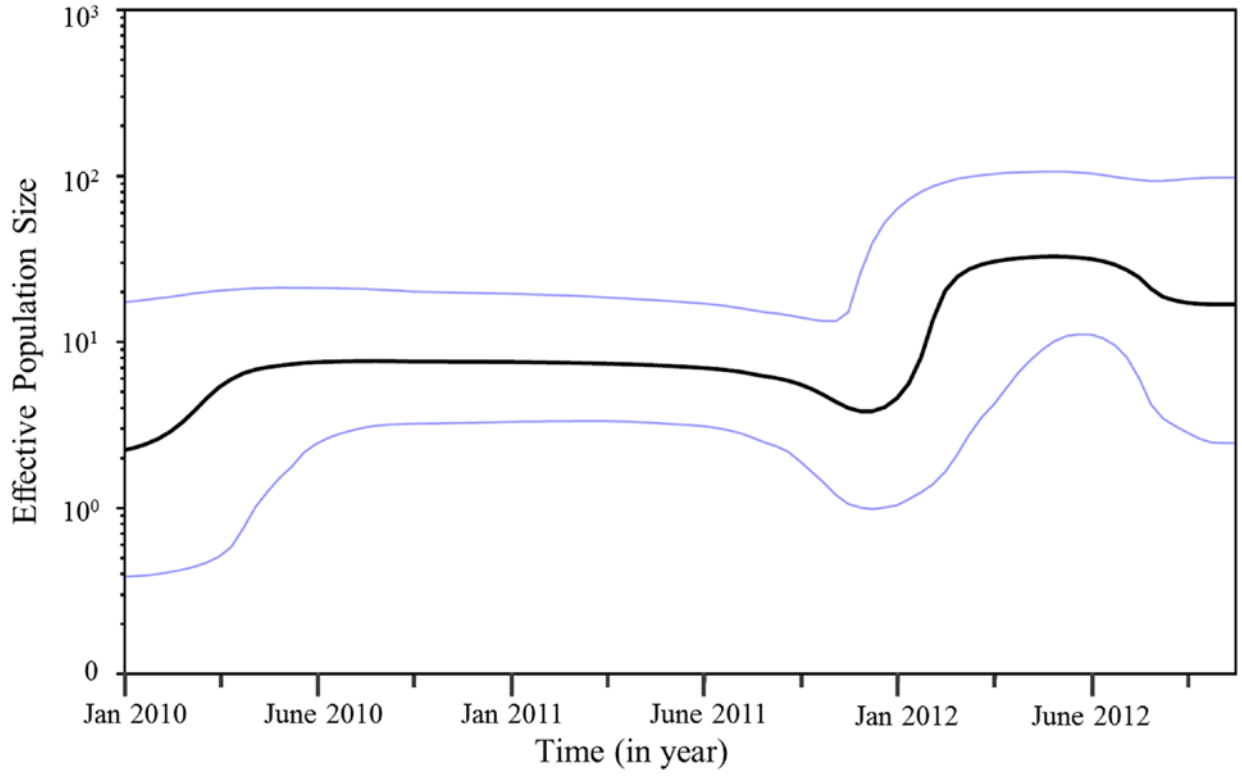


Figure 4. Phylodynamics of central Vietnam RSV-A strains *G* gene hypervariable region with Bayesian Skyline Plot (BSP)



Tables

Table 1. Summary demographic and clinical information of RSV-A paediatric ARI hospitalization cases in central Vietnam

	Year of sample collection (January 2010 - December 2012)		
	(Jan - Dec 2010)	(Jan - Dec 2011)	(Jan - Dec 2012)
Paediatric ARI cases (n)	542	513	797
RSV-A ARI cases (%)	66 (12.2%)	34 (6.6%)	169 (21.2%)
Genotype NA1 (n)	66	34	38
Genotype ON1 (n)	0	0	123
Demographic information			
Sex			
Male (%)	310 (57.2%)	294 (57.3%)	448 (56.2%)
Female (%)	232 (42.8%)	219 (42.7%)	349 (43.8%)
Age (in month)[a]	17 (IQR: 8 - 27)	18 (IQR: 8 - 31)	15 (IQR: 7 - 25)
Clinical symptom(s)			
URTI (%)[b]	443 (81.7%)	460 (89.7%)	551 (69.1%)
LRTI (%)[c]	99 (18.3%)	53 (10.3%)	246 (30.9%)

[a] Data are presented in median (IQR (Interquartile Range): 1st - 3rd).

[b] URTI: upper respiratory tract infection

[c] LRTI: lower respiratory tract infection

Table 2. Prevalence of RSV-A genotypes NA1 and ON1 lineages in central Vietnam during 2010 - 2012

RSV-A genotype(s)	Genotype NA1 (n=138)						Genotype ON1 (n=123)	
Year of sample collection	(Jan - Dec 2010)		(Jan - Dec 2011)		(Jan - Dec 2012)		(Jan - Dec 2012)	
Number of strains (n)	(n=66)		(n=34)		(n=38)		(n=123)	
Lineage 1	0		6	(17.7%)	26	(68.4%)	120	(97.6%)
Lineage 2	56	(84.9%)	7	(20.6%)	2	(5.3%)	1	(0.8%)
Lineage 3	1	(1.5%)	2	(5.9%)	8	(21.1%)	1	(0.8%)
Lineage 4	9	(13.6%)	19	(55.9%)	0		0	
p-value[a]			**< 0.001**[b]		**< 0.001**[c]			
(Undetermined lineage)	0		0		2	(5.3%)	1	(0.8%)

[a] Lineage proportion comparisons between years were performed with two-tailed Fisher's exact test. Values of *p*-value less than 0.05 were considered to be statistically significant and indicated as bold style.

[b] NA1 lineage proportion comparison between (Jan - Dec 2010) and (Jan - Dec 2011).

[c] NA1 lineage proportion comparison between (Jan - Dec 2011) and (Jan - Dec 2012).

10 **Table 3. RSV-A molecular evolutionary rate comparison between central Vietnam and global strains**

11

Study site / reference study	RSV-A categorization (subgroup / genotype)	Molecular evolution rate (substitutions/site/year) (95%HPD)[a]	Year of divergence (95%HPD)[a]
Central Vietnam paediatric ARI surveillance	Overall RSV-A (with global reference strains)	5.99×10^{-3} (5.00×10^{-3} - 7.03×10^{-3})	1954 (1951 - 1956)
(Yoshihara et al., 2016)	Overall RSV-A (only from central Vietnam) (Jan 2010 - Dec 2012)	**8.67×10^{-3} (6.27×10^{-3} - 1.16×10^{-2})**[b]	N/A
	Genotype ON1 (Jan 2012 - Dec 2012)	**2.55×10^{-2} (3.98×10^{-3} - 5.04×10^{-2})**[c]	2009 (2007 - 2011)
	Genotype NA1 (Jan 2010 - Dec 2012)	**7.12×10^{-3} (4.24×10^{-3} - 1.02×10^{-2})**[d]	1998 (1996 - 2000)
RSV-A global strains	Overall RSV-A (global reference strains)	5.36×10^{-3} (4.42×10^{-3} - 6.39×10^{-3})	1953 (1949 - 1956)
(Hirano et al., 2014)	Genotype ON1 (2010 - Aug 2013)	6.03×10^{-3} (3.43×10^{-3} - 9.10×10^{-3})	2005 (2000 - 2010)
	Genotype NA1 (Aug 2004 - Nov 2013)	4.61×10^{-3} (3.33×10^{-3} - 5.98×10^{-3})	2000 (1997 - 2002)

[a] 95%HPD stands for 95% highest probability density.

[b] Mean molecular evolution rates comparison between overall RSV-A (only from central Vietnam) and RSV-A global reference strains.

[c] Mean molecular evolution rates comparison between central Vietnam genotypes ON1 and ON1 global reference strains.

[d] Mean molecular evolution rates comparison between central Vietnam genotypes NA1 and NA1 global reference strains.

Statistically significant values were indicated in bold style.

12

Table 4. N-glycosylation pattern of central Vietnam RSV-A NA1 and ON1 *G* gene hypervariable regions

RSV-A genotype(s)	Genotype NA1				Genotype ON1
Year of sample collection	(Overall)	(Jan - Dec 2010)	(Jan - Dec 2011)	(Jan - Dec 2012)	(Jan - Dec 2012)
Number of strains (n)	(n=138)	(n=66)	(n=34)	(n=36)	(n=122)
N-glycosylation site(s)	n (%)				n (%)
237-240 (NTTK / NTSK)	33 (23.9%)	0	6 (17.7%)	27 (75.0%)	122 (100%)
251-254 (NTTG)	124 (89.9%)	64 (97.0%)	32 (94.1%)	28 (77.8%)	1 (0.8%)
273-276 (NLSP)	79 (57.3%)	47 (71.2%)	26 (76.5%)	6 (16.7%)	0
309-312 (NLSQ)	1 (0.7%)	0	0	1 (2.8%)	0
318-321 (NTTK / NTTE)	63 (45.7%)	10 (15.2%)	26 (76.5%)	27 (75.0%)	121 (99.2%)
p-value[a]			**< 0.001**[b]	**< 0.001**[c]	**< 0.001**[d]

[a] Statistical analyses for the proportion comparison of N-glycosylation pattern between two independent groups were performed with two-tailed Fisher's exact test. Values of *p*-value less than 0.05 were considered to be statistically significant and indicated as bold style.

[b] N-glycosylation pattern comparison between NA1 (Jan - Dec 2010) and NA1 (Jan - Dec 2011).

[c] N-glycosylation pattern comparison between NA1 (Jan - Dec 2011) and NA1 (Jan - Dec 2012).

[d] N-glycosylation pattern comparison between overall NA1 (Jan 2010 - Dec 2012) and ON1 (Jan 2012 - Dec 2012).

Table 5. Positively selected amino acid substitution sites of central Vietnam RSV-A NA1 and ON1 *G* gene hypervariable regions

RSV-A genotype(s)	Positively selected site(s)	SLAC[a]	FEL[b]	IFEL[c]	MEME[d]
Genotype ON1	D237N		*	*	
(Jan - Dec 2012)	E262K	*	*	*	*
(n=123)	P314L			*	
Genotype NA1	L215P		*	*	*
(Jan - Dec 2010)	N237D		*	*	
(n=66)	S250F	*			*
	K262E	*	*	*	*
	P274L	*	*	*	*
	T320I		*	*	*
Genotype NA1					
(Jan - Dec 2011)	(No positive selection site detected)				
(n=34)					
Genotype NA1	F250S	*			*
(Jan - Dec 2012)	N273Y		*	*	*
(n=38)	I320T	*	*	*	*

Values of *p*-value less than 0.05 were considered to be statistically significant and indicated as the asterisk (*)

[a] SLAC: Single Likelihood Ancestor Counting

[b] FEL: Fixed Effects Likelihood

[c] IFEL: Internal Fixed Effects Likelihood

[d] MEME: Mixed Effects Model for Episodic Diversifying Selection