Infection, Genetics and Evolution 45 (2016) 437–446

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Molecular evolution of respiratory syncytial virus subgroup A genotype NA1 and ON1 attachment glycoprotein (*G*) gene in central Vietnam

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], Lay-Myint Yoshida [a,*]

[a] Department of Paediatric Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, Nagasaki 852-8523, Japan
[b] 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

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form 11 October 2016
Accepted 12 October 2016
Available online 14 October 2016

Keywords:
ARI
RSV
ON1
Molecular evolution
Vietnam
Paediatric infectious diseases

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

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.

* Corresponding author at: Department of Paediatric Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki-city, Nagasaki 852-8523, Japan.
E-mail address: lmyoshi@nagasaki-u.ac.jp (L.-M. Yoshida).

http://dx.doi.org/10.1016/j.meegid.2016.10.010
1567-1348/

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 336 bp for genotype ON1 and 264 bp 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 previous study (Hirano et al., 2014). The nucleotide sequences of RSV-A partial *G* gene from central Vietnam used in the current study have been submitted to GenBank under accession numbers: KX946220 – KX946477.

2.2. Phylogenetic and molecular evolutionary analyses with Bayesian Markov Chain Monte Carlo (MCMC) and Maximum Likelihood (ML) methods

Nucleotide sequences of RSV-A *G* gene 3rd hypervariable region were aligned and edited using ClustalW within MEGA ver.6.0.6 (Tamura et al., 2013). KAKUSAN4 (http://www.fifthdimension.jp/products/kakusan/) was utilized for the selection of best-fit nucleotide substitution model (Tanabe, 2011). Phylogenetic and molecular evolutionary analyses were performed with Bayesian Markov Chain Monte Carlo (MCMC) method using BEAST ver.1.8.0 (Drummond and Rambaut, 2007; Nagasawa et al., 2015; Tsukagoshi et al., 2013). In the current study, four clock models (Strict clock, Uncorrelated lognormal relaxed clock, Uncorrelated exponential clock and Random local clock) and four demographic models (Constant size, Exponential growth, Logistic growth and Expansion growth) were compared to select the best-fit model for each sequence dataset based on the value of Akaike's Information Criterion through MCMC (AICM) (Suchard et al., 2001) using Tracer ver.1.6 (http://tree.bio.ed.ac.uk/software/tracer/). The model with the lowest AICM value was selected to be the best-fit model in each sequence dataset and used for analysis (Kimura et al., 2015; Nagasawa et al., 2015) (Supplementary Table 1). The detailed condition for each Bayesian MCMC analysis was summarized in Supplementary Table 2. The MCMC chains were run for 200,000,000 steps for all the analyses to achieve convergence with sampling every 2000 steps. The convergence was assessed using Tracer ver.1.6, and the parameters with effective sample sizes (ESS) of 200 or greater after 10% burn-in were accepted (Kushibuchi et al., 2013). The time-scaled Maximum Clade Credibility (MCC) trees were generated by TreeAnnotator ver.1.8.0 after removing the first 10% of trees as burn-in. The time-scaled MCC 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 1000 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 2000 steps under the uncorrelated exponential relaxed clock model and HKY85-gamma substitution model (Supplementary Table 2).