Infection, Genetics and Evolution 44 (2016) 76–84

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

A molecular epidemiological study of human respiratory syncytial virus in Croatia, 2011–2014

Anamarija Slovic [a,b,*], Jelena Ivancic-Jelecki [a,b], Sunčanica Ljubin-Sternak [c,d], Gordana Mlinarić Galinović [d,e], Dubravko Forcic [a,b]

[a] Centre for Research and Knowledge Transfer in Biotechnology, University of Zagreb, Rockefellerova 10, 10000 Zagreb, Croatia
[b] Center of Excellence for Viral Immunology and Vaccines, CERVirVac, Croatia
[c] Teaching Institute of Public Health "Dr. Andrija Štampar", Mirogojska 8, 10000 Zagreb, Croatia
[d] School of Medicine University of Zagreb, Šalata 3, 10000 Zagreb, Croatia
[e] Croatian National Institute of Public Health, Rockefellerova 12, 10000 Zagreb, Croatia

ARTICLE INFO

Article history:
Received 15 February 2016
Received in revised form 17 June 2016
Accepted 18 June 2016
Available online 23 June 2016

Keywords:
Human respiratory syncytial virus
Genetic variability
Molecular epidemiology
Phylogenetic analysis

ABSTRACT

Human respiratory syncytial virus (HRSV) causes common respiratory tract infections in infants, young children and the elderly. The diversity of HRSV strains circulating in Croatia was investigated throughout a period of four consecutive years from March 2011–March 2014. The analysis was based on sequences from the second hypervariable region of the G gene. A predominance of HRSV group A was observed in the first three years of the study, while group B became slightly predominant during the first few months of 2014. Overall, 76% of viruses belonged to group A including the genotypes NA1, ON1 and GA5. NA1 was by far the most common genotype within group A in 2011–2013; however, only ON1 and a few GA5 viruses were detected in the first three months of 2014. The majority of group B strains were of genotype BA9 (97%), and a few BA10 genotypes were detected. BA9 had the highest substitution rate of all the detected genotypes, followed by ON1.

Multiple analyses showed that HRSV group A strains were more diverse than group B strains. Gly at residue 232 (previously described to be specific for ON1) was also detected in three NA1 strains, which were phylogenetically placed on separate branches within the NA1 genotype. For all genotypes, the diversity was higher at the amino acid level than at the nucleotide level, although positive selection of mutations was shown for only a few sites using four different methods of codon-based analysis of selective pressure. More codons were predicted to be negatively selected.

The complexity of the HRSV pools present during each epidemic peak was determined and compared to previous epidemiological data. In addition to presenting genetic versatility of HRSV in this geographic region, the collected sequences provide data for further geographical and temporal comparative analyses of HRSV and its evolutionary pathways.

1. Introduction

Human respiratory syncytial virus (HRSV) is a paramyxovirus that causes common respiratory tract infections in infants, young children and among the elderly (Falsey and Walsh, 2000; Glezen et al., 1986; Henderson et al., 1979). The disease manifestation ranges from mild nonspecific respiratory symptoms to severe illness, such as bronchiolitis or pneumonia (Glezen et al., 1986). Almost all children become infected by the age of two (Glezen et al., 1986), and it has been estimated that over 30 million cases of annual worldwide infections occur by this viral agent (Nair et al., 2010). Reinfections are very common throughout life, although subsequent infections are usually associated with milder disease (Henderson et al., 1979). Due to the lack of a licensed vaccine for this virus, a substantial cost is associated with HRSV epidemics due to hospitalizations and treatment of HRSV infection-related illnesses.

HRSV belongs to the *Pneumovirus* genus of the *Paramyxoviridae* family and has RNA of negative polarity as its genome. The genome is approximately 15,200 nucleotides long, comprising 10 genes that encode 11 proteins. There are three proteins on the surface of a mature virion called SH, G and F (Collins and Karron, 2013). Proteins G and F are important for attachment and fusion with the target cell, respectively (Levine et al., 1987; Walsh and Hruska, 1983), and they can elicit the production of neutralizing antibodies (Johnson et al., 1997; Martinez and Melero, 1998). Variability among strains is responsible for repeated

* Corresponding author at: Centre for Research and Knowledge Transfer in Biotechnology, University of Zagreb, Rockefellerova 10, 10000 Zagreb, Croatia.
E-mail address: aslovic@unizg.hr (A. Slovic).

http://dx.doi.org/10.1016/j.meegid.2016.06.036
1567-1348/

infections and yearly HRSV outbreaks, with major antigenic differences located in the G protein (Johnson et al., 1987). The HRSV strains are divided into groups A and B based on antigenic variation (Mufson et al., 1985). Both groups co-circulate during epidemics, although group A usually dominates (Tan et al., 2012).

Both HRSV groups are further divided into genotypes based on sequences of the second hypervariable region (HVR2) of the G gene, located at the protein's *C*-terminus. This region is a part of the G protein ectodomain, which consists of two hypervariable regions flanking four conserved cysteines (Collins and Karron, 2013). Phylogenetic studies have demonstrated a rather complex circulation pattern of HRSV, as numerous genotypes within both groups have been identified and may circulate within the same community. Furthermore, relative proportions of genetic variants within the community may differ between epidemics (Cane, 2001; Freymuth et al., 1991), with predominant genotypes being replaced in successive years as a consequence of local factors, such as the level of herd immunity to certain strains (Anderson et al., 1991; Peret et al., 1998). The dynamics of HRSV circulation is further demonstrated with the emergence of new genotypes that quickly spread worldwide and replace previously circulating genotypes. This was shown after the emergence of the BA genotype, which derives from the group B genotype with a 60-nt, in-frame duplication in HVR2 (Trento et al., 2003). Among the group B viruses, almost all strains detected after 2005 belong to the BA genotype (Trento et al., 2010). Recently, a similar duplication event in HRSV group A led to the emergence of genotype ON1 (Eshaghi et al., 2012), which has a 72-nt, in-frame duplication in HVR2. A trend in epidemiological prevalence has also been reported for ON1 strains (Kim et al., 2014; Pierangeli et al., 2014).

In this study, we obtained the HVR2 sequences of HRSV from 486 patients (ages 0–16 years) hospitalized due to acute respiratory infections (ARI) in Croatia during the period of four consecutive years from 2011–2014. Aiming to investigate the diversity of HRSV in Croatia during this period, we performed a molecular epidemiological study, which included monitoring the seasonal epidemic progression and detection of subgroup patterns. Selective pressure, evolutionary rates and genotype dynamics were also analysed, demonstrating changes in HRSV populations.

2. Materials and methods

2.1. Clinical samples

Clinical samples were obtained from inpatients (ages 0–16 years, median 8 months, interquartile range 2.9–18.8 months) with ARI, hospitalized mostly due to bronchiolitis or pneumonia, and were archived at the Croatian National Institute of Public Health. Samples included in this study were collected from March 2011 to March 2014. HRSV was confirmed in 584 inpatients' nasopharyngeal secretions using a direct fluorescence assay (DFA Light Diagnostics, Chemicon International, Temecula, CA, USA) (Mlinaric-Galinovic et al., 1987). Based on the availability of material, 486 samples were chosen for PCR and sequencing. The study was approved by the Ethics Committees of the Croatian National Institute of Public Health, University Children's Hospital Zagreb, and University Hospital for Infectious Diseases and Medical School University of Zagreb.

2.2. Reverse transcription, PCR and sequencing

Total RNA was extracted from 500 μL of clinical samples by the method reported by Chomczynski and Mackey (1998). Reverse transcription was performed at 42 °C for 20 min, in a reaction mix containing 10 μL of isolated RNA, 1 × PCR buffer (GE Healthcare, UK), 0.1 mM of each dNTP, 20 U of RNase inhibitor (Thermo Fisher Scientific, USA), 1.25 mM $MgCl_2$, 2.5 mM of random hexanucleotide primers and 50 U of MuLV reverse transcriptase (Thermo Fisher Scientific, USA) in a final volume of 20 μL.

Nested PCR was carried out with two sets of primers to amplify the HVR2. For the first amplification forward SH1 (5′ CACAGTKACT GACAAYAAAGGAGC 3′) and reverse F164 (5′ GTTATGACACTGGTATA CCAACC 3′) primers were used, followed by a second amplification with ABG490 (5′ ATGATTWYCAYTTTGAAGTGTTC 3′) and F9AB (5′ CAACTCCATKRTTATTTGCC 3′) primer pair.

Ten microliters of the reverse transcription reaction mix was used for the first amplification, whereas two microliters of the first amplification mixture was used for the second PCR. PCR reaction mixtures contained 1 × PCR buffer (GE Healthcare, UK), 10 μM of each dNTP, 0.25 mM $MgCl_2$, 0.25 μM of each primer, and 5 U of *Taq* polymerase (GE Healthcare, UK). PCR conditions for the first amplification were: 95 °C for 5 min, 40 cycles of 95 °C/30 s, 50 °C/30 s, 72 °C/2 min, followed by final extension at 72 °C for 7 min. The second PCR was performed under the same conditions, except for a shorter extension step (1 min). The amplified products of 501 nt for group A non-ON1 strains, 573 nt for ON1 strains, and 564 nt for group B strains, were separated on a 1.5% agarose gel, excised and purified by centrifugation through glass wool, as reported by Sun et al. (2012).

Sequencing reactions were set up with purified DNA, one of the specific primers used in the second PCR, and a BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) according to the manufacturer's protocol. Sequencing and sequence analysis were performed on a 3130 Genetic Analyser (Thermo Fisher Scientific, USA).

2.3. Phylogenetic analysis

Nucleotide sequences of HRSV A and B strains of the genotypes previously described in the literature were obtained from GenBank and used to construct alignments and phylogenetic trees. Sequences spanned bases 5274–5543 of strain A2 (accession number (acc. no.) M11486) and 652–981 of BA4128/99B (acc. no. AY333364.1), prototype strains of group A and B, respectively. Alignments were performed using ClustalX v2.1 (Larkin et al., 2007) and BioEdit v7.2.5 software (Hall, 1999). Phylogenetic trees were generated using the maximum likelihood method with Molecular Evolutionary Genetics Analyses (MEGA) software v6.06. (Tamura et al., 2013), under the most appropriate model of nt substitution determined with jModeltest v2.1.4 (Darriba et al., 2012). Bootstrap probabilities for 1000 iterations were calculated to evaluate confidence estimates. Evolutionary distances (*p*-distances) within and between genotypes were estimated using MEGA *p*-distance method, which calculates proportion (*p*) of nucleotide sites at which two sequences being compared are different.

The nt sequences of HRSV strains obtained in this study were deposited in GenBank under acc. nos. KT371536–KT371609, KT371681–KT371705 and KU375579–KU375588.

2.4. Analysis of deduced amino acid sequences

Deduced amino acid sequences of HVR2 were generated using the standard genetic code implemented in MEGA software. NetOGlyc 4.0 server (http://www.cbs.dtu.dk/services/NetOGlyc/) was used to predict potential *O*-glycosylation sites (Steentoft et al., 2013) and NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) for *N*-glycosylation prediction (Gupta et al., 2004).

2.5. Selective pressure analysis

Codon-based analysis of selective pressure was performed using the HyPhy package available on the Datamonkey server (http://www.datamonkey.org/) (Delport et al., 2010; Kosakovsky Pond and Frost, 2005; Kosakovsky Pond et al., 2005). Four different methods were used: single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), internal branch fixed-effects likelihood (IFEL) and fast