Virus Research 189 (2014) 293–302

Contents lists available at ScienceDirect

Virus Research

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

Phylogeny and population dynamics of respiratory syncytial virus (Rsv) A and B

Marianna Martinelli[a], Elena Rosanna Frati[a], Alessandra Zappa[a], Erika Ebranati[b], Silvia Bianchi[a], Elena Pariani[a,c], Antonella Amendola[a,c], Gianguglielmo Zehender[b], Elisabetta Tanzi[a,c,*]

[a] Department of Biomedical Sciences for Health, University of Milan, Via C. Pascal 36, 20133, Milan, Italy
[b] "L. Sacco" Department of Biomedical and Clinical Sciences, University of Milan, Via G. B. Grassi 74, 20157 Milano, Italy
[c] CIRI-IT, Department of Health Sciences, University of Genoa, Via A. Pastore 1, 16132 Genoa, Italy

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 6 June 2014
Accepted 11 June 2014
Available online 19 June 2014

Keywords:
Respiratory syncytial virus
G protein
Phylogenetic analysis
Selective pressure
Phylodynamic analysis
Evolutionary rate

ABSTRACT

Respiratory syncytial virus (RSV) is a major cause of lower respiratory tract infections in infants and young children. RSV is characterised by high variability, especially in the G glycoprotein, which may play a significant role in RSV pathogenicity by allowing immune evasion. To reconstruct the origin and phylodynamic history of RSV, we evaluated the genetic diversity and evolutionary dynamics of RSV A and RSV B isolated from children under 3 years old infected in Italy from 2006 to 2012. Phylogenetic analysis revealed that most of the RSV A sequences clustered with the NA1 genotype, and RSV B sequences were included in the Buenos Aires genotype. The mean evolutionary rates for RSV A and RSV B were estimated to be 2.1×10^{-3} substitutions (subs)/site/year and 3.03×10^{-3} subs/site/year, respectively. The time of most recent common ancestor for the tree root went back to the 1940s (95% highest posterior density—HPD: 1927–1951) for RSV A and the 1950s (95%HPD: 1951–1960) for RSV B. The RSV A Bayesian skyline plot (BSP) showed a decrease in transmission events ending in about 2005, when a sharp growth restored the original viral population size. RSV B BSP showed a similar trend. Site-specific selection analysis identified 10 codons under positive selection in RSV A sequences and only one site in RSV B sequences. Although RSV remains difficult to control due to its antigenic diversity, it is important to monitor changes in its coding sequences, to permit the identification of future epidemic strains and to implement vaccine and therapy strategies.

1. Introduction

Respiratory syncytial virus (RSV) is a major cause of lower respiratory tract infections (LRTIs) in infants and young children, accounting for 0.1–2% of hospital admissions annually for infants under 1 year of age (Nielsen et al., 2003). RSV is also responsible for a significant proportion of LRTIs in immunocompromised adults and the elderly (Cane, 2001). In temperate climates, epidemics of RSV occur annually, typically peaking in the winter (November) and extending into the spring (April).

RSV, an enveloped virus, is a member of the genus *Pneumovirus* in the *Paramyxoviridae* family with a negative-sense single-stranded RNA genome. RSV is classified into two groups, A and B, based on antigenic and genetic variability (Anderson et al., 1985; Mufson et al., 1985). These two groups have been described based on the reactions of monoclonal antibodies against the two major glycoproteins, G and F, which facilitate viral attachment to cell receptors and mediate cell fusion, respectively (Levine et al.,

Abbreviations: AA, amino acid; BA, Buenos Aires; BF, Bayes factor; bp, base pairs; BSP, Bayesian skyline plot; FEL, fixed-effects likelihood; GARD, genetic algorithm recombination detection; GTR, general time reversible; HPD, highest posterior density; IFEL, internal fixed-effects likelihood; LRTIs, lower respiratory tract infections; MCC, maximum clade credibility; MCMC, Markov Chain Monte Carlo; MEME, mixed effects model of evolution; MMLV, Moloney murine leukemia virus; NA, Niigata; ON, Ontario; RSV, respiratory syncytial virus; SA, South Africa; SLAC, single likelihood ancestor counting; SBP, single break point; Subs, substitutions; tMRCA, time of the most recent common ancestor.

* Corresponding author at: Department of Biomedical Sciences for Health, University of Milan, Via C. Pascal 36, 20133 Milan, Italy. Tel.: +39 0250315139; fax: +39 0250315120.

E-mail addresses: marianna.martinelli@unimi.it (M. Martinelli), elena.frati@unimi.it (E.R. Frati), alessandra.zappa@unimi.it (A. Zappa), erika.ebranati@unimi.it (E. Ebranati), silvia.bianchi@unimi.it (S. Bianchi), elena.pariani@unimi.it (E. Pariani), antonella.amendola@unimi.it (A. Amendola), gianguglielmo.zehender@unimi.it (G. Zehender), elisabetta.tanzi@unimi.it (E. Tanzi).

http://dx.doi.org/10.1016/j.virusres.2014.06.006
0168-1702/

1987; Walsh and Hruska, 1983). Most RSV epidemiological and evolutionary studies have been based on the analysis of changes observed in part of the G glycoprotein, which is the most variable of the 11 proteins encoded in the 15.2-kb RSV genome.

The mature G glycoprotein is a 289- to 299-amino acid (AA) type II glycoprotein consisting of three unique regions: a cytoplasmic tail (AA 1–38), transmembrane domain (AA 38–66), and ectodomain (AA 66–298) (Melero et al., 1997). The G glycoprotein ectodomain is variable between groups A and B, as only one strictly conserved 13-AA region in the central ectodomain has been noted (AA 164–176) (Sullender, 2000).

Flanking this conserved region are two protein segments that have a high level of sequence variation and an AA composition similar to those of mucins secreted by epithelial cells. These segments contain a high percentage of serine and threonine residues, which are potential acceptor sites for *O*-linked sugars. *N*- and *O*-linked oligosaccharides contribute to the antigenic structure of the G glycoprotein and to viral infectivity (García-Beato et al., 1996). Although the G glycoprotein is highly glycosylated with *N*- and *O*-linked sugars, the positions for glycosylation are poorly conserved (Johnson et al., 1987).

To date, several genotypes for the A and B groups have been reported from different geographical regions. RSVA genotypes include GA1–GA7 (Peret et al., 1998, 2000), SAA1 (South Africa) (Venter et al., 2001), NA1, NA2 (Niigata) (Shobugawa et al., 2009), and ON1 (Ontario), a novel genotype with a 72-nucleotide (nt) G glycoprotein gene duplication (Eshaghi et al., 2012). RSV B is divided into eight genotypes: GB1–GB4 (Peret et al., 1998), SAB1–SAB3 (Venter et al., 2001), and Buenos Aires (BA) (Trento et al., 2003). The last genotype is characterised by a 60-nt duplication starting after nucleotide 792 of the G glycoprotein gene (Trento et al., 2006), and it is subdivided into 10 genotypes (BA1–BA10) (Dapat et al., 2010).

The aims of this study were: to analyse the genetic variability of RSV A and B in Italy and to reconstruct the origin and phylodynamic history of RSV including in the analysis also other sequences obtained from NCBI GenBank Database. We evaluated the genetic diversity and evolutionary dynamics of RSV A and RSV B isolated from children under 3 years old infected during six epidemic seasons, from 2006 to 2012, in Italy.

2. Materials and methods

2.1. Study samples

RSV A ($n = 23$) and RSV B ($n = 12$) sequences obtained from oropharyngeal swabs of RSV-infected children in six successive seasons, from November 2006 through April 2012, were used for molecular characterisation by sequence analysis of the G glycoprotein gene. Samples were collected from hospitalised children who were routinely tested for respiratory illness during the study period. Written informed consent was obtained from parents or legal guardians for each child included in the study in order to store their samples for further anonymous research testing.

2.2. Nucleic acids extraction and amplification

Nucleic acids were extracted by a commercial method (NucliSENS® easyMAG®, bioMerieux, Lyon, France). For RNA virus detection, cDNA was synthesised using pd(N)6 random hexamer primers (Amersham Biosciences, Little Chalfont, UK) with MMLV reverse transcriptase (Invitrogen Tech-Line, Carlsbad, CA, USA). Viral detection was performed by PCR. The RSV A and RSV B viral genomes were identified by multiplex nested PCR (fusion gene, 336- and 582-base pairs [bp], respectively) using specific primer sets (Coiras et al., 2003).

2.3. Molecular characterisation by nucleotide sequencing

Molecular characterisation of RSV A was performed by analysing the sequence of a 916-bp G gene amplicon (nt 4919–5835). For characterisation of RSV B, an 800-bp G gene amplicon (nt 4858–5658) was used. Partial RSV A and RSV B G gene sequences were amplified with two primer pairs (Zlateva et al., 2004; Zlateva et al., 2005). Two partially overlapping fragments for each virus were amplified. Fragment amplification was performed using 1 μg of amplified DNA as a template. Besides the PCR primer set, an additional forward primer and reverse primer, both located in the G glycoprotein gene were used to obtain the two overlapping fragments (Zlateva et al., 2004; Zlateva et al., 2005). Reactions were carried out in 50 μl reaction volumes containing 5X PCR Buffer, 10 mM dNTPs, 30 pmol of each primer, and 1 U Taq (GoTaq® DNA Polymerase 5 U/μl, Promega, Madison, WI). The cycling conditions were as follows: 95 °C for 5 min followed by 35 cycles at 95 °C for 1 min, 60 °C for 1 min (RSV A) or 65 °C for 1 min (RSV B), and 72 °C for 1 min. A final 72 °C extension for 7 min followed the cycles.

Amplicons were visualised on 2% agarose gels stained with ethidium bromide. Subsequently, amplicons were purified using NucleoSpin® Extract II (Macherey-Nagel GmbH, Germany), and nucleotide sequences were obtained using an ABI PRISM 3100 genetic analyser for automated DNA sequencing (Applied Biosystem, CA, USA).

2.4. Phylogenetic analysis and genotype characterisation

Viral sequences obtained in this study were aligned with reference viral strains representative of the main RSV A ($n = 78$) (Zlateva et al., 2004; Rebuffo-Scheer et al., 2011; Eshaghi et al., 2012) and B ($n = 50$) (Peret et al., 1998; Nagai et al., 2004; Zlateva et al., 2005; Trento et al., 2006) genotypes, building two specific data sets: RSV A including a total of 101 sequences and RSV B, encompassing 62 sequences.

Multiple G gene nucleotide sequences were aligned using ClustalX (Langedijk et al., 1998), version 2.0. Phylogenetic trees of RSV A and RSV B sequences were constructed using the Neighbour-Joining method (Saitou and Nei, 1987) and the Kimura 2-Parameter model (Kimura, 1980) with the MEGA5 package, version 5.05 (Tamura et al., 2011). A bootstrap re-sampling analysis was performed (1000 replicates) to test tree robustness (Felsenstein, 1985). The RSV A and RSV B G gene sequences from the viral strains studied were deposited into the NCBI GenBank Database (http://www.ncbi.nlm.nih.gov/GenBank/index.html), under accession numbers: (Accession Numbers: KJ961514-KJ961548).

2.5. Evolutionary rate estimates, time-scaled phylogeny reconstruction

Bayesian analysis was carried out by using the same datasets of sequences built for phylogenetic analysis and genotype characterisation, as described in Section 2.4. The sampling years of the sequences ranged from 1956 to 2012 for RSV A and from 1960 to 2012 for RSV B. All of these sequences matched the following criteria: (1) they had been previously published in peer-reviewed journals (except for those characterised in this study); (2) there was no uncertainty about the subtype assignment of each sequence, and they were classified as non-recombinant; (3) the sequences were not epidemiologically linked by direct donor-recipient transmission; (4) only one sequence per individual could be randomly selected; and (5) the sampling dates of each sequence were known.