Infection, Genetics and Evolution 39 (2016) 85–91

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Phylogenetic and molecular analyses of human parainfluenza type 3 virus in Buenos Aires, Argentina, between 2009 and 2013: The emergence of new genetic lineages

Stephanie Goya [a,b], Alicia Susana Mistchenko [a,c], Mariana Viegas [a,b,*]

[a] Virology Laboratory, Ricardo Gutierrez Children's Hospital, Buenos Aires City, Argentina, Gallo 1330 2°, (1425) Ciudad Autónoma Buenos Aires, Argentina
[b] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
[c] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 18 December 2015
Accepted 4 January 2016
Available online 9 January 2016

Keywords:
Human parainfluenza virus 3
Genotypes
Seasonality
Emerging lineages
Molecular characterization
B-cell epitopes

ABSTRACT

Despite that human parainfluenza type 3 viruses (HPIV3) are one of the leading causes of acute lower respiratory tract infections in children under five, there is no licensed vaccine and there is limited current information on the molecular characteristics of regional and global circulating strains. The aim of this study was to describe the molecular characterization of HPIV3 circulating in Buenos Aires. We performed a genetic and phylogenetic analysis of the HN glycoprotein gene. Between 2009 and 2013, 124 HPIV3-positive samples taken from hospitalized pediatric patients were analyzed. Four new genetic lineages were described. Among them, C1c and C3d lineages showed local circulation patterns, whereas C3e and C3f comprised sequences from very distant countries. Despite the diversity of the described genotypes, C3a and C3d predominated over the others, the latter was present during the first years of the study and it was progressively replaced by C3a. Molecular analyses showed 28 non-synonymous substitutions; of these, 13 were located in potentially predicted B-cell epitopes. Taken together, the emergence of genetic lineages and the information of the molecular characteristics of HN protein may contribute to the general knowledge of HPIV3 molecular epidemiology for future vaccine development and antiviral therapies.

1. Introduction

Human parainfluenza viruses (HPIV) are one of the leading causes of acute lower respiratory tract infections (LRTI) in children less than 5 years old, immunocompromised or chronically ill patients, and the elderly; their incidence is surpassed only by human respiratory syncytial virus (RSV) (Henrickson, 2003).

HPIVs belong to the family *Paramyxoviridae* and they have been divided into types 1 to 4, among which HPIV1 and HPIV3 are classified as members of the genus *Respirovirus*, while HPIV2 and HPIV4 are members of the genus *Rubulavirus* (Karron and Collins, 2013).

HPIV3 is the most frequent type of HPIV (Liu et al., 2013) and is associated with the most severe clinical presentations (bronchiolitis and pneumonia) (Henrickson, 2003; Karron and Collins, 2013); it is one of the leading causes of hospitalization in pediatric patients during spring and summer, with a high impact on public health (Ministerio de Salud. Boletín Integrado de Vigilancia N199. Argentina, 2013; Mao et al., 2012; Viegas et al., 2004). HPIV3 is an enveloped, single-stranded negative sense RNA virus. Its genome has 15,462 nucleotides and it encodes eight proteins, among them the fusion protein (F) and the hemagglutinin–neuraminidase protein (HN) are associated with the viral envelope (Henrickson, 2003; Karron and Collins, 2013).

The HN protein is a transmembrane glycoprotein of 572 aminoacids. It consists of a cytoplasmic domain, a membrane spanning region, a stalk region, and a globular head, and it is responsible for the binding and cleavage to the host cell sialic acid, stabilization, and activation of the F protein.The stalk region and the globular head protrude the virus envelope (ectodomain) (Karron and Collins, 2013; Palermo et al., 2009).

The HN and F proteins are the major target antigens of humoral immune response, inducing neutralizing antibodies, and the HN glycoprotein has the largest antigenic and genetic variation (Schmidt et al., 2011; Spriggs et al., 1987). Accordingly, the HN gene has been used for genotyping HPIV3 strains in molecular epidemiological studies (Mao et al., 2012; van Wyke Coelingh et al., 1987; Villaran et al., 2014; Almajhdi, 2015). In 2012, Mao et al. carried out the first study in order to unify and define a genetic classification for HPIV3, and it was recently reclassified by Almajhdi (Mao et al., 2012; Almajhdi, 2015). In these phylogenetic studies using HN gene and based on evolutionary divergence values (calculated as genetic distances), they described the existence of clusters (A, B, and C), subclusters (C1–C5), and genetic lineages (C1a–C1b, C3a–C3c) among HPIV3 strains. The authors

* Corresponding author at: Gallo 1330 2°, (1425) Ciudad Autónoma Buenos Aires, Argentina.
E-mail addresses: goyastephanie@gmail.com (S. Goya), asmistchenko@hotmail.com (A.S. Mistchenko), viegasmariana@conicet.gov.ar (M. Viegas).

http://dx.doi.org/10.1016/j.meegid.2016.01.002
1567-1348/

determined that the minimum genetic distance that must exist between two sequences of HPIV3 for classification in clusters is 0.045 and differences in the range of 0.019–0.045 define subclusters. Sequences with genetic distance under 0.010 belong to the same genetic lineage, and those with genetic distance between 0.010 and 0.019 belong to the same subcluster but to different genetic lineages (Almajhdi, 2015).

In Argentina, HPIV3 infections are the second leading cause of hospitalization for acute LRTI in patients less than 2 years old (Ministerio de Salud. Boletín Integrado de Vigilancia N199. Argentina, 2013). However, while the universal influenza vaccination in children is recommended, there is no a licensed vaccine against HPIV3 (Weinberg et al., 2009). Nevertheless, in recent years, intranasal live attenuated vaccines were generated by a reverse genetic system; they are well tolerated and immunogenic, and Phase I clinical trials with these vaccines are underway (Schmidt et al., 2011).

Taking into account that there is limited current information on the molecular characteristics of regional and global circulating HPIV3 strains, the aim of this study is to describe the molecular characterization of HPIV3 circulating in Buenos Aires, in a period of 5 years (2009–2013) through the genetic and phylogenetic analysis of the HN glycoprotein gene.

2. Materials and methods

2.1. Clinical samples

Nasopharyngeal aspirates (NPA) were taken from pediatric patients hospitalized in neonatology and pediatric wards of Public Hospitals from Buenos Aires city and Greater Buenos Aires and they were referred to the Virology Laboratory of the Ricardo Gutiérrez Children's Hospital (HNRG) for viral diagnosis. The period studied included 5 consecutive years (2009–2013).

A rapid detection of respiratory viruses by immunofluorescence assay (IFA) with monoclonal antibodies against HPIV 1, 2, and 3, human RSV, human adenovirus (AdV), and influenza A and B (Flu A and B, respectively) was performed according to manufacturer's instructions (EMD Millipore, Darmstadt, Germany) (Gardner and McQuilin, 1968).

Samples that were positive for HPIV3 by IFA were preserved at −70 °C until the molecular study was performed. All samples molecularly analyzed in this study belonged to children hospitalized at the HNRG, and they were randomly selected from the total HNRG samples. The samples and sequences were coded using the acronym HNRG, a serial number, country of origin and the year in which they were obtained.

2.2. RNA extraction and RT-PCR amplification of partial HN gene

Total RNA was extracted from 200 μl of NPA samples using PureLink® Viral RNA/DNA Mini Kit (Invitrogen, Massachusetts, USA) according to the manufacturer's instructions. RNA was eluted in 50 μl of RNase and DNase-free distilled water, and stored at −70 °C.

The partial HN gene (globular head) was amplified with the OneStep RT-PCR Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions and using the following previously published primers: HPIV3-HN.f1: 5′-ATGATCTAATACARTCAGGAGTRAATACAAG-3′ (nt 7,068–7,099, sense) and HPIV3-HN.r1: 5′-TATCTCGAGTTATGATTAACTG CAGC-3′ (nt 8,514–8,540, anti-sense) at a final concentration of 0.6 μM (Mao et al., 2012; Mizuta et al., 2014). The reaction mix was carried out in a 25 μl of reaction volume containing 7 μl of total RNA. Thermal cycling conditions were 50 °C for 30 min and then 95 °C for 15 min followed by 35 cycles of 94 °C for 30 seg, 50 °C for 30 seg, and 72 °C for 96 seg, and a final extension step at 72 °C for 10 min. The amplification products were 1.5% agarose gel purified with Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, California), and the product was eluted in 15 μl of RNase and DNase-free distilled water.

2.3. DNA sequencing

The purified PCR products were labeled with the BigDye Terminator v3.1 sequencing kit (Applied Biosystems, Massachusetts, USA) and electrophoresis was performed in an ABI3500 genetic analyzer (Applied Biosystems, Massachusetts, USA). The sequencing reactions were performed with the PCR primers and the internal primers: HPIV-HN.f2: 5′-AACTGTGTTCAACTCCHAAAG-3′ (nt 7,605–7,625, sense), HPIV-HN.f3: 5′-CAAGTTGGCAYAGCAAGTTAC-3′ (nt 8,082-8,102 sense), HPIV-HN.r2: 5′-CTGAATTGTAAGAAGCCTTGT-3′ (nt 7,094–7,114, anti-sense) (Mao et al., 2012), and HPIV-HN.r3: 5′-ATCTTGTT GTYGAGATTGAGCCA-3′ (nt 7,715–7,738, anti-sense), HPIV-HN.r4: 5′-TCAATTATTCCTAATTGTAACTTGCT-3′ (nt 8,093–8,119, anti-sense) designed for this study.

The SeqScape Software v2.7 (Applied Biosystems, Massachusetts, USA) was used to analyze, assemble, and generate the consensus nucleotide sequences obtained for each of the analyzed viral strains.

2.4. Phylogenetic analyses

The HNRG sequences were aligned with 50 previously published HPIV3 sequences downloaded from GenBank with Clustal W v2 software (Thompson et al., 1994). The jModelTest v0.1.1 software was used to determine the most suitable evolutionary model for the set of the analyzed sequences (Darriba et al., 2012).

The phylogenetic analyses were performed by ML (PhyML v.3.1 software, Guindon et al., 2010), Bayesian criteria (MrBayes v.3.2.1 software, Ronquist and Huelsenbeck, 2003), and distance methods (Neighbor-Joining, MEGA v6 software, Tamura et al., 2013). Initial random trees were used to infer the ML tree. The branch support for ML and distance methods were evaluated by non-parametric bootstrapping with 1000 pseudo-replica. The convergence of the Monte Carlo Markov Chains (MCMC) implemented in the Bayesian criteria was evaluated in TRACER v.1.5 with an effective sample size (ESS) >200; the initial 10% of the run length was discarded as burn-in. Consensus trees were visualized with FigTree v.1.4.0.

2.5. Molecular characterization and adaptive evolutionary analyses

The MEGA v6 software was used to define genetic lineages by estimating genetic distances (p-distances) within and among sequences in phylogenetic clades. Standard error (SE) estimates were obtained by the bootstrap method (1000 replicates) (Tamura et al., 2013).

The DNAsp v5 software was used for the analysis of the polymorphisms and divergences between the set of HNRG sequences and the prototype HPIV3 strain Washington/1957 (GenBank accession number: JN089924) (Librado and Rozas, 2009). The amino acid sequences were inferred using the universal genetic code.

Natural selection on the HNRG HN sequences was estimated from the ratio of non-synonymous (dN) to synonymous (dS) substitutions per site (dN/dS) at every codon in the alignment and the overall $\Omega = dN/dS$. The analysis was performed using the procedures available in the HyPhy package and accessed through the Datamonkey web server (Kosakovsky Pond and Frost, 2005).

N-glycosylation sites were predicted using N-Glycosite server (Zhang et al., 2004).

Potential conformational B-cell epitopes were mapped from the molecular model of the HN ectodomain region of the reference strain (PDB: 4MZA) with “DiscoTope: Structure-based antibody prediction” tool (IEDB Analysis Resource server) with a specificity of 75% and a sensitivity of 47% (Haste Andersen et al., 2006). The PDBePISA service of the EMBL-EBI website was used to identify sites on the surface of the protein (Krissinel and Henrick, 2007). Potential linear B-cell epitopes were predicted with “Bepipred: Linear Epitope Prediction” tool (IEDB Analysis Resource server) with a specificity of 75% and a sensitivity of 49% (Larsen et al., 2006).

2.6. Ethical statement

This study was evaluated and approved by the Teaching and Research Committee, and the Research Ethics Committee (N° CEI 14.17) of the HNRG in Buenos Aires city. Informed consent was not obtained, as patient information was anonymized and de-identified prior to analysis.

2.7. GenBank accession numbers

The accession numbers for the nucleotide sequences determined in this study are KT765887–KT766010.

3. Results

3.1. Epidemiological analysis

A total of 19,435 NPA samples from hospitalized pediatric patients with acute LRTI were referred to the Virology Laboratory of the HNRG from 2009 to 2013. Most of the patients were under 2 years of age (91.36%) and were males (55.87%). Of the total analyzed samples, 6,264 were positive for respiratory viruses by routine IFA. RSV was the most frequently detected virus in the analyzed period (83.64% of total positive cases), followed by HPIV (7.26%), FluA (5.89%), AdV (2.55%), and FluB (0.66%). HPIV3 constituted the 95.45% of the total HPIV cases. The most frequent clinical diagnoses for HPIV3 cases were bronchiolitis and pneumonia (75%).

The seasonality analysis of the total HPIV3 incidence showed that during 2009 and 2011, the positive cases occurred between June and October, while during 2012 and 2013, cases occurred between April and December extending to the first months of next year. In 2010, the epidemiological distribution began in August and continued until the first months of 2011 (Fig. 1A).

3.2. Phylogenetic analyses

In order to genotype Argentine HPIV3 strains from hospitalized patients in HNRG, the partial HN gene was sequenced (1,146 bp). A total of 124 sequences were obtained. Of these, 45 resulted identical sequences. It should be noted that only two sequences from 2010 were obtained. Unfortunately, due to a technical problem, it was not possible to molecularly characterize more strains; however, the epidemiological data from that year were available.

To genotype the HNRG strains, phylogenetic analyses were performed with HNRG sequences, reference sequences of known genotype and recently reported sequences. The most fit nucleotide substitution model for the aligned sequences was TIM1 plus invariant sites and gamma distributed (TIM1 + I + G). Similar tree topologies and statistical support were obtained with the three phylogenetic methods applied. Fig. 2 shows the Bayesian phylogenetic reconstruction (other trees under request).

From the phylogenetic analysis, it could be determined that all HNRG sequences corresponded to cluster C, and within this, the sequences were distributed in subclusters C1, C3, and C5. Table 1 shows that C3 was the most predominant subcluster during the analyzed period.

The exhaustive analysis of the phylogenetic tree also allowed us to identify that several HNRG sequences were grouped in individual well-defined phylogenetic clades within the subclusters C1 and C3. All of them were statistically well-supported. Thus, to determine if these HNRG phylogenetic clades corresponded to new genetic lineages, the genetic distances within and among these clades were calculated. The results are shown in Table 2. A new local genetic lineage in subcluster

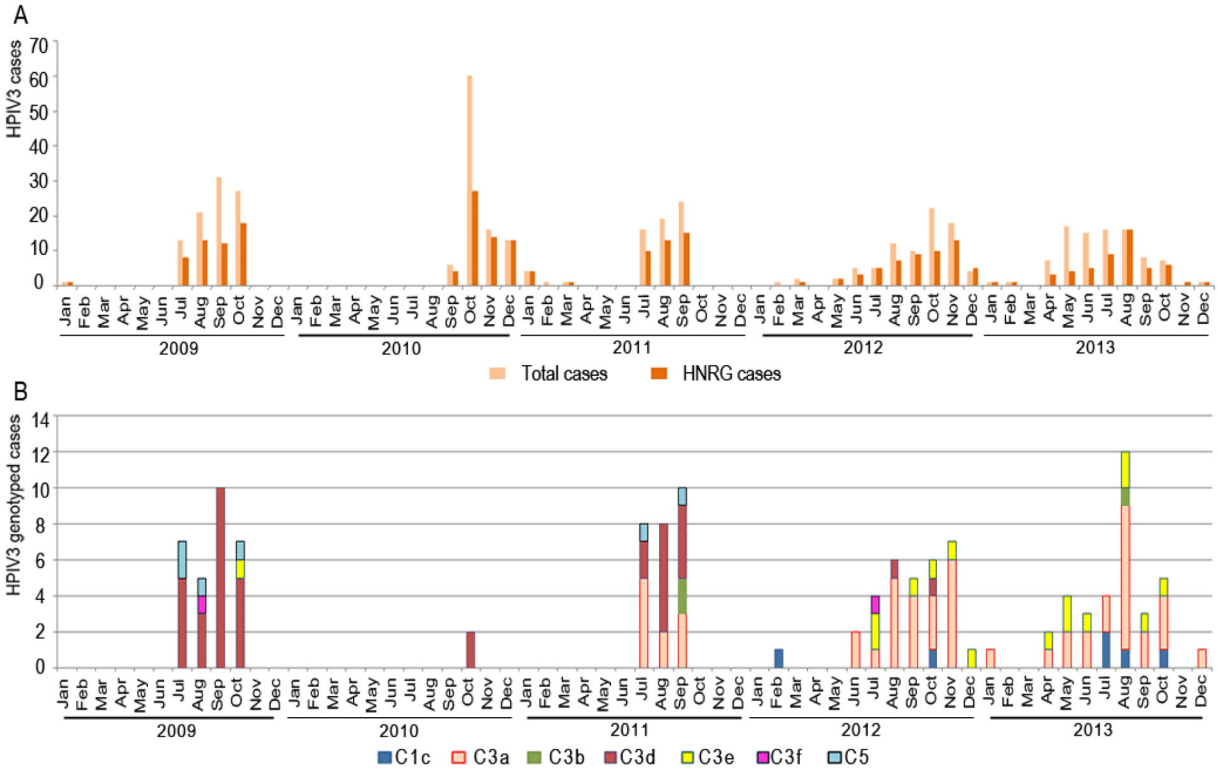

Fig. 1. Seasonality of HPIV3-positive cases per month between 2009 and 2013. A) Total HPIV3-positive cases referred to the Virology Laboratory of HNRG for viral diagnosis and HPIV3-positive cases from patients hospitalized in HNRG. B) Number of genotyped HNRG strains according to the lineage, per month per year.

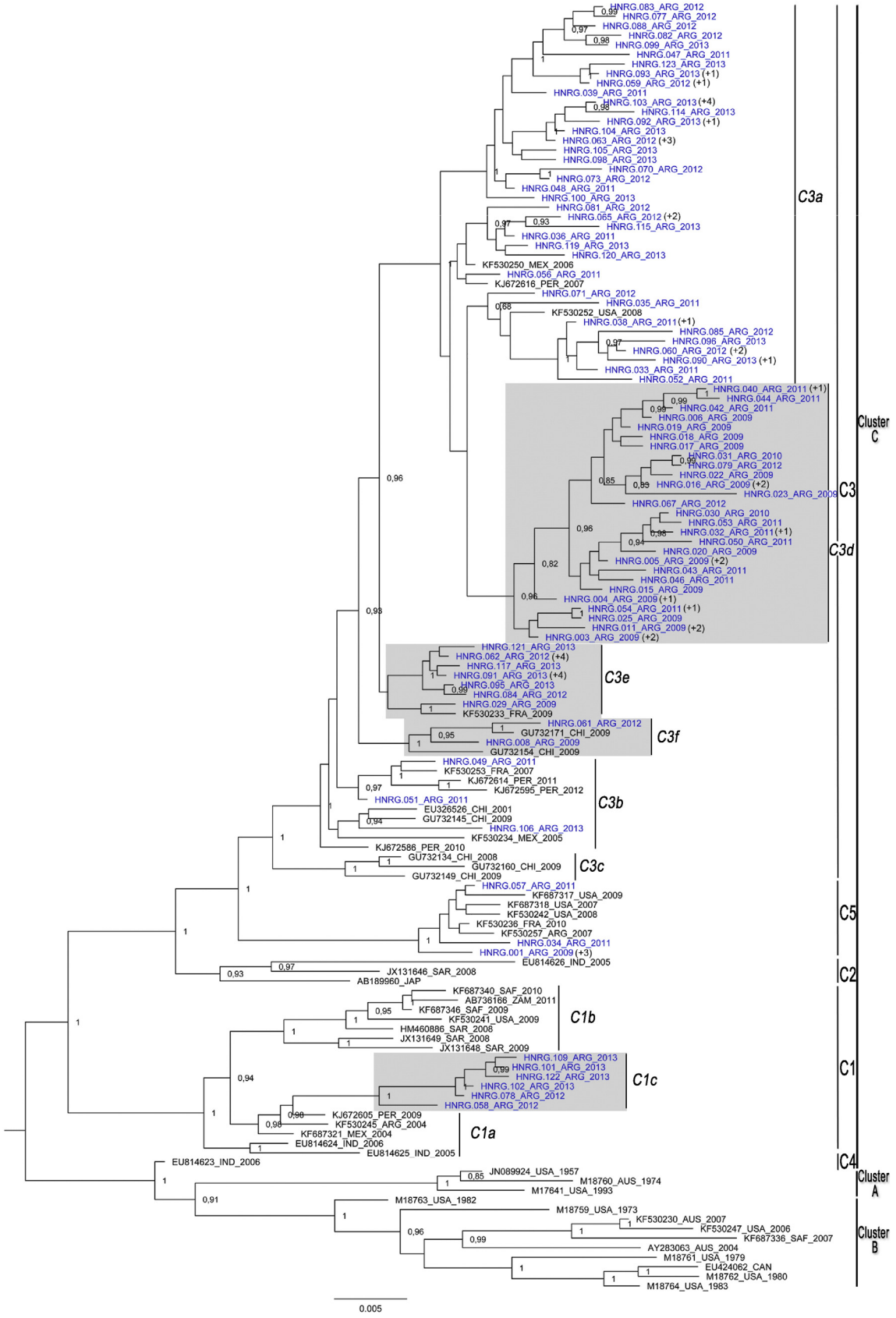

Fig. 2. Bayesian phylogenetic tree. The evolutionary model TIM + I + G was used for Bayesian reconstruction. Bayesian Markov Chain Monte Carlo (MCMC) chains were run for 10 million ngen, 10,000 sample freq to reach convergence (ESS > 200), and 10% burnin. The pool sequences include 85 HNRG sequences (with the number of identical sequences detailed in brackets), 45 references sequences from different geographic origins with previous genetic classification, and five previously published unclassified sequences. Only posterior probabilities above 0.6 are shown. Cluster, subcluster, and genetic lineage classification is remarked with vertical lines. New defined genetic lineages are highlighted in grey color. HNRG sequences are highlighted in blue.

Table 1
Number of HNRG HPIV3 strains according to the subclusters assigned by phylogenetic analyses, per year.

Subcluster	2009	2010	2011	2012	2013	Total
C1				2	4	6
C3	25	2	24	30	31	112
C5	4		2			6
Total	29	2	26	32	35	124

C1, defined as C1c, and three new genetic lineages within subcluster C3, defined as C3d–C3f, with genetic distances greater than 0.010 were determined. The genetic lineage C3d was strictly of local circulation during the analyzed period. In addition, three reference strains, from France and China, previously classified within genetic lineage C3a were reclassified within the new defined genetic lineages C3e (KF530233) and C3f (GU732154 and GU732171) (Fig. 2).

The seasonality analysis of the genotyped HPIV3 HNRG strains shows that there was cocirculation of subclusters during the analyzed period. Subcluster C3 circulated during all the period, while subclusters C1 (lineage C1c) and C5 circulated in 2012/2013 and 2009/2011, respectively (Fig. 1B).

Regarding the seasonality of genetic lineages of subcluster C3, C3a and C3d were the most prevalent lineages found during the analyzed period. C3d circulated from 2009 to 2011, and there has been a gradual replacement of C3d by C3a since 2011. In 2012, C3a prevailed over C3d and remained as the most prevalent lineage until 2013. Although there were lineages of strictly local circulation (C1c and C3d), there were also other genetic lineages which cocirculated in Argentina and other parts of the world, such as C3a and C3b.

From the phylogenetic tree, there is also an interesting association of sequences supported by a high posterior probability. Within lineage C3e, there is a phylogenetic clade composed of a French sequence of March 2009 (KF530233) and an Argentine sequence of October 2009 (HNRG.029_ARG_2009).

In order to validate the genotyping of HPIV3 described above versus the previous classification of genotypes made by Mao et al. and Almajhdi, a preliminary phylogenetic analysis was performed. Only the partial HN gene (globular head) was used instead of the complete HN gene of the reference sequences reported by them. The same genetic clades, statistically well-supported, were obtained (data not shown).

3.3. Molecular characterization and adaptive evolutionary analyses

A total of 195 nucleotide substitutions were found in the HNRG HN sequence alignment taking the HPIV3 prototype strain Washington/1957 as reference. Of these, 184 were polymorphic sites, with 11 sites presenting two nucleotide variants.

Of the total nucleotide substitutions, 167 were synonymous substitution and 28 were non-synonymous substitutions (Table 3).

At the amino acid level, the changes were distributed evenly throughout the globular head. The amino acids changes H295Y, I391V, and D556N were found in all the analyzed HNRG sequences. The selection pressure analysis revealed no positive selection and limited negative selection for HNRG sequences (18 negative selected codons were confirmed by three methods), the overall Ω ratio was 0.042 ($p < 0.05$).

The glycosylation analysis predicted four N-glycosylation sites in all the analyzed sequences (positions N308, N351, N485, and N523) and a new N-glycosilation site in three HNRG sequences from 2011 (D441N).

In order to assess whether non-synonymous substitutions occur in antigenic regions, a prediction of the conformational B-cell epitopes by using the molecular model of the globular head region of the reference strain (PDB: 4MZA) was performed. There were 74 sites predicted as potential B-cell epitopes with a specificity of 75%. Seven non-synonymous changes, found in HNRG sequences, were placed in the predicted epitopes. In addition, by using the PDBePISA service, it was determined that all of these epitopes were located in solvent-accessible areas on the surface of the protein globular head. Besides, 24 linear B-cell epitopes were predicted with a specificity of 75%, and eight non-synonymous changes, found in HNRG sequences, were placed within these epitopes. Two non-synonymous changes were found as B-cell epitopes by two methods. The results of these analyses are summarized in Table 3.

4. Discussion

In this paper, the molecular epidemiology of HPIV3 in hospitalized children was described for the first time in Argentina. Although there are several studies on the epidemiology of respiratory virus in Latin America and a study from Villaran et al. about molecular epidemiology of HPIV in this region, the information regarding the detailed molecular characteristics of the local circulating genetic lineages of HPIV3 is scarce (Villaran et al., 2014).

The results of this study arose from the analysis of a total of 19,435 clinical samples from pediatric patients hospitalized in a public hospital. Of the total cases, 6,264 were positive for respiratory viruses by IFA. This technique, which detects active infection, is recommended for routine diagnosis of a large number of clinical samples in a pediatric hospital. IFA enables physicians to obtain results within few hours after the sample was taken thus reducing overall antibiotic use and controlling nosocomial transmission (Pavia, 2011). The results obtained by this method showed that HPIV3 was the second detected virus after RSV. However, it cannot be ruled out that by using molecular diagnosis methods with higher sensitivity and a wider range of detected viruses (rhinovirus, metapneumovirus, etc), these results could be modified. Molecular diagnosis studies on these samples will be necessary to determine this impact, as well as the presence of viral co-detections.

During all the analyzed period, no seasonality pattern was observed. The number of IFA-positive cases did not show an incidence distributed between spring and summer as is commonly reported, except in 2010. In 2009 and 2011, positive cases began in winter and extended to spring, while in 2012 and 2013, they started in autumn and extended to late spring and early summer (Villaran et al., 2014). Interestingly, although the total number of HPIV3-positive cases per year was similar,

Table 2
Estimates of evolutionary divergence (genetic distances) over sequence pairs among and within genetic lineages. Genetic distances among new genetic lineages defined in this work and other genetic lineages of the same subcluster are denoted in bold.

	C1a	C1b	C1c	C3a	C3b	C3c	C3d	C3e	C3f
C1a	0.0090								
C1b	0.0147	0.0100							
C1c	**0.0150**	**0.0219**	0.0040						
C3a	0.0366	0.0393	0.0420	0.0070					
C3b	0.0322	0.0348	0.0378	0.0163	0.0130				
C3c	0.0333	0.0359	0.0387	0.0196	0.0177	0.0080			
C3d	0.0366	0.0386	0.0416	**0.0107**	**0.0181**	**0.0227**	0.0050		
C3e	0.0315	0.0342	0.0368	**0.0104**	**0.0109**	**0.0154**	**0.0123**	0.0020	
C3f	0.0343	0.0361	0.0389	**0.0164**	**0.0151**	**0.0191**	**0.0185**	**0.0103**	0.0080

Table 3
Molecular characterization and B-cell epitopes prediction analyses. Non-synonymous changes found in HNRG sequences and correlation with their assigned genetic lineages and the predicted B-cell epitopes.

Amino acid position[#]	Amino acid substitution[#]	HNRG strains	Genotype	Conformational B-cell epitope prediction	Linear B-cell epitope prediction
191	V → I	HNRG.049_ARG_2011	C3b		TVDGC**V**
193	T → A	HNRG.114_ARG_2013	C3a		
197	V → A	C1c strains	C1c		
260	L → I	HNRG.053_ARG_2011	C3d		
268	L → P	HNRG.122_ARG_2013 HNRG.123_ARG_2013	C1c and C3a, respectively		
269	C → Y	HNRG.046_ARG_2011	C3d		
271	T → I	HNRG.109_ARG_2013	C1c		**T**PKVDERSDYASSG
274	V → I	HNRG.081_ARG_2012	C3a		TPK**V**DERSDYASSG
295	H → Y	all strains			
306	N → D	HNRG.065_ARG_2012 HNRG.069_ARG_2012HNRG.072_ARG_2012	C3a		
314	P → S	HNRG.117_ARG_2013	C3e		**P**
343	P → L	HNRG.077_ARG_2012	C3a	+	LEH**P**IN
346	E → D	HNRG.043_ARG_2011	C3d	+	
369	S → I	HNRG.085_ARG_2012	C3a		TGCPGKTQRDCNQASH**S**PWF
370	P → S	HNRG.114_ARG_2013	C3a	+	TGCPGKTQRDCNQASHS**P**WF
384	V → I	HNRG.082_ARG_2012	C3a		
387	G → S	HNRG.099_ARG_2013	C3a	+	
391	I → V	all strains		+	
410	G → R	HNRG.106_ARG_2013	C3b		
426	T → S	HNRG.020_ARG_2009	C3d		**T**S
430	S → T	HNRG.047_ARG_2011	C3a		
441	D → N	HNRG.040_ARG_2011 HNRG.041_ARG_2011 HNRG.044_ARG_2011	C3d	+	
450	T → A	HNRG.044_ARG_2011	C3d		
522	R → Q	HNRG.015_ARG_2009	C3d	+	
524	K → T	HNRG.115_ARG_2013	C3a		
542	G → R	HNRG.047_ARG_2011	C3a		
556	D → N	all strains			

+ corresponds to amino acid substitution that occurs in amino acid site predicted as conformational B-cell epitope.
In the "Linear B-cell epitope prediction" column, the amino acid that was substituted in the epitope is denoted in bold.
Abbreviation: HNRG, Ricardo Gutierrez Children's Hospital.
[#] In reference to the HPIV3 prototype strain Washington/1957 (JN089924).

during 2009 and 2011, the annual distribution was sharper than in 2012 and 2013. This observation could be due to the predominance of some genetic lineages over others (see below).

The phylogenetic analysis showed that all HNRG sequences were classified in cluster C, which is the most recent globally circulating cluster as described in previous studies of Mao et al. (2012) and Almajhdi (2015). Within cluster C, the sequences were distributed in subclusters C1, C3, and C5, with C3 as the most prevalent subcluster found during the analyzed period. The analysis of genetic distances within these subclusters allowed us to describe the emergence of four novel genetic lineages highly supported in the phylogenetic analyses (called C1c, C3d, C3e, and C3f). The lineages C1c and C3d were of purely local circulation during the analyzed period. Besides, other HNRG sequences were associated with previously defined genetic lineages (C3a and C3b) that included sequences from distant countries.

Interestingly, there was a progressive replacement of C3d by C3a between 2011 and 2012, the two most prevalent genetic lineages found during the analyzed period. In addition, subcluster C5 circulated during the first years of the analysis while subcluster C1 (lineage C1c) circulated during the last years, supporting the idea that old lineages have been replaced by emerging genetic lineages. These observations could explain the different shapes of the annual distribution plots described before. However, considering that in this study only strains from hospitalized patients were analyzed, the possibility that other underestimated genetic lineages could be circulating in the general population cannot be ruled out.

The phylogenetic analyses showed that the transmission chains of HPIV3 could be explained at three levels: local, regional, and global. Two genetic lineages of strictly local circulation were found during the analyzed period which showed local evolution in consecutive years (C3d and C1c), whereas the remaining genetic lineages showed no geographical restriction, such as C3f, which has sequences from China (2009) and Argentina (2009 and 2012). The latter denote the ability for extensive and rapid spread of viral strains, probably related to the form of respiratory transmission and global international communications. Nevertheless, an exhaustive phylogeographic analysis to confirm this hypothesis will be necessary.

Although a great number of nucleotide polymorphisms were found in HNRG sequences, only 14% of them corresponded to non-synonymous changes. Considering the selection pressure analysis, only codons negatively selected were found and the overall Ω ratio was significantly lower than one ($dN/dS < 1$), suggesting that a strong purifying selection might be acting on these viruses probably due to viral protein structural constraints of the HN, attributable in part to neuraminidase activity and receptor binding.

van Wyke Coelingh et al. described amino acid residues important for the structures of the neutralization epitopes of the HN by sequence analysis of neutralization-resistant mutants selected with MAbs to sites A, B, or C, previously defined (van Wyke Coelingh et al., 1987).These sites inhibit hemagglutination and neuraminidase activities. Two non-synonymous substitutions, S369I and P370S, found in HNRG strains (HNRG.085_ARG_2012 and HNRG.114_ARG_2013, respectively), were located in antigenic site A, and they were experimentally demonstrated as neuraminidase inhibitors. In addition, both amino acids positions were predicted as B-cell epitopes in this study. Interestingly, the strain HNRG.114_ARG_2013 also presented another non-synonymous substitution described in a strain which exhibited a markedly fusogenic plaque morphology (T193A) (Porotto et al., 2007). Although an in vitro and in vivo analysis is necessary to confirm the functionality of these findings with these strains, this information may

contribute to the general knowledge of the HN protein as candidate for future prophylactic therapies or vaccine development.

5. Conclusions

In conclusion, the emergence of four genetic lineages with local and global circulation patterns was described. Besides, non-synonymous substitutions of the HN protein globular head as part of potential B-cell epitopes were detected through bioinformatic analyses, two of which were experimentally confirmed as part of neutralizing epitopes in previous works (van Wyke Coelingh et al., 1987). These mutations were located in the HN sequence of HNRG strains, clustered in the redefined C3a genetic lineage, which turned to a regional lineage (USA, Peru, Mexico, and Argentina). Notably, all substitutions found and identified as potential epitopes occurred in sequences that were part of the C3 subcluster. All this information together is important not only to understand the evolution of HPIV3 but also to evaluate whether the strain composition of a vaccine that is currently in Phase I clinical trials will be effective to protect pediatric patients against viruses circulating in our region (Schmidt et al., 2011).

Acknowledgments

We thank to M. Campal, M. A. Márques, P. M. Riveiro, and J. Caruso from the laboratory staff. We appreciate the excellent technical support of S. B. Lusso and M. Natale. We thank Dr. G. Parisi for providing information of the bioinformatic tools for detecting epitopes. We acknowledge Dr. L. E. Valinotto and MSc E. Tittarelli for the critical reading of the manuscript.

This work was partially supported by Consejo Nacional de Investigaciones Científicas y Tecnológicas, Argentina (grant PIP112-201101-00562 to MV). This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Almajhdi, F.N., 2015. Hemagglutinin-neuraminidase gene sequence-based reclassification of human parainfluenza virus 3 variants. Intervirology 58, 35–40. http://dx.doi.org/10.1159/000369208.

Karron, R.A., Collins, P.L., 2013. Parainfluenza viruses. In: Knipe, D., Howley, P. (Eds.), Fields Virology, sixth ed. 1. LWW, Philadelphia, pp. 996–1023.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772. http://dx.doi.org/10.1038/nmeth.2109.

Gardner, P.S., McQuilin, J., 1968. Viral diagnosis by immunofluorescence. Lancet 1, 597–598.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. http://dx.doi.org/10.1093/sysbio/syq010.

Haste Andersen, P., Nielsen, M., Lund, O., 2006. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 15, 2558–2567 (http://tools.immuneepitope.org/stools/discotope/discotope.do).

Henrickson, K.J., 2003. Parainfluenza viruses. Clin. Microbiol. Rev. 16, 242–264.

Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222 (http://www.datamonkey.org/).

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (http://www.ebi.ac.uk/pdbe/pisa/).

Larsen, J.E.P., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2 (http://tools.immuneepitope.org/bcell/).

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. http://dx.doi.org/10.1093/bioinformatics/btp187.

Liu, W., Liu, Q., Chen, D., Liang, H., Chen, X., Huang, W., Qin, S., Yang, Z., Zhou, R., 2013. Epidemiology and clinical presentation of the four human parainfluenza virus types. BMC Infect. Dis. 13, 28. http://dx.doi.org/10.1186/1471-2334-13-28.

Mao, N., Ji, Y., Xie, Z., Wang, H., Wang, H., An, J., Zhang, X., Zhang, Y., Zhu, Z., Cui, A., Xu, S., Shen, K., Liu, C., Yang, W., Xu, W., 2012. Human parainfluenza virus-associated respiratory tract infection among children and genetic analysis of HPIV-3 strains in Beijing, China. PLoS One 7, e43893. http://dx.doi.org/10.1371/journal.pone.0043893.

Mizuta, K., Tsukagoshi, H., Ikeda, T., Aoki, Y., Abiko, C., Itagaki, T., Nagano, M., Noda, M., Kimura, H., 2014. Molecular evolution of the haemagglutinin-neuraminidase gene in human parainfluenza virus type 3 isolates from children with acute respiratory illness in Yamagata prefecture, Japan. J. Med. Microbiol. 63, 570–577. http://dx.doi.org/10.1099/jmm.0.068189-0.

Palermo, L., Porotto, M., Yokoyama, C.C., Palmer, S.G., Mungall, B.A., Greengard, O., Niewiesk, S., Moscona, A., 2009. Human parainfluenza virus infection of the airway epithelium: viral hemagglutinin-neuraminidase regulates fusion protein activation and modulates infectivity. J. Virol. 83, 6900–6908. http://dx.doi.org/10.1128/JVI.00475-09.

Pavia, A.T., 2011. Viral infections of the lower respiratory tract: old viruses, new viruses, and the role of diagnosis. Clin. Infect. Dis. 52, S284–S289. http://dx.doi.org/10.1093/cid/cir043.

Porotto, M., Fornabaio, M., Kellogg, G.E., Moscona, A., 2007. A second receptor binding site on human parainfluenza virus type 3 hemagglutinin-neuraminidase contributes to activation of the fusion mechanism. J. Virol. 81, 3216–3228.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schmidt, A.C., Schaap-Nutt, A., Bartlett, E.J., Schomacker, H., Boonyaratanakornkit, J., Karron, R.A., Collins, P.L., 2011. Progress in the development of human parainfluenza virus vaccines. Expert Rev. Respir. Med. 5, 515–526. http://dx.doi.org/10.1586/ers.11.32.

Spriggs, M.K., Murphy, B.R., Prince, G.A., Olmsted, R.A., Collins, P.L., 1987. Expression of the F and HN glycoproteins of human parainfluenza virus type 3 by recombinant vaccinia viruses: contributions of the individual proteins to host immunity. J. Virol. 61, 3416–3423.

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. http://dx.doi.org/10.1093/molbev/mst197.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

VAN Wyke Coelingh, K.L., Winter, C.C., Jorgensen, E.D., Murphy, B.R., 1987. Antigenic and structural properties of the hemagglutinin-neuraminidase glycoprotein of human parainfluenza virus type 3: sequence analysis of variants selected with monoclonal antibodies which inhibit infectivity, hemagglutination, and neuraminidase activities. J. Virol. 61, 1473–1477.

Viegas, M., Barrero, P.R., Maffey, A.F., Mistchenko, A.S., 2004. Respiratory viruses seasonality in children under five years of age in Buenos Aires, Argentina: a five-year analysis. J. Infect. 49, 222–228.

Villaran, M.V., García, J., Gomez, J., Arango, A.E., Gonzales, M., Chicaiza, W., Alemán, W., Lorenzana de Rivera, I., Sanchez, F., Aguayo, N., Kochel, T.J., Halsey, E.S., 2014. Human parainfluenza virus in patients with influenza-like illness from Central and South America during 2006-2010. Influenza Other Respir. Viruses 8, 217–227. http://dx.doi.org/10.1111/irv.12211.

Weinberg, G.A., Hall, C.B., Iwane, M.K., Poehling, K.A., Edwards, K.M., Griffin, M.R., Staat, M.A., Curns, A.T., Erdman, D.D., Szilagyi, P.G., 2009. Parainfluenza virus infection of young children: estimates of the population-based burden of hospitalization. J. Pediatr. 154, 694–699. http://dx.doi.org/10.1016/j.jpeds.2008.11.034.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., Korber, B., 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology 14, 1229–1246 (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html).