Accepted Manuscript

Sixteen years of evolution of human respiratory syncytial virus subgroup A in Buenos Aires, Argentina: GA2 the prevalent genotype through the years

M. Viegas, S. Goya, A.S. Mistchenko

Please cite this article as: Viegas, M., Goya, S., Mistchenko, A.S., Sixteen years of evolution of human respiratory syncytial virus subgroup A in Buenos Aires, Argentina: GA2 the prevalent genotype through the years, (2016), doi: 10.1016/j.meegid.2016.04.034

Sixteen years of evolution of human respiratory syncytial virus subgroup A in Buenos Aires, Argentina: GA2 the prevalent genotype through the years.

Viegas, M;[a,b]# Goya, S;[a,b] and Mistchenko, A. S.[a,c]

[a]Virology Laboratory, Ricardo Gutiérrez Children`s Hospital; Gallo 1330 2º,

(1425) Ciudad Autónoma Buenos Aires, Argentina

[b]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina;

[c]Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), Argentina.

e-mail addresses: goyastephanie@conicet.gov.ar; asmistchenko@hotmail.com

#Corresponding author:

Mariana Viegas, viegasmariana@conicet.gov.ar, Gallo 1330 2º, (1425) Ciudad Autónoma Buenos Aires, Argentina

Abstract

Human respiratory syncytial virus (HRSV) is the main viral cause of acute lower respiratory tract infections (LRTI) in children worldwide. In recent years, several preclinical trials with vaccine candidates have been reported. It is in this sense that molecular epidemiological studies become important. Understanding viral dispersion patterns before and after the implementation of a vaccine can provide insight into the effectiveness of the control strategies. In this work we analyzed the molecular epidemiology of HRSV-A over a period of sixteen years (1999-2014) in Buenos Aires. By bioinformatic tools we analyzed 169 sequences of the G glycoprotein gene from hospitalized pediatric patients with LRTI. We found that GA2 was the most prevalent genotype (73.35%). GA5 genotype co-circulated in our region until 2009 when it was no longer detected, except in 2011. The recently globally emerging ON1 lineage with a 72-nt duplication increased its frequency to become the only lineage detected in Buenos Aires in 2014. By discrete phylogeographic analysis of global ON1 strains we could determine that Panama could be the location of the MRCA dated June 20, 2010; and this lineage could be introduced in Argentina from Spain in April, 2011. This analysis also showed temporary and geographical clustering of ON1 strains observed as phylogenetic clades with strains exclusively associated from a single country, nevertheless among our 44 ON1 strains from three outbreaks (2012-2014) we could also detect posterior reintroductions and circulation from United States, Cuba, South Korea, and Spain. The continuous phylogeographic analysis of one sublineage of Argentine ON1 strains allowed us to establish that there could be a local clustering of some strains even in neighborhoods. This work shows the potential of this type of bioinformatic tools in the context of a future vaccine surveillance network to trace the spread of new genetic lineages in human populations.

Keywords
HRSV; genotypes; ON1 genetic lineage; phylogeography; circulation patterns.

1. Introduction

Human respiratory syncytial virus (HRSV) is the main viral cause of acute lower respiratory tract infections (LRTI) in children under five years of age, being the major cause of pediatric hospital admissions during autumn and winter in developing countries. In Argentina, HRSV has contributed yearly to approximately 78.8% of all viral LRTI among the cited age group (Carballal et al., 2001; Videla et al., 1998; Viegas et al., 2004).

HRSV is an enveloped virus with a non-segmented negative-strand RNA genome. It was classified as a member of the *Pneumovirus* genus of the *Paramyxoviridae* family (Collins and Karron, 2013)). HRSV exists as a single serotype, but has two antigenic subgroups A and B (Coates et al., 1966). Among the proteins of HRSV, the G protein stands out for its high degree of both sequence and antigenic diversity among different strains, and it has been used to classify the circulating viral strains in different genotypes (Peret et al., 1998). The diversity resides mainly in the ectodomain, which has only 44% of amino acid identity between both HRSV subgroups A and B, compared with 83% for the transmembrane and cytoplasmic domain. This great diversity is located in the two variable regions on both sides of a conserved central domain rich in Cys (Collins and Karron, 2013).

Epidemics are commonly produced by both antigenic subgroups, classified in more than one genotype. Thirteen HRSV subgroup A (HRSV-A) genotypes have been described from different geographical areas, and they have been designated from GA1 to GA7 (Peret et al., 2000, 1998), SAA1 (South Africa A1) (Venter et al., 2001) and most recently, NA1-NA4 (Cui et al., 2013; Shobugawa et al., 2009) and ON1 with a 72-nt duplication (Eshaghi et al., 2012). HRSV subgroup B (HRSV-B) genotypes include GB1 to GB4 (Peret et al., 1998), SAB1 to SAB3 (Venter et al., 2001), URU1-2 (Blanc et al., 2005) and BA1 to BA10 (Buenos Aires) with a 60-nt duplication (Dapat et al., 2010; Trento et al., 2006, 2003).

Abbreviations: BA Buenos Aires, BF Bayes factor; ESS effective sample size; IFA immunofluorescence assay; LRTI lower respiratory tract infection; MCCT maximum clade

credibility tree; MCMC Markov Chain Monte Carlo; MRCA most recent common ancestor; NPA nasopharyngeal aspirates; SE standard error.

Although several studies have reported that both subgroups have shown an alternating dominance among outbreaks, in recent years subgroup A has prevailed over subgroup B (Duvvuri et al., 2015; Pretorius et al., 2013). Recently, Trento et al. reported a reclassification of HRSV-A genotypes on the basis of the p-distance method, and they defined the maximum p-distance within the GA1 genotype as a cut-off value which defines genotypes. As a consequence, they determined that NA1, NA2, NA4 and ON1 which were previously classified as emerging genotypes corresponded to emerging genetic lineages within the GA2 genotype (Trento et al., 2015).

Although numerous new genotypes and genetic lineages of HRSV have been reported and some emerged recently and others have stopped circulating in the population, during the last years, GA2 was the most frequent genotype among the studied HRSV-A strains (Trento et al., 2015). The new genetic lineages NA1-NA4 and the strain recently described in Canada, ON1probably descended from GA2 (Agoti et al., 2014; Cui et al., 2013; Eshaghi et al., 2012; Trento et al., 2015).

The emergence in 2010 of the ON1 strain with such a drastic molecular event as a 72-nt duplication and the bioinformatic tools based on the Coalescent Theory developed in recent years, can allow to describe the population dynamics of this new genetic lineage and may provide insights into parameters, such as evolutionary rate, location and time of the most recent common ancestor (MRCA) and demographic histories that are difficult to assess using traditional surveillance analyses. These tools also allow to describe the temporospatial diffusion at local and global levels of the ON1 genetic lineage, thus describing the transmission chains of HRSV in a simplest way. Our laboratory has been conducting the molecular characterization of HRSV for sixteen years, and we have found that GA2 (HRSV-A) was the most frequent genotype (Viegas and Mistchenko, 2005), and since 2012 we have detected almost entirely ON1 strains within HRSV-A. In this context, the aim of this study was to perform phylodynamic and phylogeographic analyses of HRSV-A over a period of sixteen years at local and global levels by taking advantage of the

emergence of the new genetic lineage ON1 to reconstruct viral dispersion patterns in space and time.

2. Materials and methods

2.1 Ethical statements

The study was reviewed and approved by the Medical Ethics and Research Committees of Hospital de Niños "Ricardo Gutiérrez", Buenos Aires, Argentina (IRB N° 10.27). Informed consent was not obtained, as patient information was anonymized and de-identified prior to analysis.

2.2 Clinical samples

The clinical samples were nasopharyngeal aspirates (NPA) which resulted positive for HRSV by an indirect immunofluorescence assay (IFA) with monoclonal antibodies (Light Diagnostics, Chemicon Int. Inc., USA) (Gardner and McQuilin, 1968).They belonged to pediatric patients (<5 years of age) hospitalized with acute LRTI at Hospital de Niños "Ricardo Gutiérrez", in Buenos Aires, Argentina, during 1999-2014.

The HRSV analyzed in this study corresponds to subgroup A. The subgroup was determined on randomly-selected clinical samples by multiplex nested RT-PCR previously described by Stockton et al. (Stockton et al., 1998). Samples which proved positive for HRSV-A were kept frozen at -70ºC until the molecular analysis was performed.

2.3 RNA extraction, G gene amplification and sequencing

Nucleic acids were extracted directly from NPA samples with a PureLink viral RNA/DNA minikit (Invitrogen Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions. RNA was reverse-transcribed and the ectodomain of G glycoprotein gene amplified using the Qiagen OneStep RT-PCR kit (Qiagen, GmbH, Hilden, Germany) following the manufacturer's instructions with the primers RSVABG10 5'-CGCAATGATAATCTCAACCTC-3' and RSVABF1 5'-CAACTCCATKGTTATTTGCC-3'. The retrotranscription and amplification conditions for the ectodomain of G glycoprotein gene were: 50ºC for 30 min; 95ºC for 15 min followed by 34 cycles of 94ºC for 60 sec, 59ºC for 60 sec, 72ºC for 120 sec, and a final extension of 72°C for 7 min.

PCR products were electrophoresed in a 1.5% agarose gel stained with ethidium bromide and purified with the Zymoclean™ Gel DNA Recovery Kit (Applied Biosystems, Foster City, CA, USA). The purified PCR products were labelled with the BigDye Terminator v3.1 sequencing kit (Applied Biosystems, Foster City, CA, USA) and electrophoresed in an ABI3500 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

The consensus nucleotide sequences were analyzed, assembled and generated with the SeqScape Software v2.7 (Applied Biosystems, Massachusetts, USA) and aligned with CLUSTAL W v2 software (Thompson et al., 1994). Most of the sequences corresponding to the period between 1999 and 2004 were previously published by our group in 2005 (Viegas and Mistchenko, 2005). All new sequences were submitted to GenBank (GenBank accession numbers: KU350758-KU350870).The viral strains nomenclature indicates the country of isolation (Argentina, ARG), the number of the isolate (internal laboratory coded number) and the date of sample collection. All reference sequences of known genotype were downloaded from GenBank.

2.4 Phylogenetic analyses

For the phylogenetic and evolutionary analyses, the general time reversible plus gamma distribution (GTR+G) was the most suitable model for the alignment and was selected with jModelTest v0.1.1 software (Darriba et al., 2012). The Bayesian inference was performed with MrBayes v.3.2.1 software (Ronquist and Huelsenbeck, 2003), whereas PhyML v.3.1 software was used for Maximum Likelihood (ML) (Guindon et al., 2010) and MEGA v6 software (Tamura et al., 2013) for Distance Methods (Neighbor joining, NJ) . Consensus trees were visualized with FigTree v.1.4.0.

2.5 Molecular characterization

Estimates of mean genetic distances (the number of base differences per site from averaging over all sequence pairs, p-distances) were used to calculate the divergence among sequences within and between genotypes and genetic lineages by using MEGA v6. Standard error (SE) estimates were obtained by the bootstrap method (1000 replicates) (Tamura et al., 2013).

2.6 Evolutionary rates, population dynamics and phylogeographic analyses

Evolutionary rates, time of the MRCA, demographic histories, and discrete and continuous phylogeographic analyses were estimated from the sequences stamped with date and location (latitude and longitude) using the Bayesian approach with the BEAST v1.8.2 package (Drummond et al., 2012). The uncorrelated lognormal relaxed clock model and the Bayesian skyline demographic model were the most suitable implemented models evaluated with Bayes Factor (BF)among all molecular clock and demographic models available. The convergence of the Markov Chain Monte Carlo (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. Furthermore, Tracer v1.5 was used to determine the time of the MRCA, evolutionary rates and to plot the demographic histories. The latter were used to estimate the changes in the effective population size (Ne.т). These results were also analyzed using Tree-Annotator v1.8.2 to infer a maximum clade credibility trees (MCCT). The trees were visualized with FigTree v1.4.0. The SPREAD v1.0.6 program (Spatial Phylogenetic Reconstruction of Evolutionary Dynamics) (Bielejec et al., 2011) was used to summarize the phylogeographic analyses in the diffusion dynamic plots and to generate the interactive virtual global animations (KML archives) which were played by Google Earth (https://earth.google.com). In addition, SPREAD was used to calculate the BF to provide substantial strength to migration events between two locations (cut-off value BF>10).

3. Results

This study includes a period of sixteen years with a total of 578 HRSV IFA-positive samples. A total of 371 samples tested positive for subgroup A (64.18%) by multiplex nested RT-PCR. Of these, 169 sequences of the ectodomain of G glycoprotein gene (593-665 bp) were obtained in our laboratory (Table 1). Twenty-six sequences were identical. Fifty-five sequences covering the period 1999-2004 had been previously reported by our group (Viegas and Mistchenko, 2005) plus a total of 114 new sequences from 2004 to 2014.

In order to genotype all Argentine sequences, phylogenetic analyses were performed by Bayesian inference, ML and NJ. All phylogenetic analyses were carried out using an alignment consisting of all Argentine sequences plus 59 sequences of known genotype that had been downloaded from GenBank. All the phylogenetic analyses were congruent and gave the same association of sequences (only Bayesian analysis is shown in Fig. 1, other phylogenetic analyses under request). The phylogenetic analyses showed that Argentine strains are associated with two clearly defined and well-supported genetic clades which correspond with the previously described GA2 and GA5 genotypes. From the 169 analyzed Argentine sequences, GA2 was the most frequent genotype (124/169) and circulated almost all the studied period, except during 2002, when only GA5 strains were found, and during 2005 when no HRSV-A sequences were detected in Buenos Aires. Both genotypes co-circulated between 1999 and 2001, 2003 and 2004, 2006 and 2008 and during 2011. After 2008, GA2 circulated almost exclusively, except in 2011 when GA5 was also detected (Table 1). It should be noted that only two sequences from 2007 and three sequences from 2010 were obtained. Unfortunately, due to a technical problem, it was not possible to molecularly characterize more strains during these years.

In order to determine the demographic history of both HRSV-A genotypes in Buenos Aires, a population dynamics analysis was performed with the Argentine sequences. Fig. 2 shows that both genotypes alternated their effective populations during the sixteen consecutive years, with a stationary

interval during the first years of analysis (1999-mid 2001). After that, there was a marked decrease of GA5 between the end of 2000 and the end of 2001, and also accompanied by a slight decrease in the effective population of GA2. After 2002, there was an exponential growth in the population of GA5 which surpassed that of GA2 in 2003. Both genotype populations have remained almost constant until the period between 2010 and 2011, when the effective population of GA5 totally fell. During such period when GA5 was no longer detected among the HRSV studied population, a new genetic lineage ON1 had emerged globally.

In the phylogenetic tree of Fig. 1, within GA2 genotype, Argentine sequences are distributed throughout the entire clade, associated with sequences reported from several parts of the world, except for those belonging to the genetic lineages NA3 and NA4. The reference-sequences previously assigned to genetic lineages NA1and NA2 do not represent monophyletic clusters in the tree, thus the Argentine sequences could not be associated with these lineages. A different scenario is observed with the Argentine sequences with 72-nt duplication which are associated with reference sequences from ON1 strains in a well-defined genetic clade in the tree. In order to determine if this clade corresponds to a new genotype, the genetic distances (p-distances) between and within genetic clades were calculated for GA2 and ON1. The genetic distance was 0.016 (SE 0.002) within ON1, 0.0035 (SE 0.003) within GA2, and 0.040 (SE 0.005) between GA2 and ON1. Considering that 0.049 was defined by Trento et al. as the cut-off value to determine a genotype, ON1 constituted a new genetic lineage within GA2 genotype instead of a new genotype (Trento et al., 2015). Nevertheless, the frequency of this genetic lineage increased from 42% (8/19) in 2012, 94.4% in 2013 (17/18) to 100% (19/19) of all HRSV-A detected in our laboratory in 2014.

In order to infer the probable location and the time of the MRCA of ON1 strains and how they have been spread throughout the world from its origin, phylodynamic and phylogeographic analyses were performed.

The phylodynamic analysis of 177 ON1 sequences of 665-nt from Argentina plus those downloaded from GenBank until June 01,2015, showed that the time

of the MRCA for Argentine strains was April 22, 2011 (95% HPD interval; January 02, 2011, August 02, 2011) one year before the first ON1 strain was detected in our laboratory (April 10, 2012). At the global level the time of the MRCA for ON1 strains was June 20, 2010 (95% HPD interval; January 28, 2010, October 16, 2010), the first ON1 strain analyzed in this study dated from October 14, 2010 and was reported in Panama (GenBank accession number KF300973.1).The estimated mean evolutionary rate of ON1 strains was 4.87E-03 subs/site/yr (95% HPD interval: 3.89E03-5.81E-03).

The discrete phylogeographic analysis depicted by the MCCT (Fig. 3) and summarized in the diffusion dynamic plot (Fig. 4), shows that Panama could be the location of the MRCA with a probability value of 0.95. It also reveals how the different ON1 strains were globally disseminated from Panama. The interactive animation (Supplementary KML1) also shows how ON1 strains spread from Panama to Canada (BF=54), Spain (BF=29.9), Thailand (BF=18) and China. Then, one lineage from Spain disseminated to Argentina (BF=2206.5), and from Argentina to Cuba (BF=64.8) and USA (BF=104). It is interesting to note that after the first introduction to our country, there were multiple independent entries from USA, Cuba, South Korea, and Spain. Another interesting observation is how Argentine strains spread to Paraguay and strains from Spain entered this neighboring country simultaneously. Nevertheless, if sequences published during the preparation of this work from different parts of the world had been added to this analysis, it might have been different, however the high probability values and BF for each country supports the described dispersions.

Taking into account the multiple introductions observed in the discrete analysis, we really wondered if all ON1 strains have a common ancestor, thus, two phylogenetic analyses were performed with an alignment consisted of all Argentine sequences from GA2 genotype with and without the duplication. Supplementary Fig. 1 shows that the two analyses gave the same association of sequences, and all ON1 sequences remained associated on the same cluster regardless of not having duplication, supporting the idea that these viruses could be generated in a single genetic event.

In addition, evaluating the circles around each analyzed country in Fig.4, it is possible to see that there are many sampled lineages in our country, supporting the idea of multiple reintroductions (determined by the diameter of the circles and the color scale). However, it can also be explained by the diversity generated in our country. Despite the rapid global distribution of the ON1 strains, some of these have been circulating only in our region. In the MCCT there is a well-supported clade of Argentine strains sampled from 2012 to 2014, which displays only local circulation (Fig. 3). In order to analyze the spreading of the ON1 strains associated in this Argentine clade exclusively, we performed a continuous phylogeographic analysis. The diffusion dynamic plot is shown in Fig. 5A (Supplementary KML2). In this figure, it can be seen how these strains would have been dispersed in a circumscribed geographical region northwest of Buenos Aires . Somehow all these strains would have a geographical connection, while the other ON1 Argentine strains without distinction of association in the MCCT are distributed homogenously in Buenos Aires (Fig. 5B). Both the discrete and the continuous phylogeographic analyses can explain the circulation of HRSV ON1 strains at the global and local levels.

4. Discussion

Since its discovery in 1955, the HRSV has been the leading cause of acute LRTI in children younger than five years of age and older adults, representing a public health problem in every studied country. Despite this, there is still no approved vaccine for prevention. However, in recent years, several preclinical trials have been reported and most are coming to an end, thus the possibility of preventing this viral infection seems promising (Karron et al., 2015, 2013). It is in this sense that epidemiological studies become important before and after the approval of a vaccine. Understanding the phylodynamic shapes of a virus spreading before and after the implementation of a vaccine would help to quantify the change of the evolutionary rate over time and therefore provide insight into the effectiveness of control strategies (Volz et al., 2013).

In the Virology Laboratory of Hospital de Niños "Ricardo Gutiérrez", in Buenos Aires we have been conducting an active surveillance of acute LRTI in children less than five years of age since 1999 (Maffey et al., 2010; Viegas et al., 2004; Viegas and Mistchenko, 2005). In this scenario we analyzed the molecular epidemiology of HRSV-A over a period of sixteen years (1999-2014). We found that HRSV-A has prevailed over the subgroup B during the study, as it was reported in many studies (Pretorius et al., 2013; Shobugawa et al., 2009; Trento et al., 2015).

Despite the long period of time analyzed, we found only two genotypes among all HRSV-A virus detected between 1999 and 2014, GA2 and GA5. It is important to mention that a lack of detection of any given HRSV genotype does not mean that the lineage may have been extinct, since it may be experiencing circulation in a different population, such as older people, or it may be still circulating in other areas where we did not obtain samples. Thus, we are currently conducting a molecular epidemiology analysis of HRSV expanded into other areas of our region. Nevertheless, during the last years, many researchers have reported that these two genotypes of HRSV-A are almost the currently circulating genotypes in many parts of the world (Pretorius et al., 2013; Trento et al., 2015). Even more, in bordering countries, such as Brazil, Katzov-

Eckert et al. reported that these two genotypes were the main responsible for HRSV-A outbreaks between 1996 and 2006 (Katzov-Eckert et al., 2012).

Regarding demographic history of both HRSV-A genotypes of Fig. 2, it is interesting to note that the stationary interval of effective populations during the first years of the analysis coincides with the globally emergence of BA genotype of HRSV-B (Trento et al., 2010, 2003). It is possible that the emergence of this new genotype with a drastic genetic event such as a 60-nt duplication could affect the diversity of HRSV-A genotypes. Then the marked decrease of GA5 between the end of 2000 and the end of 2001 coincides with the appearance and fast dissemination of new BA lineages as was described by Trento et al. (2010). After this, the exponential growth of GA5 occurred in a period in which they reported that the effective population of BA genotype showed a depression. In addition, during 2002 GA5 was the only genotype detected among all the HRSV-A analyzed strains in our laboratory. After 2004, the effective population of both genotypes (GA2 and GA5) remained stationary and this was coincident with an exponential growth of BA genotype worldwide due to the replacement of all circulating BA lineages by the BA-IV lineage (Trento et al., 2010).As it was described by Trento et al. (2015), two events of diversification have occurred throughout the history of HRSV-A: one that gave rise to GA1, which apparently would be extinct today, and the other that gave origin to the currently circulating genotypes (GA2-GA7). While the force that gives rise to the phylogenetic tree branching is not entirely known, it is clear that we need to unify criteria for designation of genotypes and genetic lineages or variants. In the context of a nearby vaccine, global molecular surveillance of HRSV should be unified in order to compare data. Some studies, such as those of Venter in 2001 (Venter et al., 2001) and Trento in 2015, defined the assignment of new genotypes on the basis of genetic distances (p-distances). Thus, in their last work, Trento et al. determined a cut-off value of genetic distance to define genotype, as the highest intragenotypic p-distance within GA1, the oldest genotype far described. This value was obtained analyzing 2,167 sequences downloaded from GenBank resulting in a value of 0.049. Therefore, all the previously defined genotypes NA1-NA4 and ON1 corresponded to genetic lineages within the GA2 genotype (Trento et al., 2015).

Most of the analyzed sequences of HRSV-A in this work were associated with GA2 genotype (73.37%). Within this genotype, 44 sequences presented a 72-nt duplication in their terminal third and were associated with the previously defined ON1 lineage in a well-defined, well-supported genetic clade. It was speculated that large epidemics occurred due to the emergence of new genetic lineages, as it was observed with NA1 and NA2 in Japan between 2005 and 2007 (Shobugawa et al., 2009). Since the first detection of ON1 strains in our laboratory in 2012, ON1 frequency has been increasing to be the only genetic lineage detected within genotype GA2 of HRSV-A in 2014. This result is in contrast with the report of Duvvuri et al., where a low prevalence of 13% was found in 2011-2012 in Canada, however they explained that it could be due to the lack of active surveillance in that country (Duvvuri et al., 2015). In addition, the increase in global HRSV surveillance has demonstrated that ON1 lineage is disseminating very fast and diversifying in different sublineages as reported in Kilifi, Kenya, where two waves of ON1 strains were detected in 2012 with a prevalence of 74% (Agoti et al., 2014), as well as an increasing prevalence ranging from 60% in 2011-2012 to 89% in 2012-2013, in Texas, USA (Avadhanula et al., 2015) and 94.6% in Korea in 2012-2013 (Kim et al., 2014).

The ON1 genetic lineage, like the BA genotype for HRSV-B, has spread all over the world in a very short period of time and it has almost completely replaced the other genotypes of HRSV-A, as shown in the phylodynamic analysis of Fig. 2, where the effective population of GA5 totally fell between 2010 and 2011concomitant with the emergence of ON1lineage within GA2 genotype (Trento et al., 2010). To date we do not know if this drastic duplication provides some advantage over related viruses lacking this duplication. However, it is interesting that this advantage has not modified the occurrence of the subgroup B during the last outbreaks (Table 1). Although subgroup A has prevailed over subgroup B during the last three years of this study, a higher percentage of subgroup B was detected in 2015 in our laboratory (63.3%) despite this, all strains of HRSV-A were ON1 (study currently conducted, data not shown). The mechanism by which the 72-nt duplication provides a selective advantage to ON1 strains is unknown; but recently Hotard et al. suggested a mechanism explaining how BA strains became predominant after their emergence, and a

comparable hypothesis could be applied to ON1 strains. They showed that the 60-nt duplication in a recombinant HRSV strain with a consensus G protein from BA genotypes provides advantages in attachment and in a *in vitro* competitive infection assay, rather than the escape from antibody neutralization. They hypothesized that the duplication in BA strains which harbor 10 potential O-glycosylation sites is responsible for the enhanced binding and could lead to greater transmission of HRSV in humans. Thereby, the 72-nt duplication of ON1 strains encodes 11 additional potential glycosylation sites to the existing in the previously circulating strains without the duplication (Hotard et al., 2015).

The discrete phylogeographic analysis with an alignment of ON1 strains (177 sequences of 665-nt) allowed us to infer that the time of the MRCA for Argentine strains dated one year before we detected the first ON1 strain, it is important to note that this analysis is restricted to Buenos Aires city and it is possible that ON1 strains might be previously circulating in other non-sampled locations in Argentina. This analysis also allowed us to infer with a high probability that Panama could be the location of the MRCA, dated June, 2010 with a very narrow HPD interval (January, 2010 to October, 2010). Taking into account that the first ON1 strain reported in GenBank is from Panama and dated October 14, 2010, it is possible that this place could be the real location of the ancestor or a location very close to this country. Duvvuri et al. in 2015 compared the time of the MRCA among their analyses with multiple set of sequences with different lengths and other reports and they found that the estimated time of the MRCA ranged from 2007 to 2010, with a median date of 2008 (Agoti et al., 2014; Duvvuri et al., 2015), we detected a more accurately date as seen by the narrow HPD range. It is possible that the longer the sequences used in the alignment, the greater the accuracy obtained. In this sense, Agoti et al. have demonstrated that the analysis of whole genome sequences provided more precise estimates of times of the MRCA and rates of nucleotide substitutions (Agoti et al., 2015). Further analyses with whole genomes of ON1 strains will help to confirm this observation.

The discrete phylogeographic analysis also showed temporary and geographical clustering of ON1 strains observed as phylogenetic clades with strains exclusively associated from a single country (Fig. 3); however the

studies reported to date analyzed one or two outbreaks, which could narrow the interpretation. Because among our 44 ON1 strains from three outbreaks (2012-2014) there was geographical clustering since the first introduction from Spain, but also posterior reintroductions from USA, Cuba, South Korea, and Spain. As more strains will be reported, more interconnections will be observed in the future, as it has been found for the other HRSV-A genotypes in previous studies (Katzov-Eckert et al., 2012; Pretorius et al., 2013). In addition, we found that some GA5 Argentine strains from 2002 to 2004 were closely related to one American strain from 2006, and Brazilian strains from 2005 and 2007 were closely related with Argentine strains from 2006 to 2008. For GA2 genotype (non-lineage ON1), we found a phylogenetic clade with sequences from Paraguay, Canada, Saudi Arabia and Argentina, very distant countries, but during the same period of time (2009-2011) (Fig. 1).

In our region, the persistence of one sublineage of ON1 strains allowed us to perform a continuous phylogeographic analysis and to establish that there could be a local clustering of some strains even in neighborhoods. This study shows the potential of this type of bioinformatic tools to trace the spread of a new genetic lineage in human populations and to determine precisely in what locations will emerge (Holmes, 2008). It is in this sense that we should take advantage of this type of genetic events to increase the understanding of the evolutionary patterns of HRSV not only for historical significance, but also in the context of a future vaccine surveillance network.

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. doi:10.3201/eid2006.131438

Agoti, C.N., Otieno, J.R., Munywoki, P.K., Mwihuri, A.G., Cane, P.A., Nokes, D.J., Kellam, P., Cotten, M., 2015. Local Evolutionary Patterns of Human Respiratory Syncytial Virus Derived from Whole-Genome Sequencing. J. Virol. 89, 3444–3454. doi:10.1128/JVI.03391-14

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. doi:10.1093/infdis/jiu473

Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinforma. Oxf. Engl. 27, 2910–2912. doi:10.1093/bioinformatics/btr481

Blanc, A., Delfraro, A., Frabasile, S., Arbiza, J., 2005. Genotypes of respiratory syncytial virus group B identified in Uruguay. Arch. Virol. 150, 603–609. doi:10.1007/s00705-004-0412-x

Carballal, G., Videla, C.M., Espinosa, M.A., Savy, V., Uez, O., Sequeira, M.D., Knez, V., Requeijo, P.V., Posse, C.R., Miceli, I., 2001. Multicentered study of viral acute lower respiratory infections in children from four cities of Argentina, 1993-1994. J. Med. Virol. 64, 167–174.

Coates, H.V., Alling, D.W., Chanock, R.M., 1966. An antigenic analysis of respiratory syncytial virus isolates by a plaque reduction neutralization test. Am. J. Epidemiol. 83, 299–313.

Collins PL, Karron R, 2013. Respiratory Syncytial Virus and Metapneumovirus, in: Knipe, D, Howley, P (Eds.), FIELDS Virology. LWW, Philadelphia, USA, pp. 1086–1123.

Cui, G., Zhu, R., Qian, Y., Deng, J., Zhao, L., Sun, Y., Wang, F., 2013. Genetic variation in attachment glycoprotein genes of human respiratory syncytial virus subgroups a and B in children in recent five consecutive years. PloS One 8, e75020. doi:10.1371/journal.pone.0075020

Dapat, I.C., Shobugawa, Y., Sano, Y., Saito, R., Sasaki, A., Suzuki, Y., Kumaki, A., Zaraket, H., Dapat, C., Oguma, T., Yamaguchi, M., Suzuki, H., 2010. New genotypes within respiratory syncytial virus group B genotype BA in Niigata, Japan. J. Clin. Microbiol. 48, 3423–3427. doi:10.1128/JCM.00646-10

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

David M. Knipe, Peter M. Howley, 2013. Fields Virology, Sixth Edition, 2-volume set edition. ed. LWW, Philadelphia, PA.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. doi:10.1093/molbev/mss075

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. doi:10.1038/srep14268

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. doi:10.1371/journal.pone.0032807

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. doi:10.1093/sysbio/syq010

Holmes, E.C., 2008. Evolutionary history and phylogeography of human viruses. Annu. Rev. Microbiol. 62, 307–328. doi:10.1146/annurev.micro.62.081307.162912

Hotard, A.L., Laikhter, E., Brooks, K., Hartert, T.V., Moore, M.L., 2015. Functional Analysis of the 60-Nucleotide Duplication in the Respiratory Syncytial Virus Buenos Aires Strain Attachment Glycoprotein. J. Virol. 89, 8258–8266. doi:10.1128/JVI.01045-15

Karron, R.A., Buchholz, U.J., Collins, P.L., 2013. Live-attenuated respiratory syncytial virus vaccines. Curr. Top. Microbiol. Immunol. 372, 259–284. doi:10.1007/978-3-642-38919-1_13

Karron, R.A., Luongo, C., Thumar, B., Loehr, K.M., Englund, J.A., Collins, P.L., Buchholz, U.J., 2015. A gene deletion that up-regulates viral gene expression yields an attenuated RSV vaccine with improved antibody responses in children. Sci. Transl. Med. 7, 312ra175. doi:10.1126/scitranslmed.aac8463

Katzov-Eckert, H., Botosso, V.F., Neto, E.A., Zanotto, P.M. de A., VGND consortium, 2012. Phylodynamics and dispersal of HRSV entails its permanence in the general population in between yearly outbreaks in children. PloS One 7, e41953. doi:10.1371/journal.pone.0041953

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. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 26, 103–112. doi:10.1016/j.meegid.2014.05.007

Maffey, A.F., Barrero, P.R., Venialgo, C., Fernández, F., Fuse, V.A., Saia, M., Villalba, A., Fermepin, M.R., Teper, A.M., Mistchenko, A.S., 2010. Viruses and atypical bacteria associated with asthma exacerbations in hospitalized children. Pediatr. Pulmonol. 45, 619–625. doi:10.1002/ppul.21236

Peret, T.C., Hall, C.B., Hammond, G.W., Piedra, P.A., Storch, G.A., Sullender, W.M., Tsou, C., Anderson, L.J., 2000. Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. J. Infect. Dis. 181, 1891–1896. doi:10.1086/315508

Peret, T.C., Hall, C.B., Schnabel, K.C., Golub, J.A., Anderson, L.J., 1998. Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. J. Gen. Virol. 79 (Pt 9), 2221–2229. doi:10.1099/0022-1317-79-9-2221

Pretorius, M.A., van Niekerk, S., Tempia, S., Moyes, J., Cohen, C., Madhi, S.A., Venter, M., SARI Surveillance Group, 2013. Replacement and positive evolution of subtype A and B respiratory syncytial virus G-protein genotypes from 1997-2012 in South Africa. J. Infect. Dis. 208 Suppl 3, S227-237. doi:10.1093/infdis/jit477

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

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. doi:10.1128/JCM.00115-09

Stockton, J., Ellis, J.S., Saville, M., Clewley, J.P., Zambon, M.C., 1998. Multiplex PCR for Typing and Subtyping Influenza and Respiratory Syncytial Viruses. J. Clin. Microbiol. 36, 2990–2995.

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. doi: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.

Trento, A., Ábrego, L., Rodriguez-Fernandez, R., González-Sánchez, M.I., González-Martínez, F., Delfraro, A., Pascale, J.M., Arbiza, J., Melero, J.A., 2015. Conservation of G-Protein Epitopes in Respiratory Syncytial Virus (Group A) Despite Broad Genetic Diversity: Is Antibody Selection Involved in Virus Evolution? J. Virol. 89, 7776–7785. doi:10.1128/JVI.00467-15

Trento, A., Casas, I., Calderón, A., Garcia-Garcia, M.L., Calvo, C., Perez-Breña, P., Melero, J.A., 2010. Ten years of global evolution of the human respiratory syncytial virus BA genotype with a 60-nucleotide duplication in the G protein gene. J. Virol. 84, 7500–7512. doi:10.1128/JVI.00345-10

Trento, A., Galiano, M., Videla, C., Carballal, G., García-Barreno, B., Melero, J.A., Palomo, C., 2003. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. J. Gen. Virol. 84, 3115–3120. doi:10.1099/vir.0.19357-0

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. doi:10.1128/JVI.80.2.975-984.2006

Venter, M., Madhi, S.A., Tiemessen, C.T., Schoub, B.D., 2001. Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new subgroup A and B genotypes. J. Gen. Virol. 82, 2117–2124.

Videla, C., Carballal, G., Misirlian, A., Aguilar, M., 1998. Acute lower respiratory infections due to respiratory syncytial virus and adenovirus among hospitalized children from Argentina. Clin. Diagn. Virol. 10, 17–23.

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.
doi:10.1016/j.jinf.2003.10.006

Viegas, M., Mistchenko, A.S., 2005. Molecular epidemiology of human
respiratory syncytial virus subgroup A over a six-year period (1999-2004)
in Argentina. J. Med. Virol. 77, 302–310. doi:10.1002/jmv.20457

Volz, E.M., Koelle, K., Bedford, T., 2013. Viral phylodynamics. PLoS Comput.
Biol. 9, e1002947. doi:10.1371/journal.pcbi.1002947

Table 1.

Number of HRSV positive samples according to its genotypic characteristics per year

Year	No. of HRSV samples*	No. of HRSV-B samples (%)	No. of HRSV-A samples (%)	No. of HRSV-A G sequences	HRSV-A Genotype	
					GA2	GA5
1999	56	35 (62.50)	21 (37.50)	11	7	4
2000	44	0 (0)	44 (100)	10	3	7
2001	41	7 (17.07)	34 (82.93)	11	5	6
2002	63	45 (71.43)	18 (28.57)	10	0	10
2003	92	24 (26.09)	68 (73.91)	9	7	2
2004	57	10 (17.54)	47 (82.46)	15	8	7
2005	31	31 (100)	0 (0)	0	0	0
2006	13	3 (23.08)	10 (76.92)	6	5	1
2007	10	6 (60)	4 (40)	2	1	1
2008	22	6 (27.27)	16 (72.73)	11	9	2
2009	12	0 (0)	12 (100)	8	8	0
2010	12	6 (50)	6 (50)	3	3	0
2011	28	11 (39.28)	17 (60.72)	17	12	5
2012	30	0 (0)	30 (100)	19	19	0
2013	35	10 (28.57)	25 (71.43)	18	18	0
2014	32	13 (40.62)	19 (59.38)	19	19	0
Total	578	207	371	169	124	45

* Number of HRSV IFA-positive samples grouped by multiplex RT-PCR (Stockton et al., 1998). The relative percentage of HRSV-A and HRSV-B respect to the total of the positive samples is shown in brackets.
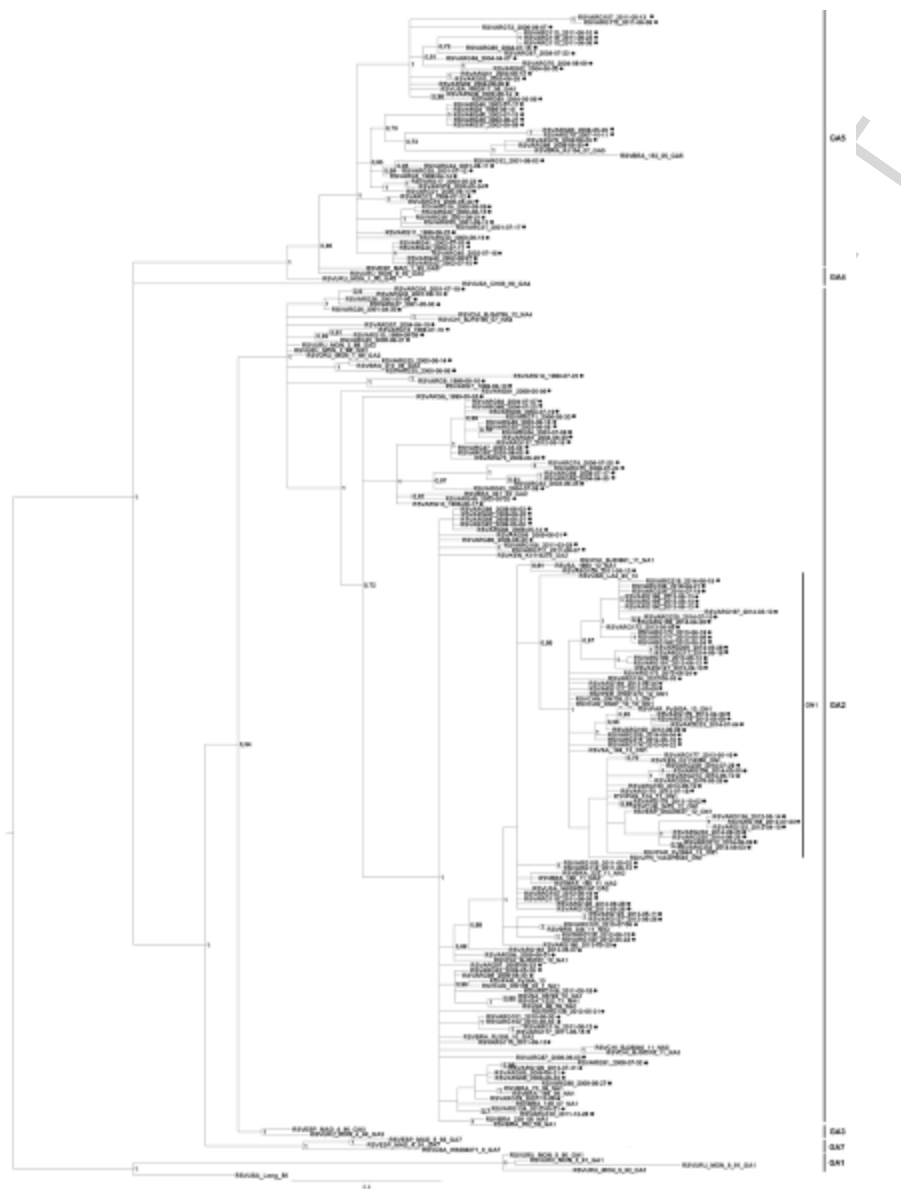
Abbreviation: No., Number.

Fig. 1. Bayesian phylogenetic tree. The evolutionary model GTR+G was used for Bayesian reconstruction. Bayesian Markov Chain Monte Carlo (MCMC) chains were run for 10 million ngen, 10,000 samplefreq to reach convergence (split value<0.01; ESS>200) and 10% burnin. The pool sequences include 114 new sequences from Buenos Aires, 45 sequences from Buenos Aires previously reported (Viegas and Mistchenko, 2005) and 59 sequences of known genotype from other countries downloaded from GenBank. Node labels

represent posterior probabilities (only values above 0.7 are shown). The scale bar represents nucleotide substitutions per site. Genotypes GA1-7 and the ON1 lineage are denoted with vertical lines. Sequences from Buenos Aires are remarked with a black circle.
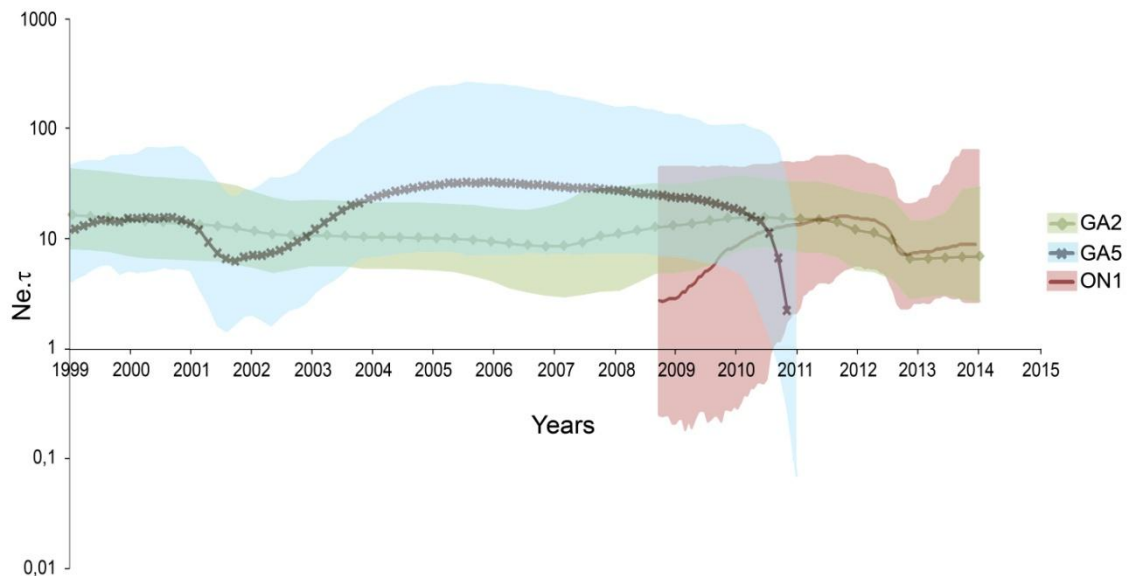

Fig. 2. Bayesian skyline plot of HRSV-A genotypes circulating in Buenos Aires. Population dynamics of genetic diversity in the genotypes GA2 and GA5, and the recent GA2 genetic lineage ON1. The horizontal axis represent time in years, the vertical axis represent the population sizes (Ne.τ). The line with diamonds represents GA2, the line with crosses represents GA5 and the smooth line represents ON1. The shaded areas show the 95% highest probability density interval of these estimates.
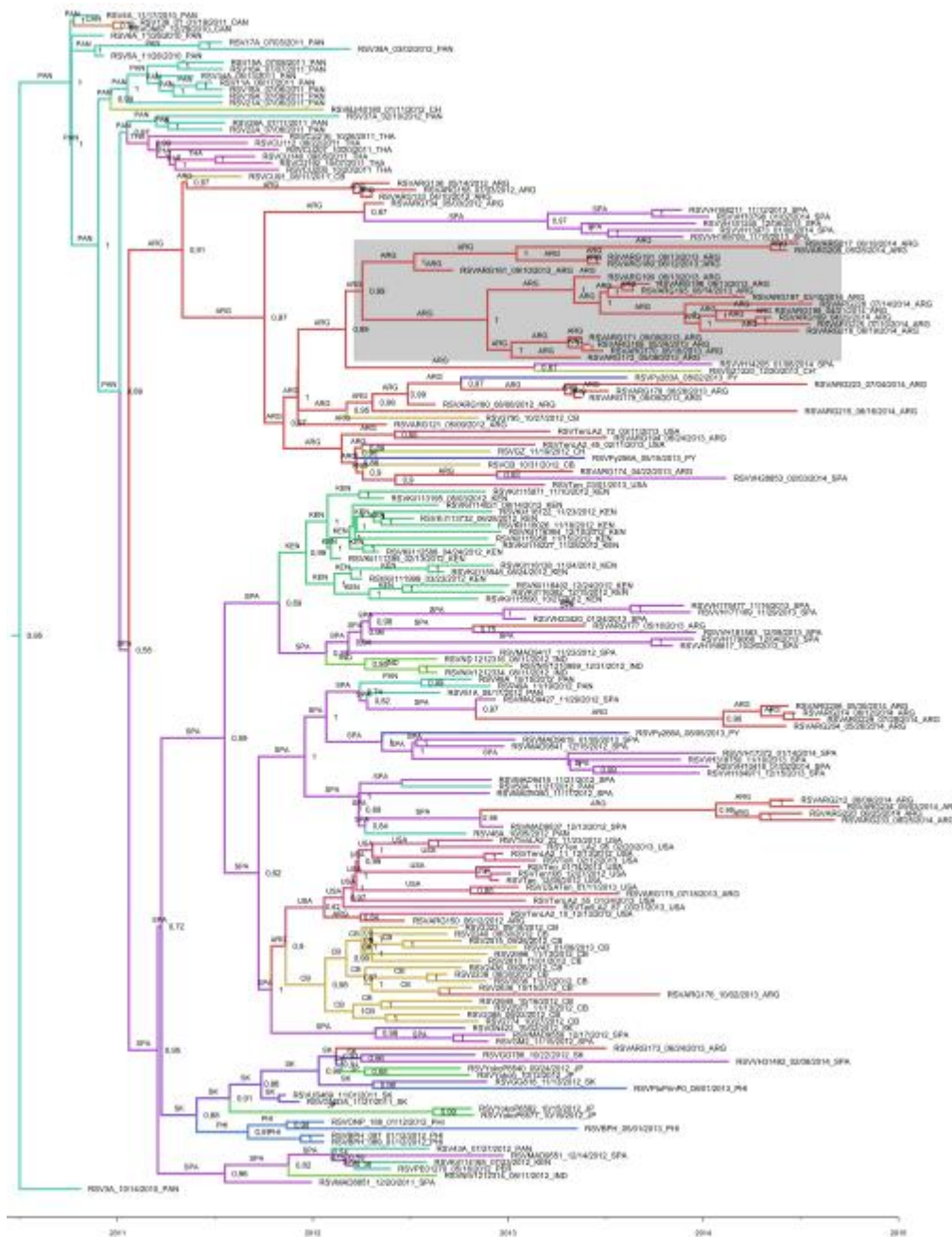
Fig. 3. Maximum Clade Credibility Tree (MCCT) from the discrete phylogeography analysis of ON1 sequences. The branches are in color by location of viral samples. The acronym above the branch indicates the country of origin for the common ancestor for each clade, and the node labels indicate the probability support for this inference. The scale axis is represented inyear. A local sublineage is highlighted in grey.

Acronyms: ARG Argentina, CAN Canada, CB Cuba, CH China, IND India, JP Japan, KEN Kenya, PAN Panama, PER Peru, PHI Philippines, PY Paraguay, SK South Korea, SPA Spain, THA Thailand ,USA United States.
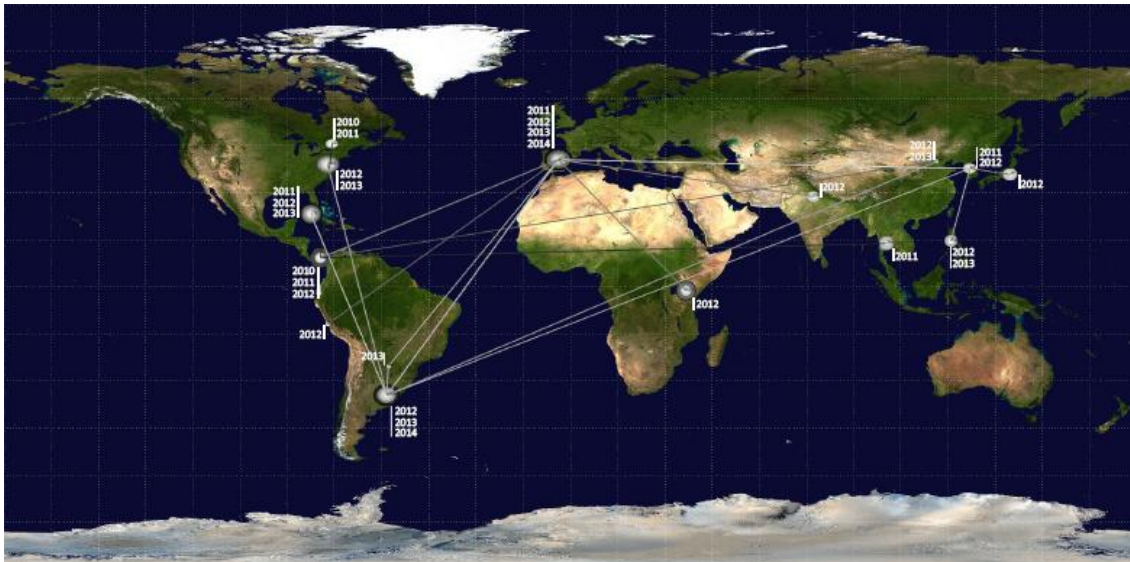
Fig. 4. Diffusion dynamic plot of the discrete phylogeographic analysis of HRSV-A ON1 genetic lineage. The lines represent the significant connections by Bayes Factor (BF) over than 10. White line corresponds to maximum BF and black line corresponds to minimum BF. Dates in each location correspond to the year of isolation of analyzed strains. The KML file for visualization in Google Earth is available as a supplementary file (KML file 1).
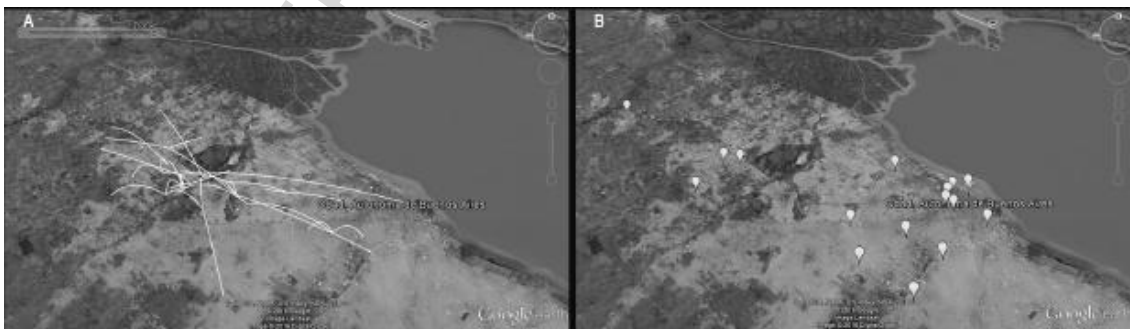

Fig. 5A. Diffusion dynamic plot of the continuous phylogeographic analysis of sequences from local ON1 sublineage (highlighted in MCCT in Fig.3). The KML file for visualization in Google Earth is available as a supplementary file 2 (KML file 2)

Fig. 5B. Location of the ON1 Argentine strains without distinction of association in the MCCT and showed in Fig.5A. The map is based on satellite images made available in Google Earth (http://earth.google.com).

Highlights

- HRSV-A was the most prevalent subgroup during 1999-2014 in BA, Argentina.
- GA2 was the most prevalent genotype and co-circulated with GA5.
- First ON1 strain emerged in BA in 2012, its frequency increased to 100% in 2014.
- Panama could be the location of the MRCA of globally ON1 strains, dated June 20, 2010
- ON1 strains showed temporary and geographical clustering even in neighborhoods.