Peer-reviewed version available at Viruses 2018, 10, 476; doi:10.3390/v10090476

Article

Evolution of human respiratory syncytial virus (RSV) over multiple seasons in New South Wales, Australia

Francesca Di Giallonardo[1,2], Jen Kok[3], Marian Fernandez[3], Ian Carter[3], Jemma L. Geoghegan[4], Dominic E. Dwyer[3], Edward C. Holmes[1] and John-Sebastian Eden[1,5]*

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia. (E.C.H.) edward.holmes@sydney.edu.au

[2]The Kirby Institute, University of New South Wales, Randwick, NSW 2052, Australia. (F.D.G.) fdigiallonardo@kirby.unsw.edu.au

[3]Institute for Clinical Pathology and Medical Research, NSW Health Pathology, Westmead Hospital and University of Sydney, Sydney, NSW 2145, Australia. (J.K.) jen.kok@health.nsw.gov.au; (M.F.) marian.fernandez@health.nsw.gov.au; (I.C.) ian.carter@health.nsw.gov.au; (D.E.D) dominic.dwyer@sydney.edu.au

[4]Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia. (J.L.G.) jemma.geoghegan@mq.edu.au

[5]Centre for Virus Research, Westmead Institute for Medical Research, Westmead, NSW 2145, Australia. (J-S.E) js.eden@sydney.edu.au

*Correspondence: js.eden@sydney.edu.au; Tel.: +61 2 8627 1817

Abstract: There is an ongoing global pandemic of human respiratory syncytial virus (RSV) infection that results in substantial annual morbidity and mortality. In Australia, RSV is the major cause of acute lower respiratory tract infections (ALRI). Nevertheless, little is known about the extent and origins of genetic diversity of RSV in Australia, nor the factors that shape this diversity. We conducted a genome-scale analysis of RSV infections in New South Wales (NSW). RSV genomes were successfully sequenced for 144 specimens collected between 2010-2016. Of these, 64 belonged to the RSVA and 80 to the RSVB subtype. Phylogenetic analysis revealed a wide diversity of RSV lineages within NSW and that both subtypes evolved rapidly in a strongly clock-like manner, with mean rates of approximately 6-8 x 10^{-4} nucleotide substitutions per site per year. There was only weak evidence for geographic clustering of sequences, indicative of fluid patterns of transmission within the infected population, and no evidence of any clustering by patient age such that viruses in the same lineages circulate through the entire host population. Importantly, we show that both subtypes circulated concurrently in NSW with multiple introductions into the Australian population in each year, and only limited evidence for multi-year persistence.

Keywords: respiratory syncytial virus; phylogenetics; evolution; multi-year persistence

Peer-reviewed version available at Viruses 2018, 10, 476; doi:10.3390/v10090476

1. Introduction

Respiratory syncytial virus (RSV) is a major cause of acute respiratory tract infections (ARTI) in humans [1]. The burden of RSV disease is greatest in specific vulnerable populations, including young children and the elderly, particularly those with pre-existing medical comorbidities or who are immunocompromised. Outbreaks in hospitals and closed environments such as aged care facilities have also been documented [2]. Importantly, the annual global hospitalization rate of RSV infection in young children is nearly 10% and is associated with approximately 59,600 deaths [3, 4]. The health and economic burden of RSV in young children surpasses that of influenza virus with total annual direct healthcare costs estimated to be between $24 – 50 million [5]. In Australia, Indigenous children living in remote communities also experience a high prevalence of RSV, particularly in comparison to non-Indigenous groups [6, 7]. Morbidity and mortality is also high in elderly adults, with approximately 14,000 deaths annually due to RSV in the USA (Centre for Disease Control, [8]).

While the factors that contribute to the prevalence of RSV are yet to be fully defined; Immunity, re-infection rates, and climate may play a role. For example, laboratory reports highlight the seasonality in temperate regions in Australia, with a peak in RSV activity typically occurring in the early winter (May/June) period and preceding the seasonal peak in influenza virus (New South Wales Health, Influenza Monthly Surveillance Reports). Climate and rainfall differences in the tropical north of Australia are also likely to be important drivers of RSV disease patterns, resulting in a seasonality distinct from that observed in temperate regions, with a correlation between peak RSV and peak rainfall levels around January [6].

RSV is a negative-sense single-stranded RNA virus (family *Pneumoviridae*) with a 15 kb genome that encodes 10 proteins [9]. Two distinct antigenic subgroups have been identified, subtypes A and B (RSVA and RSVB, respectively) that show clear phylogenetic divergence [10, 11]. The glycoprotein (G), responsible for attachment to the host cell, exhibits the greatest genetic diversity within and between subtypes [11]. This is thought to reflect strong immune pressure and the subsequent generation of escape variants, in a process analogous to antigenic drift in the hemagglutinin (HA) protein in influenza A virus [12, 13]. Hence, reinfection with RSV is commonplace [14, 15]. There are currently no effective vaccines against RSV, although a number of novel vaccines are entering clinical trials [16]. Similarly, there are novel antivirals targeting the viral polymerase and fusion protein that are in clinical trials [17].

Despite the clinical significance and the burden of RSV infection worldwide, we lack understanding of the patterns of virus emergence, evolution and spread. Phylogenetic studies of global RSV evolution are compromised due to the limited availability of gene sequence data and strongly asynchronous sampling in time and space. Most evolutionary analyses have focused on the G gene because of its high genetic diversity and utility as a phylogenetic marker. The G gene is also characterized by premature stop codons in the case of RSVB [18, 19], and duplications in the G gene are described in both subtypes [20, 21]. Less is known about the genome-scale evolution of RSV [21-25], although this is necessary for defining fine-scale phylodynamic and epidemiological processes that may assist in targeted interventions including vaccine design and implementation [20, 22, 26].

There have been several studies of RSV evolution in specific geographic regions, including South Africa [25], the Netherlands [23], Argentina [21], Italy [24], and Kenya [22]. These studies highlight the global distribution of predominant RSV variants during each season, alongside the establishment

and co-circulation of local endemic sub-lineages. There is, however, limited data exploring the genetic diversity of RSV in Australia. Similarly, the evolution and spread of RSV within specific communities, and hence how long individual lineages of RSV are able to persist in single populations, is not well understood. To address these issues, we provide the first large, genome-scale analysis of RSV in Australia, focusing on infections identified through a major clinical diagnostic laboratory that services a population over 1.57 million people. In particular, we sought to determine the extent and pattern of genetic diversity circulating within the culturally diverse region of western Sydney as well as the rural region of western New South Wales (NSW), how this relates to the global diversity of the virus, what epidemiological factors act to shape genetic diversity at the local level, and to what extent RSV transmission persists between seasonal outbreaks.

2. Materials and Methods

2.1 Ethics

This study was approved by local ethics and governance committees (LNR/17/WMEAD/128; SSA/17/WMEAD/129). Samples were de-identified with basic demographic information collected including age, sex and location (city, region, hospital).

2.2. Sample collection

This study utilized residual RSV-positive specimens collected for routine diagnostic testing at the Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, NSW, Australia between May 2010 and December 2016. Viral nucleic acid that was previously extracted (NucliSENS® easyMAG®, bioMérieux) during routine diagnostic testing was stored at -80°C prior to the commencement of this study.

2.3. Whole genome sequencing

We employed an overlapping RT-PCR strategy to amplify viral genomes (four ~4kb amplicons), targeting both RSVA and RSVB subtypes with previously published primers [27]. Briefly, viral RNA was first reverse transcribed using a pool of the four forward primers (RSVS1_01F, RSVS2_3905F, RSVS3_7215F and RSVS4_10959F) and SuperScript™ IV cDNA synthesis system (Invitrogen, ThermoFisher Scientific). The resultant cDNA was then split across four parallel PCR reactions to amplify the genome using Platinum SuperFi (Invitrogen). For successfully amplified samples, the four amplicons were pooled equally and then prepared as libraries using Nextera XT before MiSeq (Illumina) sequencing. Each sequencing run contained between 41 and 60 indexed samples, which generated at least 2000X per base coverage per genome. Raw sequence reads were quality trimmed with Trimmomatic [28], and then *de novo* assembled using Trinity [29] and SPAdes [30]. The trimmed reads were re-mapped to draft RSV genome contigs with BowTie2 [31]. The mapping alignment quality was checked manually particularly around known G gene duplications before extracting the final majority consensus genome sequence for each sample.

2.4. Data Availability

All sequence reads generated in this project are available on NCBI GenBank (Submission ID 2142716). Accession numbers will be available and listed in supplementary table 1.

2.5. Phylodynamic analysis

To place our sample set into a global context, complete and near complete genome sequences of RSVA and RSVB were obtained from GenBank. Sequences with no geographic association or sampling date were excluded. RSVA and RSVB sequences were aligned separately using the multiple-sequence alignment tool, MAFFT, using the L-INS-I algorithm followed by a visual inspection [32]. Sequences resulting from passaging experiments, or potential recombinants identified using RDP4 [33] were removed from the alignment. After this data pruning, the final data set consisted of 849 RSVA of 15,062 nt length and 500 RSVB genome sequences of 15,033 nt length (Table S1). Intergenic regions were removed for phylogenetic tree estimates as they contained single nucleotide insertions and deletions.

Maximum likelihood (ML) trees where estimated in RAxML [34, 35] employing a GTR gamma (Γ) nucleotide substitution model and 1,000 bootstrap replications. To determine the extent of temporal structure in a data a root-to-tip regression of genetic distance against year of sampling was performed using TempEst v.1.5 utilizing the separate ML trees for RSVA and RSVB [36]. As both RSVA and RSVB exhibited strong temporal structure (i.e. clock-like evolution; see Results), we estimated their evolutionary rates more accurately using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.8.2 [37], using a HKY + gamma (Γ) substitution model. A strict clock was used for the evolutionary rates estimates and a constant population size was implemented as a tree prior (although no significant differences in evolutionary rate were observed when we compared rate estimates from the uncorrelated log-normal relaxed clock to those from the strict clock). All analyses were run for at least 100 million steps and sampling every 10,000 steps to ensure convergence of all parameters. The first 10% of the posterior was removed as burn-in. Mean rates and 95% highest posterior density (HPD) were compared and values with HDP non-overlapping with mean rate values are significantly different. However, because we were unable to achieve consistent statistical convergence for the global data set, evolutionary rates were instead estimated by implementing a least-square dating algorithm (LSD) which is suitable for large data sets [38, 39]. To obtain significance, 1,000 parametric bootstraps were conducted on the branch lengths.

2.6. Phylogenetic analysis of clustering patters

We used a phylogenetic approach to determine whether there is more clustering by geography and age within the NSW RSV data set than might be expected from chance alone. Specifically, Bayesian posterior trees were used for evaluating geographic and age structure in the RSVA and RSVB trees using the Bayesian Tip-association Significance (BaTS) program, which compares parsimony score (PS), association index (AI), and maximum clade size (MC) statistics [40]. Estimates were repeated 1,000 times to infer significance. The traits investigated were age, the hospital facility where patients first presented, and state electorate. Because of large sampling biases in the data, with few sequences obtained from most years (Fig. 1A), this analysis was only performed on the samples from 2016 as it was by far the most densely sampled year (RSVA = 36 sequences, RSVB = 57).

3. Results & Discussion

3.1. Demographic characteristics of RSV in NSW

We attempted whole genome sequencing on 241 archived RSV-positive viral nucleic acid extracts held and tested by ICPMR. Virus genomes were successfully sequenced for 144 specimens,

of which 64 belonged to RSVA and 80 to RSVB. The number of samples was skewed with the majority collected in 2016, comprising 36 RSVA and 57 RSVB sequences, respectively (Figure 1A). Despite this limited sampling, it is evident that RSVA and RSVB co-circulated in every season, which is consistent with other molecular epidemiological studies [41-44]. A distinct seasonality was apparent with peaks typically occurring in the early winter period (May to July) (Figure 1A). This pattern is consistent with aggregated data from state-wide testing across NSW for RSV, influenza, and other respiratory viruses (Figure 1B) [45]. For RSVA, 54.7% of the samples were derived from female patients, while 43.8% RSVB samples were from female patients (Table 1). Forty-five per cent of RSVA sequences were obtained from patients under the age of two, and 27% from patients greater than 65 years of age. These numbers were slightly lower for RSVB, with 43% being infants and only 20% older patients. While all testing and sequencing was performed at Westmead Hospital, the cases were not limited to the locations surrounding the hospital, but also included samples from western and north-western rural NSW and hence some distance from metropolitan Sydney (Figure 2); as a consequence, the samples collected here will be referred to as from NSW. The electorates with the most sequences sampled were Orange (n = 10) and Mount Druitt (n = 11). Despite the wide geographic range of sampling, coverage was low, hence, sequences were sampled only from a small number of geographic locations.

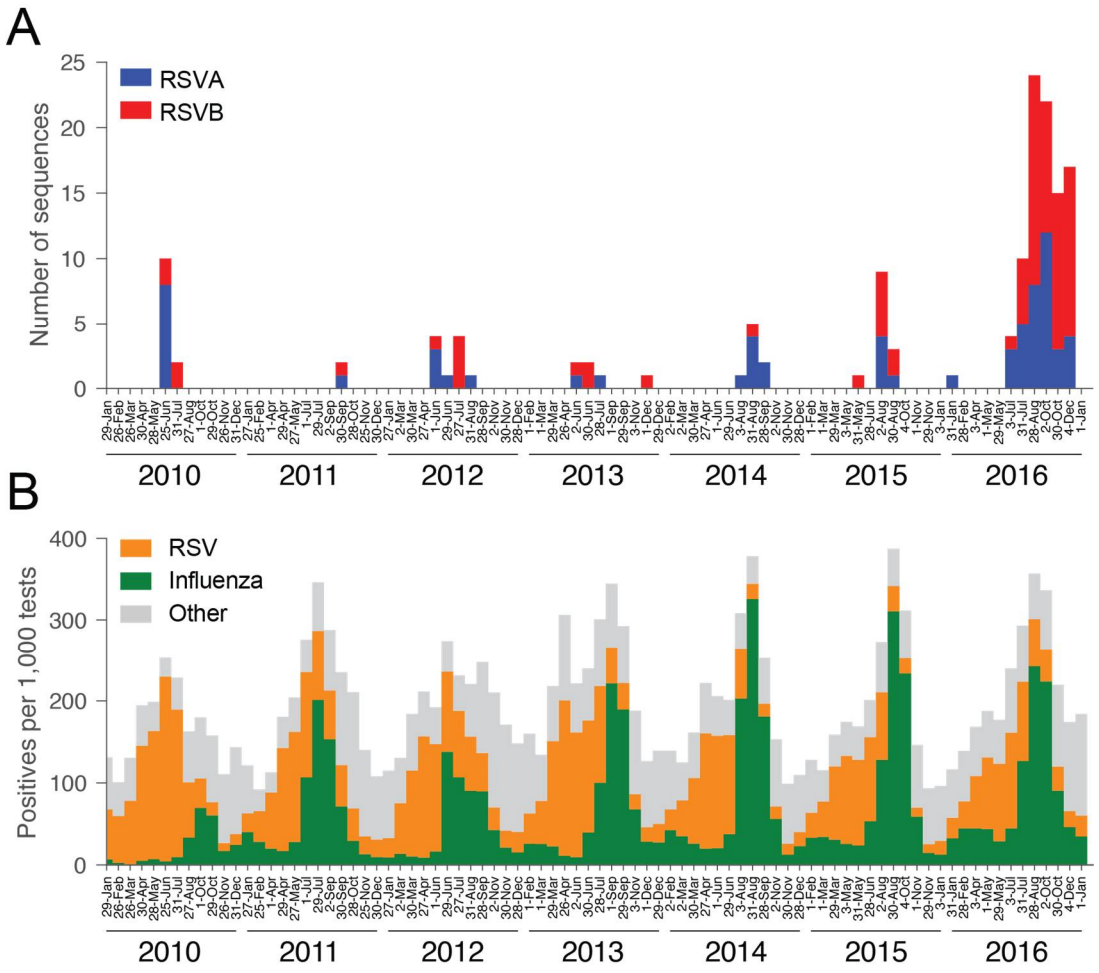
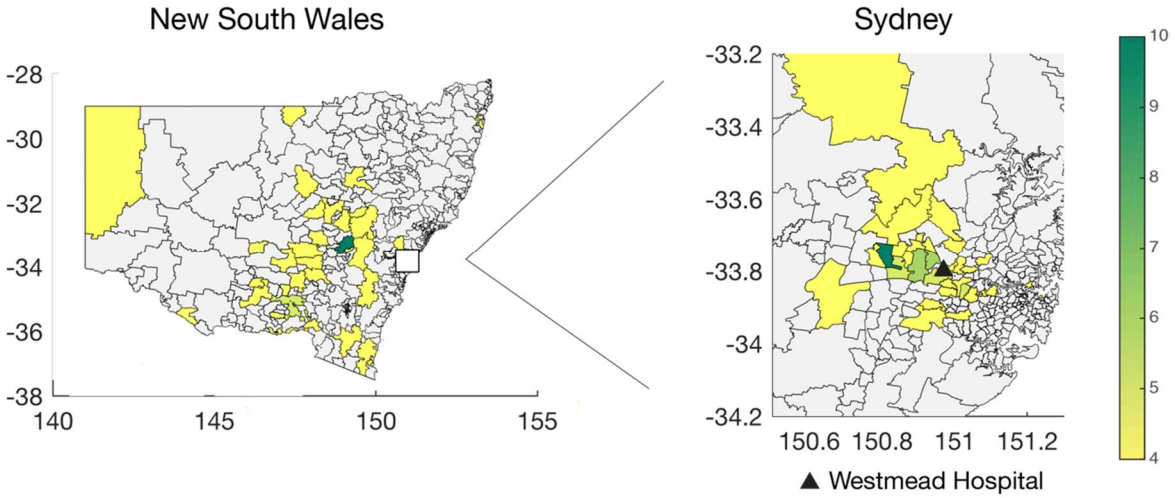


Figure 1. Incidence of RSV in New South Wales. (A) Number of RSV genome sequences per four-week period: RSVA = blue, RSVB = red. (B) Number of positive samples per 1,000 specimens reported

Peer-reviewed version available at Viruses 2018, 10, 476; doi:10.3390/v10090476

189 across NSW per four-week period: RSV = orange, Influenza = green, Other, including parainfluenza
190 virus, adenovirus, and human metapneumovirus, = grey.
191



192
193 **Figure 2. Geographic distribution of RSV in New South Wales.** Number of sequences per postcode
194 in NSW (left) and greater Westmead area (right). The location for Westmead Hospital is indicated.
195
196 **Table 1.** Demographic of the patient data used in this study. Ratios for gender and age categories are
197 shown for RSVA and RSVB. Total numbers are shown in brackets.

	RSVA (*n* = 64)		**RSVB (*n* = 80)**	
	Male	**Female**	**Male**	**Female**
Overall	0.453 (29)	0.547 (37)	0.550 (44)	0.438 (35)
By Age groups				
6 months or younger	0.094 (6)	0.078 (5)	0.163 (13)	0.075 (6)
7 months to 1 year	0.078 (5)	0.125 (8)	0.063 (5)	0.038 (3)
1 - 2 years	0.031 (2)	0.047 (3)	0.025 (2)	0.063 (5)
2 - 5 years	0.016 (1)	0.031 (2)	0.025 (2)	0.038 (3)
6 - 15 years	0.016 (1)	0.016 (1)	0.000 (0)	0.000 (0)
16 - 25 years	0.016 (1)	0.016 (1)	0.025 (2)	0.013 (1)
26 - 49 years	0.047 (3)	0.031 (2)	0.063 (5)	0.063 (5)
50 - 65 years	0.016 (1)	0.078 (5)	0.088 (7)	0.050 (4)
66 years or older	0.141 (9)	0.125 (8)	0.100 (8)	0.100 (8)

198
199 *3.2. Evolutionary history of RSV and spread within NSW*
200 To place our Australian RSV strains in the context of global RSV diversity, we performed an
201 evolutionary analysis using our genome sequences with global reference genomes sourced from
202 GenBank (Table S1). These sequences were sampled from 21 different countries across multiple
203 continents and spanned a time-span of 40 years ranging from 1977 to 2017. The final combined data
204 set consisted of 849 and 500 RSVA and RSVB genomes, respectively. While 21 countries were
205 represented in the data, there was a clear over-representation (n = 628) of viral genomes from the
206 USA, which comprised 46% of the data in this study. Other relatively well represented countries were
207 Peru (n = 122), the Netherlands (n = 61), Kenya (n = 61), Jordan (n = 85), Viet Nam (n = 53), New

Zealand (n = 92), and the sequences sampled here in NSW (n = 144). The extensive sampling biases precluded detailed phylogeographic analyses. To simplify the geographic distribution analysis, sequences were grouped according to their continent of sampling (Figures 3 and 4).

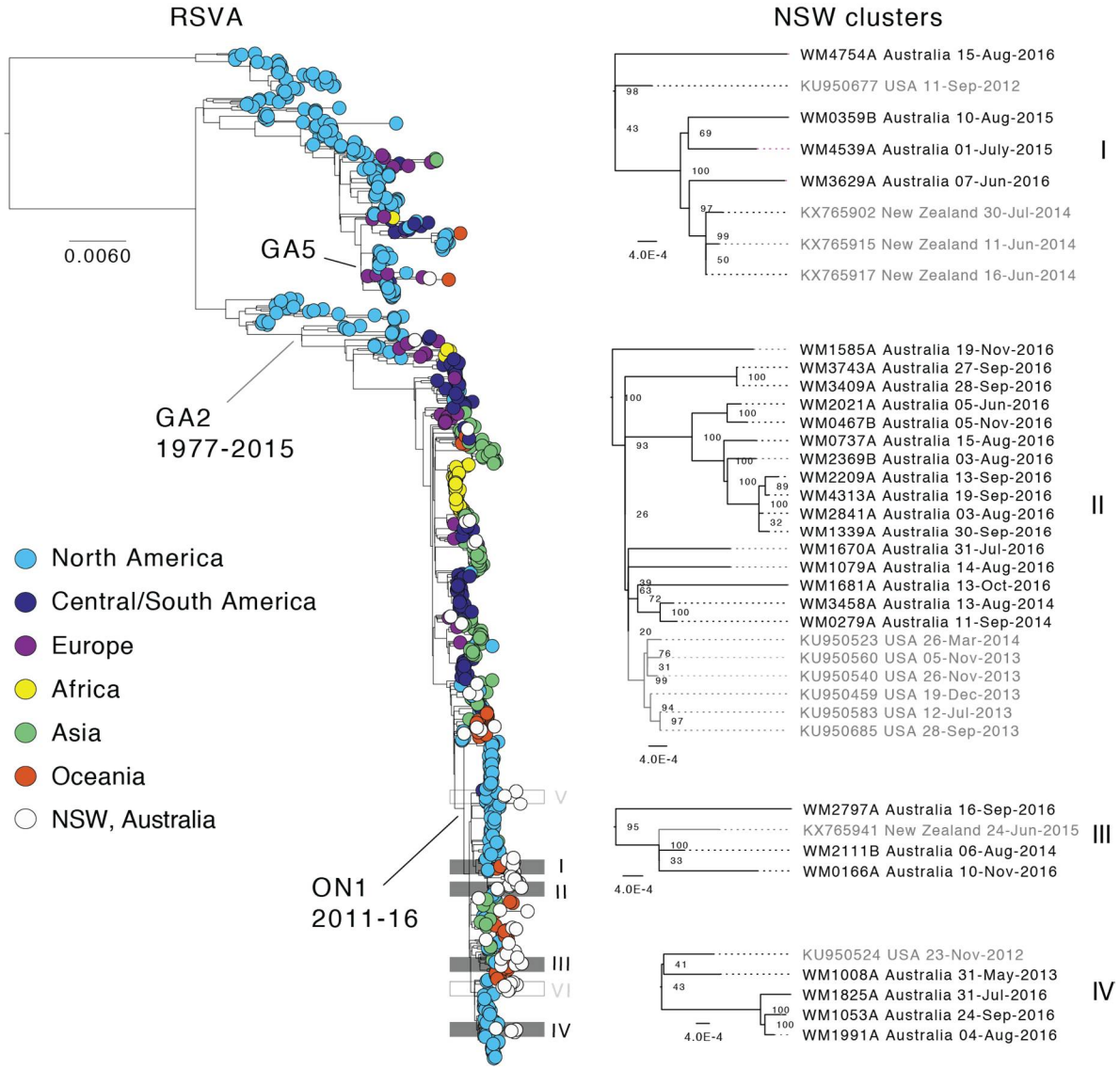


Figure 3. Global phylogeny of RSVA and local clustering within NSW. The maximum likelihood tree shown was estimated using complete RSV genome sequences. Known genotypes and the sequence time range is indicated. The tree is rooted using an RSVB outgroup (not shown), respectively. Tree tips are colored according to the geographic region and sequences from this study are shown in white. Local clusters comprising NSW sequences are marked within the global tree, and the four clusters with potential multi-season transmission events are colored grey and enlarged on the right side. Node supports are indicated, and branch lengths are scale according to the number of substitutions per site.

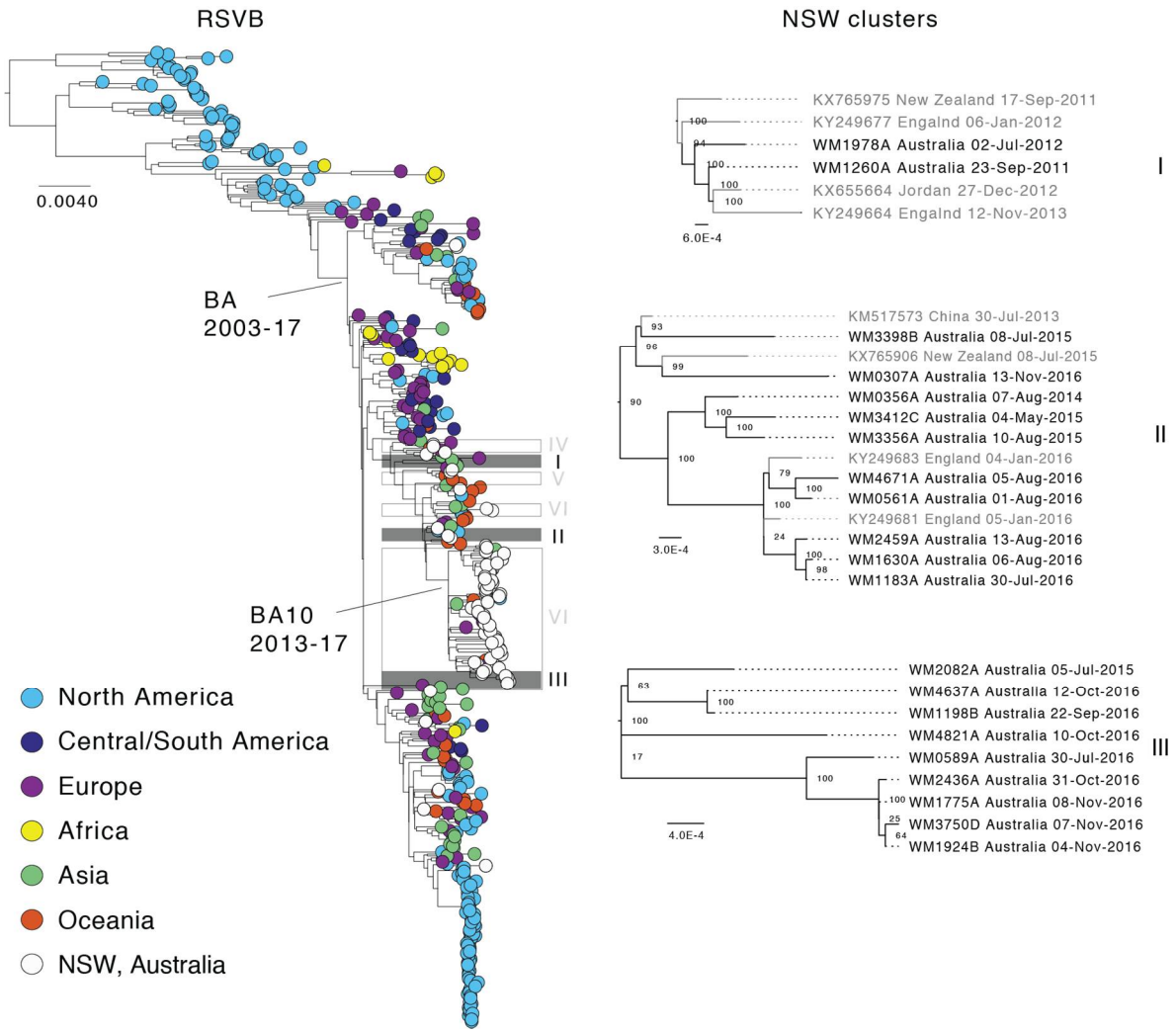


Figure 4. Global phylogeny of RSVB and local clustering within NSW. The maximum likelihood tree shown was estimated using complete RSV genome sequences. The tree is rooted using RSVA as an outgroup and the BA and BA10 genotypes are marked. Tips colors represent sampling location and sequences from this study are shown in white. Local clusters comprising NSW sequences are marked within the global tree, and the three clusters with potential multi-season transmission events are colored grey and enlarged on the right side. Node supports are indicated, and branch lengths are scale according to the number of substitutions per site.

In both the RSVA and RSVB phylogenies, the earliest described RSV genomes were derived solely from North America, and it is difficult to comment on the global distribution, diversity and genetic sources until the early-mid 2000s when sampling became more evenly distributed. Since this time, the global RSVA phylogeny has been dominated by viruses of the GA2 lineage, and more recently, the ON1 sub-lineage (Figure 3), which is defined by a 72-nt duplication in the G gene [21]. In the global RSVB phylogeny, three sub-lineages of BA viruses have co-circulated, with the exception of the most recent samples in which BA10 viruses appear to be dominant, although this could again reflect sampling biases (Figure 4). In both phylogenies, distinct geographical clusters are clearly visible. These sequences, sampled often from the same country and within a short time frame, are seemingly indicative of local outbreaks following the importation of a globally predominant variant. For example, there is a distinct cluster (Figure 3, yellow) of sequences from Kenya collected during

2010-2012, and a large cluster (Figure 4, light blue) of sequences sampled in Tennessee (USA) in 2013-2014.

The sequences from this study fell across the global RSVA and RSVB phylogenies, indicative of multiple entries of virus into NSW, both within and between individual RSV seasons (Figures 3 and 4). We defined NSW-specific sequence clusters, as nodes with a majority of NSW sequences compared to the background of global sequences. Accordingly, we identified six and seven such clusters for RSVA and RSVB, respectively. For RSVA, all six NSW clusters were from the ON1 genotype, five of which harbored sequences from 2016 (Figures 3). Due to the bias toward 2016 it was difficult to determine the extent of off-season RSV transmission (i.e. 'over-summering'). However, some evidence for the persistence of virus within Australia between RSV seasons was observed in 2015 and 2016 (Figure 3, clusters I and II), and perhaps over multiple seasons (Figure 3, clusters II, III, and IV). However, the genetic distances between the sequences are large and node support is low, so that the clustering of sequences from different years is perhaps more likely to be due to limited sampling rather than actual virus persistence.

In the case of RSVB, six sequence pairs in individual clusters and the large BA10 genotype cluster were identified (Figure 4). Potential multiple entries might have occurred in 2012 and 2016 for the BA genotype, excluding the BA10 cluster, although this inference is again based on only a small number of sequences. Interestingly, there is some evidence for multi-season persistence from 2011 to 2012 and 2015 to 2016, although this will clearly need to be confirmed with larger data sets (Figure 4, clusters I and II). Similarly, within the BA10 genotype there is some evidence of multi-year persistence from 2015 to 2016 (Figure 4, cluster III, WM2082A). The BA10 genotype contains Australian sequences sampled between 2013 and 2016 sequences from USA, China, New Zealand, England, and Japan, with the latter the most recent sequence sampled in 2017, and thus, the BA10 genotype likely represents the most recent global circulating RSVB variant.

3.3. Geographic and age structure of RSV infections in NSW

For the most comprehensive sampled year, 2016, we assessed the extent to which the phylogenetic structure in the data reflected patient age or geographic structure. Accordingly, the facility (hospital or clinic) where the patient first presented, patient electorate, and patient age were mapped across the tree (Table S2). Thirteen different facilities were associated with RSVA, although six facilities were associated with one sequence only. Similarly, RSVB was represented by 18 facilities, nine of which were represented by one sequence only. Nevertheless, one facility, Young Hospital (located in rural NSW) contained four RSVA sequences and exhibited significant clustering with a *p*-value of 0.001 (Table S2, RSVA). These sequences were sampled in August and September and most likely represent a distinct local outbreak as genetic diversity was low (Figure 3, cluster II, WM1339A, WM2841A, WM2209A & WM4313A). All four sequences were sampled from patients residing in the Cootamundra electorate, which is also the electorate with the lowest *p*-value (0.002). The significant clustering for facility and electorate is also supported by the low parsimony score (PS) and association index (AI) values, i.e. 0.00 and 0.006 for facility and electorate, respectively. Surprisingly, no significant clustering was observed in the case of RSVB despite different geographic locations and age categories being well represented in the data (Table S2). Notably, however, both the AI and PS statistics are highly conservative [46], as the null model assumes complete panmixis, and thus weak significance may in fact indicate relatively frequent virus movement. This is supported by the

Peer-reviewed version available at Viruses 2018, 10, 476; doi:10.3390/v10090476

observation that the best sampled localities often exhibited the least geographic clustering, arguing against *in situ* transmission. For example, Mount Druitt Hospital and Westmead Hospital were the best sampled facilities (13 and 11 sequences for RSVA and RSVB, respectively) but exhibited non-significant geographic clustering (p = 0.730, and 1.00, respectively). Finally, we also investigated the extent of phylogenetic clustering by patient age group, particularly as RSV mainly infects infants. Notably, none of the age groups showed significant clustering in RSVA or RSVB (Table S2). Hence, these data indicate that the same virus lineages were able to infect and circulate within multiple age groups.

3.3. Evolutionary dynamics of RSV

Previous studies have reported a difference in evolutionary rates between the two subtypes, particularly that the G gene had significantly higher rates in RSVB than RSVA [20, 47]. As the global RSV data is highly biased in time and space, we examined evolutionary dynamics at both the global scale and within the NSW sequences alone. To assess the extent of clock-like structure in the data, we first performed a simple regression of genome-scale root-to-tip genetic distances against year of sampling with TempEst v.1.5 [36] using the RSVA and RSVB ML trees. This provided evidence for a very strong molecular clock signal, with R^2 values of 0.951 and 0.927 for RSVA and RSVB, respectively, for the NSW sequences, and 0.978 and 0.990 for the global sequences (Figure 5A&B). Under this regression method, the mean rates of nucleotide substitution were also very similar at 7.97 and 7.62 x 10^{-4} substitutions per site per year (subs/site/year) for global RSVA and RSVB, respectively, and 7.29 and 7.56 x 10^{-4} subs/site/year for RSVA and RSVB sampled in NSW in this study, respectively (Figure 5A).

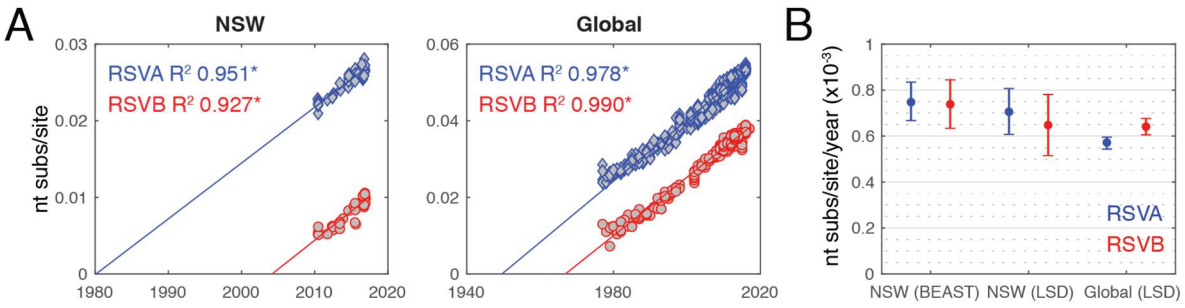


Figure 5. Evolutionary rates in RSV. (A) Linear regressions of root-to-tip genetic distances against sampling date based on maximum likelihood trees. The R^2 value for each regression is indicated and corresponding p values < 0.001 are indicated with asterisk. Sequences from NSW in this study are shown on the left and global sequences on the right. (B) Estimates of nucleotide substitution rate per site, per year are shown for sequences from NSW (BEAST and LSD estimates) and globally (LSD estimates). Rates are show as mean values (circles) and the 95% HPD (error bar). (RSVA: blue; RSVB: red).

Given this strong clock-like structure, we investigated evolutionary rates more carefully using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST for the NSW data set, and the least square dating (LSD) method for the global data set. In the case of the NSW data, the mean evolutionary rates were 6.48 x 10^{-4} (confidence interval HPD 5.15 – 7.81 x 10^{-4}) and 7.06 x 10^{-4} (confidence interval 6.07 – 8.07 x 10^{-4}) subs/site/year for RSVA and RSVB, respectively, using LSD,

and 7.48 x 10^{-4} (95% HPD 6.67 – 8.34 x 10^{-4}) and 7.39 x 10^{-4} (95% HPD 6.34 – 8.45 x 10^{-4}) subs/site/year for RSVA and RSVB, respectively, using BEAST (Figure 5B). Hence, there was no significant difference in rate between RSVA and RSVB. For the global data set LSD rate estimates were 5.72 x 10^{-4} (confidence interval 5.43 – 5.95 x 10^{-4}) and 6.41 x 10^{-4} (confidence interval 6.02 – 6.78 x 10^{-4}) subs/site/year for RSVA and RSVB, respectively. These rates are within the range reported previously for paramyxoviruses [48] and other single-stranded RNA viruses [49]. Notably, the substitution rates for the global data set were significantly different between RSVA and RSVB, and RSVB exhibited a higher rate, as previously described [20]. These global rates were also consistently lower than those observed within NSW (for both BEAST and LSD), and there was no difference between RSVA and RSVB in the NSW data set. This difference between the local (NSW) and global rates may reflect the fact that the NSW data were sampled more recently sample, and that rates are elevated towards the present because of time-dependent evolution, itself reflecting incomplete purifying selection, that is commonly observed in RNA viruses [50, 51].

5. Conclusions

We report a wide diversity of RSV lineages co-circulating in a small geographic region, reflecting a combination of continual virus entry and some sustained *in situ* transmission within NSW. Despite these fine-scale epidemiological insights, this study also highlighted the highly biased global sampling of RSV that hinders extensive analysis on the global distribution and transmission dynamics of RSV. We stress that increased targeted surveillance with more extensive virus sampling, particularly during suspected outbreaks, is essential to improve both our understanding of RSV ecology and evolution, and assist with vaccine design.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Accession numbers; Table S2: Phylogeny-trait association test for RSV in Australia.

Author Contributions: Conceptualization, J-S.E., J.K., and E.C.H.; Methodology, J-S.E., M.F, and F.D.G.; Resources, J.K, I.C., and D.E.D.; Data Curation, J-S.E.; Formal analysis; F.D.G., J.G., and J-S.E.; Visualization, F.D.G., and J-S.E; Writing-Original Draft Preparation, F.D.G., J-S.E., and E.C.H.; Supervision, J-S.E, J.K., D.E.D., and E.C.H.; Project Administration, J.K., and E.C.H.; Funding Acquisition, E.C.H.

Funding: This research was funded by the Australian Research Council grant number FL170100022 to E.C.H, and support through the Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney.

Acknowledgments: The authors acknowledge the Sydney Informatics Hub and the University of Sydney's high-performance computing cluster Artemis for providing the high-performance computing resources that have contributed to the research results reported within this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nair, H.; Nokes, D. J.; Gessner, B. D.; Dherani, M.; Madhi, S. A.; Singleton, R. J.; O'Brien, K. L.; Roca, A.; Wright, P. F.; Bruce, N.; *et al.* Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet* **2010,** *375*, 1545-1555 10.1016/S0140-6736(10)60206-1
2. Caram, L. B.; Chen, J.; Taggart, E. W.; Hillyard, D. R.; She, R.; Polage, C. R.; Twersky, J.; Schmader, K.; Petti, C. A.; Woods, C. W. Respiratory syncytial virus outbreak in a long-term care facility detected using reverse transcriptase polymerase chain reaction: an argument for real-time detection methods. *J. Am. Geriatr. Soc.* **2009,** *57*, 482-485 10.1111/j.1532-5415.2008.02153.x

3. Shi, T.; McAllister, D. A.; O'Brien, K. L.; Simoes, E. A. F.; Madhi, S. A.; Gessner, B. D.; Polack, F. P.; Balsells, E.; Acacio, S.; Aguayo, C.; *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* **2017,** *390*, 946-958 10.1016/S0140-6736(17)30938-8

4. Nolan, T.; Borja-Tabora, C.; Lopez, P.; Weckx, L.; Ulloa-Gutierrez, R.; Lazcano-Ponce, E.; Kerdpanich, A.; Weber, M. A. R.; de Los Santos, A. M.; Tinoco, J. C.; *et al.* Prevalence and incidence of respiratory syncytial virus and other respiratory viral infections in children aged 6 months to 10 years with influenza-like illness enrolled in a randomized trial. *Clin. Infect. Dis.* **2015,** *60*, E80-E89 10.1093/cid/civ065

5. Ranmuthugala, G.; Brown, L.; Lidbury, B. A. Respiratory syncytial virus - the unrecognised cause of health and economic burden among young children in Australia. *Commun. Dis. Intell.* **2011,** *35*, 177-184

6. Fagan, P.; McLeod, C.; Baird, R. W. Seasonal variability of respiratory syncytial virus infection in the Top End of the Northern Territory (2012-2014). *J. Paediatr. Child Health* **2017,** *53*, 43-46 10.1111/jpc.13303

7. Whitehall, J. S.; Bolisetty, S.; Whitehall, J. P.; Francis, F.; Norton, R.; Patole, S. K. High rate of indigenous bronchiolitis and palivuzumab. *J. Paediatr. Child Health* **2001,** *37*, 416-417

8. CDC Respiratory syncytial virus infection (RSV). https://www.cdc.gov/rsv/index.html (accessed on 05 April 2018),

9. Collins, P. L.; Dickens, L. E.; Buckler-White, A.; Olmsted, R. A.; Spriggs, M. K.; Camargo, E.; Coelingh, K. V. Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83*, 4594-4598

10. Anderson, L. J.; Hierholzer, J. C.; Tsou, C.; Hendry, R. M.; Fernie, B. F.; Stone, Y.; McIntosh, K. Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *J. Infect. Dis.* **1985,** *151*, 626-633

11. Johnson, P. R.; Spriggs, M. K.; Olmsted, R. A.; Collins, P. L. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84*, 5625-5629

12. Bouvier, N. M.; Palese, P. The biology of influenza viruses. *Vaccine* **2008,** *26*, D49-53

13. Cane, P. A.; Pringle, C. R. Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein. *J. Virol.* **1995,** *69*, 2918-2925

14. Bont, L.; Versteegh, J.; Swelsen, W. T.; Heijnen, C. J.; Kavelaars, A.; Brus, F.; Draaisma, J. M.; Pekelharing-Berghuis, M.; van Diemen-Steenvoorde, R. A.; Kimpen, J. L. Natural reinfection with respiratory syncytial virus does not boost virus-specific T-cell immunity. *Pediatr. Res.* **2002,** *52*, 363-367 10.1203/00006450-200209000-00009

15. Henderson, F. W.; Collier, A. M.; Clyde, W. A., Jr.; Denny, F. W. Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children. *N. Engl. J. Med.* **1979,** *300*, 530-534 10.1056/NEJM197903083001004

16. Graham, B. S. Vaccine development for respiratory syncytial virus. *Curr. Opin. Virol.* **2017,** *23*, 107-112 10.1016/j.coviro.2017.03.012

17. Shook, B. C.; Lin, K. Recent advances in developing antiviral therapies for respiratory syncytial virus. *Top. Curr. Chem.* **2017,** *375*, 10.1007/s41061-017-0129-4

18. Martinez, I.; Valdes, O.; Delfraro, A.; Arbiza, J.; Russi, J.; Melero, J. A. Evolutionary pattern of the G glycoprotein of human respiratory syncytial viruses from antigenic group B: the use of alternative

termination codons and lineage diversification. *J. Gen. Virol.* **1999,** *80*, 125-130 10.1099/0022-1317-80-1-125

19. Sullender, W. M.; Mufson, M. A.; Anderson, L. J.; Wertz, G. W. Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. *J. Virol.* **1991,** *65*, 5425-5434
20. Schobel, S. A.; Stucker, K. M.; Moore, M. L.; Anderson, L. J.; Larkin, E. K.; Shankar, J.; Bera, J.; Puri, V.; Shilts, M. H.; Rosas-Salazar, C.; *et al.* Respiratory syncytial virus whole-genome sequencing identifies convergent evolution of sequence duplication in the C-terminus of the G gene. *Sci. Rep.* **2016,** *6*, 26311 10.1038/srep26311
21. Trento, A.; Casas, I.; Calderon, A.; Garcia-Garcia, M. L.; Calvo, C.; Perez-Brena, P.; Melero, J. A. 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.* **2010,** *84*, 7500-7512 10.1128/JVI.00345-10
22. Agoti, C. N.; Otieno, J. R.; Munywoki, P. K.; Mwihuri, A. G.; Cane, P. A.; Nokes, D. J.; Kellam, P.; Cotten, M. Local evolutionary patterns of human respiratory syncytial virus derived from whole-genome sequencing. *J. Virol.* **2015,** *89*, 3444-3454 10.1128/JVI.03391-14
23. Brandenburg, A. H.; van Beek, R.; Moll, H. A.; Osterhaus, A. D.; Claas, E. C. G protein variation in respiratory syncytial virus group A does not correlate with clinical severity. *J. Clin. Microbiol.* **2000,** *38*, 3849-3852
24. Martinelli, M.; Frati, E. R.; Zappa, A.; Ebranati, E.; Bianchi, S.; Pariani, E.; Amendola, A.; Zehender, G.; Tanzi, E. Phylogeny and population dynamics of respiratory syncytial virus (Rsv) A and B. *Virus Res.* **2014,** *189*, 293-302 10.1016/j.virusres.2014.06.006
25. Pretorius, M. A.; van Niekerk, S.; Tempia, S.; Moyes, J.; Cohen, C.; Madhi, S. A.; Venter, M.; Group, S. S. Replacement and positive evolution of subtype A and B respiratory syncytial virus G-protein genotypes from 1997-2012 in South Africa. *J. Infect. Dis.* **2013,** *208*, S227-237 10.1093/infdis/jit477
26. Tan, L.; Lemey, P.; Houspie, L.; Viveen, M. C.; Jansen, N. J.; van Loon, A. M.; Wiertz, E.; van Bleek, G. M.; Martin, D. P.; Coenjaerts, F. E. Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology. *PLoS One* **2012,** *7*, e51439 10.1371/journal.pone.0051439
27. Bose, M. E.; He, J.; Shrivastava, S.; Nelson, M. I.; Bera, J.; Halpin, R. A.; Town, C. D.; Lorenzi, H. A.; Noyola, D. E.; Falcone, V.; *et al.* Sequencing and analysis of globally obtained human respiratory syncytial virus A and B genomes. *PLoS One* **2015,** *10*, e0120098 10.1371/journal.pone.0120098
28. Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014,** *30*, 2114-2120 10.1093/bioinformatics/btu170
29. Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011,** *29*, 644-652 10.1038/nbt.1883
30. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012,** *19*, 455-477 10.1089/cmb.2012.0021
31. Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012,** *9*, 357-359 10.1038/nmeth.1923
32. Kuraku, S.; Zmasek, C. M.; Nishimura, O.; Katoh, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* **2013,** *41*, W22-W28 10.1093/nar/gkt389

33. Martin, D. P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015,** *1,* 10.1093/ve/vev003
34. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006,** *22,* 2688-2690 10.1093/bioinformatics/btl446
35. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014,** *30,* 1312-1313 10.1093/bioinformatics/btu033
36. Rambaut, A.; Lam, T. T.; Carvalho, L. M.; Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2016,** *2,* 10.1093/ve/vew007
37. Drummond, A. J.; Suchard, M. A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012,** *29,* 1969-1973 10.1093/molbev/mss075
38. To, T. H.; Jung, M.; Lycett, S.; Gascuel, O. Fast dating using least-squares criteria and algorithms. *Syst. Biol.* **2016,** *65,* 82-97 10.1093/sysbio/syv068
39. Duchene, S.; Duchene, D. A.; Geoghegan, J. L.; Dyson, Z. A.; Hawkey, J.; Holt, K. E. Inferring demographic parameters in bacterial genomic data using Bayesian and hybrid phylogenetic methods. *BMC Evol. Biol.* **2018,** *18,* 10.1186/s12862-018-1210-5
40. Parker, J.; Rambaut, A.; Pybus, O. G. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* **2008,** *8,* 239-246 10.1016/j.meegid.2007.08.001
41. Fall, A.; Dia, N.; Cisse el, H. A.; Kiori, D. E.; Sarr, F. D.; Sy, S.; Goudiaby, D.; Richard, V.; Niang, M. N. Epidemiology and molecular characterization of human respiratory syncytial virus in senegal after four consecutive years of surveillance, 2012-2015. *PLoS One* **2016,** *11,* e0157163 10.1371/journal.pone.0157163
42. Rodriguez-Fernandez, R.; Tapia, L. I.; Yang, C. F.; Torres, J. P.; Chavez-Bueno, S.; Garcia, C.; Jaramillo, L. M.; Moore-Clingenpeel, M.; Jafri, H. S.; Peeples, M. E.; *et al.* Respiratory syncytial virus genotypes, host immune profiles, and disease severity in young children hospitalized with bronchiolitis. *J. Infect. Dis.* **2018,** *217,* 24-34 10.1093/infdis/jix543
43. Tabatabai, J.; Prifert, C.; Pfeil, J.; Grulich-Henn, J.; Schnitzler, P. Novel respiratory syncytial virus (RSV) genotype ON1 predominates in Germany during winter season 2012-13. *PLoS One* **2014,** *9,* e109191 10.1371/journal.pone.0109191
44. Thongpan, I.; Mauleekoonphairoj, J.; Vichiwattana, P.; Korkong, S.; Wasitthankasem, R.; Vongpunsawad, S.; Poovorawan, Y. Respiratory syncytial virus genotypes NA1, ON1, and BA9 are prevalent in Thailand, 2012-2015. *PeerJ* **2017,** *5,* e3970 10.7717/peerj.3970
45. Geoghegan, J. L.; Saavedra, A. F.; Duchene, S.; Sullivan, S.; Barr, I.; Holmes, E. C. Continental synchronicity of human influenza virus epidemics despite climatic variation. *PLoS Pathog.* **2018,** *14,* e1006780 10.1371/journal.ppat.1006780
46. Slatkin, M.; Maddison, W. P. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **1989,** *123,* 603-613
47. Tan, L.; Coenjaerts, F. E.; Houspie, L.; Viveen, M. C.; van Bleek, G. M.; Wiertz, E. J.; Martin, D. P.; Lemey, P. The comparative genomics of human respiratory syncytial virus subgroups A and B: genetic variability and molecular evolutionary dynamics. *J. Virol.* **2013,** *87,* 8213-8226 10.1128/JVI.03278-12
48. Pomeroy, L. W.; Bjornstad, O. N.; Holmes, E. C. The evolutionary and epidemiological dynamics of the paramyxoviridae. *J. Mol. Evol.* **2008,** *66,* 98-106 10.1007/s00239-007-9040-x
49. Duffy, S.; Shackelton, L. A.; Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **2008,** *9,* 267-276 10.1038/nrg2323

493 50. Duchene, S.; Ho, S. Y. W.; Holmes, E. C. Declining transition/transversion ratios through time reveal
494 limitations to the accuracy of nucleotide substitution models. *BMC Evol. Biol.* **2015,** *15,* 10.1186/s12862-
495 015-0312-6
496 51. Duchene, S.; Holmes, E. C.; Ho, S. Y. Analyses of evolutionary dynamics in viruses are hindered by a
497 time-dependent bias in rate estimates. *Proc. Biol. Sci.* **2014,** *281,* 10.1098/rspb.2014.0732

498