The spatial-temporal dynamics of respiratory syncytial virus infections across the east-west coasts of Australia during 2016-17

Mark Robertson[1#], John-Sebastian Eden[2,3#], Avram Levy[4,5], Ian Carter[1], Rachel L Tulloch[2,3], Elena J Cutmore[2,3], Bethany A Horsburgh[2,3], Chisha T Sikazwe[4,5], Dominic E Dwyer[1,3], David W Smith[4,6] and Jen Kok[1]

[1]NSW Health Pathology - Institute for Clinical Pathology and Medical Research, NSW Health Pathology, Westmead Hospital, Westmead, NSW 2145, Australia

[2]Centre for Virus Research, Westmead Institute for Medical Research, Westmead, NSW 2145, Australia

[3]Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia

[4]PathWest Laboratory Medicine WA, Department of Microbiology, Nedlands, WA 6009, Australia

[5]School of Biological Sciences, The University of Western Australia, Crawley, WA 6009, Australia

[6]School of Medicine, The University of Western Australia, Crawley, WA 6009, Australia

[#]These authors contributed equally

Corresponding authors

Dr John-Sebastian Eden

Email: js.eden@sydney.edu.au

Phone: +61 2 8627 1817

Dr Jen Kok

Email: jen.kok@health.nsw.gov.au

Phone: +61 2 8890 6255

Abstract

Respiratory syncytial virus (RSV) is an important human respiratory pathogen. In temperate regions a distinct seasonality is observed, where peaks of infections typically occur in early winter, often preceding the annual influenza season. Infections are associated with high rates of morbidity and mortality, and in some populations exceeds that of influenza. Two subtypes, RSV-A and RSV-B, have been described, and molecular epidemiological studies have shown that both viruses mostly co-circulate. This trend also appears to be the case for Australia, however previous genomic studies have been limited to cases from one Eastern state - New South Wales. As such, the broader spatial patterns and viral traffic networks across the continent are not known. Here, we conducted a whole genome study of RSV comparing strains across eastern and western Australia during the period January 2016 to June 2017. In total, 96 new RSV genomes were sequenced, compiled with previously generated data, and examined using a phylodynamic approach. This analysis revealed that both RSV-A and RSV-B strains were circulating, and each subtype was dominated by a single genotype, RSV-A/ON1-like and RSV-B/BA10-like viruses. Some geographical clustering was evident in strains from both states with multiple distinct sub-lineages observed and relatively low mixing across jurisdictions suggesting that endemic transmission was likely seeded from imported, unsampled locations. Overall, the RSV phylogenies reflected a complex pattern of interactions across multiple epidemiological scales from fluid virus traffic across global and regional networks to fine-scale local transmission events.

Keywords

respiratory syncytial virus, molecular epidemiology, Australia, whole-genome sequencing, phylogenetics

Introduction

Respiratory syncytial virus (RSV) is a major cause of acute respiratory tract infections in patients of all ages, producing significant morbidity and mortality (Shi et al., 2017). The greatest burden of disease is in children under 1 year old, where it is the most common cause of acute respiratory tract infection, and in this age group is second only to malaria as a cause of death globally (Griffiths et al., 2017; Hall et al., 2009). Importantly, RSV infection in young children may also lead to long-term sequelae such as asthma, chronic bronchitis and obstructive pulmonary disease (Beigelman & Bacharier, 2016; Griffiths et al., 2017). Other populations particularly impacted by RSV include those over 65, and immunosuppressed patients, such as solid organ and bone marrow transplant recipients (Beigelman & Bacharier, 2016; Griffiths et al., 2017).

Despite being discovered in the 1950s, the burden of RSV disease has only recently been appreciated, which is in part due to more reliable methods of diagnosis and detection. The laboratory diagnosis of RSV was initially reliant on viral isolation and the visualisation of characteristic syncytial cytopathic effects, for which its name is derived (Henrickson & Hall, 2007). These techniques were slow and required technical expertise. The most commonly used modalities now include rapid antigen (such as lateral flow immunochromatography and fluorescent immunoassays) and nucleic acid amplification tests (NATs). Molecular assays, including both commercial and in-house NATs, offer increased sensitivity and specificity, as well as the ability to be multiplexed to detect other respiratory pathogens such as influenza, parainfluenza, rhinovirus and human metapneumovirus (Mahony et al., 2007). Rapid test assays can also offer the additional benefit of early diagnosis, which allows for appropriate infection control interventions, rationalisation of unnecessary antibiotic therapy and shorter hospitalisation periods (Henrickson & Hall, 2007). More generally, improved diagnostics and reporting have begun to shed light on the true incidence, seasonal patterns and peaks of activity of RSV. In temperate regions such as the southern major metropolitan areas of Australia, the seasonal peak typically occurs in the late autumn to early winter period (late-March to mid-August) in the months leading into the influenza season (Di Giallonardo et al., 2018; Henrickson & Hall, 2007; Yeoh et al., 2020).

RSV can be divided into two antigenically and genetically distinct subtypes, RSV-A and RSV-B. These may be further divided into genetic groups based on the viral glycoprotein (G gene), termed genotypes, with at least 11 and 23 for RSV-A and RSV-B, respectively. However,

recent work has proposed shifting to a genotype classification based on whole genome sequencing (WGS) in order to increase phylogenetic resolution (Ramaekers et al., 2020). Serological studies have shown that the majority of people are infected by age two, and whilst primary infections are typically more severe they are not protective against repeated infection (Griffiths et al., 2017). The spectrum of disease severity associated with RSV infection remains an important yet controversial topic. Studies have made associations between increased severity and particular subtypes/genotypes (Vandini et al., 2017); however these are complicated by both host (Tal et al., 2004) and viral factors (DeVincenzo et al., 2005), as well as their interactions. Molecular epidemiological studies have shown that both RSV-A and RSV-B co-circulate during a season, and often at similar levels (Cattoir et al., 2019; James R. Otieno et al., 2017; J. R. Otieno et al., 2018; Pangesti et al., 2018; Park et al., 2017). Furthermore, for each of these subtypes, a single genotype will tend to predominate, such as with the recent RSV-A ON1-like and RSV-B BA10-like viruses (Di Giallonardo et al., 2018; Eshaghi et al., 2012; Pretorius et al., 2013). Our understanding of the basic molecular epidemiology of RSV has been strengthened by WGS, which, similar to other pathogens, is becoming increasingly common in its application in infectious disease surveillance (Agoti et al., 2014; Dapat et al., 2010; Di Giallonardo et al., 2018). The added resolution from WGS has been particularly useful for elucidating transmission networks at both local (Agoti, Otieno, Munywoki, et al., 2015) and epidemiological scales (Di Giallonardo et al., 2018), as well as for identifying and classifying the introduction and spread of new genotypes (Agoti et al., 2014; Ramaekers et al., 2020).

We recently performed the first genome-scale study of RSV molecular epidemiology in Australia (Di Giallonardo et al., 2018), demonstrating a wide diversity of co-circulating RSV lineages, with limited evidence of strong age and geographical clustering. However, this study was limited to cases obtained from eastern Australia in New South Wales (NSW) through the Western Sydney Local Health District. Here, we expand on these initial investigations to perform a transcontinental study of RSV genomic epidemiology in Australia. We compare RSV strains obtained from Western Australia (WA) to those from NSW on the east coast over an equivalent time period to examine the phylogenetic distribution of strains, and to describe viral traffic between these regions.

Materials and Methods

Sample collection and processing

WGS was conducted on residual RSV-positive specimens collected and tested through routine testing at diagnostic laboratories in two major Australian diagnostic laboratories. Specimens were de-identified with basic demographic information collected including age, sex and postcode, as per protocols approved by local ethics and governance committees (LNR/17/WMEAD/128). One hundred and three RSV-A and RSV-B positive cases from WA with a RT-PCR cycle threshold <30 in original screen were collected between January 2016 and May 2017. Sample aliquots (typically, nasopharyngeal swabs in viral transport medium) were transferred to the Institute of Clinical Pathology and Medical Research (ICPMR), NSW, and subsequently extracted using the Qiagen EZ1 Advance XL extractor with the EZ1 Virus Mini Kit v2.0 (Qiagen, Germany). For the eastern Australian RSV cases, 45 previously untyped RSV-positive viral extracts from cases collected during early 2017 at the ICPMR were obtained from storage at -80°C archive. Additional cases from 2016 in NSW were sequenced in a previous study (Di Giallonardo et al., 2018).

Whole genome sequencing

A previously published approach was used to amplify both RSV-A & RSV-B genomes (Di Giallonardo et al., 2018). In short, RT-PCR was used to amplify four overlapping amplicons (each ~4kb) that together span the RSV genome. The size and yield of each RT-PCR was determined by agarose gel electrophoresis, and the four targets pooled equally. The pooled RSV amplicons were then purified with Agencourt AMpure XP beads (Beckman Coulter, USA) and quantified using the Quant-iT PicoGreen dsDNA Assay (Invitrogen, USA). The purified DNA was then diluted to 0.25 ng/µL and prepared for sequencing with the Nextera XT DNA library prep kit (Illumina, USA). Libraries were sequenced on an Illumina MiSeq using a 300 cycle v2 kit (150 nt paired end reads). Raw paired sequence reads were trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore) and then *de novo* assembled using Trinity (Grabherr et al., 2011). RSV contigs were identified by a local Blastn (Altschul et al., 1990) using a database of RSV reference genomes from NCBI RefSeq. The trimmed reads were remapped to draft genome contigs using BBMap (https://sourceforge.net/projects/bbmap/) to check the assembly, and the final majority consensus was extracted for each sample.

Phylodynamic analysis

The aim of this study was to compare the spatial and temporal dynamics of RSV infection in Australia. To do this, we used a phylogenetic approach to compare the distribution of strains circulating in two major, geographically distinct regions – NSW and WA – representing eastern and western Australia, respectively, during the period January 2016 to June 2017. All WA data, as well as NSW data from the year 2017, was generated within this study. To ensure even sampling across sites, these data were combined with NSW data from 2016 that was obtained from a previous study (Di Giallonardo et al., 2018). To provide additional context, RSV genome data was also sourced from NCBI GenBank where location (country) and collection date (year) was known. The Australian and global RSV genomes were first aligned with MAFFT (Katoh & Standley, 2013), using the FFT-NS-i algorithm followed by manual inspection of gapped regions particularly in the G gene and non-coding regions. The alignments were then trimmed to include only the coding regions, and screened for potential recombinants using RDP4 (Martin et al., 2015) with default parameters that were then removed before analysis. The RSV-A alignment included 1,190 sequences with a length of 15,747 nt, whilst the RSV-B alignment included 1,121 sequences with a length of 15,646 nt. To increase sampling resolution, we also generated comparable datasets using the G gene region only that also included available partial genome sequences from NCBI GenBank that covered the G gene region by at least 300 bp. The final G gene alignments contained 6,603 and 4,300 sequences for RSV-A and RSV-B, respectively, and were both trimmed to the G gene coding region (approximately 960 bp long). The best-fit DNA substitution model for these data was determined with jModelTest (Darriba et al., 2012). Maximum likelihood (ML) trees were then estimated for both RSV-A and RSV-B alignments using RAxML (Stamatakis, 2014) employing the best-fit model, which in both cases was found to be the General-Time Reversible model with a gamma distribution of rates (GTR+G). Support for individual nodes was determined by 1000 bootstrap replications.

Phylogeographic clustering

From our overall ML trees, it was found that all sequences from 2016 and 2017 were either RSV-A ON1-like or RSV-B BA10-like viruses, therefore we limited further analyses to these specific clusters. To evaluate the geographic structure in the RSV-A and RSV-B trees we used the Bayesian Tip-association Significance (BaTS) program (Parker et al., 2008) with 1000 replicates to analyse the genome datasets. Using BaTS we determined the parsimony score (PS), association index (AI), and maximum clade size (MC) statistics for the location associated with each sequence, specifically focusing on the comparison between eastern and western Australia. To account for other jurisdictions (countries) present in the data, such sequences

were assigned to their continent of sampling. This analysis required a posterior distribution of trees, which were obtained using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.10.4 (Drummond et al., 2012). Here, we used the best-fit DNA model (GTR+G) with a strict clock and a constant population size, as shown to be appropriate previously (Di Giallonardo et al., 2018). All analyses were run for 50 million steps with sampling every 500 steps with 10-20% burn-in. To ensure convergence, three independent runs were conducted and merged to obtain the final set of trees used in the clustering analysis. The maximum clade credibility tree from the Bayesian analyses using BEAST were generated with heights scaled to mean values.

Local transmission events

In order to identify potential local transmission events (for instance, an outbreak at an institution), we examined both ML and time-scaled Bayesian trees using the genome datasets, for monophyletic groups containing near identical sequences (~99.9% nucleotide identity, less than 10 base pairs different across the genome) sampled up to two weeks apart. Patient demographics and sampling location were then mapped to determine traits associated with these fine-scale phylogenetic groups. The G gene ML trees were then used to validate clustering and to consider additional sources (sampling locations) for local RSV cases.

Data availability

All sequences generated in this study have been submitted to NCBI GenBank (MW160744 - MW160839). Furthermore, data and material relevant to this study are available from https://github.com/jsede/RSV_NSW_WA.

Results and Discussion

Whole genome sequencing of RSV strains from NSW and WA in 2016 and 2017

In order to compare the spatial distribution of RSV strains across eastern (NSW) and western (WA) Australia during the period January 2016 to June 2017, we performed WGS on stored samples collected from community and hospitalised persons presenting with an influenza-like illness. The demographic details of the cohort have been summarised in Table 1. For the NSW cases, no specific sample selection criteria were used except for sample availability from an archived collection of RSV positive residual diagnostic specimens. Most samples from NSW were obtained from children ≤5 years age (51.6%, n=65/126) with a further 20.6% of samples collected from patients ≥65 years (n=26/126). In contrast, samples from WA were selected to represent all age groups. Despite this, similar to NSW, most available samples were collected from the young (62.1%, n=64/103 from patients ≤ 5 years age) and the elderly (17.5%, n=18/103 from patients ≥65 years), reflecting the distribution of infection at both extremes of age. The distribution between sexes was approximately equal for both study sites (Table 1).

Table 1. Demographic details of NSW and WA RSV cohorts.

Age	NSW	WA	Sex	NSW	WA
	Number of samples (%)			Number of samples (%)	
≤5	65 (51.6%)	64 (62.1%)	Male	66 (52.4%)	48 (46.6%)
5 - 65	35 (27.8%)	21 (20.4%)	Female	60 (47.6%)	55 (53.4%)
≥65	26 (20.6%)	18 (17.5%)	*Total*	*126 (100%)*	*103 (100%)*
Total	*126 (100%)*	*103 (100%)*			

Virus genomes were successfully extracted and sequenced from 63 of the 103 respiratory samples collected in WA during 2016 and 2017, and 32 of the 45 NSW cases collected in 2017 (total new genomes n=96, where one case from NSW was an RSV-A/B mixed infection). These data were combined with existing genomes from NSW during 2016 (n=93) generated previously (Di Giallonardo et al., 2018), bringing the total number of Australian genomes for analysis across both states in 2016 and 2017 in the current study to 189. The breakdown between sampling location and RSV subtypes determined by WGS shows that both

RSV-A and RSV-B subtypes were present in both NSW and WA populations (Figure 1). In NSW, there was a higher proportion of RSV-B strains (63%, n=80/126), which was consistent across the entire study period. Furthemore, in NSW, the peak in RSV activity typically occurs in May-June each year (NSW Health influenza surveillance data), however most of our genome sequences in 2016 were from isolates in August to October that year; that is, they were from the later part of the RSV season. In WA, there was even representation of RSV-A and RSV-B in the WGS data across 2016 with an increase in the relative proportion of RSV-A during 2017. While our sampling for WA was intentionally even with regards to time, subtype, age and setting, the predominance of RSV-A during 2017 was also observed by the initial diagnostic testing where the used resolves RSV subtypes. Specifically, with this lab testing data a transition from RSV-B to RSV-A was observed between the 2016 and 2017 seasons in WA (data not shown). These results are consistent with other molecular epidemiological studies globally that show both RSV-A and RSV-B subtypes co-circulate with shifting predominance across seasons (Di Giallonardo et al., 2018; Luo et al., 2020; Razanajatovo Rahombanjanahary et al., 2020; Yun et al., 2020)

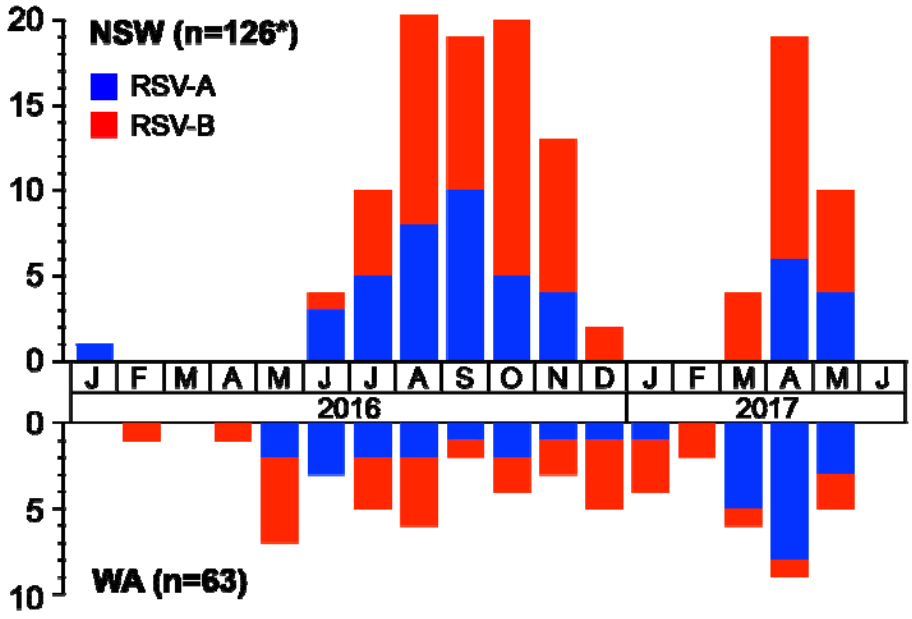

Figure 1. Whole-genome sequencing of respiratory syncytial virus (RSV) in Western Australia and New South Wales between January 2016 and June 2017 by month. The y-axis represents counts of genomes sequenced. The x-axis represents months when the samples were collected, with year shown underneath. RSV-A and RSV-B subtypes have been coloured blue and red respectively, as per key provided. In total, over the period of investigation, 189 RSV genomes were compared including 63 from WA and 126 from NSW. WA samples were selected pre-sequencing to cover both subtypes during seasonal and inter-seasonal periods and therefore do not necessarily reflect the subtype or temporal

248 distribution of RSV in WA. The asterisk indicates that of the 126 NSW genomes, 93 genomes from 2016
249 were generated in a previous study (Di Giallonardo et al., 2018).

250

251 *Phylodynamics of RSV infections in NSW and WA in a global context*

252 The WGS data from NSW and WA were aligned and compared to global genome references to
253 provide genetic context for local strains. Phylogenetic analysis using a maximum likelihood
254 approach was then employed to examine the diversity of RSV-A and RSV-B strains (Figures 2 &
255 3). This analysis revealed that for each subtype, regardless of location, a single genotype was
256 predominant. For RSV-A, recent viruses from both NSW and WA were derived from the ON1
257 lineage (Eshaghi et al., 2012), which has been the predominant RSV-A genotype globally since
258 it first emerged in 2011 (Agoti et al., 2014; Di Giallonardo et al., 2018; Eshaghi et al., 2012;
259 Tabatabai et al., 2014)(Figure 2). Similarly, the RSV-B phylogeny showed that circulating
260 viruses were mostly of the BA10 lineage (Figure 3). From the limited available sequence data
261 on GenBank, it appears that the BA10-like viruses are circulating globally, however, there are
262 few published molecular epidemiology studies to support the suggestion they are predominant.
263 It is clear however, that in Australia, in both NSW and WA, that this has been the major RSV-B
264 genotype since at least 2014 (Di Giallonardo et al., 2018).

265 Across both RSV-A and RSV-B phylogenies there does not appear to be strict spatial
266 clustering when comparing viruses from NSW and WA (Figures 2 & 3). That is, the viruses from
267 WA and NSW do not form monophyletic groups. Such monophyletic clustering would not be
268 expected based on what is known for other respiratory pathogens - such as influenza - which
269 are characterised by high levels of gene flow and viral traffic at global scales (Rambaut et al.,
270 2008; Vijaykrishna et al., 2015). As such, we observed multiple co-circulating sub-lineages
271 distributed across the entire diversity of RSV-A/ON1-like and RSV-B/BA10-like viruses (Figures
272 2 & 3). However, within these sub-lineages the viruses often formed clusters based on sampling
273 location, therefore, local spatial structure was apparent in our analysis. To explore this formally,
274 we conducted a clustering analysis with BaTS, and a posterior set of trees estimated using a
275 time-scaled phylogenetic analysis in BEAST. We limited the analysis to the RSV-A/ON1 and
276 RSV-B/BA10-like viruses, and the maximum clade credibility tree for each RSV subtype were
277 found to be congruent with the ML trees (Figures 2 & 3). Using the posterior set of trees we then
278 measured the degree of clustering based on sampling location (Table 2).

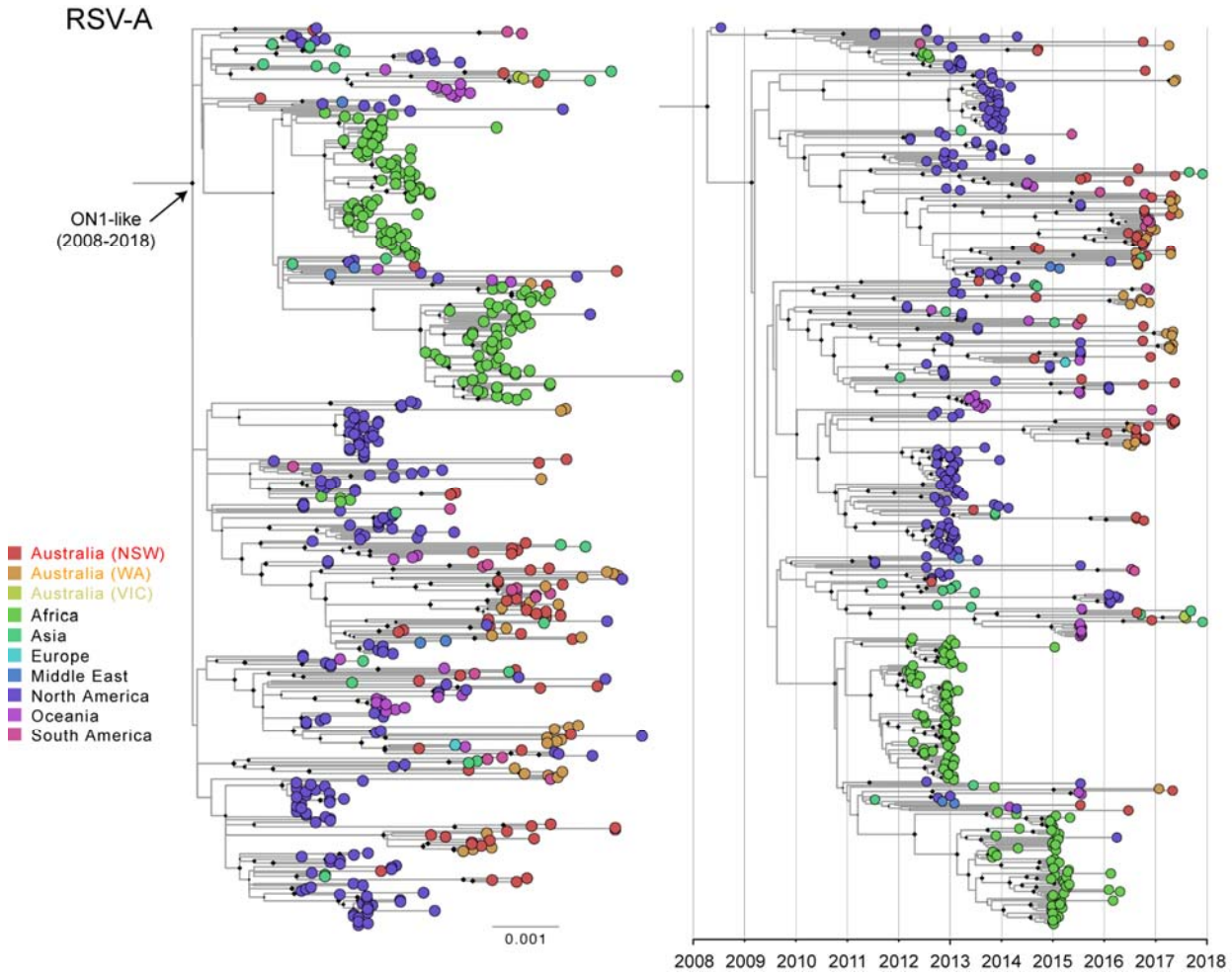

Figure 2. Phylogenetic analysis of respiratory syncytial virus (RSV) A strains circulating in Western Australia and New South Wales between January 2016 and June 2017. The global phylogenies were first estimated using alignments of complete genome sequences of RSV-A (n=1,190) strains circulating since the late 1970's, however for clarity, only the recently circulating RSV-A ON1 lineage has been shown (n=499), which is defined by the specific branch marked with an arrow. The panel on the left shows the tree estimated using a maximum likelihood (ML) approach, while the tree on the right shows the same phylogeny estimated using the time-scaled Bayesian approach. The taxa, shown as small circles, have been coloured by sampling location as per the key provided. The red and orange coloured circles sequences sampled in NSW and WA, respectively. Diamonds at nodal positions indicate branching support with bootstrap replicate values >70% and posterior probability values >0.9 for the ML and time-scaled Bayesian trees, respectively. The left sided scale bar represents the number of substitutions per site, while the right is scaled to time (years).

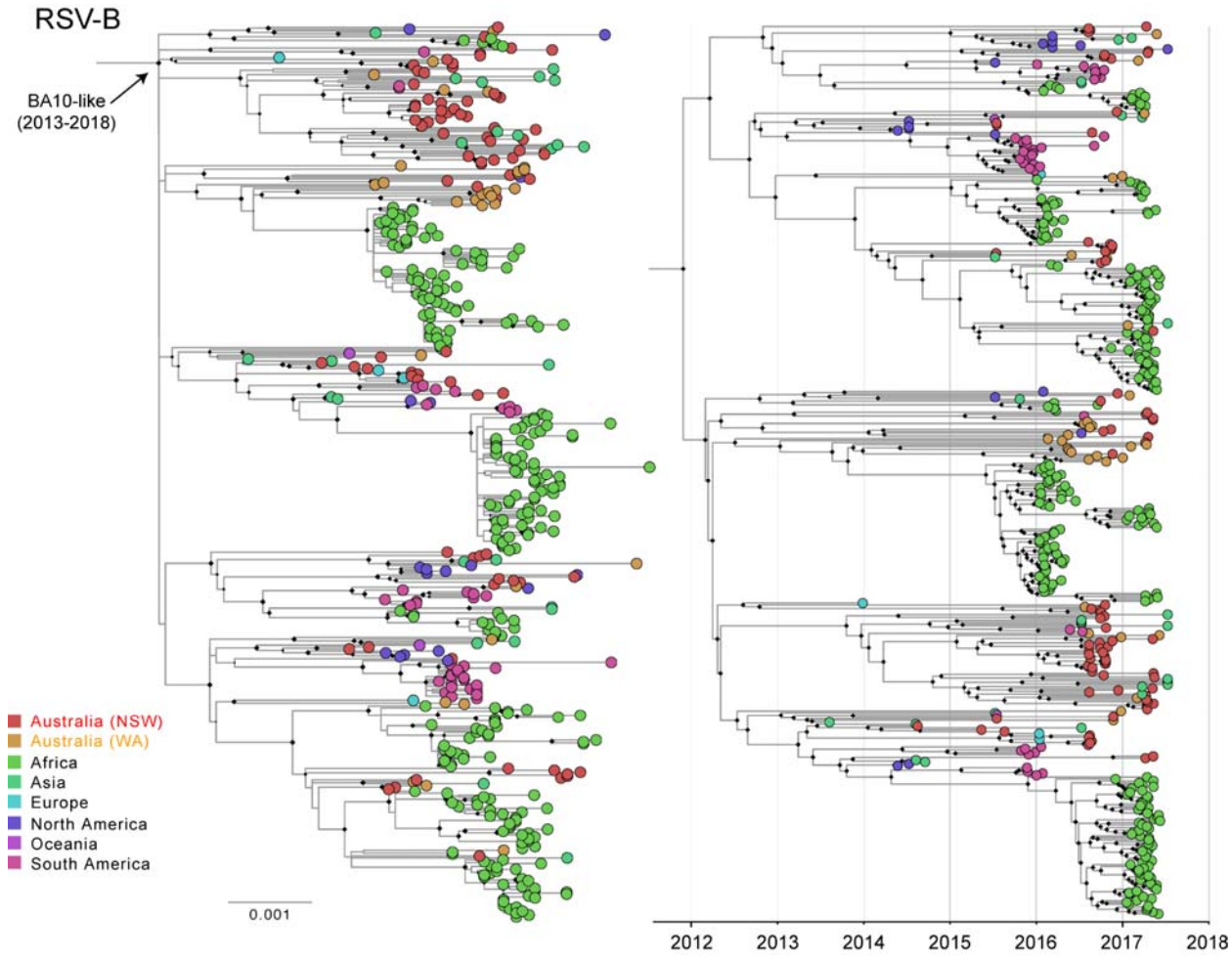

Figure 3. Phylogenetic analysis of respiratory syncytial virus (RSV) B strains circulating in Western Australia and New South Wales between January 2016 and June 2017. The global phylogenies were first estimated using alignments of complete genome sequences of RSV-B (n=1,121) strains circulating since the 1960's, however for clarity, only the recently circulating RSV-B BA10 lineage has been shown (n=470), which is defined by the specific branch marked with an arrow. The panel on the left shows the tree estimated using a maximum likelihood (ML) approach, while the tree on the right shows the same phylogeny estimated using the time-scaled Bayesian approach. The taxa, shown as small circles, have been coloured by sampling location as per the key provided. The red and orange coloured circles sequences sampled in NSW and WA, respectively. Diamonds at nodal positions indicate branching support with bootstrap replicate values >70% and posterior probability values >0.9 for the ML and time-scaled Bayesian trees, respectively. The left sided scale bar represents the number of substitutions per site, while the right is scaled to time (years).

Table 2. Phylogeny-trait association test for RSV in Australia and other regions globally.

Location	RSV-A		RSV-B	
	Seqs (n)	p-value	Seqs (n)	p-value
Overall clustering[1]	*499*	*<0.001*	*470*	*<0.001*
Asia	24	0.002	26	0.006
South America	16	0.001	42	0.001
North America	185	0.001	17	0.001
Middle East	5	0.008	0	~
Oceania[2]	28	0.001	2	1.000
Africa	147	0.001	263	0.001
Australia (NSW)	60	0.006	84	0.001
Australia (VIC)	2	0.002	0	~
Australia (WA)	31	0.001	32	0.001
Europe	1	1.000	4	1.000

[1]Determined by Association Index (AI) and Parsimony Score (PS).

[2]New Zealand only.

When sequences were grouped by broad geographical regions, specifically, continents and the two Australian states NSW & WA (and VIC for RSV-A data), phylogeny-trait association tests indicated a significant pattern of overall geographical clustering (p-values <0.001) as measured by AI and PS scores (Table 2). For individual locations, most were found to cluster with significant scores. This includes Asia, the Americas, the Middle East, Africa, Oceania, NSW, WA and VIC for RSV-A viruses (all p-values <0.008) and then similarly all for RSV-B where the sampling was sufficient (n >4 sequences). This analysis demonstrates that, at larger epidemiological scales, viral lineages are often imported and established locally. We note however, that this analysis is biased for regions that have been highly sampled, as highlighted by the American and Kenyan viruses, which are mostly from one specific city and/or hospital

(Supplementary Figures 1 & 2) (J. R. Otieno et al., 2018). Our sampling of NSW and WA strains is similarly constrained, except that our sampling sites are both major diagnostics hubs which perform testing across wide geographic regions that are representative of large and diverse populations (Di Giallonardo et al., 2018), and in our comparison the period of sampling was equivalent. Despite this and any apparent geographical clustering, the phylogenies both reflect a lack of data from many jurisdictions globally, and our insights into global and local viral traffic will remain hampered until these genome sequences become available. To address this, we also performed ML phylogenetic analysis on a more well-sampled dataset derived from sequences of the G gene region only (Supplementary Figures 3, 4 & 5). While this analysis improved the overall location sampling, particularly for European and Asian sequences, there was less phylogenetic resolution, therefore a trade-off exists between using the less sampled, higher resolution whole genome data versus the more well sample partial genome (G gene) sequences available on NCBI GenBank (Supplementary Figure 3). Regardless, as shown below, this data can shed light on the possible sources and global transmission networks for RSV.

Lineages and local clusters

A detailed examination of individual lineages (Supplementary Figures 1 & 2) confirmed that most only contained viruses from either state. There were some instances where individual viruses from NSW were nested within diversity solely comprised of WA strains, and vice-versa, and which may represent viral migration events between the jurisdictions. The overall pattern suggests low levels of mixing between NSW and WA, perhaps due to the large geographic area of each state and the distance that separates them. Moreover, we had suggested previously that some evidence of viral persistence across seasons was present in NSW (Di Giallonardo et al., 2018). Here, we found no direct link between the 2016 and 2017 seasons in both jurisdictions and that viral diversity is likely maintained and seeded from imported yet unsampled locations. An examination of the better-sampled G gene phylogenies (Supplementary Figures 4 & 5) did not clarify potential additional sources, except for sporadic examples where NSW or WA were clustered genetically and temporally with Asian or European strains. Importantly, we also found examples where two Australian sequences clustered in the G gene phylogeny but were then sufficiently different at a genome scale to not be classified as clusters such as seen with the NSW and WA strains WM1079A/2016-08-14 (MH760625) and PW3375488K/2016-08-09 (MW160759) (Supplementary Figures 1 & 4). This is consistent with other studies that have shown that local persistence makes a minimal contribution to the

355 diversity of RSV strains in any given area (Agoti, Otieno, Ngama, et al., 2015; Zou et al., 2016).
356 Rather, viral importation and global mixing of strains seem to be the major driver of RSV
357 diversity and their sources. Despite the data suggesting limited east-west viral traffic across the
358 Australian continent, the Sydney-Perth air route remains one of the busiest in the country, nor
359 can we discount other regions within Australia such as the tropical North, where persistence of
360 viral lineages may occur as important sources of viral diversity. Indeed, from our understanding
361 of influenza virus in Australia, the synchronized dissemination of viral strains across the country
362 is driven by both multiple introductions from the global population and strong domestic
363 connectivity (Geoghegan et al., 2018). Increasing the breadth of WGS data for RSV cases
364 nationally would assist in confirming what domestic viral traffic networks exist, particularly
365 between tropical and temperate regions, as well as, between major population centres on the
366 east coast such as Sydney and Melbourne.

367 Next, we examined genetic clusters defined by high genetic identity (<10 bp difference
368 across the genome) and similar sampling periods (collection dates within two weeks) to
369 consider what features such as patient age and sampling location define them. For RSV-A, we
370 identified 8 clusters (Supplementary Figure 1 & Table 1) and for RSV-B we identified 10 clusters
371 (Supplementary Figure 2 & Table 2). In this study 75% (n=6/8) and 80% (n=8/10) of RSV-A and
372 RSV-B clusters, respectively could be linked to common localities and/or institutions, including
373 individual hospital wards or emergency departments. While two of the RSV-B clusters were
374 found to be repeat samples of the same patient (Supplementary Table 2), overwhelmingly,
375 these clusters of genetically and temporal related sequences most likely represent fine-scale
376 transmission events.

377 **Conclusions**

378 In summary, our analysis has identified a number of features of RSV epidemiology and patterns
379 of spread including i) the co-circulation of both major subtypes: RSV-A and RSV-B; ii) a single
380 genotype for each subtype predominates each season; iii) multiple distinct sub-lineages of each
381 genotype will co-circulate and which are associated with regional and local clustering/outbreaks,
382 iv) little viral mixing across the east-west coasts of Australia despite apparent overall geographic
383 clustering, v) that genetically and temporal related sequences most likely represent fine-scale
384 transmission events such as institutional outbreaks and vi) that whole genome sequence is
385 required and encouraged over partial G gene sequencing for elucidating clusters and
386 transmission pathways. Taken together, this presents a complex phylodynamic pattern with

387 globally circulating diversity with viral mixing across different regions, yet, finer-scale patterns
388 revealing multiple endemic sub-lineages and clusters consistent with local transmission events
389 and outbreaks. This highlights the connections of genomic data across multiple epidemiological
390 scales and further strengthens the need for much greater sampling of RSV, not just here in
391 Australia, but globally.
392

Acknowledgments

We thank all the members of the Virology and Microbiology teams at NSW Health Pathology - ICPMR in Westmead, NSW and the PathWest QE2 Medical Centre laboratories in Perth, WA for all their contributions towards processing the diagnostics specimens used in this study.

Funding

Funding was provided through the ICPMR Private Practice Trust fund, the National Health and Medical Research Council Centre of Research Excellence in Emerging Infectious Diseases (1102962) and the Marie Bashir Institute for Infectious Diseases and Biosecurity at the University of Sydney.

Conflicts of Interest

None

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. *Emerging Infectious Diseases*, *20*(6), 950–959.

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. *Journal of Virology*, *89*(7), 3444–3454.

Agoti, C. N., Otieno, J. R., Ngama, M., Mwihuri, A. G., Medley, G. F., Cane, P. A., & Nokes, D. J. (2015). Successive Respiratory Syncytial Virus Epidemics in Local Populations Arise from Multiple Variant Introductions, Providing Insights into Virus Persistence. *Journal of Virology*, *89*(22), 11630–11642.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Beigelman, A., & Bacharier, L. B. (2016). Early-life respiratory infections and asthma development: role in disease pathogenesis and potential targets for disease prevention. *Current Opinion in Allergy and Clinical Immunology*, *16*(2), 172–178.

Cattoir, L., Vankeerberghen, A., Boel, A., Van Vaerenbergh, K., & De Beenhouwer, H. (2019). Epidemiology of RSV and hMPV in Belgium: a 10-year follow-up. *Acta Clinica Belgica*, *74*(4), 229–235.

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. *Journal of Clinical Microbiology*, *48*(9), 3423–3427.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

DeVincenzo, J. P., El Saleeby, C. M., & Bush, A. J. (2005). Respiratory syncytial virus load predicts disease severity in previously healthy infants. *The Journal of Infectious Diseases*, *191*(11), 1861–1868.

Di Giallonardo, F., Kok, J., Fernandez, M., Carter, I., Geoghegan, J. L., Dwyer, D. E., Holmes, E. C., & Eden, J.-S. (2018). Evolution of Human Respiratory Syncytial Virus (RSV) over Multiple Seasons in New South Wales, Australia. *Viruses*, *10*(9). https://doi.org/10.3390/v10090476

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with

437 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973.
438 Eshaghi, A., Duvvuri, V. R., Lai, R., Nadarajah, J. T., Li, A., Patel, S. N., Low, D. E., & Gubbay,
439 J. B. (2012). Genetic variability of human respiratory syncytial virus A strains circulating in
440 Ontario: a novel genotype with a 72 nucleotide G gene duplication. *PloS One*, *7*(3), e32807.
441 Geoghegan, J. L., Saavedra, A. F., Duchêne, S., Sullivan, S., Barr, I., & Holmes, E. C. (2018).
442 Continental synchronicity of human influenza virus epidemics despite climactic variation.
443 *PLoS Pathogens*, *14*(1), e1006780.
444 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X.,
445 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
446 Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011).
447 Full-length transcriptome assembly from RNA-Seq data without a reference genome.
448 *Nature Biotechnology*, *29*(7), 644–652.
449 Griffiths, C., Drews, S. J., & Marchant, D. J. (2017). Respiratory Syncytial Virus: Infection,
450 Detection, and New Options for Prevention and Treatment. *Clinical Microbiology Reviews*,
451 *30*(1), 277–319.
452 Hall, C. B., Weinberg, G. A., Iwane, M. K., Blumkin, A. K., Edwards, K. M., Staat, M. A., Auinger,
453 P., Griffin, M. R., Poehling, K. A., Erdman, D., Grijalva, C. G., Zhu, Y., & Szilagyi, P. (2009).
454 The burden of respiratory syncytial virus infection in young children. *The New England*
455 *Journal of Medicine*, *360*(6), 588–598.
456 Henrickson, K. J., & Hall, C. B. (2007). Diagnostic assays for respiratory syncytial virus disease.
457 *The Pediatric Infectious Disease Journal*, *26*(11 Suppl), S36–S40.
458 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
459 improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–
460 780.
461 Luo, H.-J., Huang, X.-B., Zhong, H.-L., Ye, C.-X., Tan, X., Zhou, K., Yuan, L., Zhang, S.-F., Zhu,
462 X., Lin, C.-J., Wang, W.-J., Xu, L., & Cao, K.-Y. (2020). Epidemiological characteristics and
463 phylogenic analysis of human respiratory syncytial virus in patients with respiratory
464 infections during 2011-2016 in southern China. *International Journal of Infectious Diseases:*
465 *IJID: Official Publication of the International Society for Infectious Diseases*, *90*, 5–17.
466 Mahony, J., Chong, S., Merante, F., Yaghoubian, S., Sinha, T., Lisle, C., & Janeczko, R. (2007).
467 Development of a respiratory virus panel test for detection of twenty human respiratory
468 viruses by use of multiplex PCR and a fluid microbead-based assay. *Journal of Clinical*
469 *Microbiology*, *45*(9), 2965–2970.
470 Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and

analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), vev003.

Otieno, J. R., Kamau, E. M., Agoti, C. N., Lewa, C., Otieno, G., Bett, A., Ngama, M., Cane, P. A., & Nokes, D. J. (2017). Spread and Evolution of Respiratory Syncytial Virus A Genotype ON1, Coastal Kenya, 2010-2015. *Emerging Infectious Diseases*, *23*(2), 264–271.

Otieno, J. R., Kamau, E. M., Oketch, J. W., Ngoi, J. M., Gichuki, A. M., Binter, Š., Otieno, G. P., Ngama, M., Agoti, C. N., Cane, P. A., Kellam, P., Cotten, M., Lemey, P., & Nokes, D. J. (2018). Whole genome analysis of local Kenyan and global sequences unravels the epidemiological and molecular evolutionary dynamics of RSV genotype ON1 strains. *Virus Evolution*, *4*(2), vey027.

Pangesti, K. N. A., Abd El Ghany, M., Walsh, M. G., Kesson, A. M., & Hill-Cawthorne, G. A. (2018). Molecular epidemiology of respiratory syncytial virus. *Reviews in Medical Virology*, *28*(2). https://doi.org/10.1002/rmv.1968

Park, E., Park, P. H., Huh, J. W., Yun, H. J., Lee, H. K., Yoon, M. H., Lee, S., & Ko, G. (2017). Molecular and clinical characterization of human respiratory syncytial virus in South Korea between 2009 and 2014. *Epidemiology and Infection*, *145*(15), 3226–3242.

Parker, J., Rambaut, A., & Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *8*(3), 239–246.

Pretorius, M. A., van Niekerk, S., & Tempia, S. (2013). Replacement and positive evolution of subtype A and B respiratory syncytial virus G-protein genotypes from 1997–2012 in South Africa. *The Journal of*. https://academic.oup.com/jid/article-abstract/208/suppl_3/S227/789685

Ramaekers, K., Rector, A., Cuypers, L., Lemey, P., Keyaerts, E., & Van Ranst, M. (2020). Towards a unified classification for human respiratory syncytial virus genotypes. *Virus Evolution*, *6*(2), veaa052.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature*, *453*(7195), 615–619.

Razanajatovo Rahombanjanahary, N. H., Rybkina, K., Randriambolamanantsoa, T. H., Razafimanjato, H., & Heraud, J.-M. (2020). Genetic diversity and molecular epidemiology of respiratory syncytial virus circulated in Antananarivo, Madagascar, from 2011 to 2017: Predominance of ON1 and BA9 genotypes. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, *129*, 104506.

Shi, T., McAllister, D. A., O'Brien, K. L., Simoes, E. A. F., Madhi, S. A., Gessner, B. D., Polack,

505 F. P., Balsells, E., Acacio, S., Aguayo, C., Alassani, I., Ali, A., Antonio, M., Awasthi, S.,
506 Awori, J. O., Azziz-Baumgartner, E., Baggett, H. C., Baillie, V. L., Balmaseda, A., … RSV
507 Global Epidemiology Network. (2017). Global, regional, and national disease burden
508 estimates of acute lower respiratory infections due to respiratory syncytial virus in young
509 children in 2015: a systematic review and modelling study. *The Lancet*, *390*(10098), 946–
510 958.
511 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of
512 large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.
513 Tabatabai, J., Prifert, C., Pfeil, J., Grulich-Henn, J., & Schnitzler, P. (2014). Novel respiratory
514 syncytial virus (RSV) genotype ON1 predominates in Germany during winter season 2012-
515 13. *PloS One*, *9*(10), e109191.
516 Tal, G., Mandelberg, A., Dalal, I., Cesar, K., Somekh, E., Tal, A., Oron, A., Itskovich, S., Ballin,
517 A., Houri, S., Beigelman, A., Lider, O., Rechavi, G., & Amariglio, N. (2004). Association
518 between common Toll-like receptor 4 mutations and severe respiratory syncytial virus
519 disease. *The Journal of Infectious Diseases*, *189*(11), 2057–2063.
520 Vandini, S., Biagi, C., & Lanari, M. (2017). Respiratory Syncytial Virus: The Influence of
521 Serotype and Genotype Variability on Clinical Course of Infection. *International Journal of*
522 *Molecular Sciences*, *18*(8). https://doi.org/10.3390/ijms18081717
523 Vijaykrishna, D., Holmes, E. C., Joseph, U., Fourment, M., Su, Y. C. F., Halpin, R., Lee, R. T.
524 C., Deng, Y.-M., Gunalan, V., Lin, X., Stockwell, T. B., Fedorova, N. B., Zhou, B., Spirason,
525 N., Kühnert, D., Bošková, V., Stadler, T., Costa, A.-M., Dwyer, D. E., … Barr, I. G. (2015).
526 The contrasting phylodynamics of human influenza B viruses. *eLife*, *4*, e05055.
527 Yeoh, D. K., Foley, D. A., Minney-Smith, C. A., Martin, A. C., Mace, A. O., Sikazwe, C. T., Le,
528 H., Levy, A., Blyth, C. C., & Moore, H. C. (2020). The impact of COVID-19 public health
529 measures on detections of influenza and respiratory syncytial virus in children during the
530 2020 Australian winter. *Clinical Infectious Diseases: An Official Publication of the Infectious*
531 *Diseases Society of America*. https://doi.org/10.1093/cid/ciaa1475
532 Yun, K. W., Choi, E. H., & Lee, H. J. (2020). Molecular epidemiology of respiratory syncytial
533 virus for 28 consecutive seasons (1990-2018) and genetic variability of the duplication
534 region in the G gene of genotypes ON1 and BA in South Korea. In *Archives of Virology*
535 (Vol. 165, Issue 5, pp. 1069–1077). https://doi.org/10.1007/s00705-020-04580-z
536 Zou, L., Yi, L., Wu, J., Song, Y., Huang, G., Zhang, X., Liang, L., Ni, H., Pybus, O. G., Ke, C., &
537 Lu, J. (2016). Evolution and Transmission of Respiratory Syncytial Group A (RSV-A)
538 Viruses in Guangdong, China 2008–2015. *Frontiers in Microbiology*, *7*, 1263.

539