

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,*}

¹NSW Health Pathology-Institute for Clinical Pathology and Medical Research, NSW Health Pathology, Redbank Road, Westmead Hospital, Westmead, NSW 2145, Australia, ²Centre for Virus Research, Westmead Institute for Medical Research, 176 Hawkesbury Road, Westmead, NSW 2145, Australia, ³Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia, ⁴Department of Microbiology, PathWest Laboratory Medicine WA, Hospital Ave, Nedlands, WA 6009, Australia, ⁵School of Biological Sciences, The University of Western Australia, Crawley, WA 6009, Australia and ⁶School of Medicine, The University of Western Australia, Crawley, WA 6009, Australia

[†]<https://doi.org/0000-0003-1374-3551>

[†]These authors contributed equally to this work.

*Corresponding authors: E-mail: js.eden@sydney.edu.au; jen.kok@health.nsw.gov.au

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 exceed 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.

Key words: respiratory syncytial virus; molecular epidemiology; Australia; whole-genome sequencing; phylogenetics

1. 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 (Hall et al., 2009; Griffiths, Drews, and Marchant 2017). Importantly, RSV infection in young children may also lead to long-term sequelae such as asthma, chronic bronchitis, and obstructive pulmonary disease (Beigelman and Bacharier 2016; Griffiths, Drews, and Marchant 2017). Other populations particularly impacted by RSV include those over 65 years and immunosuppressed patients, such as solid organ and bone marrow transplant recipients (Beigelman and Bacharier 2016; Griffiths, Drews, and Marchant 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 and 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 and 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 (Henrickson and Hall 2007; Di Giallonardo et al., 2018; 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 2, and whilst primary infections are typically more severe, they are not protective against repeated infection (Griffiths, Drews, and Marchant 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, Biagi, and Lanari 2017); however, these are complicated by both host (Tal et al., 2004) and viral factors (DeVincenzo, El Saleeby, and Bush 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 (Otiene et al., 2017; 2018; Pangesti et al., 2018; Park et al., 2017; Cattoir et al., 2019). 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 (Eshaghi et al., 2012; Pretorius, Van Niekerk, and Tempia 2013; Di Giallonardo et al., 2018). 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 (Dapat et al., 2010; Agoti et al., 2014; Di Giallonardo et al., 2018). The added resolution from WGS has been particularly useful for elucidating transmission networks at both local (Agoti et al., 2015a) 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.

2. Materials and methods

2.1 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 reverse transcription polymerase chain reaction (RT-PCR) cycle threshold <30 in the 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).

2.2 Whole-genome sequencing

A previously published approach was used to amplify both RSV-A and RSV-B genomes (Di Giallonardo et al., 2018). In short, RT-PCR was used to amplify four overlapping amplicons (each ~ 4 kb) 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>; last accessed 11 April 2021) 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 BMap (<https://sourceforge.net/projects/bbmap/>; last accessed 11 April 2021) to check the assembly, and the final majority consensus was extracted for each sample.

2.3 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 Australia and WA, respectively, during the period from January 2016 to June 2017. All WA data, as well as NSW data from the year 2017, were 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 (Kato and 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 data sets

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

Age (years)	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%)			

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 1,000 bootstrap replications.

2.4 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, Rambaut, and Pybus 2008), with 1,000 replicates to analyse the genome data sets. Using BaTS, we determined the parsimony score (PS), association index (AI), and maximum clade size statistics for the location associated with each sequence, specifically focusing on the comparison between eastern and WA. 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 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 per cent 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 was generated with heights scaled to mean values.

2.5 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 data sets, for monophyletic groups containing near-identical sequences (~99.9 per cent nucleotide identity, less than 10 base pairs different across the genome) sampled up to 2 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.

2.6 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 (last accessed 17 April 2021).

3. Results and discussion

3.1 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 Australia (NSW) and WA Australia during the period from 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 of age (51.6 per cent, $n = 65/126$), with a further 20.6 per cent 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 per cent, $n = 64/103$ from patients ≤5 years of age) and the elderly (17.5 per cent,

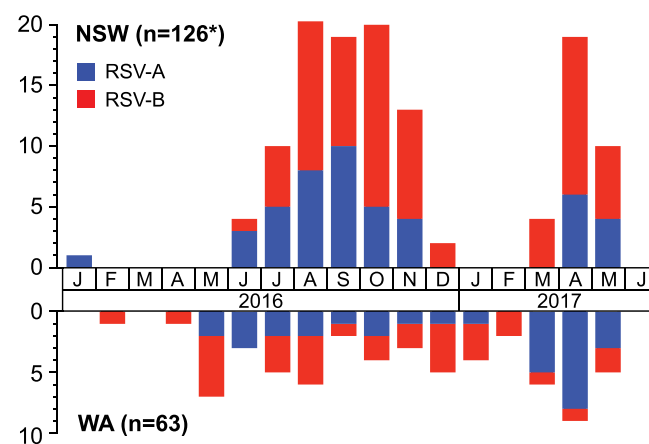


Figure 1. WGS of RSV in WA and NSW between January 2016 and June 2017 by month. The y-axis represents counts of genomes sequenced. The x-axis represents months when the samples were collected, with year shown underneath. RSV-A and RSV-B subtypes have been coloured blue and red respectively, as per the 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 distribution of RSV in WA. The asterisk indicates that of the 126 NSW genomes, 93 genomes from 2016 were generated in a previous study (Di Giallonardo et al., 2018).

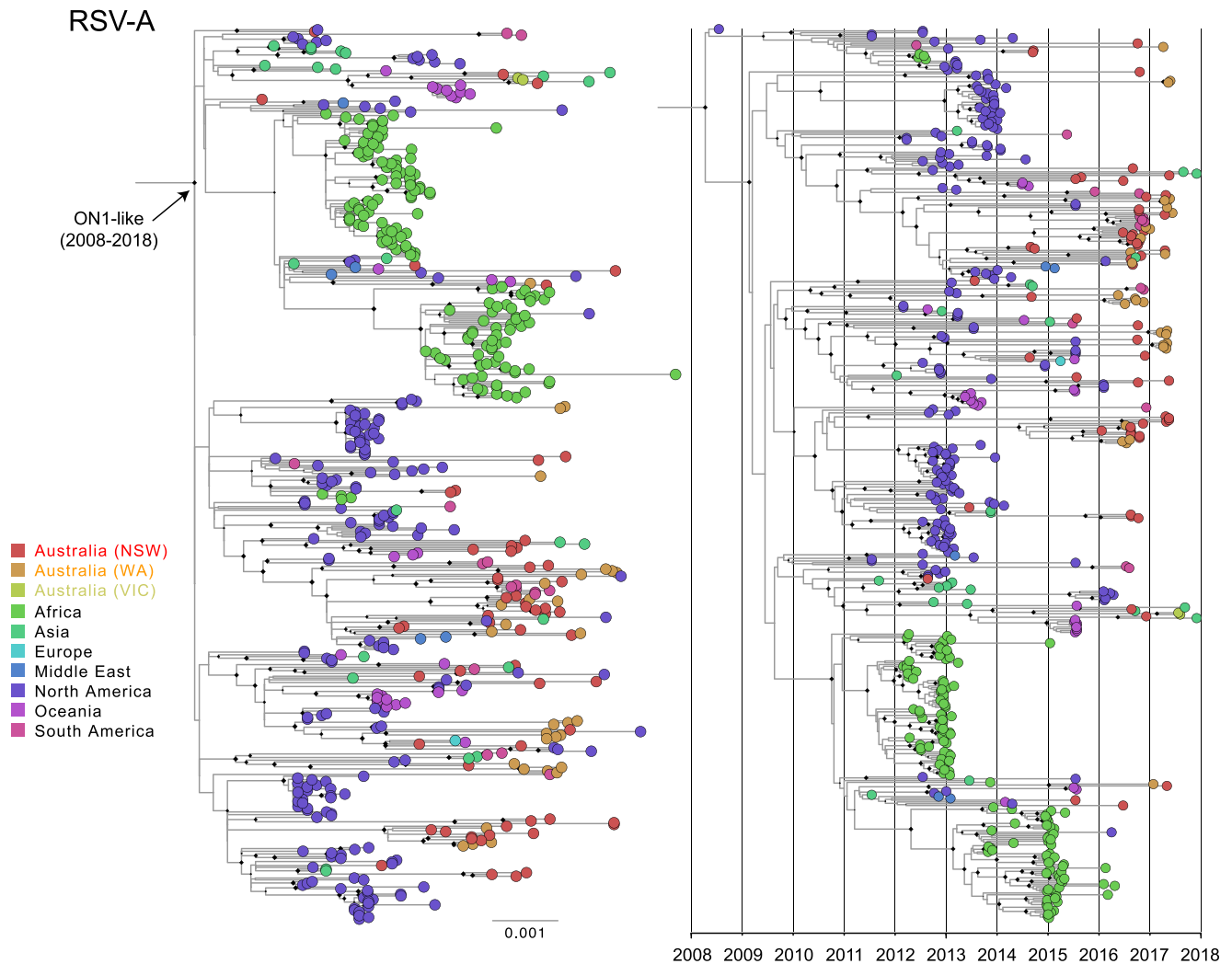


Figure 2. Phylogenetic analysis of RSV-A strains circulating in WA and NSW between January 2016 and June 2017. The global phylogenies were first estimated using alignments of complete genome sequences of RSV-A ($n = 1,190$) strains circulating since the late 1970s; however for clarity, only the recently 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 ML approach, whilst 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. Diamonds at nodal positions indicate branching support with bootstrap replicate values >70 per cent 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, whilst the right is scaled to time (years).

$n = 18/103$ 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).

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 (Fig. 1). In NSW, there was a higher proportion of RSV-B strains (63 per cent, $n = 80/126$), which was consistent across the entire study period. Furthermore, 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–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. Whilst our sampling for WA was intentionally even with regard to time, subtype, age, and setting, the predominance of RSV-A during 2017 was observed by the initial diagnostic testing where the assay 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, Choi, and Lee 2020).

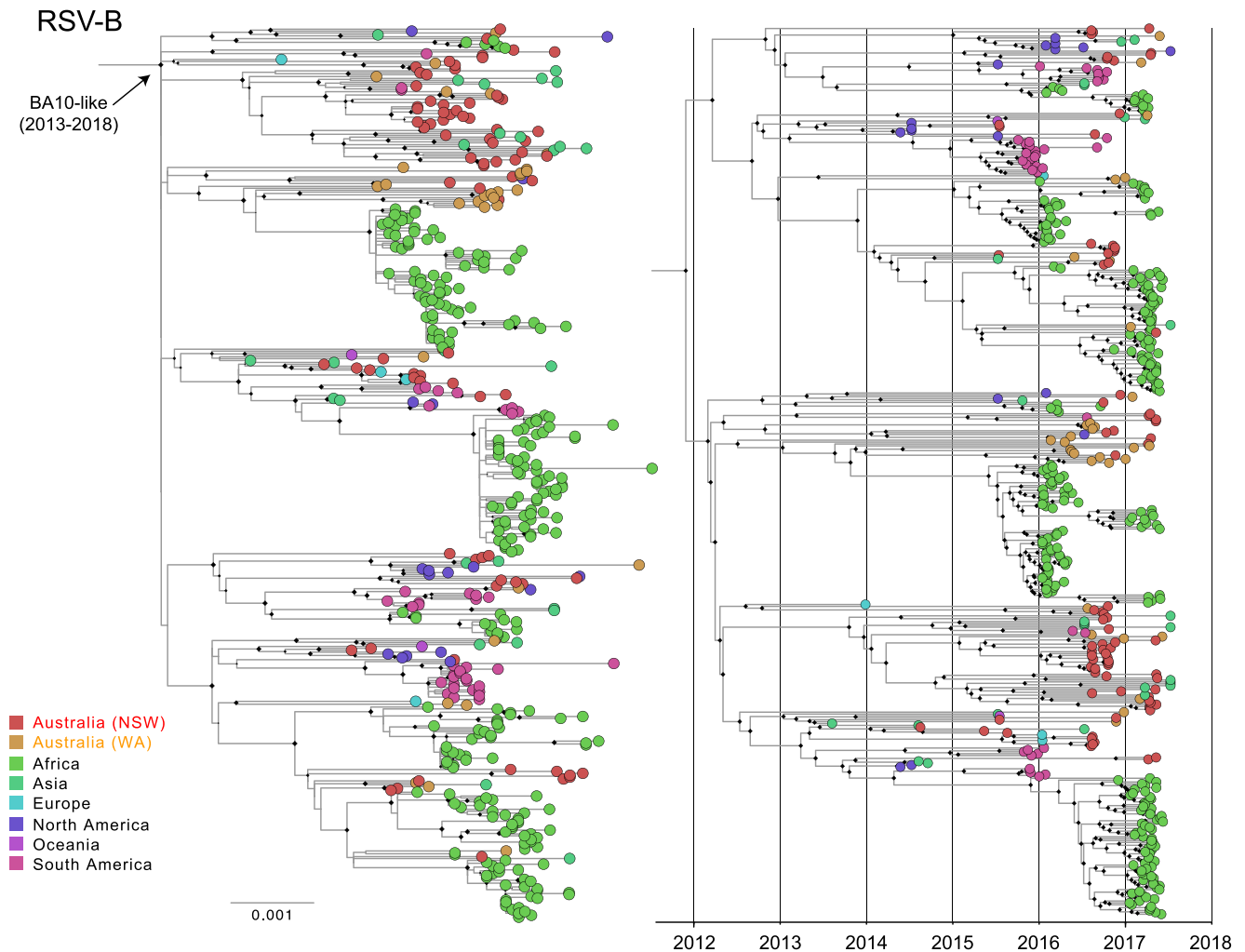


Figure 3. Phylogenetic analysis of RSV-B circulating in WA and NSW between January 2016 and June 2017. The global phylogenies were first estimated using alignments of complete genome sequences of RSV-B ($n = 1,121$) strains circulating since the 1960s; however, for clarity, only the recently circulating 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 ML approach, whilst 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. Diamonds at nodal positions indicate branching support with bootstrap replicate values >70 per cent 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, whilst the right is scaled to time (years).

3.2 Phylodynamics of RSV infections in NSW and WA in a global context

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

Australia, in both NSW and WA that this has been the major RSV-B genotype since at least 2014 (Di Giallonardo et al., 2018).

Across both RSV-A and RSV-B phylogenies, there does not appear to be strict spatial clustering when comparing viruses from NSW and WA (Figs 2 and 3), that is, the viruses from WA and NSW do not form monophyletic groups. Such monophyletic clustering would not be expected based on what is known for other respiratory pathogens—such as influenza—which are characterised by high levels of gene flow and viral traffic at global scales (Rambaut et al., 2008; Vijaykrishna et al., 2015). As such, we observed multiple co-circulating sub-lineages distributed across the entire diversity of RSV-A ON1-like and RSV-B BA10-like viruses (Figs 2 and 3). However, within these sub-lineages, the viruses often formed clusters based on sampling location; therefore, local spatial structure was apparent in our analysis. To explore this formally, we conducted a clustering analysis with BaTS and a posterior set of trees estimated using a time-scaled phylogenetic analysis in BEAST. We limited the analysis to the RSV-A ON1-like

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 ^a	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 ^b	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

^aDetermined by AI and PS.^bNew Zealand only.

and RSV-B BA10-like viruses, and the maximum clade credibility tree for each RSV subtype was found to be congruent with the ML trees (Figs 2 and 3). Using the posterior set of trees, we then measured the degree of clustering based on sampling location (Table 2).

When sequences were grouped by broad geographical regions, specifically, continents and the two Australian states NSW and 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 (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 Figs S1 and S2) (Otieno et al., 2018). Our sampling of NSW and WA strains is similarly constrained, except that our sampling sites are both major diagnostics hubs that 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 data set derived from sequences of the G gene region only (Supplementary Figs S3–S5). Whilst 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-sampled partial genome (G gene) sequences available on NCBI GenBank (Supplementary Fig. S3). Regardless, as shown below, these data can shed light on the possible sources and global transmission networks for RSV.

3.3 Lineages and local clusters

A detailed examination of individual lineages (Supplementary Figs S1 and S2) 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 Figs S4 and S5) 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/14 August 2016 (MH760625) and PW3375488K/9 August 2016 (MW160759) (Supplementary Figs S1 and S4). This is consistent with other studies that have shown that local persistence makes a minimal contribution to the diversity of RSV strains in any given area (Agoti et al., 2015b; Zou et al., 2016). Rather, viral importation and global mixing of strains seem to be the major driver of RSV diversity and their sources. Despite the data suggesting limited east–west viral traffic across the Australian continent, the Sydney–Perth air route remains one of the busiest in the country, nor can we discount other regions within Australia such as the tropical North, where persistence of viral lineages may occur as important sources of viral diversity. Indeed, from our understanding of the influenza virus in Australia, the synchronised dissemination of viral strains across the country is driven by both multiple introductions from the global population and strong domestic connectivity (Geoghegan et al., 2018). Increasing the breadth of WGS data for RSV cases nationally would assist in confirming what domestic viral traffic networks exist, particularly between tropical and temperate regions, as well as between major population centres on the east coast such as Sydney and Melbourne.

Next, we examined genetic clusters defined by high genetic identity (<10-bp difference across the genome) and similar sampling periods (collection dates within 2 weeks) to consider what features such as patient age and sampling location define them. For RSV-A, we identified 8 clusters (Supplementary Fig. S1 and Table S1), and for RSV-B, we identified 10 clusters (Supplementary Fig. S2 and Table S2). In this study, 75 per cent ($n = 6/8$) and 80 per cent ($n = 8/10$) of RSV-A and RSV-B clusters, respectively, could be linked to common localities and/or institutions, including individual hospital wards or emergency departments. Whilst two of the RSV-B clusters were found to be repeat samples of the same patient (Supplementary Table S2), overwhelmingly, these clusters of genetically and temporal related sequences most likely represent fine-scale transmission events.

4. Conclusions

In summary, our analysis has identified a number of features of RSV epidemiology and patterns of spread, including (1) the co-circulation of both major subtypes—RSV-A and RSV-B, (2) a single genotype for each subtype predominates each season, (3) multiple distinct sub-lineages of each genotype will co-circulate and which are associated with regional and local clustering/outbreaks,

(4) little viral mixing across the east–west coasts of Australia despite apparent overall geographic clustering, (5) that genetically and temporal-related sequences most likely represent fine-scale transmission events such as institutional outbreaks, and (6) that whole-genome sequence is required and encouraged over partial G gene sequencing for elucidating clusters and transmission pathways. Taken together, this presents a complex phylogenetic pattern with globally circulating diversity with viral mixing across different regions, yet, finer-scale patterns revealing multiple endemic sub-lineages and clusters consistent with local transmission events and outbreaks. This highlights the connections of genomic data across multiple epidemiological scales and further strengthens the need for a much greater sampling of RSV, not just here in Australia, but globally.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

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 Institute for Clinical Pathology and Medical Research (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.

Conflict of interest: None declared.

References

- Agoti, C. N. et al. (2014) 'Rapid Spread and Diversification of Respiratory Syncytial Virus Genotype ON1, Kenya', *Emerging Infectious Diseases*, 20: 950–9.
- et al. (2015a) 'Local Evolutionary Patterns of Human Respiratory Syncytial Virus Derived from Whole-Genome Sequencing', *Journal of Virology*, 89: 3444–54.
- et al. (2015b) 'Successive Respiratory Syncytial Virus Epidemics in Local Populations Arise from Multiple Variant Introductions, Providing Insights into Virus Persistence', *Journal of Virology*, 89: 11630–42.
- Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.
- Beigelman, A., and 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: 172–8.
- Cattoir, L. et al. (2019) 'Epidemiology of RSV and hMPV in Belgium: A 10-Year Follow-up', *Acta Clinica Belgica*, 74: 229–35.
- Dapat, I. C. et al. (2010) 'New Genotypes within Respiratory Syncytial Virus Group B Genotype BA in Niigata, Japan', *Journal of Clinical Microbiology*, 48: 3423–7.
- Darriba, D. et al. (2012) 'jModelTest 2: More Models, New Heuristics and Parallel Computing', *Nature Methods*, 9: 772.
- DeVincenzo, J. P., El Saleeby, C. M., and Bush, A. J. (2005) 'Respiratory Syncytial Virus Load Predicts Disease Severity in Previously Healthy Infants', *The Journal of Infectious Diseases*, 191: 1861–8.
- Di Giallonardo, F. et al. (2018) 'Evolution of Human Respiratory Syncytial Virus (RSV) over Multiple Seasons in New South Wales, Australia', *Viruses*, 10: 476.
- Drummond, A. J. et al. (2012) 'Bayesian Phylogenetics with BEAUti and the BEAST 1.7', *Molecular Biology and Evolution*, 29: 1969–73.
- Eshaghi, A. et al. (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: e32807.
- Geoghegan, J. L. et al. (2018) 'Continental Synchronicity of Human Influenza Virus Epidemics despite Climatic Variation', *PLoS Pathogens*, 14: e1006780.
- Grabherr, M. G. et al. (2011) 'Full-length Transcriptome Assembly from RNA-Seq Data without a Reference Genome', *Nature Biotechnology*, 29: 644–52.
- Griffiths, C., Drews, S. J., and Marchant, D. J. (2017) 'Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment', *Clinical Microbiology Reviews*, 30: 277–319.
- Hall, C. B. et al. (2009) 'The Burden of Respiratory Syncytial Virus Infection in Young Children', *The New England Journal of Medicine*, 360: 588–98.
- Henrickson, K. J., and Hall, C. B. (2007) 'Diagnostic Assays for Respiratory Syncytial Virus Disease', *The Pediatric Infectious Disease Journal*, 26: S36–40.
- Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.
- Luo, H.-J. et al. (2020) 'Epidemiological Characteristics and Phylogenetic Analysis of Human Respiratory Syncytial Virus in Patients with Respiratory Infections during 2011–2016 in Southern China', *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, 90: 5–17.
- Mahony, J. et al. (2007) 'Development of a Respiratory Virus Panel Test for Detection of Twenty Human Respiratory Viruses by Use of Multiplex PCR and a Fluid Microbead-based Assay', *Journal of Clinical Microbiology*, 45: 2965–70.
- Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.
- Otieno, J. R. et al. (2017) 'Spread and Evolution of Respiratory Syncytial Virus A Genotype ON1, Coastal Kenya, 2010–2015', *Emerging Infectious Diseases*, 23: 264–71.
- et al. (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: vey027.
- Pangesti, K. N. A. et al. (2018) 'Molecular Epidemiology of Respiratory Syncytial Virus', *Reviews in Medical Virology*, 28: e1968.
- Park, E. et al. (2017) 'Molecular and Clinical Characterization of Human Respiratory Syncytial Virus in South Korea between 2009 and 2014', *Epidemiology and Infection*, 145: 3226–42.
- Parker, J., Rambaut, A., and 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: 239–46.
- Pretorius, M. A., van Niekerk, S., and 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 Infectious Diseases*. <https://academic.oup.com/jid/article-abstract/208/suppl_3/S227/789685> Accessed 21 Jul 2021.

- Ramaekers, K. et al. (2020) 'Towards a Unified Classification for Human Respiratory Syncytial Virus Genotypes', *Virus Evolution*, 6: veaa052.
- Rambaut, A. et al. (2008) 'The Genomic and Epidemiological Dynamics of Human Influenza A Virus', *Nature*, 453: 615–9.
- Razanajatovo Rahombanjanahary, N. H. et al. (2020) 'Genetic Diversity and Molecular Epidemiology of Respiratory Syncytial Virus Circulated in Antananarivo, Madagascar, from 2011 to 2017: Prevalence of ON1 and BA9 Genotypes', *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 129: 104506.
- Shi, T. et al. (2017) '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', *The Lancet*, 390: 946–58.
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies', *Bioinformatics*, 30: 1312–3.
- Tabatabai, J. et al. (2014) 'Novel Respiratory Syncytial Virus (RSV) Genotype ON1 Predominates in Germany during Winter Season 2012–13', *PLoS One*, 9: e109191.
- Tal, G. et al. (2004) 'Association between Common Toll-like Receptor 4 Mutations and Severe Respiratory Syncytial Virus Disease', *The Journal of Infectious Diseases*, 189: 2057–63.
- Vandini, S., Biagi, C., and Lanari, M. (2017) 'Respiratory Syncytial Virus: The Influence of Serotype and Genotype Variability on Clinical Course of Infection', *International Journal of Molecular Sciences*, 18: 1717.
- Vijaykrishna, D. et al. (2015) 'The Contrasting Phylodynamics of Human Influenza B Viruses', *eLife*, 4: e05055.
- Yeoh, D. K. et al. (2020) 'The Impact of COVID-19 Public Health Measures on Detections of Influenza and Respiratory Syncytial Virus in Children during the 2020 Australian Winter', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 72: 2199–202.
- Yun, K. W., Choi, E. H., and Lee, H. J. (2020) 'Molecular Epidemiology of Respiratory Syncytial Virus for 28 Consecutive Seasons (1990–2018) and Genetic Variability of the Duplication Region in the G Gene of Genotypes ON1 and BA in South Korea', *Archives of Virology*, 165: 1069–77.
- Zou, L. et al. (2016) 'Evolution and Transmission of Respiratory Syncytial Group A (RSV-A) Viruses in Guangdong, China 2008–2015', *Frontiers in Microbiology*, 7: 1263.