[Type of the Paper (Article)

An amplicon-based approach for the whole genome sequencing of human metapneumovirus

Rachel L. Tulloch[1,2], Jen Kok[3], Ian Carter[3], Dominic E. Dwyer[2,3] and John-Sebastian Eden[1,2,*]

[1] Centre for Virus Research, Westmead Institute for Medical Research, Westmead, NSW 2145, Australia; rachel.tulloch@sydney.edu.au (R.L.T.); js.eden@sydney.edu.au (J.-S.E.)

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

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

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

Human metapneumovirus (HMPV) is an important cause of upper and lower respiratory tract disease in individuals of all ages. It is estimated that most individuals will be infected by HMPV by the age of 5 years old. Despite this burden of disease, there remains caveats in our knowledge of virus global genetic diversity due to a lack of HMPV sequencing, particularly at whole genome scale. The purpose of this study was to create a simple and robust approach for HMPV whole genome sequencing to be used for genomic epidemiological studies. To design our assay, all available HMPV full length genome sequences were downloaded from the NCBI GenBank database and used to design four primer sets to amplify long, overlapping amplicons spanning the viral genome and, importantly, specific to all known HMPV subtypes. These amplicons were then pooled and sequenced on an Illumina iSeq; however the approach is suitable to other common NGS platforms. We demonstrate the utility of this method using a representative subset of clinical samples and examine these sequences using a phylogenetic approach. Here we present an amplicon-based method for the whole genome sequencing of HMPV from clinical extracts that can be used to better inform genomic studies of HMPV epidemiology and evolution.

Keywords: human metapneumovirus; whole genome sequencing; genomic epidemiology

1. Introduction

Since its discovery in 2001, human metapneumovirus (HMPV) has become increasingly recognised as a major cause of acute respiratory illness (ARI), globally [1]. Serological studies estimate that almost all individuals will be exposed to HMPV by the age of 5 [2]. Clinically, HMPV is indistinguishable from ARI caused by other respiratory pathogens, including respiratory syncytial virus (RSV) [3]. While most infections are mild and self-limiting, HMPV has increased morbidity and mortality in high risk populations, including immunosuppressed individuals, young children and the elderly [4]. Globally, HMPV is associated with 3.9-7% of children hospitalised with lower respiratory tract infections (LRTI) with outpatient detection rates ranging from 6.2-12%,

highlighting its clinical significance as a cause of ARI in this age group [3,5-7]. HMPV is also a known cause of ARI in hospitalized adults with one study showing that the hospitalisation rates of adults >50 years of age were statistically similar to those with influenza infections in the same region [8].

HMPV is a member of *Pneumoviridae* family, with a negative-sense, single-stranded RNA genome of approximately 13.3kb in length [9,10]. HMPV is genetically similar to RSV, however it lacks two non-structural genes – NS1 and NS2. Phylogenetic analysis of the HMPV fusion (F) and glycoprotein (G) genes has led to the identification and classification of viruses into two major subtypes, HMPV A and HMPV B, which can further be subdivided into A1, A2a, A2b, B1 and B2 sublineages [11]. Epidemiological studies have revealed the cocirculation of HMPV subtypes globally, with the predominant subtype fluctuating throughout the year [12]. Historically, HMPV molecular epidemiological studies have relied on the subgenomic sequencing of partial F or G protein genes to perform genomic and evolutionary studies [13]. Indeed, only 2.3% (n=226/9795) of available sequences on the National Center for Biotechnology Information (NCBI) GenBank database are near or complete genomes (as of November 2020). Therefore, our understanding of the genomic epidemiology, genetic diversity and evolution of HPMV remains limited.

Whole genome sequencing (WGS) is a powerful tool in for public health infectious disease surveillance, it can also inform the treatment and control of viruses in the population [14,15]. WGS offers increased resolution at multiple epidemiological scales, from investigating global virus traffic networks to elucidating individual transmission events within outbreaks [15,16]. The recent SARS-CoV-2 epidemic has highlighted the utility of amplicon-based WGS methods as a cost effective, rapid method to sequence whole genomes approach [17-19]. The purpose of this study was to develop a simple and robust amplicon-based method for sequencing the HMPV full length genome with the aim to inform a better understanding of its molecular epidemiology.

2. Materials and Methods

2.1 Primer design

Our approach was based on an existing WGS workflow designed to amplify and sequence the RSV genome using four amplicons between 3528 nt and 4375 nt in length, each with an overlapping region of at least 100 nt [20]. Given the similar genome lengths between HPMV and RSV, we focused on designing HMPV specific primer sets that would also generate four overlapping amplicons of ~3.5kb each that span the viral genome. To include historical and current circulating HMPV subtypes, all available full or near full length (>1300kb) HMPV genome sequences were obtained from the NCBI GenBank database (downloaded 10th December 2019) with all sequence analysis was conducted in using Geneious Prime 2019.2.3. Sequences were excluded from the initial analysis if they contained > 20 continuous ambiguous bases. This result was a final set of 153 full or near full-length HMPV sequences which were then aligned using MAFFT v7.45 [21]. Phylogenetic analysis using a neighbour joining approach was performed to show that all known HMPV subtypes were represented in the subset of sequences (Figure 1).
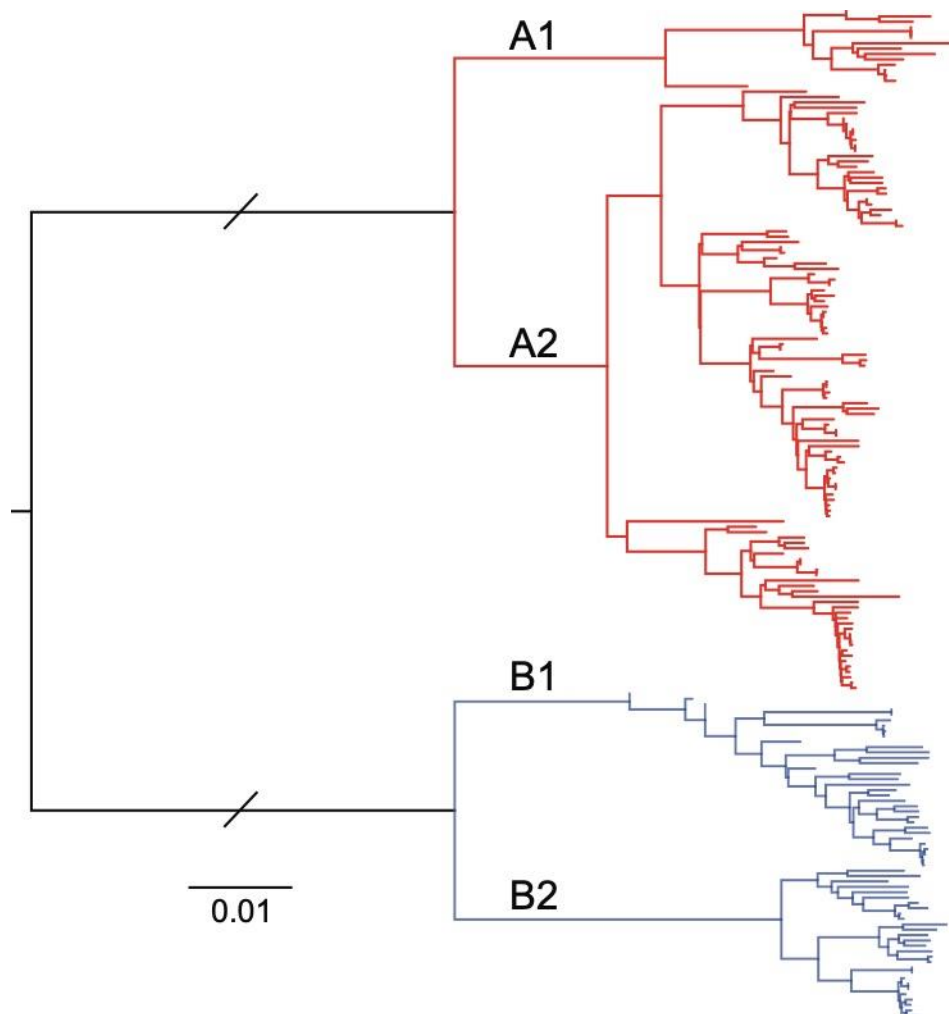
Figure 1 – Neighbor joining phylogenetic tree of 153 HMPV full length or near full length genome sequences sourced from NCBI GenBank database. The two major phylogenetic subtypes of HMPV are shown with different colored branches, with known sub-lineages further differentiated by labelled branches (A1, A2, B1 and B2). The scale bar is proportional to the number of substitutions per site. The branch lengths between the A and B subtypes have been shorten for clarity.

A sliding window approach was then used to plot to sequence identity along the viral genome alignment and identify conserved regions for targeted primer design. These primers were designed to be ~25bp in length with degeneracies where necessary to capture position variation between HMPV subtypes, as well as to have melting temperatures within 5°C of each-other and avoiding potential dimer formations [22]. We also tested an existing published primer set designed to amplify the HPMV genome using a similar overlapping amplicon scheme (~3.5kb each) from a study examining local virus traffic and genetic diversity of HMPV in Peru [23]. The final set of primer pairs were designed to be run as four separate polymerase chain reactions (PCRs) performed in parallel (Table 1) and were tested on a selection of known viral extracts from HMPV-positive clinical specimens.

Table 1 – Primer pairs used in this study to complete near whole genome sequencing of human metapneumovirus.

Assay	Primer Name	Sequence (5' - 3')	Position (nt)*	PCR amplicon size (bp)
PCR1	HMPV1_F	GGGACAAATAAAAATGTCTCTTCA	43	4148
	HMPV1_R	CTTCCTGTGCTRACYTTRCA	4154	
PCR2	HMPV2_F	ACAGCAGCRGGRATYAATGT	4082	4077
	HMPV2_R	TAGTACTGAAYTGAGCATGYTCAG	8158	
PCR3	HMPV3_F	AACTGTTAACATGGAAAGATGTGATG	7897	3229
	HMPV3_R	TAAGCTGGAACWGAWGCTG	11125	
PCR4	HMPV4_F	TCAATAGGGAGTCTRTGTCARGAA	9804	3685
	HMPV4_R	GRCAAAAAAACCGTATACATYC	13488	

* Nucleotide positions are relative to strain B2 (NCBI GenBank Accession: MF045425)

2.2 Clinical Specimens and HMPV Isolation

Residual total nucleic acid from respiratory specimens submitted to NSW Health Pathology - Institute of Clinical Pathology & Microbiology Research (ICPMR) for diagnostic testing were utilised in this study as per protocols approved by local ethics and governance committees (LNR/17/WMEAD/128 and SSA/17/WMEAD/129). Total nucleic acid was extracted from each submitted respiratory specimen using a high-throughput, magnetic bead-based extraction platform and screened against a panel of known respiratory viruses using an accredited multiplex quantitative PCR. Viruses included on the respiratory panel are Influenza A, Influenza A subtype H3N2, Influenza A H1N1 2009 Pandemic, Influenza B, Adenovirus, Parainfluenza 1,2 & 3, RSV, Rhinovirus, Enterovirus and HPMV. Archived clinical extracts which were reported as positive for HMPV were de-identified before inclusion in this study. No specific subtyping information was available on the archived samples, therefore we selected ten random HMPV-positive nucleic acid specimens collected over seven years between 2013 and 2020 to attempt to capture historic HMPV genetic diversity in NSW, Australia.

2.3 Reverse Transcription

Complementary DNA (cDNA) synthesis was performed in a 20 μL reaction containing 4 μL of 5X SuperScript IV VILO MasterMix (Invitrogen), 12 μL of nuclease-free water and 4 μL of viral RNA template. The thermal cycling protocol used was as follows, random priming was performed at 25°C for 10 minutes, followed by extension at 50°C for 20 minutes and then enzyme denaturation at 85°C for 5 minutes before holding at 4°C. All incubation steps were performed on an SimpliAmp thermocycler (Applied Biosystems). Viral cDNA was used immediately or stored at -80°C until required.

2.4 HMPV genome amplification

The viral cDNA was then split across four separate PCR reactions each amplifying one part of the HMPV genome (Table 1). Each PCR was performed in a 25 μL reaction containing 12.5 μL of 2X Platinum SuperFi MasterMix (Invitrogen,), 1.25 μL of 10 μM forward primer, 1.25 μL of 10 μM

reverse primer, 7 µL of nuclease-free water, and 3 µL of cDNA template. The reactions were then incubated at 98°C for 30 seconds, followed by 44 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 20 seconds and extension at 72°C for 2:10 minutes, with a final extension for 5 minutes at 72°C followed by holding at 4°C. Amplicon size and yield was assessed by gel electrophoresis of 5 µL of PCR reactions using an 1% E-Gel-48 Agarose Gel (Invitrogen) with 500 ng of 1 Kb Plus DNA Ladder. The gels are pre-stain with ethidium bromide for amplicon visualisation under UV light for approximate quantification and sizing. Amplicon approximate quantity was estimated using the target PCR product band intensity. To ensure even coverage across the HPMV genome, the 4 amplicons of each clinical sample were pooled equally based on target amplicon intensity into a final pooled volume of 40 µL (adjusted with Qiagen EB buffer where necessary). Where non-specific amplification was present, band intensity of only the target amplicon was taken into consideration and pooling was adjusted accordingly. The HMPV genome amplicon pools were purified using AMPure XP (Beckman Coulter) at a bead-to-sample ratio of 1X according to the manufacturer's protocol. The purified DNA was then quantified using the 1X double stranded DNA high sensitivity (1x dsDNA HS) Qubit assay (Invitrogen) and measured on the Qubit 4 fluorometer. The pooled amplicons were then volumetrically diluted to 0.2 ng/µL, the required input concentration for library preparation.

2.5 Library preparation and Sequencing

Amplicons were prepared for sequencing using the Nextera XT DNA Library Preparation kit with the v2 Set B indexing kit (Illumina). The manufacturer's protocol was followed for genomic DNA tagmentation, library amplification and clean-up except that all volumes were halved for reagent conservation. The purified DNA libraries were quantified using the 1x dsDNA HS Qubit assay and Qubit 4 fluorometer before normalization using Qubit DNA concentrations. The final library pool molarity and fragment length distribution was determined using the 4200 TapeStation System with a high sensitivity D5000 tape (Agilent) before dilution to 0.1 nM with Qiagen EB buffer for loading and sequencing on an Illumina iSeq 100 platform with a v1 300 cycle kit.

2.6 Viral assembly

Raw sequences were initially quality scored using fastqc [24] following this, reads were quality trimmed to a QC threshold of phred score 20 using bbduk [25]. The trimmed reads were then *de novo* assembled using Megahit with default parameters [26]. The trimmed reads were then re-mapped onto the draft genome using bbmap [25], before the overall the mapping alignment quality was assessed using the Geneious Prime 2019.2.3 and majority consensus genome extracted. The final genome was trimmed of terminal primer sequences and annotated using NCBI GenBank reference sequences.

### 2.7 Phylogenetic analysis

Phylogenetic analysis was performed by aligning all sequences generated in this study against a reference set obtained from NCBI GenBank using MAFFT and PhyML [27] with node supported estimated by 100 bootstrap replicates. Sequences obtained in this study were published to NCBI GenBank with the following accession IDs: MW221986-MW221995.

3. Results and Discussion

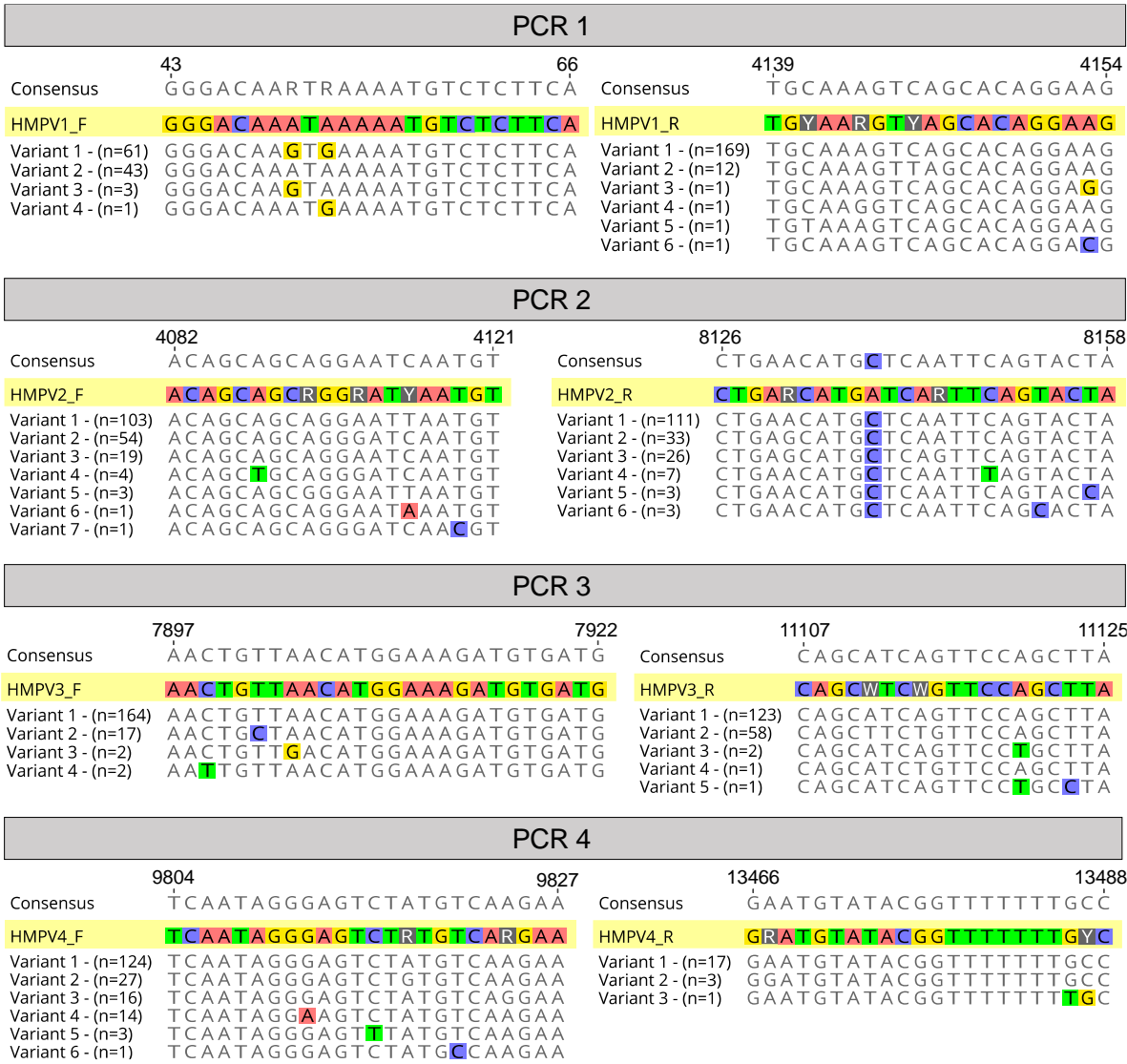*3.1 Designing primers to amplify the HMPV genome*

The aim of this study was the develop a simple and robust amplicon-based approach for amplifying and sequencing the HMPV genome. To do this, we adapted a previous approach used for RSV [20] to design four primer sets generating ~3.5kb amplicons that overlap and span the viral genome. Our primers were based on all available HMPV genomes from the NCBI GenBank database and targeted conserved regions at suitably spaced positions in the genome (Table 1 & Figure 2).
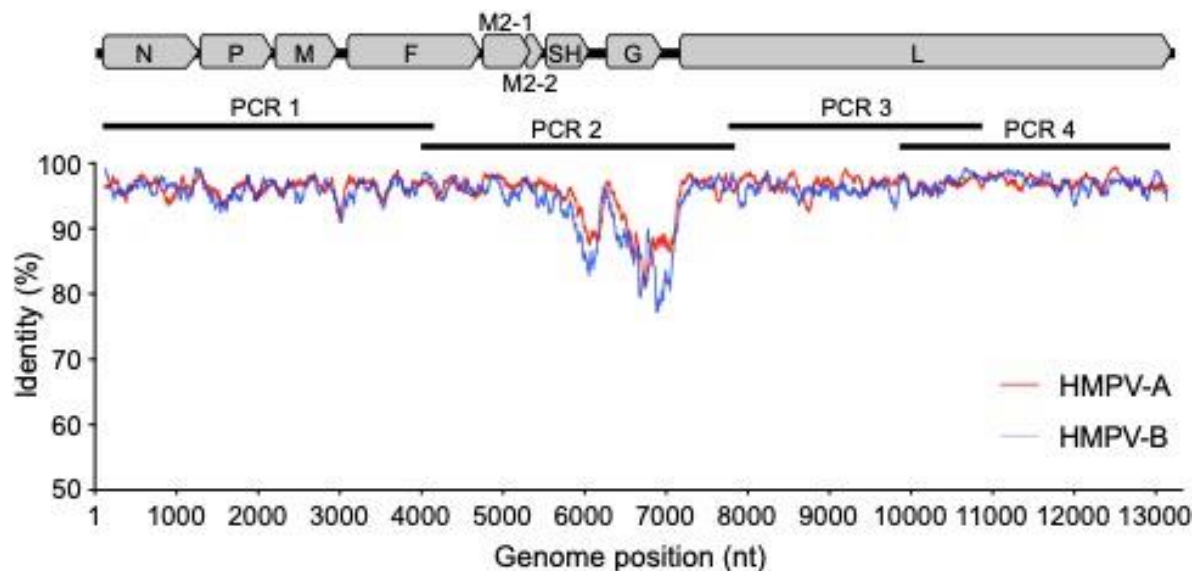
Figure 2 – Genetic variability across the HPMV genome. A graphic representation showing the percentage identity at each nucleotide position along the genome with each HMPV subtype plotted separately – A (red) & B (blue). Also included is an HMPV genome shown to scale with annotated genes to highlight primer and amplicon positions.

The final primer sets were located in the terminal regions of the genome, as well as in the fusion and large protein genes, and avoided divergent regions of the genome such as in the viral glycoprotein – G protein (Figure 2). Other amplicon-based methods for viral genome sequencing often employ shorter amplicon lengths (1kb or less) to improve performance for low viral load or low quality samples, such as with the ARTIC protocol for SARS-CoV-2 genome sequencing [19] or even for enteric virus including human norovirus [28]. Here we chose targeted amplicons to be between 3 – 4 kb based on previous performance against RSV [20], where genomes would reliably amplify from 80-90% of clinical samples. Furthermore, there is greater diversity present in HMPV compared to SARS-CoV-2 such that there are less suitable target positions across the genome to readily amplify all subtypes. Indeed, to capture this diversity a number of degenerate nucleotides were included in our primers, and based on our current understanding HMPV diversity, it would be expected our primers cover the vast majority of variants present in circulation (Figure 3).

PCR 1

```
            43                          66           4139              4154
Consensus   G G G A C A A R T A A A A T G T C T C T T C A   Consensus   T G C A A A G T C A G C A C A G G A A G

HMPV1_F     G G G A C A A A T A A A A A T G T C T C T T C A   HMPV1_R     T G Y A A R G T Y A G C A C A G G A A G
Variant 1 - (n=61)  G G G A C A A G T G A A A A T G T C T C T T C A   Variant 1 - (n=169)  T G C A A A G T C A G C A C A G G A A G
Variant 2 - (n=43)  G G G A C A A A T A A A A A T G T C T C T T C A   Variant 2 - (n=12)   T G C A A A G T T A G C A C A G G A A G
Variant 3 - (n=3)   G G G A C A A G T A A A A A T G T C T C T T C A   Variant 3 - (n=1)    T G C A A A G T C A G C A C A G G A G G
Variant 4 - (n=1)   G G G A C A A A T G A A A A T G T C T C T T C A   Variant 4 - (n=1)    T G C A A G G T C A G C A C A G G A A G
                                                                      Variant 5 - (n=1)    T G T A A A G T C A G C A C A G G A A G
                                                                      Variant 6 - (n=1)    T G C A A A G T C A G C A C A G G A C G
```

PCR 2

```
            4082                        4121          8126                          8158
Consensus   A C A G C A G C A G G A A T C A A T G T   Consensus   C T G A A C A T G C T C A A T T C A G T A C T A

HMPV2_F     A C A G C A G C R G G R A T Y A A T G T   HMPV2_R     C T G A R C A T G A T C A R T T C A G T A C T A
Variant 1 - (n=103)  A C A G C A G C A G G A A T T A A T G T   Variant 1 - (n=111)  C T G A A C A T G C T C A A T T C A G T A C T A
Variant 2 - (n=54)   A C A G C A G C A G G G A T C A A T G T   Variant 2 - (n=33)   C T G A G C A T G C T C A A G T T C A G T A C T A
Variant 3 - (n=19)   A C A G C A G C A G G A A T C A A T G T   Variant 3 - (n=26)   C T G A G C A T G C T C A G T T C A G T A C T A
Variant 4 - (n=4)    A C A G C T G C A G G G A T C A A T G T   Variant 4 - (n=7)    C T G A A C A T G C T C A A T T T A G T A C T A
Variant 5 - (n=3)    A C A G C A G C G G G A A T T A A T G T   Variant 5 - (n=3)    C T G A A C A T G C T C A A T T C A G T A C C A
Variant 6 - (n=1)    A C A G C A G C A G G A A T A A A T G T   Variant 6 - (n=3)    C T G A A C A T G C T C A A T T C A G C A C T A
Variant 7 - (n=1)    A C A G C A G C A G G G A T C A A C G T
```

PCR 3

```
            7897                                7922        11107              11125
Consensus   A A C T G T T A A C A T G G A A A G A T G T G A T G   Consensus   C A G C A T C A G T T C C A G C T T A

HMPV3_F     A A C T G T T A A C A T G G A A A G A T G T G A T G   HMPV3_R     C A G C W T C W G T T C C A G C T T A
Variant 1 - (n=164)  A A C T G T T A A C A T G G A A A G A T G T G A T G   Variant 1 - (n=123)  C A G C A T C A G T T C C A G C T T A
Variant 2 - (n=17)   A A C T G C T A A C A T G G A A A G A T G T G A T G   Variant 2 - (n=58)   C A G C T T C T G T T C C A G C T T A
Variant 3 - (n=2)    A A C T G T T G A C A T G G A A A G A T G T G A T G   Variant 3 - (n=2)    C A G C A T C A G T T C C T G C T T A
Variant 4 - (n=2)    A A T T G T T A A C A T G G A A A G A T G T G A T G   Variant 4 - (n=1)    C A G C A T C T G T T C C A G C T T A
                                                                          Variant 5 - (n=1)    C A G C A T C A G T T C C T G C C T A
```

PCR 4

```
            9804                        9827           13466                     13488
Consensus   T C A A T A G G G A G T C T A T G T C A A G A A   Consensus   G A A T G T A T A C G G T T T T T T T G C C

HMPV4_F     T C A A T A G G G A G T C T R T G T C A R G A A   HMPV4_R     G R A T G T A T A C G G T T T T T T T G Y C
Variant 1 - (n=124)  T C A A T A G G G A G T C T A T G T C A A G A A   Variant 1 - (n=17)  G A A T G T A T A C G G T T T T T T T G C C
Variant 2 - (n=27)   T C A A T A G G G A G T C T G T G T C A A G A A   Variant 2 - (n=3)   G G A T G T A T A C G G T T T T T T T G C C
Variant 3 - (n=16)   T C A A T A G G G A G T C T A T G T C A A G A A   Variant 3 - (n=1)   G A A T G T A T A C G G T T T T T T T T G C
Variant 4 - (n=14)   T C A A T A G G A A G T C T A T G T C A A G A A
Variant 5 - (n=3)    T C A A T A G G G A G T T T A T G T C A A G A A
Variant 6 - (n=1)    T C A A T A G G G A G T C T A T G C C A A G A A
```

Figure 3 – Primer alignments. A graphical representation of the four chosen primer pairs aligned against known variants at each binding site relative to strain B2 (NCBI GenBank Accession: MF045425)

3.3 HMPV genome RT-PCR performance

To examine the performance of our newly designed primers, we tested them against a set of HMPV positive extracts from clinical respiratory specimens. Since the subtype and sublineage classifications from our samples were unknown, we instead obtained samples across a wide time period (2013 to 2020) to attempt to capture a breadth of diversity. Initial end-point PCRs showed good levels of amplification of across the four targets (data not shown); however, we then attempted further optimization of the assay using a temperature gradient (59°C – 61.5°C) to establish the optimal annealing temperature was 60°C to ensure efficient target amplification and minimize non-specific amplification (Figure 4A). We also compared our primers to those previously published from the Peru WGS study (Figure 4B) and showed improved performance, particularly for the specific amplification of the targeted HMPV region; however, this may be partly due to our initial optimization of annealing and cycling conditions favoring our newly designed set.
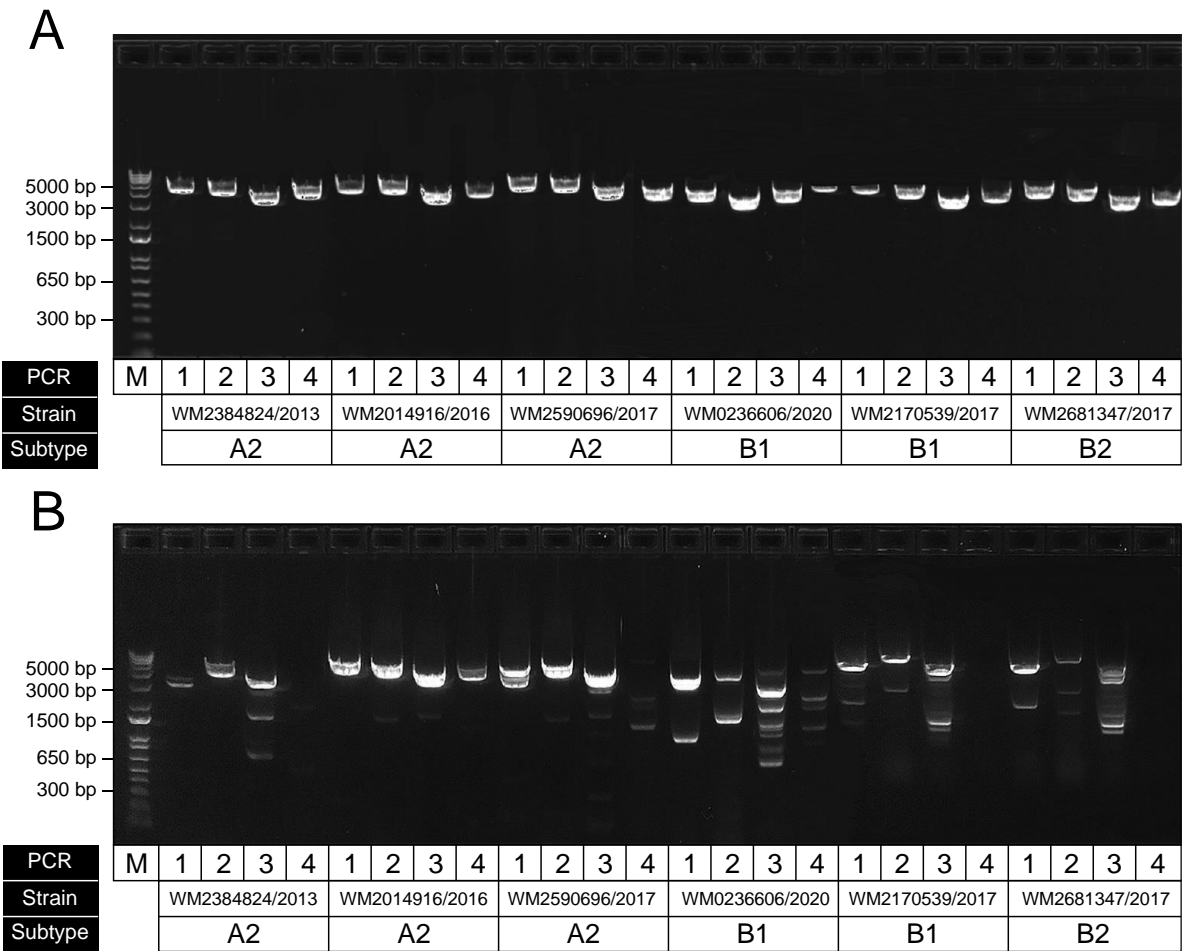
Figure 4 - Representative gel electrophoresis result showing RT-PCR performance of amplicons for HMPV genome sequencing. Panel A shows the results for our optimized assay using our newly designed primers while Panel B shows the same samples amplified using previously published primers (Pollett et al., 2018). For each panel, the four amplicons for each sample have been run in sequential order, with strain ID corresponding to the original sample accession and collected year. The samples shown here were of the subtypes A2, B1 and B2 as indicated.

3.4 Genome sequencing, assembly and analysis

Following successful amplification of all ten HMPV samples, the four amplicons from each were pooled, purified and sequenced using the Nextera XT library prep kit and an Illumina iSeq. In this study, we sequenced the 10 clinical HMPV samples along with libraries from other projects. However, we targeted 100,000 paired reads per HMPV library to achieve an expected genome coverage depth between 800-1000X, which is sufficient for calling a consensus genome. Given an Illumina iSeq run would yield a total of 5,000,000 paired reads, it would be possible to reliably multiplex up to 48 HMPV genomes per run. We then used a *de novo* assembly approach to generate the final consensus genomes for the ten HMPV samples. Reference mapping would also be an appropriate method for genome assembly, however, similar to RSV, there are notable structural variants (insertions) in the HMPV G protein [20] that may be mis-assembled when using an inappropriate reference strain for mapping such as an historical prototype. Therefore, a *de novo* approach would be recommended for both RSV and HMPV WGS. It is also important to note that the final genome sequences generating using our approach will be in-complete in the terminal regions and missing an expected 43 and 28 base pairs in the 5' and 3' ends, respectively.

We then analyzed our 10 HMPV genome sequences from clinical samples using a phylogenetic approach to determine their subtype. Of the samples presented in this study, four were identified as A2b strains, two as B1 and four B2 (Figure 5). There is limited data on the molecular epidemiology of HMPV in Australia and nothing published previously for the state of NSW, where these samples were collected. The finding of no A1 or A2a strains may be due to the under-sampling in this current study; however, one study from Queensland Australia showed declining levels of A1 over the period 2001-2004 [29], and since 2006 these subtypes have been rarely identified with A2b and B strains most commonly identified [30]. Despite this, based on our alignments, we would expect the primers and amplification to capture all subtypes including A1 and A2, as these viruses were represented in our genome alignments (Figures 1 & 3), and this approach would bd useful for on-going genomic studies here in Australia and globally.
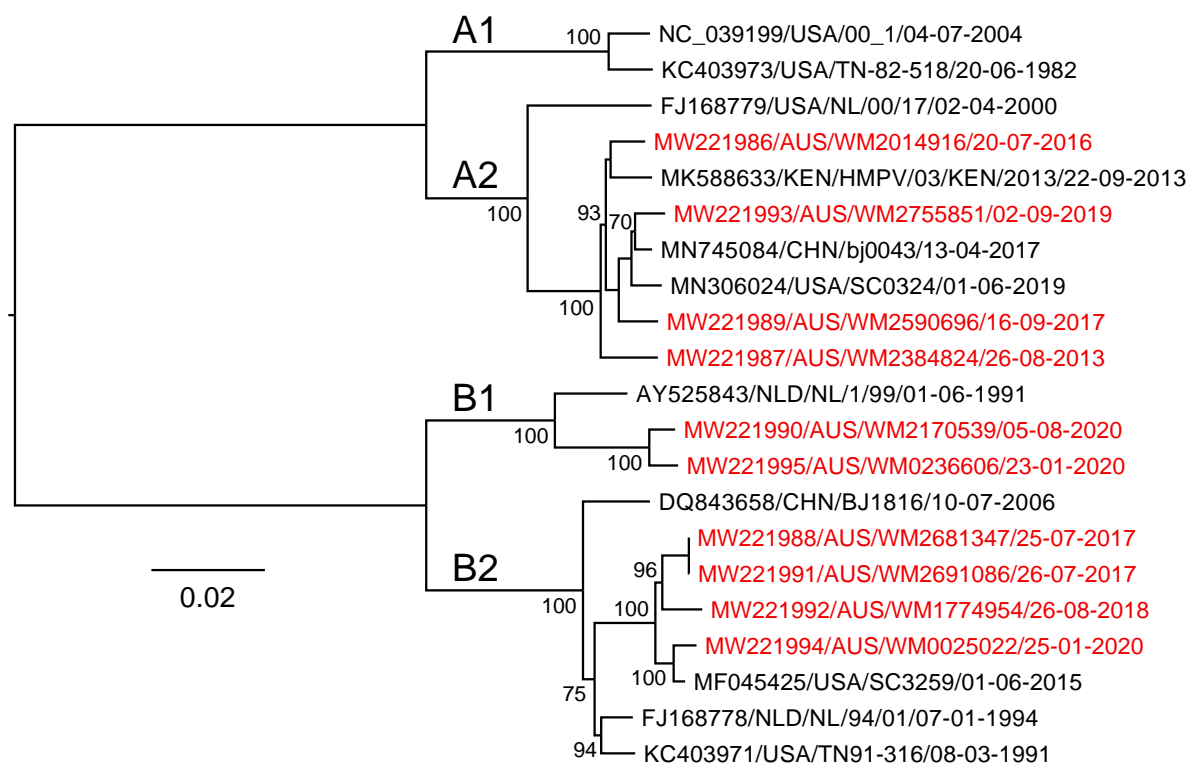
Figure 5 – A maximum likelihood tree constructed using near full length Human Metapneumovirus sequences generated in this study. Node supports are indicated, and branch lengths are scaled according to sequence divergence.

4. Conclusions

Using publicly available, genome sequence representing the full known diversity of HMPV, we designed a simple and reliable assay for amplifying and sequencing HMPV genomes from clinical samples. Ten HMPV genomes were generated from residual-diagnostic specimens using this approach to demonstrate multiple subtypes circulating in NSW, Australia since 2013. This work highlights the utility of amplicon-based sequencing for genomic epidemiological studies of respiratory viruses to inform public health investigations and understand the patterns of evolution and spread.

References

1. van Den Hoogen, B.G.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Analysis of the Genomic Sequence of a Human Metapneumovirus. *Virology* **2002**, *295*, 119-132, doi:10.1006/viro.2001.1355.

2. Kaslow, R.A.; Stanberry, L.R.; Le Duc, J.W. *Viral Infections of Humans Epidemiology and Control*, 5th ed. 2014. ed.; Springer US: Boston, MA, 2014.

3. Biacchesi, S.; Skiadopoulos, M.H.; Boivin, G.; Hanson, C.T.; Murphy, B.R.; Collins, P.L.; Buchholz, U.J. Genetic diversity between human metapneumovirus subgroups. *Virology* **2003**, *315*, 1-9, doi:10.1016/S0042-6822(03)00528-2.

4. Falsey, A.R.; Erdman, D.; Anderson, L.J.; Walsh, E.E. Human metapneumovirus infections in young and elderly adults. (Major Article). *Journal of Infectious Diseases* **2003**, *187*, 785, doi:10.1086/367901.

5. Bernadette, G.V.D.H.; Jan, C.D.J.; Jan, G.; Thijs, K.; Ronald De, G.; Ron, A.M.F.; Albert, D.M.E.O. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nature Medicine* **2001**, *7*, 719, doi:10.1038/89098.

6. McAdam, A.J.; Hasenbein, M.E.; Feldman, H.A.; Cole, S.E.; Offermann, J.T.; Riley, A.M.; Lieu, T.A. Human metapneumovirus in children tested at a tertiary-care hospital.(Major Article). *Journal of Infectious Diseases* **2004**, *190*, 20, doi:10.1086/421120.

7. Williams, J.V.; Harris, P.A.; Tollefson, S.J.; Halburnt-Rush, L.L.; Pingsterhaus, J.M.; Edwards, K.M.; Wright, P.F.; Crowe, J.E. Human Metapneumovirus and Lower Respiratory Tract Disease in Otherwise Healthy Infants and Children. *The New England Journal of Medicine* **2004**, *350*, 443-450, doi:10.1056/NEJMoa025472.

8. Widmer, K.; Zhu, Y.; Williams, J.V.; Griffin, M.R.; Edwards, K.M.; Talbot, H.K. Rates of hospitalizations for respiratory syncytial virus, human metapneumovirus, and influenza virus in older adults.(VIRUSES)(Report). *Journal of Infectious Diseases* **2012**, *206*, 56, doi:10.1093/infdis/jis309.

9. Sutherland, K.A.; Collins, P.L.; Peeples, M.E. Synergistic Effects of Gene-End Signal Mutations and the M2-1 Protein on Transcription Termination by Respiratory Syncytial Virus. *Virology* **2001**, *288*, 295-307, doi:10.1006/viro.2001.1105.

10. El Najjar, F.; Cifuentes-Munoz, N.; Chen, J.; Zhu, H.; Buchholz, U.J.; Moncman, C.L.; Dutch, R.E. Human metapneumovirus Induces Reorganization of the Actin Cytoskeleton for Direct Cell-to-Cell Spread.(Research Article). *PLoS Pathogens* **2016**, *12*, e1005922, doi:10.1371/journal.ppat.1005922.

11. Lim, Y.K.; Kweon, O.J.; Kim, H.R.; Kim, T.-H.; Lee, M.-K. Clinical Features, Epidemiology, and Climatic Impact of Genotype-specific Human Metapneumovirus Infections: Long-term Surveillance of Hospitalized Patients in South Korea. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2020**, *70*, 2683-2694, doi:10.1093/cid/ciz697.

12. Jain, B.; Singh, A.K.; Dangi, T.; Agarwal, A.; Verma, A.K.; Dwivedi, M.; Singh, K.P.; Jain, A. High prevalence of human metapneumovirus subtype B in cases presenting as severe acute respiratory illness: an experience at tertiary care hospital. *The Clinical Respiratory Journal* **2014**, *8*, 225-233, doi:10.1111/crj.12064.

13.   Reiche, J.; Jacobsen, S.; Neubauer, K.; Hafemann, S.; Nitsche, A.; Milde, J.; Wolff, T.; Schweiger, B. Human Metapneumovirus: Insights from a Ten-Year Molecular and Epidemiological Analysis in Germany. *PLoS One* **2014**, *9*, e88342, doi:10.1371/journal.pone.0088342.

14.   Gardy, J.L.; Naus, M.; Amlani, A.; Chung, W.; Kim, H.; Tan, M.; Severini, A.; Krajden, M.; Puddicombe, D.; Sahni, V., et al. Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *The Journal of Infectious Diseases* **2015**, *212*, 1574-1578, doi:10.1093/infdis/jiv271.

15.   Houlihan, C.F.; Frampton, D.; Ferns, R.B.; Raffle, J.; Grant, P.; Reidy, M.; Hail, L.; Thomson, K.; Mattes, F.; Kozlakidis, Z., et al. Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak. *The Journal of Infectious Diseases* **2018**, *218*, 1485-1489, doi:10.1093/infdis/jiy335.

16.   Gymoese, P.; Sorensen, G.; Litrup, E.; Olsen, J.E.; Nielsen, E.M.; Torpdahl, M. Investigation of Outbreaks of Salmonella enterica Serovar Typhimurium and Its Monophasic Variants Using Whole-Genome Sequencing, Denmark.(RESEARCH). *Emerging Infectious Diseases* **2017**, *23*, 1631, doi:10.3201/eid2310.161248.

17.   Eden, J.-S.; Rockett, R.; Carter, I.; Rahman, H.; de Ligt, J.; Hadfield, J.; Storey, M.; Ren, X.; Tulloch, R.; Basile, K., et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. **2020**.

18.   Freed, N.E.; Vlková, M.; Faisal, M.B.; Silander, O.K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods & Protocols* **2020**, *5*, doi:10.1093/biomethods/bpaa014.

19.   Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D., et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* **2020**, doi:10.1101/2020.09.04.283077.

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

21.   Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **2013**, *30*, 772-780, doi:10.1093/molbev/mst010.

22.   Chuang, L.-Y.; Cheng, Y.-H.; Yang, C.-H. Specific primer design for the polymerase chain reaction. *Biotechnology Letters* **2013**, *35*, 1541-1549, doi:10.1007/s10529-013-1249-8.

23.   Pollett, S.; Trovao, N.S.; Tan, Y.; Eden, J.-S.; Halpin, R.A.; Bera, J.; Das, S.R.; Wentworth, D.; Ocana, V.; Mendocilla, S.M., et al. The transmission dynamics and diversity of human metapneumovirus in Peru.(Report). *Influenza and Other Respiratory Viruses* **2018**, *12*, 508, doi:10.1111/irv.12537.

24.   Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **2017**, *33*, 3137-3139, doi:10.1093/bioinformatics/btx373.

25.   Bushnell. BBMap short-read aligner, and other bioinformatics tools. **2016**.

26. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674-1676, doi:10.1093/bioinformatics/btv033.

27. Silva, J.O.; Orellana, E.T.V.; Torres, M. Development of a Parallel Version of PhyML 3.0 Using Shared Memory. *Revista IEEE América Latina* **2017**, *15*, 959-967, doi:10.1109/TLA.2017.7912593.

28. Cotten, M.; Petrova, V.; Phan, M.V.T.; Rabaa, M.A.; Watson, S.J.; Ong, S.H.; Kellam, P.; Baker, S. Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. *Journal of virology* **2014**, *88*, 11056-11069, doi:10.1128/JVI.01333-14.

29. Sloots, T.P.; Mackay, I.M.; Bialasiewicz, S.; Jacob, K.C.; McQueen, E.; Harnett, G.B.; Siebert, D.J.; Masters, I.B.; Young, P.R.; Nissen, M.D. Human Metapneumovirus, Australia, 2001–2004. *Emerging infectious diseases* **2006**, *12*, 1263-1266, doi:10.3201/eid1708.051239.

30. Yi, L.; Zou, L.; Peng, J.; Yu, J.; Song, Y.; Liang, L.; Guo, Q.; Kang, M.; Ke, C.; Song, T., et al. Epidemiology, evolution and transmission of human metapneumovirus in Guangzhou China, 2013-2017. *Scientific reports* **2019**, *9*, 14022-14029, doi:10.1038/s41598-019-50340-8.