# Nanopore metagenomic sequencing of full length human metapneumovirus (HMPV) within a unique sub-lineage

Yifei Xu[1,2*], Kuiama Lewandowski[3], Sheila Lumley[4,5], Nicholas Sanderson[1,2], Ali Vaughan[1,2], Richard Vipond[3], Miles Carroll[3], Katie Jeffery[5], Dona Foster[1,2], A Sarah Walker[1,2], Timothy Peto[1,2], Derrick Crook[1,2], Steven T Pullan[3], Philippa C Matthews[2,4,5]

[1] Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, [2] NIHR Oxford Biomedical Research Centre, University of Oxford, United Kingdom, [3] Public Health England, National Infection Service, Porton Down, Salisbury, United Kingdom, [4] Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom, [5] Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, United Kingdom

*Address correspondence to Yifei Xu, yifei.xu@ndm.ox.ac.uk.

## Abstract

Human metapneumovirus (HMPV) has been recognized as an important pathogen which can cause a spectrum of respiratory tract disease. Here, we report Nanopore metagenomic sequencing of the first full length HMPV genome directly from a throat swab from a UK patient with complex lung disease and immunocompromise. We found a predominance (26.4%) of HMPV reads in the metagenomic sequencing data and consequently assembled the full genome at a high depth of coverage (mean 4,786). Through phylogenetic analyses, we identified this HMPV strain to originate from a unique genetic group in A2b, showing the presence of this group in the UK. Our study demonstrated the effectiveness of Nanopore metagenomic sequencing for diagnosing infectious diseases and recovering complete

sequences for genomic characterization, highlighting the applicability of Nanopore sequencing in clinical settings.

## Importance

Nanopore metagenomic sequencing has the potential to evolve as a point-of-care test for a range of infectious diseases. Here, we report the first full length human metapneumovirus (HMPV) genome in the UK sequenced by Nanopore from a non-invasive sample from an immunocompromised patient. We demonstrate the presence of HMPV from a unique genetic group not previously reported from the UK. Our study demonstrates the effectiveness of Nanopore sequencing for diagnosing an infection that was not detected by routine first-line tests in the clinical microbiology laboratory. We report sufficient genomic data to provide insight into the epidemiology of infection and with the potential to inform treatment decisions.

## Keywords

Human metapneumovirus, HMPV, nanopore sequencing, metagenomics, throat swab, unique genetic group, respiratory tract infection, cystic fibrosis, lung transplant, microbiome

## Manuscript text

Human metapneumovirus (HMPV) is a negative-sense, single-stranded RNA virus of approximately 13kb and belongs to the family Paramyxoviridae [1]. Since it was first described in 2001, HMPV has been recognized as an important pathogen which can cause respiratory tract diseases, ranging from mild upper respiratory tract infections to severe bronchiolitis and pneumonia [2]. HMPV can also cause severe disease in immunocompromised patients and those with underlying medical conditions, including lung

52 transplant recipients [3]. Two main genetic lineages (A and B) and five sublineages (A1, A2a,
53 A2b, B1, and B2) have been described [4].

54 The Nanopore sequencing platform (Oxford Nanopore Technology, ONT) is capable of
55 generating real-time sequencing data, with the potential to evolve as a point-of-care test for a
56 range of infectious diseases [5,6]. In this report, we describe recovery of full length HMPV
57 genome directly from a throat swab through the application of Nanopore metagenomic
58 sequencing.

59 A male in his 40's with cystic fibrosis (CF) and a previous lung transplant presented with
60 breathlessness, thick sputum and low oxygen saturations. His condition was further
61 complicated by CF-related diabetes mellitus and bronchiolitis obliterans. To our knowledge,
62 he had not travelled outside the UK. As he presented to hospital during the peak of the
63 influenza season, a throat swab was taken to test for respiratory viruses in a clinical
64 diagnostic laboratory; this sample was negative by PCR for influenza A, influenza B, and
65 respiratory syncytial virus. Given his previous confirmed colonisation with *Pseudomonas*
66 *aeruginosa*, he was treated with broad spectrum intravenous antibiotics, and discharged from
67 hospital after two weeks.

68 We performed Nanopore metagenomic sequencing and generated 168,811 reads from this
69 throat swab. We identified 44,580 (26.4%) HMPV reads and 5,393 (3.1%) human reads
70 (which were discarded and not retained). The remaining reads mostly comprised bacteria
71 representing oral flora (predominantly Lactobacilli (20%), Actinobacteria (7%), and
72 Proteobacteria (6%)) (Fig. S1). Mapping results showed that HMPV reads covered 99.8%
73 (13,291/13,319) of the reference sequence (USA/NM009/2016, accession number KY474539)
74 at a high mean depth of coverage (4,786). The mean alignment length was 1,534bp and 25%
75 of the alignments were longer than 2,000bp (Fig. 1). We used an alignment-based approach

to recover a HMPV genomic sequence of 12,893bp, referred to as JR001 (accession number xxx). The sequence is nearly complete excepting 205bp at the start of the coding region.

To determine the relationship between JR001 and previously published HMPV genomes, we constructed phylogenetic trees for the full length genome and eight genes (N, P, M, F, M2, SH, G, and L). JR001 clustered within genetic sublineage A2b on the basis of the full length genome and individual genes (Fig. 2 and Fig. S2). Seven HMPV strains from the United States and one strain from China were closely related to JR001, and formed a unique genetic group separated from other strains in A2b, strongly supported by a bootstrap value of 100. The pair-wise nucleotide sequence identities between JR001 and the eight related genomes ranges from 98.3% to 99.2%. This subgroup has been recently identified based on phylogenetic analysis of fusion and attachment genes [7], and comprises sequences originating from East and Southeast Asian countries, including Malaysia, Vietnam, Cambodia, China, and Japan, between 2006 and 2012 [7,8], and Croatia between 2011 and 2014 [9]. Our study provided evidence supporting the presence of HMPV from this unique group in the UK. While we found JR001 shared high nucleotide sequence identities with HMPV strains from the US, its source remains unclear. Further studies are needed to investigate the geographical distribution of this unique genetic group of HMPV and its contribution to respiratory disease in the population.

We conducted time-scale phylogenetic analyses for the HMPV genome to estimate the time of emergence of this group. The topology of the time-scale phylogeny was consistent with that from the maximum-likelihood phylogenetic analyses. HMPV strains within the group were estimated to share a common ancestor originating in 2003 (95% highest posterior density [HPD], 1994 to 2008).

The extent to which the virus is a pathogen in this context is uncertain, as the patient was also at high risk of acute exacerbations of bacterial infection arising from Pseudomonas

colonisation. However, the recovery of the complete genome and the predominance of HMPV reads from the metagenome suggest active infection which could have been completely or partly responsible for the acute clinical deterioration. It is not uncommon to observe co-infection of HMPV with other respiratory viral pathogens, especially respiratory syncytial virus [10]; however we did not detect sequencing reads likely to represent other significant pathogens in this case (Fig. S1).

The case is the first full length HMPV genome in the UK sequenced by Nanopore technology directly from a non-invasive sample without the need for enrichment or viral isolation, diagnosing a potentially relevant pathogen that was not detected by routine first-line tests in the clinical microbiology laboratory, and producing data that can inform treatment as well as providing insights into the epidemiology of infection. Characterisation of the microbiome of patients with complex underlying lung disease, both during periods of clinical stability and in the setting of lower respiratory tract infections, could be valuable in informing intervention and supporting antimicrobial stewardship.

## Methods

### Sample collection, preparation, and Nanopore sequencing

A throat swab was collected in viral transport media from a patient presenting to our tertiary referral teaching hospital in Oxford, United Kingdom. The sample was tested for respiratory viruses using Xpert Xpress Flu/RSV assay (Cepheid, Sunnyvale, CA, USA) in a clinical diagnostic laboratory. The sample was frozen for retrospective Nanopore sequencing. The sample was thawed and passed through a 0.45 μm filter prior to RNA extraction and DNase treatment. cDNA was prepared and amplified using a Sequence-Independent-Single-Primer-Amplification method as described previously [11]. cDNA was used as input for a

SQK-LS108 library preparation and sequencing on a R9.4.1 flow cell using a MinION device (ONT).

**Genomic analysis**

Nanopore reads were basecalled using Albacore v2.1.7 (ONT). Metagenomic classification and mapping were used to identify HMPV reads. Reads were first taxonomically classified against RefSeq database using Centrifuge v1.0.3 [12]. *De novo* assembly was then performed with HMPV-like reads using Canu v1.7 [13]. The resulting contigs were BLASTed against GenBank nt database to determine the reference HMPV sequence. Reads were mapped against the selected reference (USA/NM009/2016, accession number KY474539) using Minimap2 [14]. HMPV reads were defined as those assigned to HMPV by centrifuge and confirmed by mapping. Consensus sequence for the HMPV strain was built using Nanopolish v0.9.2 [15].

Phylogenetic analyses were conducted using an integrated dataset that comprised the HMPV sequence from this study and 154 complete HMPV genomic sequences from NIAID Virus Pathogen Database and Analysis Resource (ViPR) and NCBI Genbank [16]. Maximum-likelihood phylogenies were generated using RAxML v8.2.10 [17]. Time scale phylogenies were built for genomic sequences with complete sampling dates (month, day, and year) using BEAST v1.10.1 [18]. The SRD06 partitioned substitution model, uncorrelated lognormal relaxed clock model, and Bayesian skyline coalescent tree prior were used in the analyses. Multiple independent runs were performed with a chain length of 200 million steps and sampled every 10,000 steps. These runs were combined to ensure an adequate effective sample size (>200) for relevant parameters.

**Ethics statement**

149 This sample that was surplus to diagnostic requirements was sequenced as part of a larger
150 study with Research Ethics Committee approval (17/LO/1420).

151

152 **Accession number**

153 The sequencing data was deposited in the xxx under accession no. xxx.

154

155 **Funding**

156 This work was supported by NIHR Oxford Biomedical Research Centre.

157

158 **References**

159 1. Kahn JS. Epidemiology of human metapneumovirus. Clin Microbiol Rev. Am Soc
160 Microbiol; 2006;19:546–57.

161 2. Van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RAM, et al. A
162 newly discovered human pneumovirus isolated from young children with respiratory tract
163 disease. Nat Med. Nature Publishing Group; 2001;7:719.

164 3. Hopkins P, McNeil K, Kermeen F, Musk M, McQueen E, Mackay I, et al. Human
165 metapneumovirus in lung transplant recipients and comparison to respiratory syncytial virus.
166 Am J Respir Crit Care Med. American Thoracic Society; 2008;178:876–81.

167 4. Kim J Il, Park S, Lee I, Park KS, Kwak EJ, Moon KM, et al. Genome-wide analysis of
168 human metapneumovirus evolution. PLoS One. Public Library of Science; 2016;11:e0152962.

169 5. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time,
170 portable genome sequencing for Ebola surveillance. Nature. Nature Publishing Group;
171 2016;530:228.

172 6. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid
173 metagenomic identification of viral pathogens in clinical samples by real-time nanopore
174 sequencing analysis. Genome Med. BioMed Central; 2015;7:99.

175 7. Chow WZ, Chan YF, Oong XY, Ng LJ, Nor'E SS, Ng KT, et al. Genetic diversity,
176 seasonality and transmission network of human metapneumovirus: identification of a unique
177 sub-lineage of the fusion and attachment genes. Sci Rep. Nature Publishing Group;
178 2016;6:27730.

179 8. Nidaira M, Taira K, Hamabata H, Kawaki T, Gushi K, Mahoe Y, et al. Molecular
180 epidemiology of human metapneumovirus from 2009 to 2011 in Okinawa, Japan. Jpn J Infect
181 Dis. National Institute of Infectious Diseases, Japanese Journal of Infectious Diseases
182 Editorial Committee; 2012;65:337–40.

183 9. Jagušić M, Slović A, Ljubin-Sternak S, Mlinarić-Galinović G, Forčić D. Genetic diversity

of human metapneumovirus in hospitalized children with acute respiratory infections in Croatia. J Med Virol. Wiley Online Library; 2017;89:1885–93.

10. Paranhos-Baccalà G, Komurian-Pradel F, Richard N, Vernet G, Lina B, Floret D. Mixed respiratory virus infections. J Clin Virol. Elsevier; 2008;43:407–10.

11. Kafetzopoulou LE, Efthymiadis K, Lewandowski K, Crook A, Carter D, Osborne J, et al. Assessment of Metagenomic MinION and Illumina sequencing as an approach for the recovery of whole genome sequences of chikungunya and dengue viruses directly from clinical samples. bioRxiv. Cold Spring Harbor Laboratory; 2018;355560.

12. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. Cold Spring Harbor Lab; 2016;

13. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. Cold Spring Harbor Lab; 2017;27:722–36.

14. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. Oxford University Press; 2018;1:7.

15. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. Nature Publishing Group; 2015;12:733.

16. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, et al. ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res. Oxford University Press; 2011;40:D593–8.

17. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. Oxford University Press; 2014;30:1312–3.

18. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. Oxford University Press; 2018;4:vey016.

## Figure legend

**Figure 1.** Results of Nanopore sequencing of an HMPV isolate from a throat swab. (A) Histogram of alignment length derived by mapping Nanopore reads to HMPV reference sequence (USA/NM009/2016, accession number KY474539). The mean alignment length is 1,534bp and 25% of the alignment are longer than 2,000bp. (B) Plot of depth of coverage. HMPV reads cover the full reference genome (99.8%) at a high depth of coverage (mean 4,786). Five HMPV reads, indicated by red lines, are nearly able to cover the full reference genome.

**Figure 2.** Maximum-likelihood (ML) phylogenetic trees of HMPV isolates from this study and public databases. (A) ML tree for the full HMPV genome, (B) ML tree for G gene, (C) ML tree for F gene. Five known genetic sublineages, A1, A2a, A2b, B1, and B2, are indicated by blue boxes and grey triangles. Numbers at the nodes indicate bootstrap support evaluated by 1,000 replicates. The complete phylogenies, showing name of all strains included in the analyses, are shown in supplementary Fig. S2. HMPV strain from this study and eight strains from US and China formed a unique group within A2, indicated by a red box.

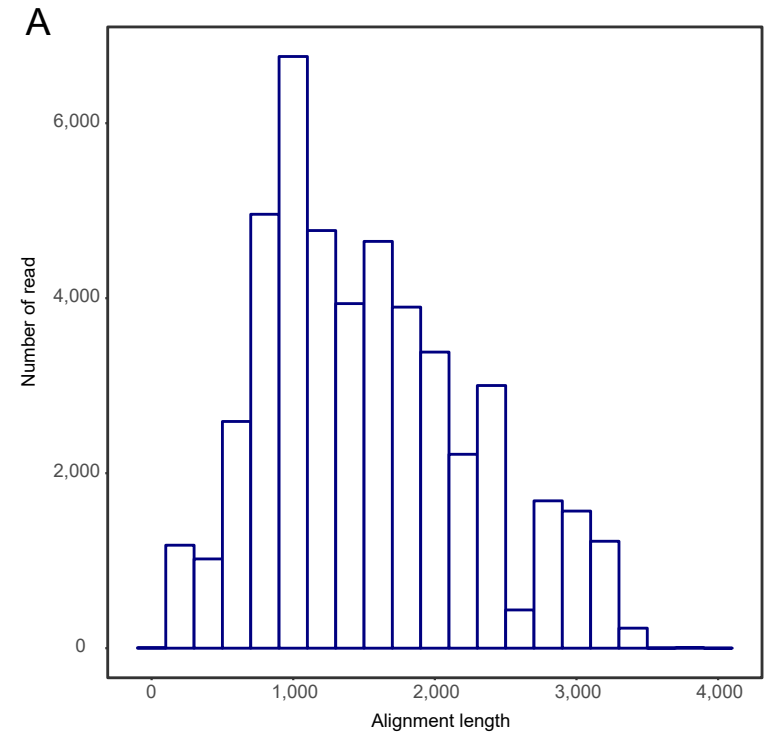
228 **Supplemental materials**

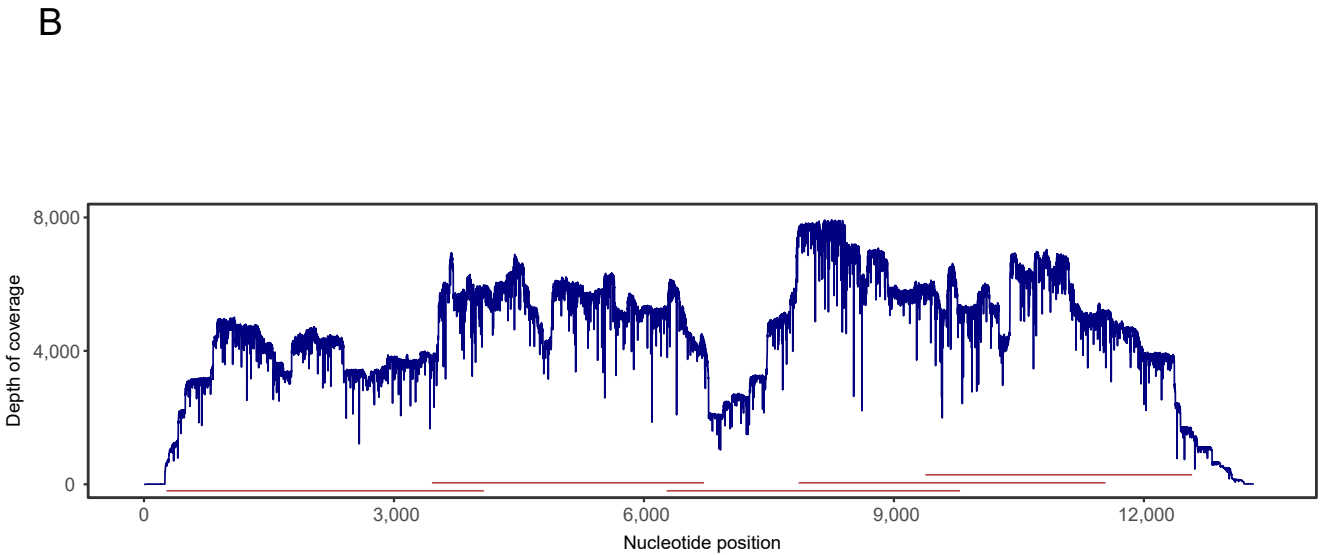
229 **Figure S1.** Taxonomic assignment of Nanopore sequencing reads of a throat swab from a
230 patient with complex lung disease and immunocompromise. HMPV reads accounted for 26%
231 of the total reads.

232

233 **Figure S2.** Maximum-likelihood (ML) phylogenetic trees for the full length genome and
234 gene of HMPV isolates from this study and public databases. Numbers at the nodes indicate
235 bootstrap support evaluated by 1,000 replicates. Five known genetic sublineages, A1, A2a,
236 A2b, B1, and B2, are indicated by blue and grey boxes. HMPV strain from this study and
237 eight strains from US and China formed a unique group within A2, indicated by a red box.

Figure 1


A
6,000
4,000
2,000
0
Number of read
0
1,000
2,000
3,000
4,000
Alignment length


B
8,000
4,000
0
Depth of coverage
0
3,000
6,000
9,000
12,000
Nucleotide position

# Figure 2