

Eight coding-complete genomes of human metapneumovirus recovered by virus metagenomics in coastal Kenya, 2021–2024

Samuel K. Mwasya,¹ Dorcas Okanda,¹ Samuel Odoyo,¹ Esther Nyadzua Katama,¹ Clement Lewa,¹ Arnold W. Lambisia,¹ George Githinji,^{1,2} Charles N. Agoti^{1,3}

AUTHOR AFFILIATIONS See affiliation list on p. 4.

ABSTRACT Few human metapneumovirus coding-complete genomes are available from Africa despite significance in optimizing local molecular diagnostics and global phylogeographic analysis. We report eight genomes recovered following nanopore metagenomic sequencing of samples collected in coastal Kenya (2021–2024). These fell into sub-lineages A2b/A2.2.1 ($n = 3$), A2c-wt/A2.2.2 ($n = 4$), and B1 ($n = 1$).

KEYWORDS human metapneumovirus, hMPV, genomic surveillance, lineages, metagenomics, Kilifi, Kenya

Human metapneumovirus (hMPV; genus *Metapneumovirus*, family *Pneumoviridae*) is a globally ubiquitous respiratory pathogen that infects persons of all ages, causing severe disease in young children and the elderly (1). A surge of hMPV cases in late 2024 in China raised international concern about an unusually severe epidemic that might spread beyond (2). hMPV genome is a negative-sense, single-stranded RNA molecule (~13.3 kb) encoding nine proteins across eight genes (N, P, M, F, M2, SH, G, and L) (3). Clinical isolates divide into two antigenic lineages (A and B), further subdivided into sub-lineages based on phylogenetic analysis of F and/or G gene. Spatially and temporally representative genomic data are crucial for updating molecular diagnostics, revealing molecular adaptations, and tracking global spread and transmission (4). However, by December 2025, only 3.4% (67/1,969) genomes (>90% completeness) and 4.4% (148/3360) G gene sequence in GenBank originated from Africa. The paucity of hMPV genomic data from Africa impedes full understanding of its local and global epidemiology.

We report eight hMPV genomes from nasopharyngeal/oropharyngeal swabs collected between 2021 and 2024, during acute respiratory infection surveillance studies within Kilifi, Kenya (Table 1). Participants were enrolled in outpatient facility-based surveillance, inpatient pediatric pneumonia surveillance at Kilifi County Referral Hospital, and community-based respiratory infection surveillance. Samples identified as hMPV positive (quantitative PCR [qPCR] Ct < 35.0) as previously described (5) underwent metagenomic sequencing. Viral RNA was extracted using QIAamp Viral RNA Mini Kit (QIAGEN), treated with TURBO DNase (37°C, 30 min), and purified with Zymo RNA Clean & Concentrator-5 kit. Sequence-independent single-primer amplification used Sol-A primer (5'-GTTTCCCACTGGAGGATA-N9-3') for first-strand synthesis with SuperScript IV RT, and second-strand synthesis with Sequenase V2.0 complementary DNA underwent PCR amplification using primers (5'-GTTTCCCACTGGAGGATA-3') and Q5 polymerase (30 cycles: 94°C 15 s, 50°C 20 s, 68°C 2 min). Amplified products were purified with AMPure XP beads (×0.6 ratio), quantified with Qubit dsDNA HS assay, and libraries prepared using Native Barcoding V14 Kit for GridION sequencing on R10.4.1 flowcells and automated basecalling with Guppy v6.3.9 (<https://nanoporetech.com/document/Guppy-protocol>).

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Samuel K. Mwasya, smwasya@kemri-wellcome.org, or Charles N. Agoti, cnyaigoti@kemri-wellcome.org.

The authors declare no conflict of interest.

Received 15 January 2026

Accepted 27 February 2026

Published 16 March 2026

Copyright © 2026 Mwasya et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Sample information, read metrics, and associated links to GenBank and sequence read archive (SRA)

Sample ID	Collection date	Study ^a	Clinical status (Ct) ^b	Sequence length	GC content (%)	Total reads	hMPV reads	Sub-lineage	GenBank closest hit	Identity (%)	Country, year	GenBank accession	SRA accession
HF001/21	09-Jun-21	HF	S (22.0)	13,292	36.5	1,547,680	101,309	B1	PP947569.1	99.6	USA, 2020	PX120316	SRR350000059
HF002/23	18-Sep-23	HF	S (24.0)	13,321	36.4	949,219	6,277	A2c-wt/A2.2.2	OL794413.1	98.5	Netherlands, 2010	PX120317	SRR350000058
HF003/23	04-Oct-23	HF	S (24.3)	13,325	36.3	834,365	26,876	A2c-wt/A2.2.2	OL794413.1	98.5	Netherlands, 2010	PX120318	SRR350000057
HS001/23	30-Nov-23	ResViRe	A (24.6)	13,183	36.5	566,197	3,031	A2c-wt/A2.2.2	OL794413.1	98.6	Netherlands, 2010	PX120320	SRR350000056
HF004/24	09-Jan-24	HF	S (23.8)	13,305	36.3	652,679	9,716	A2c-wt/A2.2.2	OL794413.1	98.5	Netherlands, 2010	PX120319	SRR350000055
IP001/24	11-Nov-24	IP	S (24.0)	13,324	36.0	2,455,055	1,678	A2b/A2.2.1	PV252964.1	99.3	USA, 2023	PX120321	SRR350000054
IP002/24	13-Nov-24	IP	S (26.5)	13,321	36.1	1,271,351	7,587	A2b/A2.2.1	PV252964.1	99.5	USA, 2023	PX120322	SRR350000053
IP003/24	18-Nov-24	IP	S (25.3)	13,322	36.1	758,305	14,296	A2b/A2.2.1	PV252964.1	99.4	USA, 2023	PX120323	SRR350000052

^aResViRe is the community-based study that collected and analyzed respiratory swabs irrespective of symptom status; IP is the inpatient pediatric pneumonia study, with swabs collected from infants admitted to Kilifi County Hospital with respiratory illness; and HF is the outpatient acute respiratory illness study, where swabs were obtained from participants presenting with respiratory illness at health facilities across Kilifi County.

^bCt is qPCR cycle threshold; S stands for symptomatic while A stands for asymptomatic.

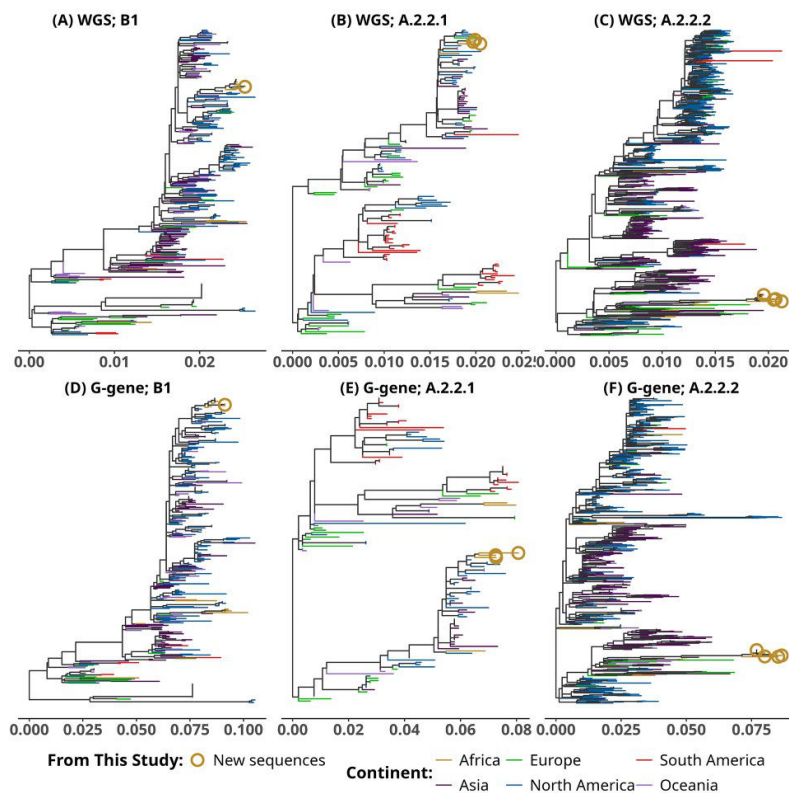


FIG 1 Maximum likelihood (ML) lineage-specific phylogenetic trees of eight hMPV genomes generated in this study, analyzed in the context of global hMPV genomes (>10 kb), and retrieved from GenBank up to 2025. All trees were reconstructed under the HKY+F+R4 nucleotide substitution model on IQ-TREE2 v2.3.5 (13). The top panel shows ML trees inferred from coding-complete genomes, while the bottom panel shows ML trees inferred from G gene sequences. (A and D) Lineage B1 ($n = 324$). (B and E) Lineage A2b/A2.2.1 ($n = 158$). (C and F) Lineage A2c-wt/A2.2.2 ($n = 911$). Tips are colored according to continent of sampling. Sequences generated in this study are indicated by orange circles.

Bioinformatic analysis removed human reads from raw reads using Kraken2 (v2.0.8-beta) (6) against prebuilt standard-16 database; <https://benlangmead.github.io/aws-indexes/k2>). *Homo sapiens* reads were excluded and adapters trimmed using Porechop v0.2.4 (7). Contigs were assembled using metaFlye v2.9.6-b1802 (8). Contigs >12 kb and >100× depth underwent BLASTn (9) against custom reference database of representative hMPV sub-lineages (accessions: [OL794375.1](#), [OL794442.1](#), [OL794455.1](#), [PP591753.1](#), [PV081663.1](#), [PV353978.1](#), [KJ627399.1](#), and [NC_039199.1](#)) to identify best reference (>90% identity). The best-hit reference for each sample was used for mapping with Minimap2 v2.22-r1101 (10), processed with Samtools v1.13 (11), and consensus sequence generated with iVar v1.3.1 (12). Lineage assignment combined phylogenetics and nextclade (v3.18.1) (<https://clades.nextstrain.org/>). All tools were run with default parameters unless otherwise specified.

The genomes had a median length of 13,321 bp (interquartile range 13,299–13,323 bp) and >100× coverage (Table 1). Sub-lineage-specific phylogenies, with available GenBank sequences up to 2025, showed our sequences clustering with sequences from other continents, except A2c-wt/A2.2.2 formed a distinct cluster with predominantly African sequences (Fig. 1).

In conclusion, we present eight coding-complete hMPV sequences from coastal Kenya (2021–2024). These data will contribute to characterizing global diversity, molecular diagnostic design, and phylogeographic analyses.

ACKNOWLEDGMENTS

The authors are grateful to the study participants who provided samples and members of the Pathogen Epidemiology and Omics group at KEMRI-Wellcome Trust Research Programme who did sample collection and laboratory processing. This work is published with permission from director KEMRI.

This work was supported by (a) Wellcome (UK) through a Career Development Award (Ref. #226002/A/22/Z and Ref. #226002/Z/22/Z) and (b) the National Institute of Health and Care Research (NIHR; Ref. NIHR156467) using UK international development funding from the UK Government to support global health research. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome or NIHR or the UK government. A.W.L. was supported by the Sub-Saharan African Network for TB/HIV Research Excellence, which is funded by the Science for Africa Foundation (Del-22-007) with support from Wellcome Trust and the UK Foreign, Commonwealth & Development Office and is part of the EDCPT2 program supported by the European Union; the Bill and Melinda Gates Foundation (INV-033558); and Gilead Sciences Inc., (19275). All content contained within is that of the authors and does not necessarily reflect positions or policies of any funder. The funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. For Open Access, the author has applied for a CC-BY public copyright license.

AUTHOR AFFILIATIONS

¹Kenya Medical Research Institute, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

²Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya

³School of Public Health, Pwani University, Kilifi, Kenya

AUTHOR ORCIDs

Samuel K. Mwasya  <http://orcid.org/0009-0006-8629-4215>

George Githinji  <https://orcid.org/0000-0001-9640-7371>

Charles N. Agoti  <http://orcid.org/0000-0002-2160-567X>

AUTHOR CONTRIBUTIONS

Samuel K. Mwasya, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft | Dorcas Okanda, Conceptualization, Data curation, Methodology, Writing – original draft | Samuel Odoyo, Formal analysis, Methodology, Visualization | Esther Nyadzua Katama, Data curation, Formal analysis, Methodology | Clement Lewa, Data curation, Methodology | Arnold W. Lambisia, Conceptualization, Formal analysis, Validation, Writing – review and editing | George Githinji, Conceptualization, Funding acquisition, Supervision, Writing – review and editing | Charles N. Agoti, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review and editing

DATA AVAILABILITY

Consensus hMPV genomes generated in this study were deposited in the GenBank database under the accession numbers [PX120316](#) to [PX120323](#). The raw sequencing reads were deposited in the NCBI Sequence Reads Archive database under the SRA accession number: [PRJNA1306485](#).

ETHICS APPROVAL

Each surveillance platform (community-based ResViRe, outpatient HF, and inpatient IP) that provided samples for this study operated under dedicated research protocols. The consenting and sample collection procedures were reviewed and approved by the Kenya Medical Research Institute Scientific and Ethics Review Unit, Nairobi, Kenya

(protocol numbers #3178, #3103, and #4724). Written informed consent was obtained from parents or legal guardians for participants aged <18 years, with assent obtained from children aged 13–18 years. Individual written informed consent was obtained from participants aged ≥18 years. All research was conducted in accordance with the principles of the Declaration of Helsinki.

REFERENCES

- Schuster JE, Williams JV. 2013. Human metapneumovirus. *Pediatr Rev* 34:558–565. <https://doi.org/10.1542/pir.34-12-558>
- Murphy F. 2025. Fact check: human metapneumovirus in China. *BMJ* 388:r68. <https://doi.org/10.1136/bmj.r68>
- Soto JA, Gálvez NMS, Benavente FM, Pizarro-Ortega MS, Lay MK, Riedel C, Bueno SM, Gonzalez PA, Kalgis AM. 2018. Human metapneumovirus: mechanisms and molecular targets used by the virus to avoid the immune system. *Front Immunol* 9:2466. <https://doi.org/10.3389/fimmu.2018.02466>
- Shen Y, Liu Y, Krafft T, Wang Q. 2025. Progress and challenges in infectious disease surveillance and early warning. *Medicine Plus* 2:100071. <https://doi.org/10.1016/j.medp.2025.100071>
- Molecular Epidemiology of Human. 2026. Wellcome open research [Internet]. Available from: <https://wellcomeopenresearch.org/articles/11-106>
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Bonenfant Q, Noé L, Touzet H. 2023. Porechop_ABI: discovering unknown adapters in oxford nanopore technology sequencing reads for downstream trimming. *Bioinform Adv* 3:vbac085. <https://doi.org/10.1093/bioadv/vbac085>
- Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Pevnikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 17:1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 20:8. <https://doi.org/10.1186/s13059-018-1618-7>
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>