Virus Research 302 (2021) 198490

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Whole genome sequencing of human metapneumoviruses from clinical specimens using MinION nanopore technology

Kevin Groen, Stefan van Nieuwkoop, Theo M. Bestebroer, Pieter L. Fraaij, Ron A.M. Fouchier, Bernadette G. van den Hoogen *

Department of Viroscience, Erasmus MC, Wijtemaweg 80, 3015 CN Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Human metapneumovirus
Whole genome sequencing
Nanopore technology
MinION sequencing

ABSTRACT

Human metapneumovirus (HMPV), a member of the *Pneumoviridae* family, is a causative agent of respiratory illness in young children, the elderly, and immunocompromised individuals. Globally, viruses belonging to two main genetic lineages circulate, A and B, which are further divided into four genetic sublineages (A1, A2, B1, B2). Classical genotyping of HMPV is based on the sequence of the fusion (F) and attachment (G) glycoprotein genes, which are under direct antibody-mediated immune pressure. Whole genome sequencing provides more information than sequencing of subgenomic fragments and is therefore a powerful tool for studying virus evolution and disease epidemiology and for identifying transmission events and nosocomial outbreaks. Here, we report a robust method to obtain whole genome sequences for HMPV using MinION Nanopore technology. This assay is able to generate HMPV whole genome sequences from clinical specimens with good coverage of the highly variable G gene and is equally sensitive for strains of all four genetic HMPV sublineages. This method can be used for studying HMPV genetics, epidemiology, and evolutionary dynamics.

1. Introduction

After its identification in 2001, human metapneumovirus (HMPV), a member of the *Pneumoviridae* family, was shown to have a large clinical impact on human health, especially in young children, the elderly, and immunocompromised individuals (Falsey et al., 2003; Van Den Hoogen et al., 2001; Williams et al., 2006). Globally, HMPV accounts for approximately 5–15% of all respiratory tract infections (RTI) with a peak of infections in infants between 6 and 12 months of age (Jain et al., 2015; O'Brien et al., 2019; Van Den Hoogen et al., 2001; Williams et al., 2004).

HMPV is a non-segmented negative-strand RNA virus with a genome of ±13,300 nucleotides containing 8 genes encoding nine open reading frames (ORF), including three surface glycoproteins: the fusion protein (F), the small hydrophobic protein (SH) and the attachment protein (G). Of these, F is the major target of neutralizing and protective antibodies (Skiadopoulos et al., 2006; Herfst et al., 2007). However, antibody titers rapidly wane over time and therefore reinfections with HMPV can occur throughout life (Ebihara et al., 2004; van den Hoogen et al., 2007). Genetically, HMPV strains cluster in two major lineages (A and B) which are subdivided in four sublineages (A1, A2, B1, B2) (Van Den Hoogen et al., 2004). Viruses from multiple HMPV sublineages can co-circulate in a single season, but in general, viruses from one sublineage predominate per season (Boivin et al., 2004; Mackay et al., 2006; Van Den Hoogen et al., 2004). Phylogenetic analyses of HMPV strains have suggested that the A2 sublineage may be further divided in two clades, A2a and A2b, or rather A2.1 and A2.2, following nomenclature as used for Influenza viruses and the international committee on taxonomy of Viruses (ICTV) (Huck et al., 2006; Lefkowitz et al., 2018; Nao et al., 2020). Additionally, two different sizes of duplications in the open reading frame (ORF) of G of the virus belonging to lineage A2 were detected of 111 or 180 nucleotides, in Japan, followed by reports of the occurrence of these duplications in viruses circulating in Spain, Vietnam and China (Saikusa et al., 2019, 2017b, 2017a). One of these strains, with a 111 nucleotide duplication in G, became the predominant circulating strain in Yokohoma city suggesting a beneficial role for the duplication (Saikusa et al., 2019). Furthermore, studies reported variability in the length of the G ORF due to changes in the position of the stop codon (Kamau et al., 2020; Van Den Hoogen et al., 2004). These evolutionary events highlight the need to carefully monitor HMPV evolutionary dynamics, which is critical towards the design of diagnostic assays and future intervention strategies that can lead towards epidemiological

* Corresponding author.
E-mail address: b.vandenhoogen@erasmusmc.nl (B.G. van den Hoogen).

https://doi.org/10.1016/j.virusres.2021.198490
Received 12 January 2021; Received in revised form 27 May 2021; Accepted 7 June 2021
Available online 17 June 2021

containment strategies of the virus.

Thus far, studying HMPV genetic variation and evolution has been mainly based on the genetic variability of F and G (Boivin et al., 2004; Saikusa et al., 2019, 2017b, 2017a; Van Den Hoogen et al., 2004). Whole genome sequences could provide additional genomic information about virus evolution that will remain undiscovered by sequencing of sub-genomic fragments (Dudas and Bedford, 2019). With rapidly developing next generation sequencing (NGS) techniques, the time and cost efficiency to sequence entire viral genomes is improving considerably. NGS technologies, such as Ion Torrent, Illumina and the relatively new Oxford Nanopore Technologies MinION sequencing, are used in a wide range of applications such as population surveillance of viral outbreaks, tracking nosocomial infections and the design of vaccine strategies (Radford et al., 2012). While all three platforms can be used for viral whole genome sequencing, MinION sequencing has the advantage that it does not require additional amplification steps during the sequencing procedure (Quick et al., 2018; Siddle et al., 2018). The cost of MinION sequencing provides an additional benefit of MinION sequencing over Ion Torrent and Illumina sequencing (Van Nimwegen et al., 2016), and the scale of MinION sequencing is very suitable for virus-sized genomes.

Here, we introduce a sensitive HMPV whole genome sequencing method by MinION sequencing of multiplexed PCR amplicons. As the G is highly variable between viruses from the different lineages (nucleotide identity: 50%−57%), and therefore used for classification, a good genomic coverage of this gene is desirable (Van Den Hoogen et al., 2004). However, this high variability makes it challenging to design primers that can detect the G gene of HMPV strains from all sublineages with the same assay. With the method described here, HMPV whole genome sequences can be generated from viruses in clinical specimens belonging to all sublineages, with good genomic coverage of the entire genome including the highly variable G gene. This rapid whole genome sequencing technique facilitates HMPV genomic and evolution studies and can be used to monitor HMPV epidemiology.

2. Methods

2.1. HMPV reference strains

HMPV strains NL/00/1 (Genbank accession number AF371337), NL/00/17 (Genbank accession number FJ168779), NL/99/1 (Genbank accession number AY525843) and NL/94/1 (Genbank accession number FJ168778) were used as reference strains for HMPV sublineages A1, A2, B1 and B2, respectively. The HMPV reference strains were isolated and cultured as described previously (Herfst et al., 2004).

2.2. Patient materials

For the purpose of this study surplus samples initially tested for routine clinical care were obtained from the diagnostics unit of the Viroscience department of the Erasmus MC. Only samples in which HMPV could be detected by routine diagnostic qRT-PCR assays at a cycle times less than 25 were selected. A volume of 200 µl was used for RNA extraction using the MagnaPureLC (Roche Diagnostics) that was eluted in a volume of 100 µl. qRT-PCR was performed as described previously (Hoek et al., 2013). Data collection and analyses were conducted in an anonymized matter, which does not require further medical ethics review as consented by our Medical ethical board (MEC-2015–306).

2.3. RNA extraction, cDNA synthesis and multiplexed PCR amplification of clinical HMPV samples

RNA extraction was performed using the High Pure RNA Isolation Kit (Roche). Briefly, 200 µl of HMPV-positive clinical material or diluted virus reference strain culture supernatant was lysed in 400 µl RNA lysis buffer. RNA was extracted according to the manufacturer's instructions and eluted in 50 µl elution buffer. For each sample, two separate cDNA reactions (1 for each primer pool) and subsequently two separate multiplexed PCR reactions were performed. cDNA synthesis was performed using the SuperScript IV First-Strand cDNA Synthesis kit (Invitrogen) in

Table 1
List of primer sequences used for the uneven amplicons (primer pool 1) and their optimized concentrations used for HMPV whole genome MinION sequencing. For degenerate primers, $Y = C$ or T, $W = A$ or T, $R = A$ or G, $M = A$ or C, $S = C$ or G, $K = G$ or T, $D = A$, G or T, $N =$ any base.

Amplicon	Primer	Nucleotide sequence (5′−3′)	µM primer per PCR reaction
1	FW 1	ACGCGAAAAAAACGCGTATA	40
	REV 1	CCAGAYTCWGGRCCCATYTC	
3	FW 3A	CMARCAACCAAAACAACAGATCC	10
	FW 3B	CAANYAAYCAAARYYATGGATCC	
	REV 3A	GAAGTACAGACATNGCWGCACC	
	E	GAAGTACMGACATNGCWGCACC	
5	FW 5A	GCAAGACTTGGAGCCATCAAGG	30
	FW 5B	GCAAGAGCTGGAGTCACCAAGG	
	REV 5	CAAGGYGTRYTATNACNCCAAAGAT	
7	FW 7A	TTGAAAACAGTCAAGCACTAGT	40
	FW 7B	TTYCCTGARGATCARTTYAATGT	
	REV 7	CATCYAATGTTWTCATTGTCAYTTATC	
9	FW 9A	GGAAAATAAGYAGAAAYCAATGCAC	50
	FW 9B	GGRAARTAAGYANAAATCAATGYAC	
	REV 9A	CATTRAGAGGATCCATTGYYATTT	
	REV 9B	CACARARNGGATCCATTGYYATTT	
11	FW 11A	AACCCACCTCAGATAACACATCAATYCT	10
	FW 11B	ACCCAACCTCAGATAACACATCAATTCT	
	REV 11A	AGTTGACTGGGGTAAYTTTTWGCTT	
	REV 11B	AGTTGGCTRGGGTAACTTTTRGCTT	
13	FW 13A	GGAAATSAAATCRGAACTTTCTTCYAYTAAAAC	5
	FW 13B	GGAAATRAAATCAGAACTTTCYTCYATTAAAAC	
	REV 13A	TCTCCYCCWCCAAAYTGCATTG	
	REV 13B	TCTCCTCCWCCAAAYTGCATCG	
15	FW 15	AATGGTAGGCTGATATGCTGYCAG	5
	REV 15A	YGATAYRAACCCRTCACCCCAGTC	
	REV 15B	TGARATGAACCCATCACCCCARTC	
17	FW 17A	AYCAGTTCRGAYTACAACAAAGGG	5
	FW 17B	TCCCYAAGATAACATTYGAAAGGCTAAAAA	
	REV 17	CTTRCTRCCNCCAACTGTTGCT	

Table 2
List of primer sequences used for the even amplicons (primer pool 2) and their optimized concentrations used for HMPV whole genome MinION sequencing. For degenerate primers, $Y = C$ or T, $W = A$ or T, $R = A$ or G, $M = A$ or C, $S = C$ or G, $K = G$ or T, $D = A$, G or T, $N =$ any base.

Amplicon	FW/RV	Nucleotide sequence (5′−3′)	μM primer per PCR reaction
2	FW 2	CWACAGGMAGCAAAGCAGAAAG	10
	REV 2	GARAGCAARTCTAGRGCATCTT	
4	FW 4	AYACAGCYGCTGTTCAAGTTGA	40
	REV 4	ACCARCCTGTYCTYARAACACT	
6	FW 6	GTGCGGCARTTTTCAGACAATG	10
	REV 6	ACACCAYTYAGCTCYGGAGG	
8	FW 8	AGTGRCATGGTCCTGTYTTCA	30
	FW 8A	GACAGTGAARGCAYTAATCAAGTGC	
	REV 8	ACYTCCATRRCYACTTGTCCCA	
10	FW 10A	YCRCAYRAGCAGCAYARGRRAAAGA	20
	FW 10B	CCGCACKAGCAGCRCAAKAAGGAGA	
	FW 10C	CAACCARAYCAGMAATGNAAGNGAGRCA	
	FW10D	CAACCAAACCAGCAATGGAAGAGAGGCG	
	FW 10E	CAACCARAYCAGMAATGNAAGNGAG	
	REV 10A	CTYAYTCTTYTRCTTTTGTTGCT	
	REV 10B	GCTYAYTCTTYTRCTTTTGTTGCT	
12	FW 12	TYGGWCAYCCDATGGTAGATGA	10
	REV 12	ACCYYTTGTTTCYGGTGGTGCA	
14	FW 14	RGGRGARAGYATRYTAGTTAGTYTGATA	5
	REV 14A	TGAGTGCTTGATCCTACCCAGG	
	REV 14B	AGGGCTYTTTGGACCTCTTTGA	
16	FW 16	ACAYTRGGRAARATGCTYATGC	10
	REV 16	CARTTWCCTGCYCCTTCTCCAA	
18	FW 18	GAYCTTGAYCAYCATTAYCCNYTRGAATAYCA	40
	REV 18	ACGGCAAAAAAACCGTATACAWTCAA	

a total volume of 20 μl containing 6 μl eluted RNA, 10 mM dNTPs and 20 U RNAse OUT (Invitrogen), 10 μM of each forward primer from primer pool 1 or primer pool 2, 4 μl 5x First Strand Buffer, 100 mM DTT, 20 U RNAse OUT and 200 U SuperScript IV Reverse Transcriptase. The cDNA reaction was carried out for 15 min at 50 °C followed by 10 min at 80 °C. The cDNA was amplified by PCR in a total volume of 50 μl containing 4 μl cDNA, 5 μl 10x PFU DNA polymerase buffer (Agilent Technologies), 12.5 mM dNTPs, 1 μl PFU DNA polymerase (Agilent Technologies) and primer concentrations as indicated in Tables 1 and 2. Thermocycling was performed with the following parameters: 95 °C for 2 min, 40 cycles of 95 °C for 20 s, 50 °C for 30 s and 72 °C for 90 s, followed by 72 °C for 3 min.

2.4. PCR purification and MinION sequencing

For each primer pool, 5 μl PCR product was analyzed on a 1% agarose gel. When bands were visible, the remaining PCR products from both primer pools were pooled and purified using Agencourt AMPure XP beads (Beckman Coulter). Upon measurement of sample concentrations with the Qubit dsDNA HS assay kit (Thermo Fisher) on a Qubit fluorometer (Thermo Fisher), 1000 ng DNA was used per sample for library preparation. A maximum of twelve purified samples were barcoded using the 1D Native barcoding genomic DNA kit (EXP-NBD104 and SQK-LSK109) according to the manufacturer's instructions and sequenced using a R9.4 FLO-MIN106 flowcell (Oxford Nanopore Technologies) on a GridION Mk1 (Oxford Nanopore Technologies) for 16 h.

2.5. Sequence data analysis

Obtained sequence data was demultiplexed using the Porechop algorithm (https://github.com/rrwick/Porechop) as described previously (Oude Munnink et al., 2019). The demultiplexed sequence data was analyzed using CLC Genomics Workbench version 12.0 (Qiagen). After trimming the 30-nt primer sequences from the ends of the reads, the samples were genotyped by mapping the reads of an individual sample against the full genome sequences of the 4 reference strains (HMPV A1, A2, B1, B2). The HMPV reference sequence to which most of the reads were mapped was considered the genotype of that sample. If an equal number of reads mapped to two different reference strains, the genotype was determined based on homology of the F and G sequences of the sample compared to the reference strains. Subsequently, all reads were mapped against one reference genome. A consensus sequence was extracted with a minimum threshold of 25 read coverage per nucleotide. Gaps in homopolymeric regions of the obtained consensus sequences where checked and resolved by consulting reference genomes. A schematic overview of the complete HMPV MinION sequencing workflow is depicted in Fig. 1.

2.6. Quantitative reverse-transcription PCR

RNA extraction was performed as described above and quantitative reverse-transcription PCR on the HMPV N gene was performed as described previously (Maertzdorf et al., 2004).

2.7. PCR and Sanger sequencing for genotyping of HMPV strains

RNA extracted from HMPV-positive clinical samples was subjected to PCR amplification of a region covering the HMPV F gene as previously described (Van Den Hoogen et al., 2004). PCR products were purified and Sanger sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's protocol. The sequencing products were purified using a Performa V3 96-well short plate (EdgeBio) according to the manufacturer's protocol. Nucleotide sequences were determined using a 3130xl Genetic Analyzer (Applied Biosystems). The HMPV genotype was determined from the sequence of the F gene as described previously (Van Den Hoogen et al., 2004).

3. Results

3.1. Primer design for HMPV whole genome sequencing

To obtain amplicons with a size of 1000 nucleotides and a 200 nucleotide overlap with neighboring amplicons, primers were designed using Primalscheme (Quick et al., 2018). A total of 183 available HMPV whole genome sequences from GenBank were aligned using MAFFT (https://mafft.cbrc.jp/alignment/server/). Two separate alignments were generated for HMPV lineage A and B strains, respectively, based on the information supplied on GenBank or, in case the information was lacking, by analysis of the sequence of the F gene as described previously (Van Den Hoogen et al., 2004). Area's containing gaps in the alignments were removed manually, the designed primers were added to the alignment and primers sequences directed against these manipulated alignments were optimized using Primer3 plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The designed primers were imported in the MAFFT alignment containing all full genome sequences for both genotype A and B, including the gaps. Only primers that matched sequences of both HMPV-A and HMPV-B viruses were selected and, if necessary, degeneracies were introduced into the primer sequences to increase primer sensitivity for viruses of both HMPV lineages. In case degenerate primers could not be designed that would align to both genotype A and B viruses, multiple primers for that position were designed. Primers were designed to yield 18 PCR fragments, each of

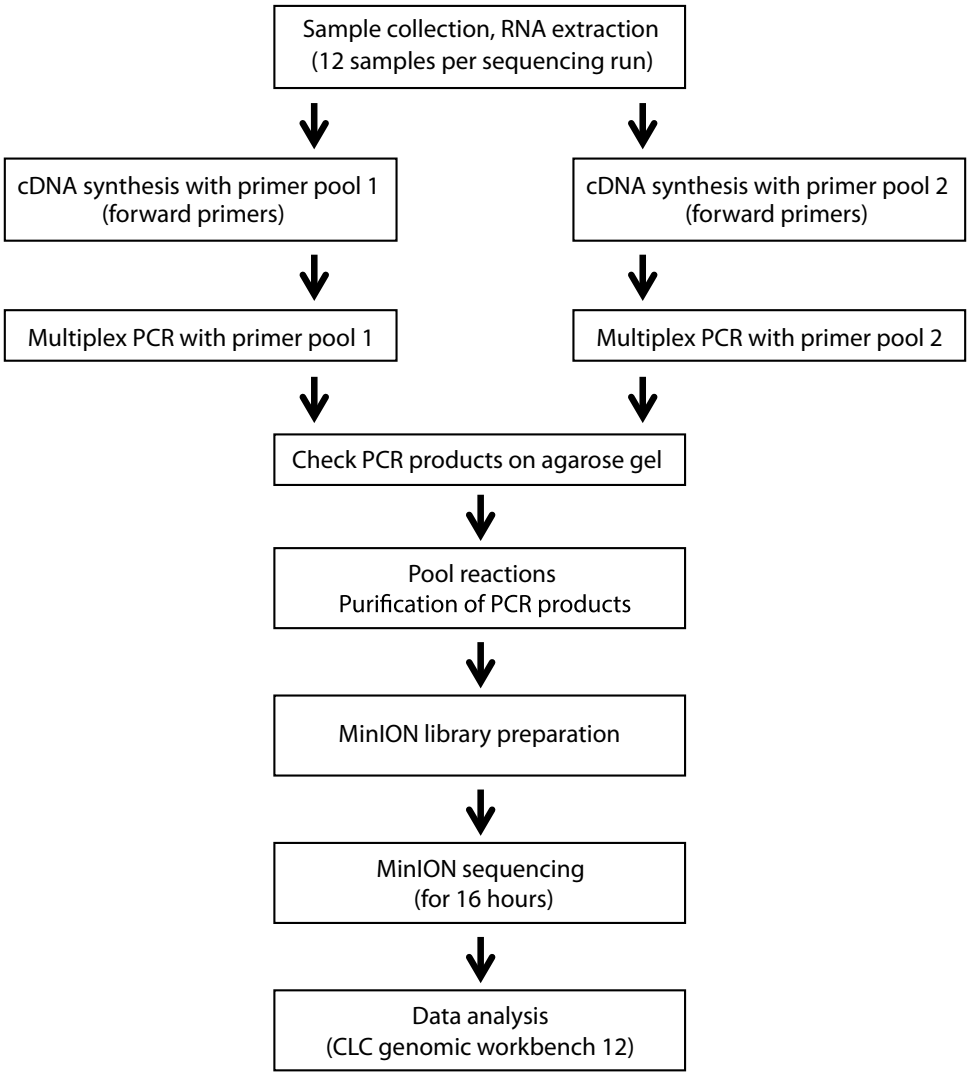

Fig. 1. Schematic representation of the HMPV whole genome MinION sequencing workflow.

approximately 1000 nucleotides in length with average overlap of approximately 260 nucleotides with neighboring amplicons (Fig. 2a). The average amplicon length was 970 nucleotides, with the exception of amplicon 9, with a length of ± 1250 nucleotides to span the entire G gene. The amplicons were divided over two pools (Tables 1 and 2). Primer pool 1 contained the primers of the uneven numbered amplicons and primer pool two contained the primers of even numbered amplicons to prevent the generation of small PCR products from the overlapping part of neighboring amplicons.

3.2. Validation and optimization of HMPV MinION whole genome sequencing

The sensitivity and specificity of primers for individual amplicons was first determined using full genome HMPV NL/00/1 (A1) and HMPV NL/99/1 (B1) plasmid DNA [30]. Dilution series ranging from 50 to 0.005 pg plasmid were used as template DNA for PCR amplification of all described amplicons. The PCR products were run on an agarose gel to confirm the presence of PCR product (data not shown). Next, the primers were used for cDNA synthesis and PCR amplification of viral RNA isolated from reference strains NL/00/1 (genotype A1), NL/00/17 (genotype A2), NL/99/1 (genotype B1) and NL/94/1 (genotype B2). To obtain equal input for all strains, viral RNA was subjected to quantitative reverse-transcription PCR (qRT-PCR) and RNA samples were diluted in order to match a Ct value of 20. From these samples, cDNA was generated using the forward primers of primer pool 1 or primer pool 2 separately, followed by a PCR reaction with primers for each individual PCR amplicon. The presence of PCR product for each amplicon was validated on an agarose gel and primer concentrations were optimized by increasing primer concentrations for those PCR products that had a low intensity. Next, the cDNA was subjected to multiplexed PCR using the primers from primer pool 1 or primer pool 2 in two separate reactions (Fig. 1). The multiplexed PCR products were then subjected to MinION sequencing and coverage plots were generated for each HMPV reference strains (data not shown). Primer concentrations from amplicons which yielded a genomic coverage <100 reads per nucleotide were increased in the multiplexed PCR reaction and the new primer concentrations were validated by MinION sequencing. This process was repeated until all amplicons had a minimum genomic coverage of at least 100 reads per nucleotide throughout the entire genome, including the highly variable G gene. The finally selected primer concentrations are listed in Table 1 (pool 1) and 2 (pool 2).

To validate the sensitivity of the HMPV whole genome MinION sequencing assay, three dilutions of the RNA extracted from each HMPV reference strain were generated. The input of viral RNA varied based on Ct values obtained by qRT-PCR, with a range of Ct values from 17.8 to 23.7. The presence of bands from the multiplexed PCR products was validated on an agarose gel (Fig. 2b). Samples were subjected to MinION

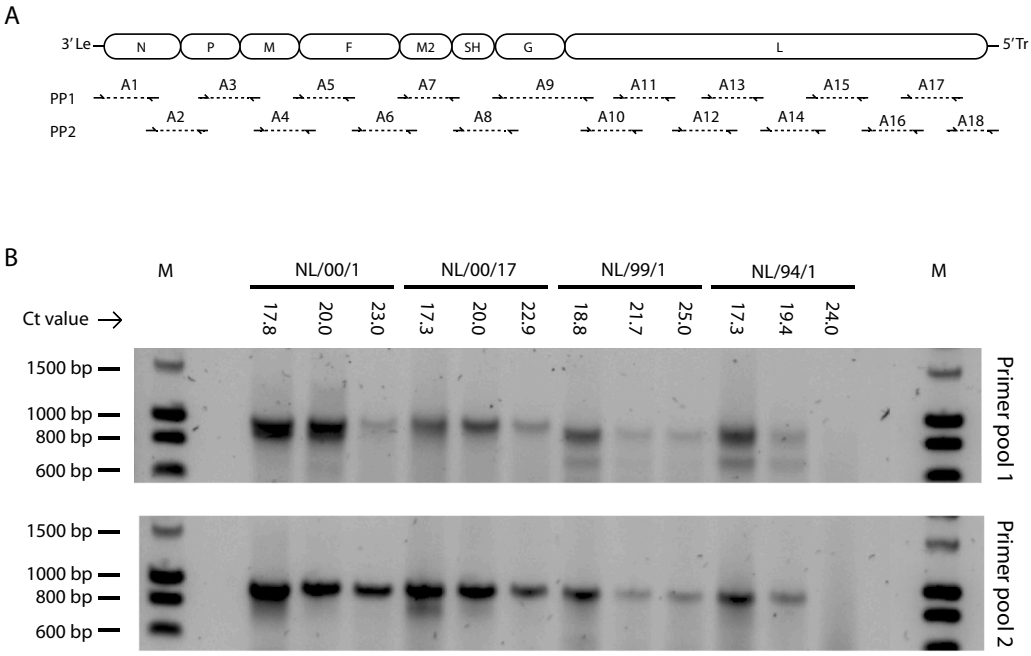

Fig. 2. (a) Schematic overview of the 18 amplicons used to cover the HMPV genome by MinION sequencing. Le: leader sequence, Tr: trailer sequence, N: nucleoprotein, P: phosphoprotein, M: matrix protein, F: fusion protein, M2: Matrix protein 2, SH: small hydrophobic protein, G: glycoprotein, L: polymerase, PP1: primer pool 1, PP2: primer pool 2, A: amplicon. (b) Gel electrophoresis of multiplexed PCR products of serially diluted RNA from four HMPV reference strains NL/00/1, NL/00/17, NL/99/1, and NL/94/1. PCR products were run on a 1% agarose gel to validate the presence of PCR products. M: Molecular marker.

sequencing using the selected primer concentrations from Tables 1 and 2. Whole genome consensus sequences were obtained from all dilutions of the reference strains (Fig. 3). The PCR product of amplicon 9 had a size of 1200 nucleotides and should run higher than the other amplicons but was not visible on the agarose gel (Fig. 2b). Despite this, good genomic coverage of this region was obtained for all sample dilutions for viruses of all lineages (Fig. 3). For the viruses belonging to the A1 and B2 lineages, the genomic coverage was lower for the samples with higher Ct values (23.0 and 24.0, respectively) (Fig. 3a and 3d). However, a consensus sequence with a minimum coverage of 100 reads per nucleotide throughout the genome, including the G gene, was obtained for all dilutions. For the viruses belonging to the A2 and B1 lineages, similar genomic coverages were obtained for the different dilutions of these samples, with the exception of the 5′ end of the genome. For all three dilutions of these viruses a coverage of at least 100 reads per nucleotide was obtained, including good coverage of the G gene (Fig. 3b and 3c). Although multiplexed PCR amplification for NL/94/01 (lineage B2) with a Ct value of 24.0 resulted in bands with lower intensities on the agarose gel (Fig. 2b), a whole genome sequence was still obtained (Fig. 3d). Thus, we were able to obtain HMPV whole genome sequences of viruses from all four HMPV lineages with good genomic coverage throughout the genome using RNA with Ct values obtained by qRT-PCR up to at least 25.0.

3.3. *MinION sequencing of HMPV-positive clinical samples*

The HMPV whole genome MinION sequencing assay was subsequently used to sequence HMPV genomes from clinical specimens. HMPV-positive clinical samples were obtained and their genotype was determined by sanger sequencing of the HMPV F gene. Three samples belonged to sublineage A2, three to B1, three to B2 and one to sublineage A1, and Ct values ranged from 18.7 to 22.7 (Table 3). A HMPV whole genome consensus sequence was obtained for all 10 selected clinical specimens with a minimum genomic coverage of 100 reads per nucleotide, with the exception of the 5′ and 3′ ends of the genome and a small region of the virus belonging to the A1 sublineage (Fig. 4). The total number of reads ranged from 110,000 to 580,000 per sample. The mean genomic coverage, expressed as the mean number of reads per nucleotide in the genome, ranged from 2500 to 18,500. The total reads and mean genomic coverage for HMPV A lineage viruses (total reads: 365,000–580,000, mean coverage in reads per nucleotide: 8100–18,500) was somewhat higher than for the HMPV B lineage viruses (total reads: 110,000–391,000, mean coverage in reads per nucleotide: 2500–12,500), a full length consensus sequence with a minimum genomic coverage of 100 reads per nucleotide was obtained for viruses from both lineage A and B. The mean Ct value of lineage A viruses was similar to that of lineage B viruses (20,75 for HMPV A, 21,35 for HMPV B), therefore the difference in total reads and genomic coverage between the two genotypes were not attributable to differences in Ct values. The genomic coverage of the highly variable G gene was above 100 reads per nucleotide for viruses from all four sublineages. This was similar to the genomic coverage of other regions of the genome, which demonstrated that the high variability of the G gene was not problematic for this assay. In conclusion, this method can be used for sensitive whole genome sequencing of viruses from all four sublineages from clinical specimens.

4. Discussion

Monitoring genetic evolution of HMPV is crucial for the design of diagnostic assays, for studying virus epidemiology, for identifying transmission events and nosocomial outbreaks, and for the rational design of future vaccines or other intervention strategies. Monitoring HMPV genetic evolution is classically done by sequencing of the F and G genes, of which the HMPV G gene is the most variable gene of the virus and is therefore used to distinguish between HMPV sublineages (Van Den Hoogen et al., 2004). In contrast, the HMPV F gene is relatively conserved but contains a number of distinct amino acids that can be used to distinguish between viruses from lineage A, B1 and B2 (Van Den Hoogen et al., 2004).

HMPV whole genome sequencing will provide additional important information about virus evolution which could be missed by partial genome sequencing (Radford et al., 2012). Here, we introduce a method to obtain whole genomes from HMPV positive clinical specimens that is equally sensitive to all four genetic sublineages. This method was optimized based on the intensity of PCR products on agarose gels and based on optimal genomic coverage obtained by MinION sequencing. However, a low band intensity of PCR products on gel did not translate to poor genomic coverage, as a whole genome consensus sequence with a 100 reads per nucleotide threshold was still obtained from samples with

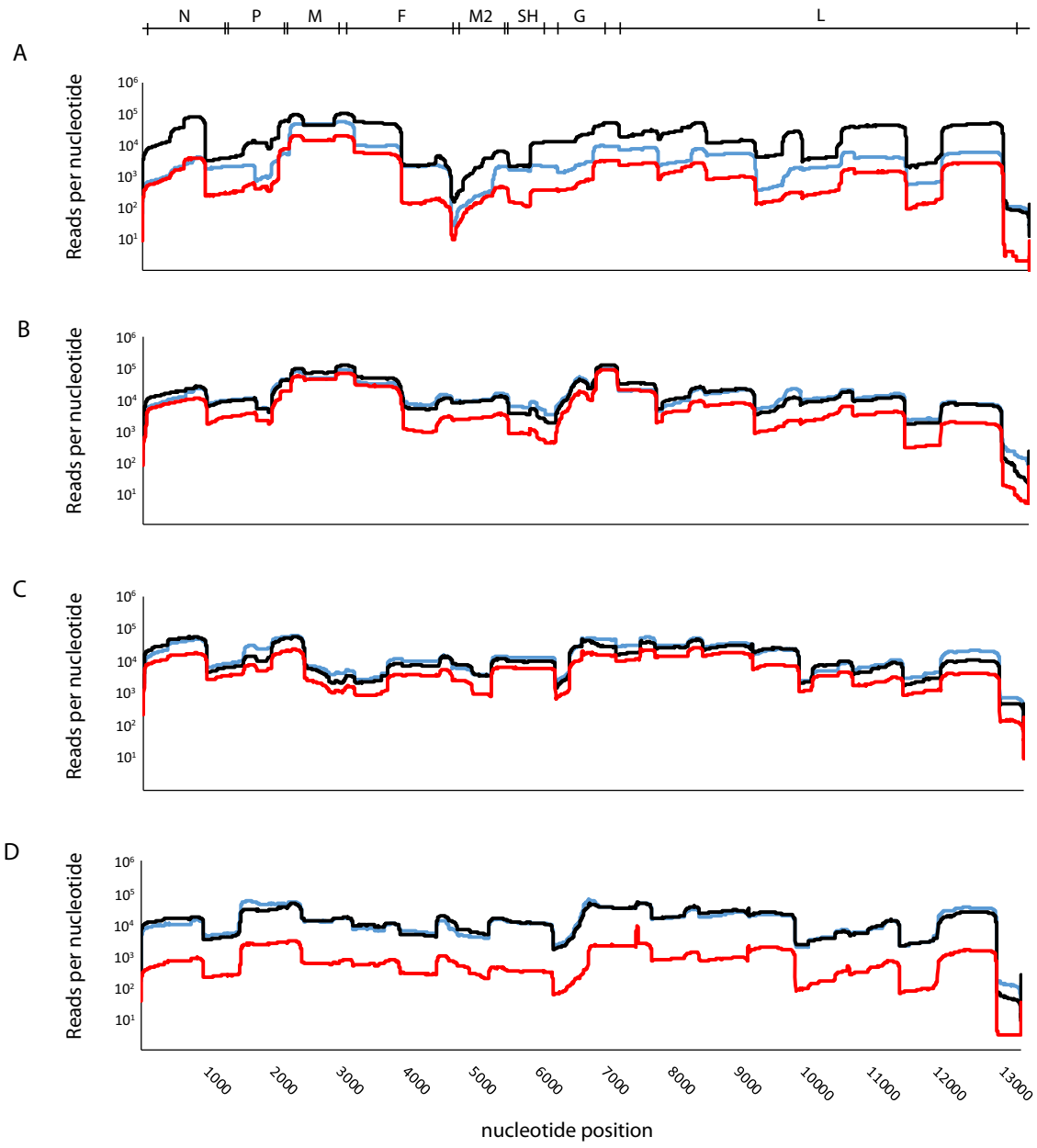

Fig. 3. Coverage plots generated by MinION sequencing of serially diluted HMPV reference strains. Three dilutions of each reference strain were sequenced, with Ct values ranging from 17.3 to 25.0 obtained by qRT-PCR. Coverage indicated on the y-axis is expressed as the number of reads per nucleotide. (a) Coverage plots of serially diluted HMPV NL/00/1. Blue line: Ct 17.8, black line: Ct 20.0, red line: Ct 23.0. (b) Coverage plots of serially diluted HMPV NL/17/00. Blue line: Ct 17.3, black line: Ct 20.0, red line: Ct 22.9. (c) Coverage plots of serially diluted HMPV NL/99/1. Blue line: Ct 18.8, black line: Ct 21.7, red line: Ct 25.0. (d) Coverage plots of serially diluted HMPV NL/94/1. Blue line: Ct 17.3, black line: Ct 19.4, red line: Ct 24.0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of MinION sequencing results of HMPV-positive clinical samples. The mean genomic coverage is expressed as the mean number of reads per nucleotide in the genome.

Clinical sample	Material	Ct value	Genotype	Total reads obtained (x1000)	mapped reads (x1000)	Mean genomic coverage (x1000)
1	Throat swab	20.1	A1	393	292 (74.3%)	18.5
2	Nasal wash	22.4	A2	580	362 (62.4%)	16.2
3	Sputum	19.5	A2	365	307 (83.8%)	8.4
4	Sputum	21	A2	400	316 (79%)	8.1
5	Nasal wash	21.1	B1	376	206 (54.9%)	12.5
6	Sputum	18.7	B1	163	133 (81.6%)	6.1
7	Throath swab	22.5	B1	110	58 (52.7%)	2.5
8	Lung Lavage	21.6	B2	208	78 (37,7%)	4.9
9	Lung Lavage	21.6	B2	208	78 (37,7%)	4.9
10	Nasal wash	22.7	B2	391	269 (68,7%)	9.0

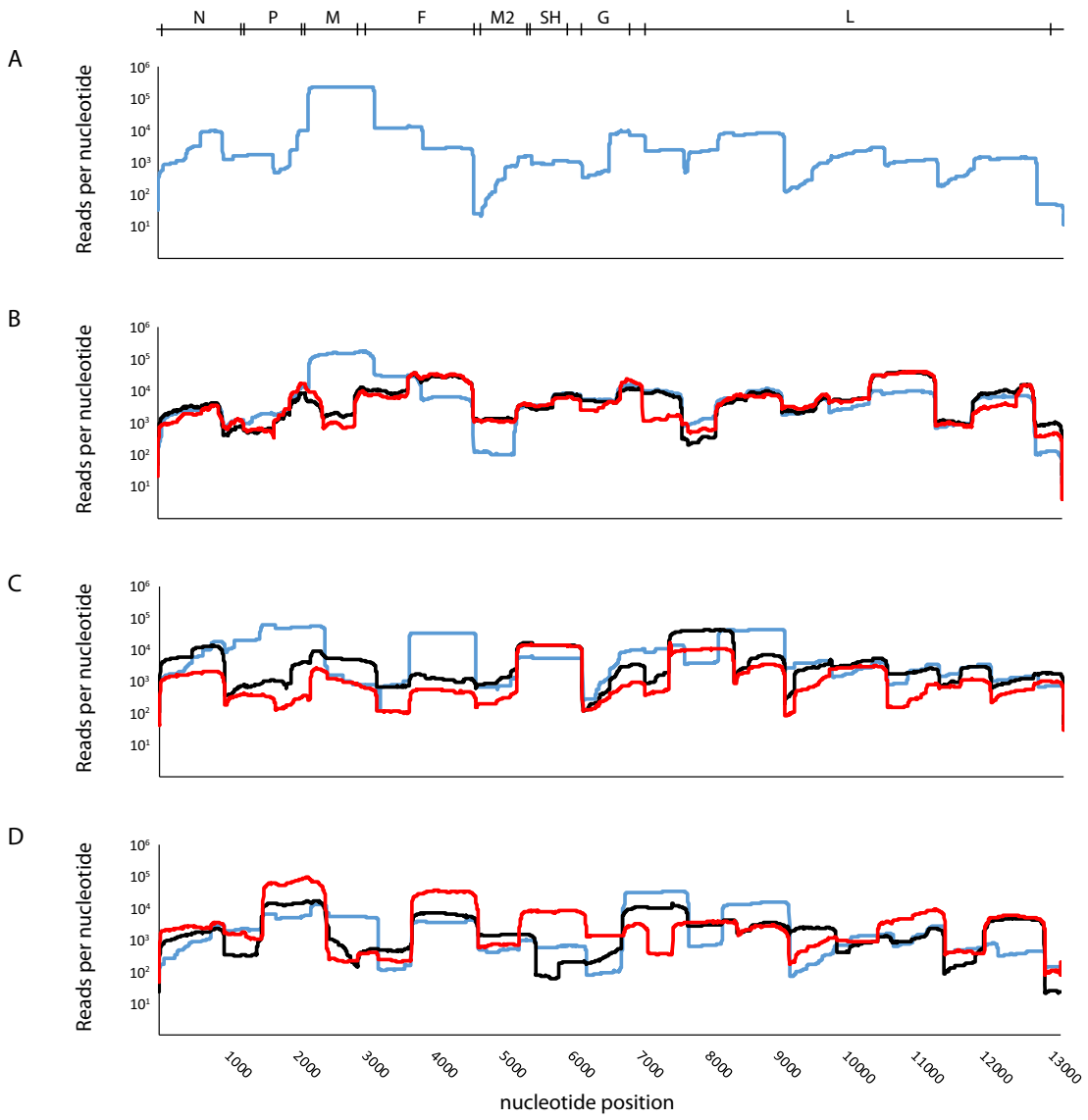

Fig. 4. Coverage plots generated by MinION sequencing of 10 HMPV-positive clinical isolates. Ct values of HMPV clinical samples ranged from 18.7 to 22.7. Coverage indicated on the y-axis is expressed as the number of reads per nucleotide. (a) Coverage plot for a virus belonging to the A1 sublineage with a Ct of 20.1 (b) Coverage plots for viruses belonging to the A2 sublineage. Blue line: sample 2 (Ct 22.4), black line: sample 3 (Ct 19.5), red line: sample 4 (Ct 21.0). (c) Coverage plots for viruses belonging to the B1 sublineage. Blue line: sample 5 (Ct 21.1), black line: sample 6 (Ct 18.7), red line: sample 7 (Ct 22.5). (d) Coverage plots for viruses belonging to the B2 sublineage. Blue line: sample 8 (Ct 21.6), black line: sample 9 (Ct 22.4), red line: sample 10 (Ct 22.7). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a low band intensity on gel. Adjustments to the primer concentrations did affect genomic coverage, as higher primer concentrations resulted in higher genomic coverage obtained by MinION sequencing. Using this method, a HMPV whole genome sequence was obtained from samples with an RNA input that equals a Ct value up to 25.0. However, the genomic coverage of the 5′ end of the genome from samples with Ct values above 23.0 was relatively low. Additionally, we used our assay to sequence one clinical specimen with a Ct value of 28.0. This sample had reduced genomic coverage as well as regions with poor genomic coverage (in comes cases even below 10 reads per nt).

A number of studies explored the option of HMPV whole genome sequencing, either by Illumina sequencing, Ion Torrent, or Nanopore technology (Kamau et al., 2020; Pollett et al., 2018; Xu et al., 2020, 2018). All three methods can be used to generate whole genome sequences, but the cost and scale of MinION sequencing is a benefit of the Nanopore technology for sequencing of viral genomes (Oude Munnink et al., 2019; Van Nimwegen et al., 2016). While metagenomics Nanopore sequencing of HMPV-positive clinical specimens can yield HMPV whole genome sequences, the sensitivity of metagenomic sequencing, as well as the coverage of the G gene, was sub-optimal in general (Xu et al., 2020, 2018). As the HMPV G gene is used for virus classification, good coverage of this gene in HMPV whole genome sequencing techniques is desired, but this can be challenging due to the high genetic variability (between 50%–57% nucleotide identity between lineages A and B) (Xu et al., 2018). We aimed to develop a whole genome sequencing method that detects HMPV from genotype A and B with equal sensitivity and with good coverage of the G gene. To this end, degeneracies were introduced in primer sequences or multiple primers were used for each PCR amplicon to improve the sensitivity of this assay. The introduction of multiple degenerate primers per amplicon resulted in an optimal coverage of the G gene for strain of all lineages.

A recent study, analyzing 2212 sequences, highlighted a notable difference between partial and whole genome sequences of respiratory syncytial virus (Ramaekers et al., 2020). Phylogenetic analyses of ten

whole genome sequences obtained in our study, combined with 70 sequences obtained from Genbank, revealed no differences with phylogenetic analysis of only the fusion protein gene of these sequences (data not shown). This could be due to the limited availability of HMPV whole-genome sequences.

In conclusion, we introduce a HMPV whole genome sequencing technique using a MinION platform with virus-specific PCR amplicons. This assay is able to generate HMPV whole genome sequences from clinical specimens of all HMPV genotypes with good genomic coverage throughout the genome, including the variable G gene. This method can be used for in-depth studies of HMPV epidemiology and virus evolution, for following transmission routes during virus outbreaks, and can provide the foundation for the design of novel intervention strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data statement

The data generated during this study are available from the corresponding author upon request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Kevin Groen: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Stefan van Nieuwkoop:** Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Theo M. Bestebroer:** Investigation, Methodology. **Pieter L. Fraaij:** Resources, Writing - review & editing. **Ron A.M. Fouchier:** Conceptualization, Funding acquisition, Resources, Supervision, Writing - review & editing. **Bernadette G. van den Hoogen:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing - review & editing.

Acknowledgements

The authors would like to thank Dr. B. B. Oude Munnink for helping with troubleshooting during the optimization process of the assay and the ErasmusMC virology diagnostic department for carrying out the qRT-PCR on clinical specimens.

References

Boivin, G., Mackay, I., Sloots, T.P., Madhi, S., Freymuth, F., Wolf, D., Shemer-Avni, Y., Ludewick, H., Gray, G.C., LeBlanc, É., 2004. Global genetic diversity of human metapneumovirus fusion gene. Emerg. Infect. Dis. 10, 1154–1157. https://doi.org/10.3201/eid1006.031097.

Dudas, G., Bedford, T., 2019. The ability of single genes vs full genomes to resolve time and space in outbreak analysis. BMC Evol. Biol. 19, 1–27. https://doi.org/10.1186/s12862-019-1567-0.

Ebihara, T., Endo, R., Kikuta, H., Ishiguro, N., Ishiko, H., Hara, M., Takahashi, Y., Kobayashi, K., 2004. Human Metapneumovirus Infection in Japanese Children. J. Clin. Microbiol. 42, 126–132. https://doi.org/10.1128/JCM.42.1.126-132.2004.

Falsey, A.R., Erdman, D., Anderson, L.J., Walsh, E.E., 2003. Human metapneumovirus infections in young and elderly adults. J. Infect. Dis. 187, 785–790. https://doi.org/10.1086/367901.

Herfst, S., de Graaf, M., Schickli, J.H., Tang, R.S., Kaur, J., Yang, C.-F., Spaete, R.R., Haller, A.A., van den Hoogen, B.G., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2004. Recovery of human metapneumovirus genetic lineages A and B from cloned cDNA. J. Virol. 78, 8264–8270. https://doi.org/10.1128/jvi.78.15.8264-8270.2004.

Herfst, S., de Graaf, M., Schrauwen, E.J.A., Ulbrandt, N.D., Barnes, A.S., Senthil, K., Osterhaus, A.D.M.E., Fouchier, R.A.M., van den Hoogen, B.G., 2007. Immunization of Syrian golden hamsters with F subunit vaccine of human metapneumovirus induces protection against challenge with homologous or heterologous strains. J. Gen. Virol. 88, 2702–2709. https://doi.org/10.1099/vir.0.83084-0.

Hoek, R.A.S., Paats, M.S., Pas, S.D., Bakker, M., Hoogsteden, H.C., Boucher, C.A.B., Van Der Eerden, M.M., 2013. Incidence of viral respiratory pathogens causing exacerbations in adult cystic fibrosis patients. Scand. J. Infect. Dis. 45, 65–69. https://doi.org/10.3109/00365548.2012.708942.

Huck, B., Scharf, G., Neumann-Haefelin, D., Puppe, W., Weigl, J., Falcone, V., 2006. Novel human metapneumovirus sublineage. Emerg. Infect. Dis. 12, 147–150. https://doi.org/10.3201/eid1201.050772.

Jain, S., Self, W.H., Wunderink, R.G., Fakhran, S., Balk, R., Bramley, A.M., Reed, C., Grijalva, C.G., Anderson, E.J., Courtney, D.M., Chappell, J.D., Qi, C., Hart, E.M., Carroll, F., Trabue, C., Donnelly, H.K., Williams, D.J., Zhu, Y., Arnold, S.R., Ampofo, K., Waterer, G.W., Levine, M., Lindstrom, S., Winchell, J.M., Katz, J.M., Erdman, D., Schneider, E., Hicks, L.A., McCullers, J.A., Pavia, A.T., Edwards, K.M., Finelli, L., 2015. Community-acquired pneumonia requiring hospitalization among U.S. adults. N. Engl. J. Med. 373, 415–427. https://doi.org/10.1056/nejmoa1500245.

Kamau, E., Oketch, J.W., De Laurent, Z.R., Phan, M.V.T., Agoti, C.N., Nokes, D.J., Cotten, M., 2020. Whole genome sequencing and phylogenetic analysis of human metapneumovirus strains from Kenya and Zambia. BMC Genomics 21, 1–10. https://doi.org/10.1186/s12864-020-6498-z.

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., Smith, D.B., 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 46, D708–D717. https://doi.org/10.1093/nar/gkx932.

Mackay, I.M., Bialasiewicz, S., Jacob, K.C., McQueen, E., Arden, K.E., Nissen, M.D., Sloots, T.P., 2006. Genetic diversity of human metapneumovirus over 4 consecutive years in Australia. J. Infect. Dis. 193, 1630–1633. https://doi.org/10.1086/504260.

Maertzdorf, J., Wang, C.K., Brown, J.B., Quinto, J.D., Chu, M., De Graaf, M., Van Den Hoogen, B.G., Spaete, R., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2004. Real-time reverse transcriptase PCR assay for detection of human metapneumoviruses from all known genetic lineages. J. Clin. Microbiol. 42, 981–986. https://doi.org/10.1128/JCM.42.3.981-986.2004.

Nao, N., Saikusa, M., Sato, K., Sekizuka, T., Usuku, S., Tanaka, N., Nishimura, H., Takeda, M., 2020. Recent molecular evolution of human metapneumovirus (HMPV): subdivision of HMPV a2b strains. Microorganisms 8, 1–14. https://doi.org/10.3390/microorganisms8091280.

O'Brien, K.L., Baggett, H.C., Brooks, W.A., Feikin, D.R., Hammitt, L.L., Higdon, M.M., Howie, S.R.C., Knoll, Deloria, M., Kotloff, K.L., Levine, O.S., Madhi, S.A., Murdoch, D.R., Prosperi, C., Scott, J.A.G., Shi, Q., Thea, D.M., Wu, Z., Zeger, S.L., Adrian, V., P., Akarasewi, P., Anderson, T.P., Antonio, M., Awori, J.O., Baillie, V.L., Bunthi, C., Chipeta, J., Chisti, M.J., Crawley, J., DeLuca, A.N., Driscoll, A.J., Ebruke, B.E., Endtz, H.P., Fancourt, N., Fu, W., Goswami, D., Groome, M.J., Haddix, M., Hossain, L., Jahan, Y., Kagucia, E.W., Kamau, A., Karron, R.A., Kazungu, S., Kourouma, N., Kuwanda, L., Kwenda, G., Li, M., Machuka, E.M., Mackenzie, G., Mahomed, N., Maloney, S.A., McLellan, J.L., Mitchell, J.L., Moore, D. P., Morpeth, S.C., Mudau, A., Mwananyanda, L., Mwansa, J., Silaba Ominde, M., Onwuchekwa, U., Park, D.E., Rhodes, J., Sawatwong, P., Seidenberg, P., Shamsul, A., Simões, E.A.F., Sissoko, S., Wa Somwe, S., Sow, S.O., Sylla, M., Tamboura, B., Tapia, M.D., Thamthitiwat, S., Toure, A., Watson, N.L., Zaman, K., Zaman, S.M.A., 2019. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. Lancet 394, 757–779. https://doi.org/10.1016/S0140-6736(19)30721-4.

Oude Munnink, B.B., Kik, M., de Bruijn, N.D., Kohl, R., van der Linden, A., Reusken, C.B. E.M., Koopmans, M., 2019. Towards high quality real-time whole genome sequencing during outbreaks using Usutu virus as example. Infect. Genet. Evol. 73, 49–54. https://doi.org/10.1016/j.meegid.2019.04.015.

Pollett, S., Trovão, N.S., Tan, Y., Eden, J.S., Halpin, R.A., Bera, J., Das, S.R., Wentworth, D., Ocaña, V., Mendocilla, S.M., Álvarez, C., Calisto, M.E., Garcia, J., Halsey, E., Ampuero, J.S., Nelson, M.I., Leguia, M., 2018. The transmission dynamics and diversity of human metapneumovirus in Peru. Influenza Other Respi. Viruses 12, 508–513. https://doi.org/10.1111/irv.12537.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-Ximenez, L.L., Jesus, J.G.de, Giovanetti, M., Hill, S., Black, A., Bedford, T., Carroll, M.W., Nunes, M., Alcantara Jr, L.C., Sabino, E.C., Baylis, S.A., Faria, N., Loose, M., Simpson, J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2018. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. 12, 1261–1276. https://doi.org/10.1038/nprot.2017.066.Multiplex.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. J. Gen. Virol. 93, 1853–1868. https://doi.org/10.1099/vir.0.043182-0.

Ramaekers, K., Rector, A., Cuypers, L., Lemey, P., Keyaerts, E., van Ranst, M., 2020. Towards a unified classification for human respiratory syncytial virus genotypes. Virus Evol. 6, 1–11. https://doi.org/10.1093/ve/veaa052.

Saikusa, M., Kawakami, C., Nao, N., Takeda, M., Usuku, S., Sasao, T., Nishimoto, K., Toyozawa, T., 2017aa. 180-nucleotide duplication in the G Gene of Human metapneumovirus A2b subgroup strains circulating in Yokohama City, Japan, since 2014. Front. Microbiol. 8, 1–11. https://doi.org/10.3389/fmicb.2017.00402.

Saikusa, M., Nao, N., Kawakami, C., Usuku, S., Sasao, T., Toyozawa, T., Takeda, M., Okubo, I., 2017ba. A novel 111-nucleotide duplication in the G gene of human

metapneumovirus. Microbiol. Immunol. 61, 507–512. https://doi.org/10.1111/1348-0421.12543.

Saikusa, M., Nao, N., Kawakami, C., Usuku, S., Tanaka, N., Tahara, M., Takeda, M., Okubo, I., 2019. Predominant detection of the subgroup a2b human metapneumovirus strain with a 111-nucleotide duplication in the g gene in Yokohama city, Japan in 2018. Jpn. J. Infect. Dis. 72, 350–352. https://doi.org/10.7883/yoken.JJID.2019.124.

Siddle, K.J., Eromon, P., Barnes, K.G., Mehta, S., Oguzie, J.U., Odia, I., Schaffner, S.F., Winnicki, S.M., Shah, R.R., Qu, J., Wohl, S., Brehio, P., Iruolagbe, C., Aiyepada, J., Uyigue, E., Akhilomen, P., Okonofua, G., Ye, S., Kayode, T., Ajogbasile, F., Uwanibe, J., Gaye, A., Momoh, M., Chak, B., Kotliar, D., Carter, A., Gladden-Young, A., Freije, C.A., Omoregie, O., Osiemi, B., Muoebonam, E.B., Airende, M., Enigbe, R., Ebo, B., Nosamiefan, I., Oluniyi, P., Nekoui, M., Ogbaini-Emovon, E., Garry, R.F., Andersen, K.G., Park, D.J., Yozwiak, N.L., Akpede, G., Ihekweazu, C., Tomori, O., Okogbenin, S., Folarin, O.A., Okokhere, P.O., MacInnis, B.L., Sabeti, P. C., Happi, C.T., 2018. Genomic analysis of lassa virus during an increase in cases in Nigeria in 2018. N. Engl. J. Med. 379, 1745–1753. https://doi.org/10.1056/nejmoa1804498.

Skiadopoulos, M.H., Biacchesi, S., Buchholz, U.J., Amaro-Carambot, E., Surman, S.R., Collins, P.L., Murphy, B.R., 2006. Individual contributions of the human metapneumovirus F, G, and SH surface glycoproteins to the induction of neutralizing antibodies and protective immunity. Virology 345, 492–501. https://doi.org/10.1016/j.virol.2005.10.016.

Van Den Hoogen, B.G., De Jong, J.C., Groen, J., Kuiken, T., De Groot, R., Fouchier, R.A. M., Osterhaus, A.D.M.E., 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat. Med. 7, 719–724. https://doi.org/10.1038/89098.

van den Hoogen, B.G., Herfst, S., de Graaf, M., Sprong, L., van Lavieren, R., van Amerongen, G., Yüksel, S., Fouchier, R.A.M., Osterhaus, A.D.M.E., de Swart, R.L., 2007. Experimental infection of macaques with human metapneumovirus induces transient protective immunity. J. Gen. Virol. 88, 1251–1259. https://doi.org/10.1099/vir.0.82663-0.

Van Den Hoogen, B.G., Herfst, S., Sprong, L., Cane, P.A., Forleo-Neto, E., De Swart, R.L., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2004. Antigenic and Genetic Variability of Human Metapneumoviruses. Emerg. Infect. Dis. 10, 658–666. https://doi.org/10.3201/eid1004.030393.

Van Nimwegen, K.J.M., Van Soest, R.A., Veltman, J.A., Nelen, M.R., Van Der Wilt, G.J., Vissers, L.E.L.M., Grutters, J.P.C., 2016. Is the 1000 genome as near as we think? A cost analysis of next-generation sequencing. Clin. Chem. 62, 1458–1464. https://doi.org/10.1373/clinchem.2016.258632.

Williams, J.V., Wang, C.K., Yang, C., Tollefson, S.J., House, F.S., Heck, J.M., Chu, M., Brown, J.B., Lintao, L.D., Quinto, J.D., Chu, D., Spaete, R.R., Edwards, K.M., Wright, P.F., Crowe Jr., J.E., 2006. The role of human metapneumovirus in upper respiratory tract infections in children: a 20-year experience. J. Infect. Dis. 193, 387–395. https://doi.org/10.1086/499274.

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., 2004. Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. N. Engl. J. Med. 350, 687–696. https://doi.org/10.1056/NEJMoa1505949.

Xu, Y., Lewandowski, K., Jeffery, K., Downs, L.O., Foster, D., Sanderson, N.D., Kavanagh, J., Vaughan, A., Salvagno, C., Vipond, R., Carroll, M., Danby, R., Peto, T., Crook, D., Walker, A.S., Matthews, P.C., Pullan, S.T., 2020. Nanopore metagenomic sequencing to investigate nosocomial transmission of human metapneumovirus from a unique genetic group among haematology patients in the United Kingdom. J. Infect. 80, 571–577. https://doi.org/10.1016/j.jinf.2020.02.003.

Xu, Y., Lewandowski, K., Lumley, S., Sanderson, N., Ali, V., Vipond, R., Carroll, M., Jeffery, K., Foster, D., Walker, A.S., Peto, T., Crook, D., Pullan, S.T., Matthews, P.C., 2018. Nanopore metagenomic sequencing of full length human metapneumovirus (HMPV) within a unique sub-lineage. bioRxiv.

Saikusa, M., Kawakami, C., Nao, N., Takeda, M., Usuku, S., Sasao, T., Nishimoto, K., Toyozawa, T., 2017ab. 180-nucleotide duplication in the G Gene of Human metapneumovirus A2b subgroup strains circulating in Yokohama City, Japan, since 2014. Front. Microbiol. 8, 1–11. https://doi.org/10.3389/fmicb.2017.00402.

Saikusa, M., Nao, N., Kawakami, C., Usuku, S., Sasao, T., Toyozawa, T., Takeda, M., Okubo, I., 2017bb. A novel 111-nucleotide duplication in the G gene of human metapneumovirus. Microbiol. Immunol. 61, 507–512. https://doi.org/10.1111/1348-0421.12543.