Journal of Infection 87 (2023) 103–110

Contents lists available at ScienceDirect

Journal of Infection

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

The emergence, impact, and evolution of human metapneumovirus variants from 2014 to 2021 in Spain

Maria Piñana [a,1,2], Alejandra González-Sánchez [a,1,3], Cristina Andrés [a,4], Michel Abanto [b,5], Jorgina Vila [c,6], Juliana Esperalba [a,7], Noelia Moral [d], Elena Espartosa [d], Narcís Saubi [a,8], Anna Creus [c], Maria Gema Codina [a,9], Dolores Folgueira [d], Jaime Martinez-Urtaza [e,10], Tomàs Pumarola [a,*,11,12], Andrés Antón [a,11,13]

[a] Respiratory Viruses Unit, Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Universitat Autònoma de Barcelona, Barcelona, Spain
[b] Genomics and Bioinformatics Unit, Scientific and Technological Bioresource Nucleus (BIOREN), Universidad de La Frontera, Temuco, Chile
[c] Paediatric Hospitalization Unit, Paediatrics Department, Hospital Universitari Maternoinfantil Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
[d] Department of Clinical Microbiology, Hospital Universitario 12 de Octubre, Instituto de Investigación Hospital 12 de Octubre (imas12), Universidad Complutense de Madrid, Madrid, Spain
[e] Department of Genetics and Microbiology, School of Biosciences, Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain

ARTICLE INFO

Article history:
Accepted 3 May 2023
Available online 11 May 2023

Keywords:
Human metapneumovirus
Duplication
Epidemiology
Evolution
Whole-genome sequencing

SUMMARY

Background: Human metapneumovirus (HMPV) is an important aetiologic agent of respiratory tract infection (RTI). This study aimed to describe the prevalence, genetic diversity, and evolutionary dynamics of HMPV.
Methods: Laboratory-confirmed HMPV were characterised based on partial-coding G gene sequences with MEGA.v6.0. WGS was performed with Illumina, and evolutionary analyses with Datamonkey and Nextstrain.
Results: HMPV prevalence was 2.5%, peaking in February-April and with an alternation in the predominance of HMPV-A and –B until the emergence of SARS-CoV-2, not circulating until summer and autumn-winter 2021, with a higher prevalence and with the almost only circulation of $A2c_{111dup}$. G and SH proteins were the most variable, and 70% of F protein was under negative selection. Mutation rate of HMPV genome was 6.95×10^{-4} substitutions/site/year.
Conclusion: HMPV showed a significant morbidity until the emergence of SARS-CoV-2 pandemic in 2020, not circulating again until summer and autumn 2021, with a higher prevalence and with almost the only circulation of $A2c_{111dup}$, probably due to a more efficient immune evasion mechanism. The F protein showed a very conserved nature, supporting the need for steric shielding. The tMRCA showed a recent emergence of the A2c variants carrying duplications, supporting the importance of virological surveillance.

* Correspondence to: Respiratory Viruses Unit – Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Passeig Vall d'Hebron 119-129, Barcelona, Spain.
E-mail address: virusrespiratoris@vallhebron.cat (T. Pumarola).
[1] These first authors contributed equally to this article.
[2] ORCID: 0000-0002-4766-2613.
[3] ORCID: 0000-0002-1809-3415.
[4] ORCID: 0000-0002-3200-0895.
[5] ORCID: 0000-0001-9136-0041.
[6] ORCID: 0000-0001-9715-0627.
[7] ORCID: 0000-0003-1326-1341.
[8] ORCID: 0000-0001-5493-1044.
[9] ORCID: 0000-0002-5313-5086.
[10] ORCID: 0000-0001-6219-0418.
[11] These two authors are co-principal investigators.
[12] ORCID: 0000-0002-5171-7461.
[13] ORCID: 0000-0002-1476-0815.

https://doi.org/10.1016/j.jinf.2023.05.004

Introduction

Human metapneumovirus (HMPV) is an important aetiologic agent of upper and lower respiratory tract infections (URTI and LRTI) in children and adults.[1] HMPV belongs to the *Pneumoviridae* family together with human respiratory syncytial virus (HRSV),[2] causing similar symptomatology.[1] HMPV is an enveloped, lineal, negative-sensed, single-stranded RNA virus, classified into HMPV-A and HMPV-B genotypes and subdivided into subgenotypes A1, A2 (A2a, A2b and A2c lineages),[1,3,4] B1 and B2 (B2a and B2b lineages).[5,6]

HMPV's genome comprise 8 genes within its 13 kb, encoding for 9 proteins in the following scheme: 3'-N-P-M-F-M2(M2–1/M2–2)-SH-G-L-5'. Recently, 180- and 111-nucleotide duplications have been described within the attachment glycoprotein's (G) ectodomain,[6–9] and have been associated with immune evasion and LRTI in the adulthood.[6]

The aims of this study were to describe the genetic diversity and the evolutionary dynamics of HMPV in paediatric and adult patients attended at a tertiary university hospital in Barcelona from the 2014–2015 to the 2020–2021 seasons.

Materials and methods

Sample collection

From October 2014 to March 2020, respiratory specimens from patients with suspicion of RTI attended at the Hospital Universitari Vall d'Hebron (HUVH) were collected. From March 2020, due to the increasing SARS-CoV-2 diagnostic activity during the pandemic, the routine performed changed, and while the study of all respiratory samples from children for surveillance purposes was intended to be maintained, in adults it was only performed under request. Also, the inclusion criteria changed, including those patients requiring hospitalisation for non-RTI causes, such as trauma patients or surgical interventions. This is due to the fact that, because of the SARS-CoV-2 pandemic, a policy was put in place in the hospital for the screening of all patients in order to cohort them as laboratory-confirmed SARS-CoV-2 or not, and the remaining respiratory viruses were also screened. Institutional Review Board approval (PR(AG)161/2016) was obtained from the hospital's Clinical Research Ethics Committee.

Respiratory viruses' laboratory-confirmation

Detection of respiratory viruses was performed by immunofluorescence (D3 Ultra 8™ DFA Respiratory Virus, Diagnostic HYBRIDS, USA, which targets Influenza A and B viruses, HRSV, HMPV, Adenovirus, and Parainfluenza viruses 1–3) until season 2015–2016 or mainly by real-time multiplex RT-PCR (Anyplex II RV16, Seegene, Korea, until 2015; Allplex Respiratory Panels 1–3, Seegene, Korea, since 2015, which targets Influenza A and B viruses, HRSV, HMPV, Adenovirus, Parainfluenza viruses 1–4, Rhinovirus, Enterovirus, Bocavirus, and Coronaviruses 229E, NL63 and OC43).

Importantly, during the HRSV and influenza epidemics (highest circulation from November to February approximately), rapid tests were performed for diagnosis of HRSV and Influenza A and B viruses in cases with a clinical emergency. These rapid tests were based on immunochromatography (Alere BinaxNOW® Influenza A&B or RSV, Alere, USA), immunofluorescence (Sofia RSV or influenza FIA, Quidel, USA) or real-time RT-PCR (GenXpert for Flu, RSV and/or SARS-CoV-2, Cepheid, USA). In case of having a positive result for HRSV or Influenza A and B viruses, no more tests were performed.

For epidemiological calculations such as the comparison of the circulation and prevalence of different viruses or that of the same virus throughout different years, epidemiological seasons for the surveillance of respiratory viruses were considered to last from week 40 to week 39 of the following year.

HMPV phylogenetic analysis

Partial *G* gene was sequenced from all HMPV laboratory-confirmed samples as previously described.[6] Nucleotide sequences were assembled using MEGA v6.0,[10] and haplotypes were collapsed with ALTER.[11] The best fit substitution model was determined by MEGA v6.0, and the lowest Bayesian information criterion score model was used for the construction of phylogenetic trees and evaluated with 1000 bootstrap resamplings.

Whole genome sequencing

Samples from HUVH were randomly selected to maintain representativeness among all the genetic clusters observed in the phylogenetic trees of G protein and per season. Since no samples before October 2014 were available at HUVH, Hospital 12 de Octubre (H12O, Madrid, Spain) contributed in the present study with older HMPV laboratory-confirmed samples.

Amplification of HMPV whole genome was performed in 4 overlapping 4-kb-amplicons with the SuperScript™ III Reverse Transcriptase (Invitrogen, USA) with the primers designed by Tulloch et al.,[12] and the following PCR conditions: 55ºC 1 h – 94ºC 2 min – 35x(94ºC 30 s – 56ºC 30 s – 68ºC 4 min 30 s) – 68ºC 10 min. PCR products were purified with AMPure Beads (Beckman Coulter, USA), and quantified with NanoDrop One (ThermoScientific, USA). Amplicons were normalised and pooled. Library preparation was performed with DNA Prep (Illumina, CA, USA), and indexed with IDT® for Illumina® DNA/RNA UD Indexes Set A-D (384 IDX) (Illumina, USA). Products were quality tested with 4200 TapeStation System (Agilent, USA). The final pool was loaded in a MiSeq Reagent Kit 600v3 cartridge (Illumina, CA, USA) and sequenced in a MiSeq platform (Illumina, CA, USA).

FASTQ files were parsed to Trimmomatic v0.39,[13] to filter bad quality reads. Sliding window trimming option was enabled, using a 10-base wide window, and cutting when the average quality within the window drops below 30. In addition, leading and trailing bases below quality 30 were removed. Two different approaches were followed:

For the reference mapping consensus assembly, resulting FASTQ files were mapped to HMPV reference genome NC_039199.1 (retrieved from NCBI) through Minimap2 v2.17-r941.[14] Variant calling over mapped reads was performed by LoFreq v2.1.5,[15] with *–call-indels* option enabled. Consensus sequences were generated by BCFtools v1.14,[16] based on the VCF files previously obtained.

Reads pre-processing for *de novo* assembly went under the same first steps. After, FASTQ files were processed by Bowtie2 v2.3.5.1[17] for host sequences removal, using as host reference sequence the human genome assembly GRCh38. Last step of reads pre-processing was performed by pTrimmer v1.3.4,[18] which consisted of trimming off primer sequences. High-quality filtered reads obtained after pre-processing were *de novo* assembled by SPAdes v3.15.2,[19] using –rnaviral option. Output scaffolds were manually revised and curated using MEGA v6.0.

Performance of both approaches was assessed by comparing G protein sequence obtained by Sanger with the G protein information contained in consensus sequences from *de novo* and reference mapping assemblies.

Evolutionary analysis

All complete genomes of HMPV available in GenBank were downloaded as reference sequences, and temporal signal of all genomes was assessed with TempEst.[20] The date of these sequences was retrieved from GenBank or the original article linked, and the 1st of July was assumed when only the year was published. References with low coverage, generated from isolates of viral culture,

without date or resulting in outliers in TempEst were not included in the analyses.

Positively and negatively selected sites were identified through Datamonkey Webserver[21] with 4 different algorithms: single likelihood ancestor counting (SLAC),[22] fixed effects likelihood (FEL),[22] mixed effects model of evolution (MEME)[23] and a fast, unconstrained Bayesian approximation for inferring selection (FUBAR).[24] The best-fit substitution model was also determined by Datamonkey Webserver. Each site was considered under positive or negative selection only when two or more methods agreed with statistical significance ($p < 0.1$ or Bayes Factor > 50). The mean dN/dS ratio was estimated using the SLAC algorithm.

Conservative blocks were obtained through GBlocks v0.91b for evolutionary analysis, to reduce possible noise caused by repetitive duplications or large insertions found in some sequences. Nucleotide substitution rate per site and the time of the most recent common ancestor (tMRCA) were estimated through Nextstrain[25] using the phylodynamic package TreeTime.[26] The resulting analysis can be visualised in nextstrain.org/groups/valldhebronvirology/human-metapneumovirus-AB.

Results

HMPV epidemiology

From October 2014 (season 2014–2015) to December 2021 (season 2021–2022), a total of 72,263 samples were tested, of which 31,368 (43%) were laboratory-confirmed for at least one respiratory virus (excluding SARS-CoV-2). HMPV was laboratory-confirmed in 1751 samples (2.4% of total samples tested; 5.6% of respiratory viruses-confirmed samples). This virus presented a stable prevalence before the SARS-CoV-2 pandemic, except at 2017–2018, when it presented an increase of prevalence, achieving a 3.6% (Table 1). Until week 11/2020, HMPV presented a marked seasonality pattern, with a high circulation between February and April (Fig. 1). In March 2020, the HMPV epidemic was interrupted by the SARS-CoV-2 pandemic, not circulating again until May 2021 and October 2021, in parallel with HRSV on both peaks (Fig. 1). The prevalence of HMPV in 2020–2021 decreased, but it achieved a 12% of prevalence in paediatric population in 2021–2022, exceeding HRSV's circulation.

The rate of co-detections with respiratory viruses was 24% (288/1192) until the pandemic started, when it increased to a 32% (178/559). The most common were rhinovirus, bocavirus, adenovirus and enterovirus, though during the pandemic seasons the order changed to rhinovirus, enterovirus, adenovirus and bocavirus.

Prior to the SARS-CoV-2 pandemic, adults represented more than 45% of cases (Table 2). The distribution of HMPV by patients' sex was balanced, standing out > 64-year-old female adults or male paediatric patients. From 2020–2021, the median age of paediatric patients increased 1 year.

Phylogenetic analyses of viral strains

Phylogenetic analyses revealed that HMPV-A (800/1363, 59%) and HMPV-B (562/1363, 41%) co-circulated, and there was 1 coinfection HMPV-A/B. Due to belonging to the same clinical episode, 77 samples were excluded. The remaining 311/1751 (18%) samples could not be sequenced due to low viral load or RNA quality.

HMPV-A sequences were classified into A2a (12, 1.5%), A2b (49, 6.1%) and A2c (739, 92.4%), of which 185/739 (25.0%) carried the 180-nucleotide duplication ($A2c_{180dup}$) and 455/739 (61.6%), the 111-nucleotide duplication ($A2c_{111dup}$), remaining 99 (13.4%) $A2c_{wt}$. HMPV-B sequences were subdivided into B1 (259, 46.1%), B2a (14, 2.5%) and B2b (289, 51.4%) (Supplementary Fig. 1). Sequences were submitted to GenBank (OM243131-OM243805).

Both HMPV-A and HMPV-B co-circulated throughout 2014–2021, showing shifts in predominance (Fig. 2). The predominant genotype always remained under 65% until the pandemic, when HMPV-A achieved 90% of predominance. Within HMPV-A, the new variants progressively replaced the previous predominant lineages, leading old circulating lineages to their extinction, while within HMPV-B, B1 and B2 co-circulated, showing shifts in the predominant subgenotype (Fig. 2).

Evolutionary analyses of whole-genome sequences

A total of 166 HMPV viral genomes from HUVH 2014–2021, and 22 from H12O 2009–2015 were successfully sequenced.

Comparison between the two approaches used for genome assembly and the sequences obtained by Sanger revealed that the 111- and 180- nucleotide duplications were only correctly assembled through *de novo* approach, while almost none duplications were obtained by reference mapping (one 111-nucleotide duplication was found, while none 180-nucleotide duplication was assembled). Consequently, all the consensus sequences used for the analysis were assembled following the *de novo* assembly pipeline designed.

Sequencing process outputted a mean value of 412,259.31 reads per sample, 0.23% of which were removed due to bad quality. Removal of host sequences left a median of 94.03% of high-quality reads, and 0.29% reads containing primer sequences were trimmed. Pre-processed reads had a median length of 207.87 nucleotides, leading to a median genome depth of all samples of 4740.6 reads per position.

From 13 seasons studied (2008–2021), 98 HMPV-A samples (6 A2a, 15 A2b and 77 A2c, of which 38 $A2c_{111dup}$ and 27 $A2c_{180dup}$) and 90 HMPV-B (48 B1, 5 B2a and 37 B2b) were selected. One sample carrying $A2c_{wt}$ and $A2c_{180dup}$ was excluded due to the probable crossing in the assembly of the reads, remaining 187 samples from HUVH and H12O.

A total of 164 complete genomes of HMPV were available on GenBank, of which 92 HMPV-A and 44 HMPV-B met the inclusion criteria.

The dN/dS analyses showed that G and SH glycoproteins were the most variable, with different profiles between genotypes, followed by M2–1, M2–2, and P (Table 3). G and SH also had the lowest percentage of sites negatively selected, while M, N, F, and L had more than 70% of sites negatively selected. Almost all the amino acids belonging to antigenic epitopes of the F protein were under purifying selection. Moreover, G, SH, M2–1, L, and P were the only proteins with positively selected sites.

A final multiple-sequence alignment containing 12,279 positions was obtained with GBlocks, discarding the duplications. The evolution and phylogeny of the HMPV variants was made publicly available at nextstrain.org/groups/valldhebronvirology/. The estimated mutation rate was 7.05×10^{-4} substitutions/site/year, higher in HMPV-A than in -B. HMPV-A and HMPV-B were estimated to diverge in 1823 [95% confidence interval (CI) 1814–1832] (Fig. 3). Divergence dates of $A2c_{180dup}$ and $A2c_{111dup}$ were 2012 (CI 2011–2012) and 2013 (CI 2013), respectively.

Discussion

This study reports the genetic diversity, molecular epidemiology and evolution of HMPV strains detected throughout eight consecutive seasons, the last two of them occurring during the SARS-CoV-2 pandemic, at a tertiary university hospital in Catalonia, Spain.

The prevalence of HMPV in pre-pandemic seasons was similar to that observed in other studies.[4,27–29] A yearly seasonal pattern with a clear peak was observed in all pre-pandemic seasons, following the concatenated epidemic peaks of HRSV and influenza virus, and lasting from February to April. However, the start of SARS-CoV-2

Table 1
Prevalence of HMPV before and during SARS-CoV-2 pandemic. Seasons start at week 40 and last until week 39 of the following natural year. Pre-SARS-CoV-2 pandemic seasons include 2014–2020 seasons, while SARS-CoV-2 pandemic seasons include 2020–2022 seasons.

Season	Total samples received	Total samples from paediatric population	Total samples from adult population	Total samples positive for at least 1 respiratory virus (%)	HMPV-positive				
					n	% over total samples	% over paediatric population	% over adult population	% over positive samples
2014–2015	6207	3509	2698	2360 (38%)	103	1.7%	1.8%	1.5%	4.4%
2015–2016	7545	3958	3587	3622 (48%)	167	2.2%	1.9%	2.5%	4.6%
2016–2017	7836	3598	4238	3978 (51%)	167	2.1%	2.7%	1.7%	4.2%
2017–2018	9164	3794	5370	5116 (56%)	333	3.6%	5.1%	2.6%	6.5%
2018–2019	8496	3912	4584	4675 (55%)	231	2.7%	3.5%	2.1%	4.9%
2019–2020	9318	4644	4674	3906 (42%)	191	2.0%	2.0%	2.1%	4.9%
Total pre-pandemic seasons	48,566	23,415	25,151	23,657 (49%)	1192	2.5%	2.8%	2.1%	5.0%
2020–2021	16,466	12,723	3743	5751 (35%)	201	1.2%	1.5%	0.2%	3.5%
2021–2022	7231	2786	4445	1960 (27%)	358	5.0%	12.4%	0.3%	18.3%
Total pandemic seasons	23,697	15,509	8188	7711 (33%)	559	2.4%	3.5%	0.2%	7.2%
OVERALL	72,263	38,924	33,339	31,368 (43%)	1751	2.4%	3.1%	1.7%	5.6%

pandemic[30] changed this pattern, disrupting its circulation. This emerging virus and the use of non-pharmaceutical measures displaced all respiratory viruses until summer 2020, when non-enveloped viruses started circulating (data not shown), but no influenza viruses, HRSV or HMPV were detected that winter nor spring, when expected. This circulation of non-enveloped viruses coincided with the end of the lockdown in Spain and thus, with the increase of social activities, which would enhance their spread. It was not until summer 2021 that HRSV and HMPV started circulating again, out of the typical pattern and drawing two different epidemic peaks, the second one starting in autumn. This could be related to the relaxation of preventive measures by the population, as government measures remained the same, or due to the lack of viral interference, both facts occurring due to a low SARS-CoV-2 incidence at that time. The prevalence of HMPV in the second peak was higher in comparison to previous seasons, probably because there were two generations who had not suffered a primary infection yet, due to the interruption of HMPV circulation.

The most common co-detections were rhinovirus, bocavirus, adenovirus and enterovirus, as previously described.[31] After the emergence of SARS-CoV-2, the co-detection rate increased, and the order of the most common co-detected viruses changed. Both facts could be related with the different seasonality of HMPV, as not only it coincided with HRSV, but also with the burden of enterovirus D68, which arrived one year later than expected due to the pandemic.[32] among other viruses in autumn 2021.

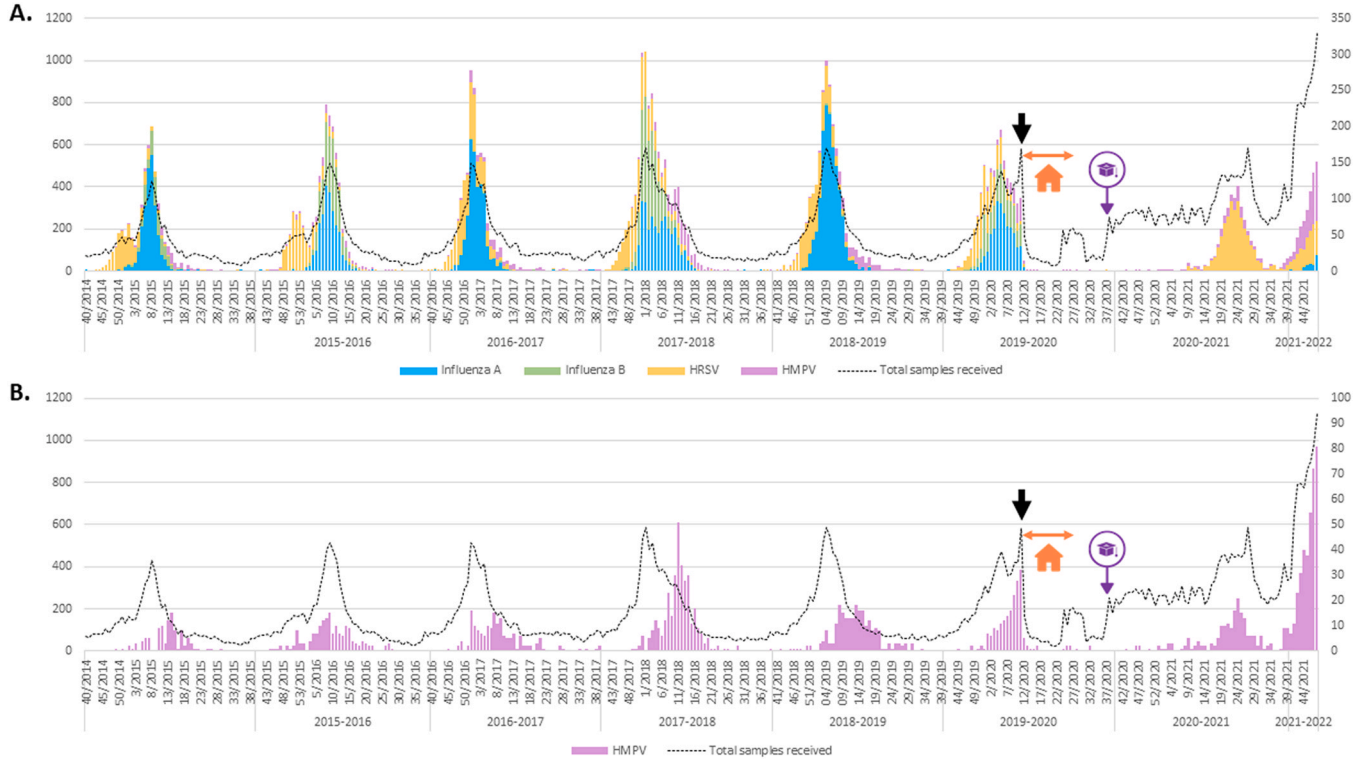

Fig. 1. Seasonality. The X-axis represents the timeline of the study, where epidemiological seasons were considered to last from week 40 to week 39 of the following year. The left Y-axis represents the number of total samples received and the right Y-axis represents the number of samples of each specific virus. The black arrow indicates the moment when the SARS-CoV-2 pandemic led to the lockdown in Spain, the orange arrow indicates the duration of the lockdown in Spain, and the purple arrow indicates the start of school period. A) shows the circulation of Influenza A, Influenza B, HRSV and HMPV, while B) only shows the circulation of HMPV.

Table 2
Characteristics of HMPV laboratory-confirmed patients. Distribution of age and sex for each cohort (paediatric and adult populations).

Season	Paediatric population				Adult population			
	n	F	M	Age (median, IQR)	n	F	M	Age (median, IQR)
2014–2015	62	40,3%	59,7%	1.5 (0.5–3.0)	41	46,3%	53,7%	66.7 (54.4–79.9)
2015–2016	76	44,7%	55,3%	1.4 (0.8–3.2)	91	54,9%	45,1%	69.9 (59.1–78.7)
2016–2017	97	37,1%	62,9%	1.2 (0.6–2.6)	70	52,9%	47,1%	69.1 (58.1–81.5)
2017–2018	192	43,8%	56,3%	1.9 (0.8–3.6)	139	54,0%	46,0%	73.9 (62.1–84.4)
2018–2019	137	51,1%	48,9%	1.6 (0.6–4.3)	92	59,8%	40,2%	64.3 (52.1–79.9)
2019–2020	93	54,8%	45,2%	2.4 (1.3–3.9)	97	54,6%	45,4%	56.0 (51.6–82.7)
Total prepandemic	**657**	**45,7%**	**54,34%**	**1.7 (0.7–3.6)**	**530**	**54,53%**	**45,47%**	**69.5 (55.6–82.0)**
2020–2021	194	45,9%	54,1%	2.8 (1.6–4.6)	7	14,3%	85,7%	36.0 (26.7–61.6)
2021–2022	344	47,7%	52,3%	2.9 (1.5–4.2)	12	25,0%	75,0%	67.0 (60.5–71.0)
Total during pandemic	**538**	**47,0%**	**53,0%**	**2.9 (1.6–4.3)**	**19**	**21,05%**	**78,95%**	**62.9 (32.5–70.8)**

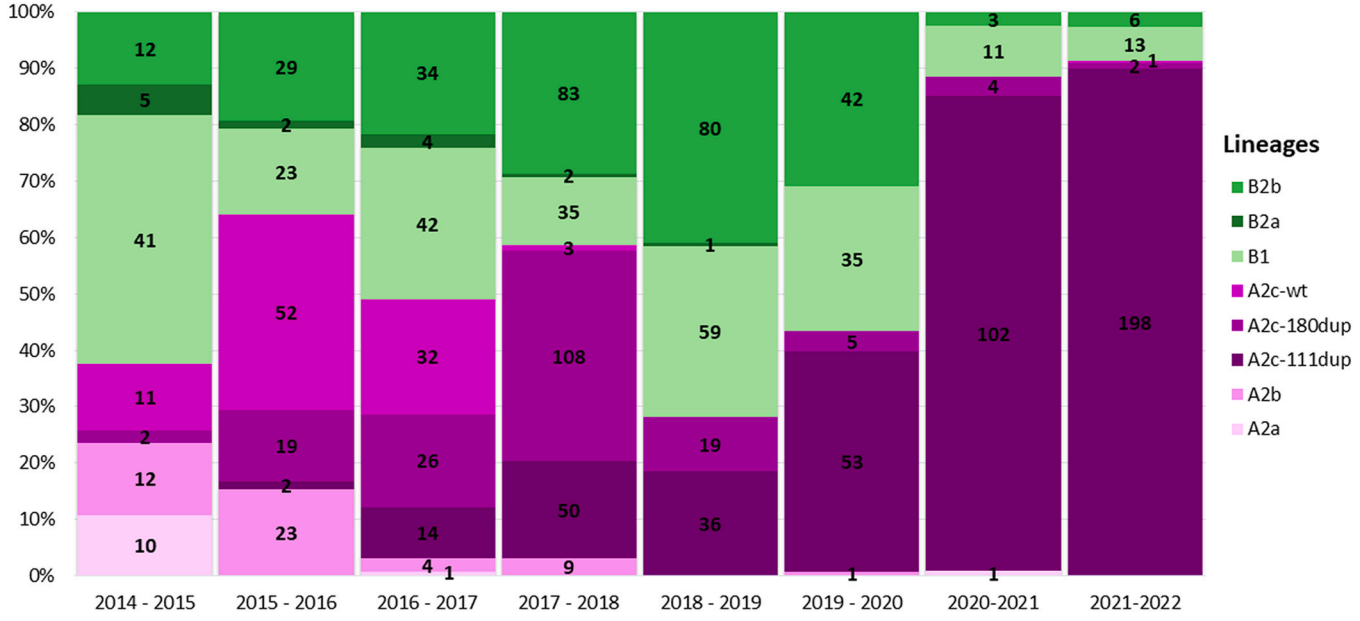

Fig. 2. Circulation of the different subgenotypes and lineages of HMPV. The upper figure compares the circulation of both HMPV-A and –B throughout the study period, while the figures below compare the subgenotypes and lineages within each genotype.

Table 3
Selection profiles for the different proteins of HMPV. Analyses were also calculated for SH and G proteins of HMPV-A and HMPV-B.

Protein	dN/dS	Sites (NO STOP)	Sites negatively selected		Sites positively selected	
			N	%	N	%
G-B	0668	340	34	10%	19	6%
G	0642	340	65	19%	15	4%
G-A	0618	340	55	16%	15	4%
SH-A	0386	186	43	23%	1	1%
SH	0377	186	82	44%	3	2%
SH-B	0367	186	47	25%	2	1%
M2–2	0162	71	36	51%	0	0%
P	0120	294	190	65%	1	0%
M2–1	0.064	187	110	59%	3	2%
L	0054	2005	1416	71%	1	0%
F	0045	539	390	72%	0	0%
N	0035	394	281	71%	0	0%
M	0021	254	187	74%	0	0%

HMPV presented a higher incidence in children under 2 years old in pre-pandemic seasons, as previously observed.[1,2] However, the median age increased after the emergence of SARS-CoV-2, consistent with the hypothesis aforementioned where there was a larger population susceptible to infection.[33,34] Moreover, there was a tendency in affecting more male infants, as described for other respiratory viruses.[35] On the other hand, in the adult population, HMPV presented more incidence in patients aging more than 50, with a tendency to affect more female adults, as observed with other respiratory viruses.[36] It is important to highlight that this study was performed in hospitalised cases, and data might differ whether mild or asymptomatic cases were included. Combined studies from community and hospitals would be valuable to measure the real impact of any respiratory virus, as well as to relate viral features to the disease outcome.

Regarding the circulation of HMPV-A and HMPV-B, a pattern of cyclic shifts of predominance was observed, similar to other studies,[29] though the achievement of a 90% of prevalence by HMPV-A had not been described before the SARS-CoV-2 pandemic occurred. An increasing prevalence of A2c lineage viruses carrying duplications has been described, first for $A2c_{180dup}$, and later for $A2c_{111dup}$. In fact, almost all HMPV-A circulating during the pandemic seasons belonged to $A2c_{111dup}$. This is consistent with what has been observed within HMPV-A, as it seems that emerging lineages replace the previous ones, leading them to extinction. In fact, A1 subgenotype, which was the first described within HMPV-A genotype, has not been detected since 2006, according to literature.[3] The A2a lineage circulated widely during the first season, as well as A2b lineage on that first season. However, A2c lineage rapidly replaced the A2b lineage, predominating on the first, second and third seasons. Though this fact suggested that the A2c lineage had a better

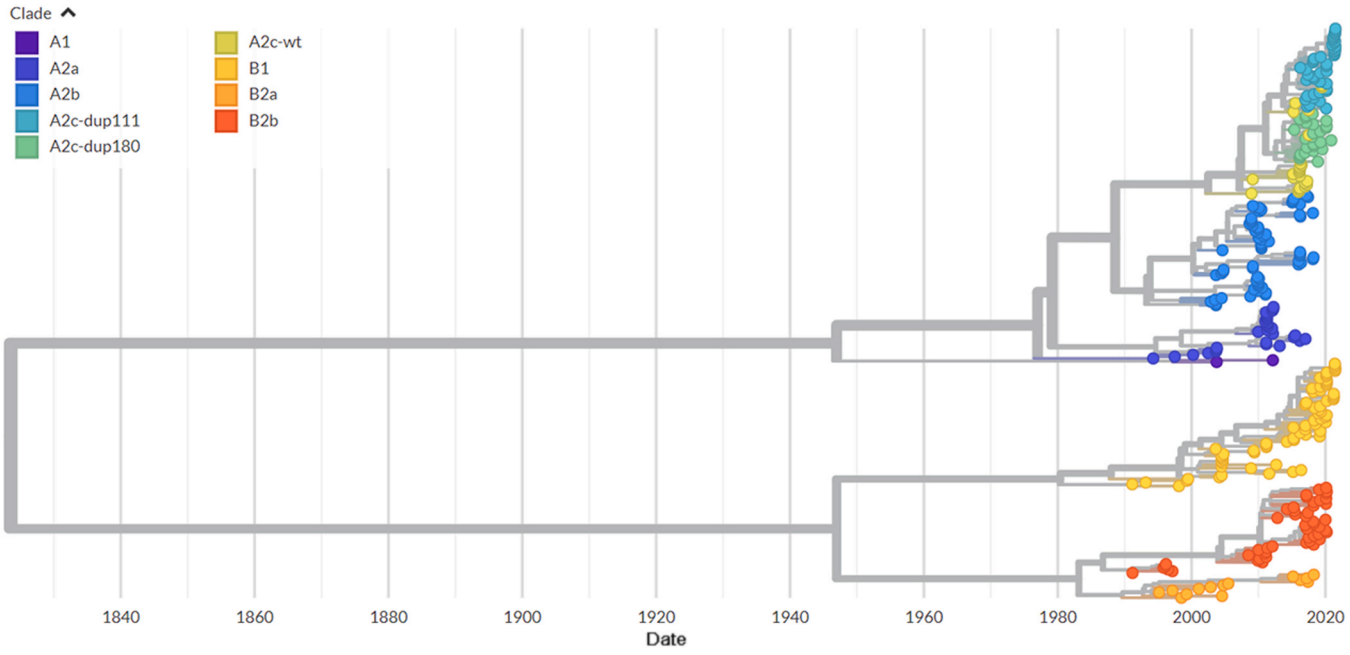

Fig. 3. Phylodynamic analysis of whole-genome sequences from HMPV. The phylogenetic tree es shown as rendered by nextstrain, with sequenced labelled by genetic group. The branch lengths correspond to the sampling date of the sample.

viral fitness, the emergence of A2c lineage viruses carrying duplications demonstrated a better performance, as they achieved a higher predominance in less time than older lineages. Moreover, the duplications within A2c lineage had been related with immune evasion by shielding the fusion protein from neutralising antibodies, $A2c_{180dup}$ protruding farther than $A2c_{111dup}$.[6] This might lead to the interpretation of $A2c_{180dup}$ having a better immune evasion mechanism and thus, becoming the predominant lineage. Instead, the predominant lineage has turned to be $A2c_{111dup}$, which suggests that the 180-nucleotide duplication might cover too much the F protein, hindering the membranes' fusion and thus, preventing proper virus replication, resulting in a worse viral fitness than that of $A2c_{111dup}$.

In line with the emergence and increasing prevalence of these A2c variants and related with their absolute predominance after the SARS-CoV-2 pandemic, an increase of HMPV's morbidity had already been described in the winter of 2010, at the end of the 2009 influenza A(H1N1)pdm09 pandemic.[37] The A2c lineage, previously named A2b2, has been dominant over more years than B1 or B2 subgenotypes, and has been described to have a higher mean genetic distance.[38] Both studies suggested that this lineage could evade more efficiently the immune system due to its more plasticity to acquire genetic changes, which would explain its aggressive predominance. The present study shows that after almost two years of SARS-CoV-2 pandemic, HMPV has circulated again, first as an epidemic with similar number as in previous seasons and immediately after in a second epidemic overcoming all numbers observed along 8 years, and with the indisputable majority of $A2c_{111dup}$, the last emerging lineage of A2 subgenotype.

Interestingly, HMPV-B lineages present a very different dynamic, where B1 and B2b co-exist and draw a pattern of shifting predominance.

Regarding the different approaches used for whole genome sequence assembly, the comparison highlighted the problematic of using reference mapping in terms of finding large structural variants. Using the dataset generated in this study, the assembly by reference mapping did not find the duplications in most cases, probably forcing the reads to align to a reference genome sequence that does not contain these duplications, losing this variant information. This should be taken into account in future studies of WGS of RNA respiratory viruses, as this kind of insertion is quite commonly found.[39,40]

As previously reported, G and SH proteins presented the highest dN/dS and consequently, the lowest amount of sites negatively selected and the highest positively selected. Curiously, they are not major antigenic determinants targeted by neutralising or protective antibodies.[41,42] Also, the differences in the positive selection profiles between HMPV-A and HMPV-B would point to a lineage-specific selection. In addition to these proteins, there was positive selection in L, M2–1 and in P, though the role of the amino acids under positive selection is still unknown. Regarding the F protein, which is the only one eliciting a neutralising immune response, two of the asparagine amino acid residues located on the apex (positions 57 and 172) resulted to be under negative selection. These asparagines are suggested to play a role in immune evasion by shielding the epitopes from the immune system.[43,44] Also, the fact that the antigenic epitopes of the F protein are under purifying selection would suggest that the immune system is not exerting a pressure on this protein, therefore supporting the need for steric shielding of G protein over F protein to modulate the recognition of this antigenic features by neutralising antibodies.[6]

There is scarce literature about the evolution of the whole genome of HMPV. Previous studies had estimated that HMPV-A and -B diverged about 400 years ago. However, this tMRCA varied according to the studied protein, differing in more than 200 years,[42] as well as presenting a confidence interval of more than 700 years.[45] This study contributed with 187 whole-genome sequences dating a range of 12 years, which might explain the more recent tMRCA along with the shorter confidence interval, though it must be noted that this estimation was performed with a maximum likelihood method instead of a Bayesian method. However, the tMRCA calculated with TreeTime were very similar to that calculated with BEAST 2 software (data not shown). Importantly, there is a lack of information about ancestral samples, as this virus was discovered only twenty years ago, although it has circulated at least since 1958.[46] Regarding the tMRCA of $A2c_{180dup}$ and $A2c_{111dup}$, these variants emerged 3–4 years before their first description, which supports the importance of virological surveillance and the need of tracking the evolution of this rapidly evolving virus.

Though HMPV-A and HMPV-B lineages and sublineages had shown different predominance patterns, they showed a directional evolution, continuously evolving to adapt, probably with the objective of evading the immune response acquired in previous seasons, aiming staying in the human population. Regarding $A2c_{180dup}$ and $A2c_{111dup}$, both phylogenetic and evolutionary analyses demonstrated that they emerged from two independent genetic events, as stated by Saikusa.[9] Thus, a new duplication event cannot be discarded, which might happen in HMPV-B, since it shows such a similar profile to HMPV-A's.

Genomic surveillance is important to monitor HMPV evolution, as it allows the detection of emerging variants and the association of this findings to the clinical features of confirmed cases, which is relevant for public health purposes. Nextstrain has proven its usefulness for this, especially during the SARS-CoV-2 pandemic, allowing the comparison of hundreds to thousands sequences.[47] In the present study Nextstrain allows a real-time view of the evolution of HMPV, becoming a model to track the evolution of other respiratory viral pathogens, as it is currently done for enterovirus D68.[48]

In conclusion, HMPV showed a significant morbidity throughout all the study period, affecting both paediatric and adult populations. This virus usually presented its epidemic peak in late winter and early spring in pre-pandemic seasons. However, HMPV epidemic in 2020 was subtly interrupted by the SARS-CoV-2 pandemic, and two epidemic peaks in summer and autumn 2021 were unexpectedly reported, the latter one increasing both the virus' prevalence and the median age of paediatric patients affected. The common shifts in predominance between HMPV-A and -B drastically changed to a vast majority of $A2c_{111dup}$ circulating, probably due to a more efficient immune evasion mechanism. WGS revealed the high diversity in G protein and the high conservation of F protein, supporting the need for steric shielding of G over F. Also, the tMRCA showed a recent emergence of the A2c variants carrying duplications, supporting the need for virological surveillance of HMPV. This study represents a nice example of virological surveillance of a non-HRSV nor influenza respiratory virus, in a tertiary university hospital, providing valuable knowledge of the real-time evolution of HMPV and its impact on health.

Funding

This study was supported by the European Regional Development Fund (ERDF) "A way to achieve Europe", Spanish Network for Research in Infectious Diseases [REIPI RD16/0016/0003], and supported by the Health Research Fund, Spanish Ministry of Economy and Competitiveness [Grant FIS PI18/00685].

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.

Acknowledgements

We would like to thank all the team from the NGS platform and the respiratory viruses unit, both from the Microbiology Department of our hospital, for their restless work since the beginning of the SARS-CoV-2 pandemic.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jinf.2023.05.004.

References

1. Panda S, Mohakud NK, Pena L, Kumar S. *Human metapneumovirus: review of an important respiratory pathogen. Int J Infect Dis* 2014;**25**:45–52.
2. Shafagati N, Williams J. *Human metapneumovirus - what we know now. F1000Research* 2018;**7**.
3. 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* 2017;**89**(11):1885–93.
4. Chow WZ, Chan YF, Oong XY, Ng LJ, Nor'E SS, Ng KT, et al. *Genetic diversity, seasonality and transmission network of human metapneumovirus: identification of a unique sub-lineage of the fusion and attachment genes. Sci Rep* 2016;**6**:1–10.
5. Carr MJ, Waters A, Fenwick F, Toms GL, Hall WW, O'Kelly E. *Molecular epidemiology of human metapneumovirus in Ireland. J Med Virol* 2008;**80**:510–6.
6. Piñana M, Vila J, Maldonado C, Galano-Frutos JJ, Valls M, Sancho J, et al. *Insights into immune evasion of human metapneumovirus: novel 180- and 111-nucleotide duplications within viral G gene throughout 2014-2017 seasons in Barcelona, Spain. J Clin Virol* 2020;**132**:104590.
7. Piñana M, Vila J, Gimferrer L, Valls M, Andrés C, Codina MG, et al. *Novel human metapneumovirus with a 180-nucleotide duplication in the G gene. Future Microbiol* 2017;**12**(7):565–71.
8. Saikusa M, Kawakami C, Nao N, Takeda M, Usuku S, Sasao T, et al. *180-nucleotide duplication in the G gene of human metapneumovirus A2b subgroup strains circulating in Yokohama city, Japan, since 2014. Front Microbiol* 2017;**8**:1–11.
9. Saikusa M, Nao N, Kawakami C, Usuku S, Sasao T, Toyozawa T, et al. *A novel 111-nucleotide duplication in the G gene of human metapneumovirus. Microbiol Immunol* 2017;**61**(11):507–12.
10. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. *MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol* 2013;**30**(12):2725–9.
11. Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D. *ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Res* 2010;**38**(Suppl. 2):S14–8.
12. Tulloch RL, Kok J, Carter I, Dwyer DE, Eden J-S. *An amplicon-based approach for the whole-genome sequencing of human metapneumovirus. Viruses* 2021;**13**(3):499.
13. Bolger AM, Lohse M, Usadel B. *Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics* 2014;**30**(15):2114–20.
14. Li H. *Minimap2: pairwise alignment for nucleotide sequences*. In: Birol I (Ed.) Bioinformatics, 34(18); 2018, p. 3094–100. [Internet] Available from: ⟨https://github.com/ruanjue/smartdenovo⟩.
15. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. *LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res* 2012;**40**(22):11189–201.
16. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. *Twelve years of SAMtools and BCFtools. Gigascience* 2021;**10**(2):1–4.
17. Langmead B, Salzberg SL. *Fast gapped-read alignment with Bowtie 2. Nat Methods* 2012;**9**(4):357–9.
18. Zhang X, Shao Y, Tian J, Liao Y, Li P, Zhang Y, et al. *PTrimmer: an efficient tool to trim primers of multiplex deep sequencing data. BMC Bioinform* 2019;**20**(1):1–6.
19. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. *Using SPAdes De Novo Assembler. Curr Protoc Bioinform* 2020;**70**(1):1–29.
20. Rambaut A, Lam TT, Carvalho LM, Pybus OG. *Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol* 2016;**2**(1):1–7.
21. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. *Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics* 2010;**26**(19):2455–7.
22. Kosakovsky Pond SL, Frost SDWW. *Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol* 2005;**22**(5):1208–22.
23. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky et al. *Detecting individual sites subject to episodic diversifying selection*. In: Malik HS (Ed.) PLoS Genet, 8(7); 2012, e1002764.
24. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. *FUBAR: a fast, unconstrained bayesian AppRoximation for inferring selection. Mol Biol Evol* 2013;**30**(5):1196–205.
25. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. *Nextstrain: real-time tracking of pathogen evolution. Bioinformatics* 2018;**34**(23):4121.
26. Sagulenko P, Puller V, Neher RA. *TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol* 2018;**4**(1):1–9.
27. Cattoir L, Vankeerberghen A, Boel A, Van Vaerenbergh K, De, Beenhouwer H. *Epidemiology of RSV and hMPV in Belgium: a 10-year follow-up. Acta Clin Belg* 2019;**74**(4):229–35.
28. García-García ML, Calvo C, Rey C, Díaz B, Molinero MDM, Pozo F, et al. *Human metapneumovirus infetions in hospitalized chlidren and comparison with other respiratory viruses. 2005-2014 prospective study. PLoS One* 2017;**12**(3):e0173504.
29. Jagusic M, Slovic A, Ivancic-Jelecki J, Ljubin-Sternak S, Vilibić-Čavlek T, Tabain I, et al. *Molecular epidemiology of human respiratory syncytial virus and human metapneumovirus in hospitalized children with acute respiratory infections in Croatia, 2014–2017. Infect Genet Evol* 2019;**76**.
30. Andrés C, Piñana M, Borras-Bermejo B, González-Sánchez A, García-Cehic D, Esperalba J, et al. *A year living with SARS-CoV-2: an epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain). Emerg Microbes Infect* 2022;**11**(1):172–81.
31. Zhang L, Liu W, Liu D, Chen D, Tan W, Qiu S, et al. *Epidemiological and clinical features of human metapneumovirus in hospitalised paediatric patients with acute respiratory illness: a cross- sectional study in Southern China, from 2013 to 2016. BMJ Open* 2018;**8**:6–12.

32. Benschop KS, Albert J, Anton A, Andrés C, Aranzamendi M, Armannsdóttir B, et al. *Re-emergence of enterovirus D68 in Europe after easing the COVID-19 lockdown, September 2021. Eurosurveillance* 2021;**26**(45).
33. Zheng Z, Pitzer VE, Shapiro ED, Bont LJ, Weinberger DM. *Estimation of the timing and intensity of reemergence of respiratory syncytial virus following the COVID-19 pandemic in the US. JAMA Netw Open* 2021;**4**(12):e2141779.
34. Fourgeaud J, Toubiana J, Chappuy H, Delacourt C, Moulin F, Parize P, et al. *Impact of public health measures on the post-COVID-19 respiratory syncytial virus epidemics in France. Eur J Clin Microbiol Infect Dis* 2021;**40**(3):2389–95.
35. Regis E, Fontanella S, Lin L, Howard R, Haider S, Curtin JA, et al. *Sex differences in innate anti-viral immune responses to respiratory viruses and in their clinical outcomes in a birth cohort study. Sci Rep* 2021;**11**(1):1–15.
36. Ursin RL, Klein SL. *Sex differences in respiratory viral pathogenesis and treatments. Annu Rev Virol* 2021;**8**:393–414.
37. Regev L, Meningher T, Hindiyeh M, Mendelson E, Mandelboim M. *Increase human metapneumovirus mediated morbidity following pandemic influenza infection. PLoS One* 2012;**7**(4):2–7.
38. Li J, Ren L, Guo L, Xiang Z, Paranhos-Baccalà G, Vernet G, et al. *Evolutionary dynamics analysis of human metapneumovirus subtype A2: genetic evidence for its dominant epidemic. PLoS One* 2012;**7**(3):1–9.
39. Schobel SA, Stucker KM, Moore ML, Anderson LJ, Larkin EK, Shankar J, et al. *Respiratory Syncytial Virus whole-genome sequencing identifies convergent evolution of sequence duplication in the C-terminus of the G gene. Sci Rep* 2016;**6**:26311.
40. Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, et al. *Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol* 2016;**23**(10):899–905.
41. Oketch JW, Kamau E, Otieno GP, Otieno JR, Agoti CN, Nokes DJ. *Human metapneumovirus prevalence and patterns of subgroup persistence identified through surveillance of pediatric pneumonia hospital admissions in coastal Kenya, 2007-2016. BMC Infect Dis* 2019;**19**(1):1–13.
42. Nao N, Saikusa M, Sato K, Sekizuka T, Usuku S, Tanaka N, et al. *Recent molecular evolution of human metapneumovirus (HMPV): subdivision of HMPV a2b strains. Microorganisms* 2020;**8**(9):1–14.
43. Huang J, Diaz D, Mousa JJ. *Antibody epitopes of pneumovirus fusion proteins. Front Immunol* 2019;**10**.
44. Xiao X, Tang A, Cox KS, Wen Z, Callahan C, Sullivan NL, et al. *Characterization of potent RSV neutralizing antibodies isolated from human memory B cells and identification of diverse RSV/hMPV cross-neutralizing epitopes. MAbs* 2019;**11**(8):1415–27.
45. Kim J Il, Park S, Lee I, Park KS, Kwak EJ, Moon KM, et al. *Genome-wide analysis of human metapneumovirus evolution*. In: Tee KK (Ed.) PLoS One, 11(4); 2016, e0152962.
46. van den Hoogen BG, de Jong JC, Groen J, Kuiken T, De Groot R, Fouchier RAM, et al. *A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat Med* 2001;**7**(6):719–24.
47. Van Poelvoorde LAE, Bogaerts B, Fu Q, De Keersmaecker SCJ, Thomas I, Van Goethem N, et al. *Whole-genome-based phylogenomic analysis of the Belgian 2016–2017 influenza A(H3N2) outbreak season allows improved surveillance. Microb Genom* 2021;**7**(9).
48. Dyrdak R, Mastafa M, Hodcroft EB, Neher RA, Albert J. *Intra- and interpatient evolution of enterovirus D68 analyzed by whole-genome deep sequencing. Virus Evol* 2019;**5**(1):1–9.