Insights into immune evasion of human metapneumovirus: novel 180- and 111-nucleotide duplications within viral G gene throughout 2014-2017 seasons in Barcelona, Spain.

AUTHORS: Maria Piñana[1], Jorgina Vila[2], Carolina Maldonado[3], Juan José Galano-Frutos[4,5], Maria Valls[2], Javier Sancho[4,5,6], Francesc Xavier Nuvials[3], Cristina Andrés[1], María Teresa Martín-Gómez[1], Juliana Esperalba[1], Maria Gema Codina[1], Tomàs Pumarola[1], Andrés Antón[1].

AFFILIATIONS:

1. Respiratory Viruses Unit, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain
2. Paediatric Hospitalization Unit, Paediatrics Department, Hospital Universitari Maternoinfantil Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
3. Intensive Care Department, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
4. Biochemistry and Molecular and Cell Biology Department, Sciences Faculty, Universidad de Zaragoza, Zaragoza, Spain
5. Biocomputation and Complex Systems Physics Institute (BIFI). Joint Units BIFI-IQFR (CSIC) and GBs-CSIC, Universidad de Zaragoza, Zaragoza, Spain
6. Aragon Health Research Institute (IIS Aragón), Zaragoza, Spain

ORCID:

- Maria Piñana: 0000-0002-4766-2613
- Juan José Galano-Frutós: 0000-0002-1896-7805
- Javier Sancho: 0000-0002-2879-9200
- Francesc Xavier Nuvials: 0000-0002-6648-2394
- Cristina Andrés: 0000-0002-3200-0895
- Maria Teresa Martín-Gómez: 0000-0001-8119-8749
- Juliana Esperalba: 0000-0003-1326-1341
- Maria Gema Codina: 0000-0002-5313-5086
- Tomas Pumarola: 0000-0002-5171-7461
- Andrés Antón: 0000-0002-1476-0815

CORRESPONDING AUTHOR:

Prof. Tomàs Pumarola

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

+34 93 274 00 00 (ext 6918)

E-mail: **virusrespiratoris@vhebron.net**

Preliminary results from this study were presented at the **ECCMID** conference; April 13-16 2019, Amsterdam, the Netherlands.

RUNNING TITLE

HMPV with duplications and its relation with immune evasion

KEYWORDS

Human metapneumovirus, duplication, G gene, F gene, steric shielding.

ABSTRACT

Background. Human metapneumovirus (HMPV) is an important aetiologic agent of respiratory tract infection (RTI). This study aimed to describe its genetic diversity and clinical impact in patients attended at a tertiary university hospital in Barcelona from the 2014-2015 to the 2016-2017 seasons, focusing on the emerging duplications in G gene and their structural properties.

Methods. Laboratory-confirmed HMPV were characterised based on partial-coding *F* and *G* gene sequences with MEGA.v6.0. Computational analysis of disorder propensity, aggregation propensity and glycosylation sites in viral G predicted protein

sequence were carried out. Clinical data was retrospectively reviewed and further associated to virological features.

Results. HMPV prevalence was 3%. The 180- and 111-nucleotide duplications occurred in A2c lineage G protein increased in prevalence throughout the study, in addition to short genetic changes observed in other HMPV lineages. The A2c G protein without duplications was calculated to protrude over F protein in 23% of cases and increased to a 39% and a 46% with the 111- and 180-nucleotide duplications, respectively. Children did not seem to be more affected by these mutant viruses, but there was a strong association of these variants to LRTI in adults.

Discussion. HMPV presents a high genetic diversity in all lineages. Novel variants carrying duplications might present an evolutionary advantage due to an improved steric shielding, which would have been responsible for the reported increasing prevalence and the association to LRTI in adults.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

Words in abstract: 231

Words in main text: 3,225

INTRODUCTION

Human metapneumovirus (HMPV) is an important aetiologic agent for acute respiratory infections in both children and adults [1]. First described in 2001 [2], this virus has become one of the most important respiratory pathogens worldwide, with primary infection during early childhood and achieving almost a 100% seroprevalence in 5-year-old children [1].

HMPV belongs to the Pneumoviridae family within Mononegavirales order together with human respiratory syncytial virus (HRSV) [3], causing an indistinguishable symptomatology [1]. HMPV is an enveloped, lineal, negative-sense, single-stranded RNA virus with a genome of approximately 13 kb, including 8 genes that encode for 9 proteins with a genomic organization similar to HRSV [1].

HMPV are classified into HMPV-A and HMPV-B genotypes, which in turn are subdivided into 4 subgenotypes known as A1, A2 (subdivided into A2a, A2b and A2c lineages) [1, 4, 5], B1 and B2 (B2a and B2b lineages) [6]. HMPV has a seasonal circulation with its incidence peak in late winter and early spring. Subgenotypes co-circulate at varying levels, being one predominant over the others [1, 3].

The fusion (F) and the attachment (G) proteins are the major envelope glycoproteins. F protein can perform the virus entry on its own by the attachment and fusion to the host cell membrane [7]. Moreover, it is recognized as the major cross-protective antigenic determinant and is highly conserved between genotypes (88%) [5]. Hence, it is the main target for most antivirals or vaccine strategies under development [8]. On the other hand, G protein is weakly immunogenic and protective [9]. It may contribute to the attachment via glycosaminoglycans, but the exact role of the protein remains unknown [9]. G protein exhibits potential glycosylation sites and a high genetic variability, suggesting to be under selective immune pressure [10]. It has about 28% genetic divergence between genotypes and 74-82% intra-genotype [5]. In addition,

180- and 111-nucleotide duplications have been recently described into the ectodomain of the G protein [11–14].

HMPV can cause upper and lower respiratory tract infections (URTI and LRTI, respectively). URTI is common in children older than 2 years old and adults, whereas LRTI is frequent in children younger than 2, in adults older than 65 and in patients with asthma, chronic obstructive pulmonary disease (COPD) or other chronic medical conditions [15]. The main clinical features of children having HMPV LRTI are recurrent wheezing, asthma, bronchopneumonia and pneumonia [15, 16]. In adults, pneumonia, asthma exacerbation and COPD reagudization are the main manifestations [1].

The aims of this study were to describe circulation pattern, genetic diversity and clinical impact of HMPV in paediatric and adult population attended at a tertiary university hospital in Barcelona from the 2014-2015 to the 2016-2017 seasons, focusing on the emergence and spread of variants carrying these two nucleotide duplications, and the effect they could produce in immune evasion.

MATERIALS AND METHODS

Sample collection

From October 2014 (2014-2015 season) to May 2017 (2016-2017 season), respiratory specimens (nasopharyngeal aspirates, nasal and pharyngeal swabs, bronchoaspirates, and bronchoalveolar washes) were received for the laboratory-confirmation of respiratory viruses from children and adults attended at the Hospital Universitari Vall d'Hebron (HUVH) with suspicion of respiratory tract infection (RTI). Institutional Review Board approval (PR(AG)161/2016) was obtained from the HUVH Clinical Research Ethics Committee.

Respiratory viruses' detection

During the HRSV and influenza epidemics, point-of-care tests (POCT) were performed for rapid diagnosis, which were based on immunocromatography (Alere BinaxNOW®

Influenza A&B/RSV, Alere, USA), inmunofluorescence (Sofia RSV FIA, Quidel, USA) or real-time RT-PCR (GenXpert Flu/RSV XC, Cepheid, USA). Samples received out of HRSV/FLUV epidemics or negative for POCT were analysed by immunofluorescence (D3 Ultra 8™ DFA Respiratory Virus, Diagnostic HYBRIDS, USA) or mainly by real-time RT-PCR (Anyplex II RV16, Seegene, Korea, until 2015; Allplex Respiratory Panels 1-3, Seegene, Korea, since 2015). Total nucleic acids were extracted using NucliSense easyMAG (BioMérieux, Marcy l´Etoile) and kept at -80ºC.

HMPV phylogenetic analysis

Both partial *F* and *G* genes were sequenced from HMPV laboratory-confirmed samples. The amplification was performed using the One-Step RT-PCR kit (Qiagen, Hilden, Germany), conditions in Table 1. PCR products were purified using Exo-SAP-IT (USB, Affymetrix Inc., Cleveland, USA) and sequenced by the ABI Prism BigDye Terminator v3.1 (Applied Biosystems, Carlsbad, USA) on the ABI PRISM 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, USA). Nucleotide sequences were edited and assembled using MEGA v6.0 [17]. A collapse of the sequences to haplotypes was done with the ALingment Transformation EnviRonment (ALTER) server [18]. The best fit substitution model was also determined by MEGA v6.0, and the model with the lowest Bayesian information criterion score was used to construct the phylogenetic tree, which was further evaluated with 1,000 bootstrap resamplings.

Computational analysis of disorder propensity, aggregation propensity and glycosylation sites in viral G predicted protein sequence

The propensity of three HMPV G sequences of different length (with or without nucleotide duplications) to adopt disordered conformations was analyzed using the MetaDisorder server [19], their propensity to aggregate using the Pasta 2.0 server [20], and the prediction of potential N- and O-glycosylation sites using NetNGlyc 1.0 [21] and NetOGlyc 4.0 [22] servers, respectively.

Generation and geometric analysis of unfolded ensembles of viral G predicted protein sequence

Ensembles consisting of 2,000 unfolded conformations were generated for each of three HMPV G sequences of different length (with or without nucleotide duplications) using the ProtSA server [23]. The PDB file of each conformation was analyzed to compute the distance between the N atom of the first extracellular residue (Asn52) and the more distant atom, as well as the radius of gyration of the particular conformation.

Clinical features

Demographic (age and sex) and clinical features (URTI/LRTI, co-morbidities, co-infections, antibiotic use, need of respiratory support, type and length of respiratory support, length of hospital stay, intensive-care unit admission or exitus) of HMPV-laboratory confirmed cases (children, <18 years-old and adults, ≥18) were retrospectively reviewed from medical records and related to viral features (genotype, lineage and duplication).

Statistical analysis

Data were analyzed with R software v3.5.1. For categorical data, Chi-squared or Fisher's exact test were performed. For numerical variables, t student, Mann-Withney, ANOVA or Kruskall-Wallis tests were performed according to the need. Statistical significance was taken at the p-value <0.05.

RESULTS

HMPV epidemiology

A total of 20,132 samples of 14,769 patients were tested, of which 9,370 (47%) were laboratory-confirmed for at least one respiratory virus. HMPV was laboratory-confirmed in 423 (2%) samples from 407 (3%) patients. Other respiratory viruses were (in descendant order): rhinovirus (RV, 13%), influenza A virus (FLUAV, 12%), HRSV (9%), influenza B virus (FLUBV, 4%), human adenovirus (HAdV, 4%), enterovirus (EV, 1%),

human parainfluenza virus 3 (HPIV-3, 1%), human coronavirus OC43 (HCoV OC43, 1%), human bocavirus (HBoV, 1%), HCoV 229E (<1%), HCoV NL63 (<1%), HPIV-4 (<1%), HPIV-1 (<1%), HPIV-2 (<1%). HMPV overall showed an important prevalence in the adult population, similar to the reported in children and even higher in the 2015-2016 season (Table 2).

From the 407 cases of HMPV, 75 (18%) presented co-infections with 88 respiratory viruses. These, in descendant order, were the following: RV (26, 30%), hAdV (15, 17%), HBoV (15, 17%), EV (11, 13%), HRSV (6, 7%), HCoV 229E (5, 6%), HCoV OC43 (5, 6%), HCoV NL63 (2, 2%), HPIV-3 (1, 1%), FLUA (1, 1%), FLUB (1, 1%).

Weekly distribution of HMPV laboratory-confirmed cases showed a higher circulation from February to April in the first two seasons, but started at mid December in the last season (Figure 1). The peaks of incidence of the first two seasons were in March (weeks 14/2015 and 09/2016), but the last season presented a different pattern with two incidence peaks, in late December and mid February (weeks 52/2016 and 07/2017).

Genetic characterisation of viral strains

Phylogenetic analyses of HMPV F and G sequences from 387 strains revealed that both genotypes, HMPV-A (201/387, 52%) and HMPV-B (185/387, 48%), co-circulated throughout the study period. One HMPV-A/B (<1%) was also observed. The remaining 20/407 (5%) HMPV could not be characterised, likely due to low RNA quality or low viral load. HMPV-B (61%) predominated during the 2014-2015 season and HMPV-A (62%) did so during 2015-2016. No difference in circulation between HMPV-A (49%) and HMPV-B (51%) was observed during the 2016-2017 season (Table 3).

Phylogenetic analysis of F (382, 94%) and G (365, 90%) sequences (Figure 2) showed congruent results between them. Overall, 11 (3%) samples belonged to A2a, 37 (10%) to A2b, 153 (40%) to A2c, 106 (27%) to B1 and 79 (20%) to B2 (Table 3).

Genetic characterisation of A2 G revealed that A2a and A2c sequences generally had a length of 220 aa and A2b of 218 due to premature stop codons. Genetic characterisation of 153 A2c strains revealed the presence of the novel 180- ($A2c_{180dup}$; 46; 30%) and 111-nucleotide ($A2c_{111dup}$; 13; 9%) duplications into G ectodomain previously described [11–14] with increasing prevalence (Table 3). The 180-nucleotide duplication results in 60 additional aa (between 154-155 aa positions, reference to JX082177), whereas the 111-nucleotide duplication results in 37 additional aa in a nearby region (between 149-150 aa positions). While all $A2c_{180dup}$ clustered together, two subgroups could be observed in the F phylogenetic tree (Figure 2). Differently, $A2c_{111dup}$ G clustered into 2 groups but their F genes clustered together (except NSVH2017-09-82477, whose F gene clustered with other sequences without duplication).

B1 G clustered into two phylogenetic groups (I and II), differing in the acquisition of a premature stop codon in the 232 aa (relative to HR475-13, accession number KU375606) in all strains belonging to group II (Figure 2), but one (NSVH2015-12-87728). In addition, two sequences presented deletions of one (NSVH2017-04-59510) or two aa (NSVH2016-03-50135) in the G protein. Genetic characterisation of 10 B2a and 69 B2b G sequences revealed the acquisition of short duplications or insertions. Whereas B2a group, represented by CAN98-75 strain (AY297748), presented the aa KE in the 160-161 positions, B2b group presented 1 or 2 duplications (KEKE and KEKEKE). Also, a group differed from the rest of B2b group presenting an R after only one KE, which could probably be a mutation of the duplication of K (Figure 2).

The sequences of the present study were submitted to GenBank with accession numbers MN617398-MN617753.

Structural biology of G protein

The ectodomain ensemble of the non-glycosylated form of G protein of NSVH2015-06-62150 ($A2c_{wt}$) was simulated [23–25], resulting in a composition of conformations with maximal length (D_{max}) of 4.5-22.2 nm and radii of gyration (R_g) of 1.9-7.4 nm (Figure 3).

The *MetaDisorder* server predicted that both 180- and 111-nucleotide duplications were as fully disordered as $A2c_{wt}$, with no self-aggregation segments. Also, the glycosylation pattern showed a similar distribution of the numerous O-glycosylation sites (Figure 4), contributing with additional potential 23-26 O-glycosylation sites in $A2c_{180dup}$ and 12-13 in $A2c_{111dup}$.

Once verified that amino acid sequences with a duplication did not differ from $A2c_{wt}$ in order, aggregation propensity or glycosylation potential, 2000-conformation unfolded ensembles were generated for NSVH2017-09-78834 ($A2c_{111dup}$) and NSVH2015-19-63118 ($A2c_{180dup}$) and analyses of their geometries were performed. $A2c_{111dup}$ increased size to D_{max} 5.1-27.3 nm and R_g 2.4-9.0 nm, while $A2c_{180dup}$ increased D_{max} 5.2-29.4 nm and R_g 2.4-9.9 nm (Figure 3).

The pre-fusion conformation of the F trimer has been previously calculated to protrude about 13 nm from the membrane [10]. According to the distance distributions of the three ensembles, the actual fraction of G protein's ectodomain protruding more than 13 nm from the membrane amounts to 23% in the $A2c_{wt}$, and it increases to 39% in $A2c_{111dup}$ and 46% and $A2c_{180dup}$.

Clinical impact of human metapneumovirus

Due to the absence of clinical information (2), non-amplification (20) or manifestation of other syndromes rather than URTI or LRTI (20), clinical features of 203 pediatric and 162 adult cases were finally studied (Table 4).

In paediatric cases, the median age (1.17, IQR 0.49-2.61) of patients presenting LRTI was significantly lower (p 0.011) than those with URTI (2.18, IQR 0.88-6.41). 42% were female, and male:female proportion was similar in URTI and LRTI cases. A2c lineage

was more associated to LRTI (p 0.032) than other lineages, but A2c with duplications were not associated with a higher risk of LRTI compared to other strains (p 0.743 and 0.202).

In adult cases, the median age (74.7, IQR 61.3-82.8) of patients presenting LRTI was significantly higher (p 0.001) than those with URTI (66.8, IQR 53.7-77.1), each year of age increasing 1.03 times the odds of having LRTI. 52% were female, but the proportion male:female was higher in the LRTI group though not being statistically significant (p 0.197). No lineages or subgenotypes were more associated to LRTI (p 0.052 and 0.246, respectively). Cardiopathy was associated to LRTI (OR 4.2, p <0.001). A2c strains with duplications were significantly more associated to LRTI, compared to all other strains (OR 2.83, p 0.018) or to $A2c_{wt}$ (OR 3.45, p 0.034).

For the severity study, only patients hospitalized due to LRTI (176) were considered, being 116 (66%) pediatric (Table 5) and 60 (34%) adult patients (Table 6). Comparison of factors related to severity was done between subgenotypes; A2c with duplications and other HMPV strains and between A2c with and without duplications. Only the difference in ICU admission in the distribution of pediatric patients between subgenotypes was statistically significant (p 0.031). No other variables were found to be significant.

DISCUSSION

This study reports recent data on prevalence, genetic diversity, structural biology of G protein and clinical features of HMPV detected at a tertiary hospital in Barcelona, Spain.

The positivity rate of HMPV was similar to other recent reports [5, 26]. Interestingly, the prevalence in adults was similar or even higher than in children, which emphasizes the importance of HMPV in the adult population. The prevalence of this virus increased throughout the three seasons, probably due to the higher implementation of more

284 sensitive molecular methods over antigen detection-based tests. As well, there might
285 be an underestimation of the prevalence of HMPV, as a large number of positive
286 samples for HRSV and influenza by POCT during their epidemics were not tested for
287 other respiratory viruses, including HMPV. Almost 20% of cases presented co-
288 infections with other respiratory viruses, mostly rhinoviruses, adenoviruses and
289 bocaviruses, as previously reported [27].

290 HMPV presented a clear seasonality, mainly at the late winter and early spring, as
291 previously described [3, 8, 27]. Interestingly, the last season presented a different
292 pattern, showing two different peaks in one epidemic season without changes among
293 circulating genotypes.

294 Genetic characterisation revealed that both genotypes co-circulated with a shift in the
295 predominant genotype, as expected [3]. However, there was an unpredicted co-
296 predominance of both genotypes in the third season, which could be due to an
297 intermediate alternation of genotypes or due to the emergence of HMPV-A viruses with
298 new antigenic features that would evade the immunity created on the previous season.

299 Congruent classification of both F and G genes was expected, as no genetic
300 recombination has been described for HMPV. All subgenotypes were detected except
301 A1, suggesting it has extinguished and been replaced by A2, according to previous
302 studies [4]. According to the data of the present study, A2c lineage appears to be
303 replacing A2a and A2b. Moreover, A2c strains with duplications might be replacing
304 $A2c_{wt}$ in the near future due to its rapid increase in prevalence these three seasons,
305 suggesting they might present an improved mechanism of immune evasion. In fact, a
306 Japanese group observed that $A2c_{111dup}$ had totally replaced the rest of A2 strains [28],
307 including $A2c_{180dup}$. Interestingly, our group has observed how both $A2c_{111dup}$ and
308 $A2c_{180dup}$ have replaced together the rest of HMPV-A viruses, being the latter more
309 prevalent [29].

Different lengths of G protein due to premature stop codons have been observed, as previously described [4]. However, this might be the first study that associates the different lengths to different (current or future) lineages. A2b and A2c lineages included viruses with G proteins of 218 and 220 aa respectively; and two different genetic groups (I and II) could be distinguished within B1 subgenotype, with a difference of 10 aa in length, which might evolve into novel lineages. Also, nucleotide duplications can lengthen the predicted G amino acid sequence, such as long duplications in A2c, and short duplications in B2. For B2 viruses, KE duplications or KER variants in 160-161 aa should be monitored next seasons to reveal whether they confer an evolutionary advantage. The deletions observed seem not to have been fixed in the viral population.

Once these $A2c_{111dup}$ and $A2c_{180dup}$ were described, one of the aims of this study was to study their structural properties. G has a heavily glycosylated pattern, enhanced by the emergence of duplications that increase the number of potential glycosylation sites. Although it is a very disordered protein and seems to have numerous random conformations, a composition of these conformations could be done. This prediction suggests that both $A2c_{180dup}$ and $A2c_{111dup}$ proteins protrude more than $A2c_{wt}$. This finding supports the hypothesis of Leyrat [10], who suggested that G protein had a shielding function towards F protein, masking its antigenic epitopes, and at the same time validates the hypothesis that these novel long duplications would enhance this immune evasion mechanism, as it would hide more efficiently F epitopes [11].

The phylogenetic analysis of the HMPV-A *G* gene also revealed an important finding. Sequences of the newly described A2c lineage [4, 5] were compared to sequences of the previously described A2b1 and A2b2 sublineages [30, 31] and clustered together; that is to say, A2b and A2c lineages are exactly the same as A2b1 and A2b2, respectively. This misunderstanding between the genetic classification used in several articles highlights the urgent need of an official classification, as well as universal criteria to define new genotypes or lineages.

Furthermore, clinical impact was also assessed. As in literature [3], LRTI is more common in children under 2 years old and adults over 65 years old. Moreover, adults have an increase of 1.03 times the odds of suffering LRTI every passing year. The presence of chronic medical conditions such as cardiopathy, more frequent in the elderly, may be responsible for the increased risk of LRTI, so this virus should be tightly surveilled in these cases.

Paediatric and adult patients underwent more antibiotic treatment when manifesting LRTI than URTI. However, only 8% of children and 30% of adults treated with antibiotics had a positive bacterial culture. Hence, over-antibiotic prescription is still reported.

Regarding infections by A2c, children seemed to be as affected by A2c with duplications as by $A2c_{wt}$ or other lineages, as it is probably a primary infection. Instead, A2c with duplications were more associated with LRTI in adults than $A2c_{wt}$ or other lineages. Although adults should have an efficient immune response (6), they have 3.45 times more odds of manifesting LRTI when infected by A2c with duplication than by $A2c_{wt}$. This suggests that it might be acting as a primary infection, which supports the hypothesis of G protein's steric shielding over F protein. Whether strains with duplication cause more severe disease could be demonstrated neither in children nor in adults.

The increasing prevalence of viral variants carrying a duplication into the ectodomain of the G protein throughout the study period, the association of $A2c_{111dup}$ and $A2c_{180dup}$ with more severe disease in adults, and the prediction of an enhancing steric shielding of the G protein masking antigenic epitopes of the F protein suggest that these duplications might confer an evolutionary advantage contributing to the immune evasion during the infection. Given that F protein is the main target for most vaccine strategies currently under development, the fact that it could be masked by G should be taken into account.

ACKNOWLEDGMENTS

This study was supported by the Spanish Ministry of Economy and Competitiveness (grants BFU2016-78232-P), Instituto de Salud Carlos III and by the European Regional Development Fund, through the Interreg V-A programme: POCTEFA 2014-2020 (grant Pirepred EFA086/15). It was also co-financed by the European Development Regional Fund (ERDF) "A way to achieve Europe", Spanish Network for Research in Infectious Diseases (REIPI RD16/0016/0003). We also would like to acknowledge the Statistics and Bioinformatics Unit (UEB) in Vall d'Hebron Research Institute (VHIR) for the statistical analyses.

REFERENCES

1. Panda S, Mohakud NK, Pena L, Kumar S (2014) Human metapneumovirus: Review of an important respiratory pathogen. Int J Infect Dis 25:45–52. https://doi.org/10.1016/j.ijid.2014.03.1394

2. van den Hoogen BG, de Jong JC, Groen J, et al (2001) A newly discovered human pneumovirus isolated from young children with respiratory tract disease. NatMed 7:719–724

3. Shafagati N, Williams J (2018) Human metapneumovirus - what we know now. F1000Research 7:135. https://doi.org/10.12688/f1000research.12625.1

4. Jagušić M, Slović A, Ljubin-Sternak S, et al (2017) Genetic diversity of human metapneumovirus in hospitalized children with acute respiratory infections in Croatia. J Med Virol 89:1885–1893. https://doi.org/10.1002/jmv.24884

5. Chow WZ, Chan YF, Oong XY, et al (2016) Genetic diversity, seasonality and transmission network of human metapneumovirus: identification of a unique sub-lineage of the fusion and attachment genes. Sci Rep 6:27730. https://doi.org/10.1038/srep27730

6. Carr MJ, Waters A, Fenwick F, et al (2008) Molecular epidemiology of human

metapneumovirus in Ireland. J Med Virol 80:510–516. https://doi.org/10.1002/jmv

7. Biacchesi S, Skiadopoulos MH, Yang L, et al (2004) Recombinant human Metapneumovirus lacking the small hydrophobic SH and/or attachment G glycoprotein: deletion of G yields a promising vaccine candidate. J Virol 78:12877–87. https://doi.org/10.1128/JVI.78.23.12877-12887.2004

8. Kumar P, Srivastava M (2018) Prophylactic and therapeutic approaches for human metapneumovirus. Virus Dis 29:434–444. https://doi.org/10.1007/s13337-018-0498-5

9. Papenburg J, Carbonneau J, Isabel S, et al (2013) Genetic diversity and molecular evolution of the major human metapneumovirus surface glycoproteins over a decade. J Clin Virol 58:541–547. https://doi.org/10.1016/j.jcv.2013.08.029

10. Leyrat C, Paesen G, Charleston J, et al (2014) Structural insights into the human metapneumovirus glycoprotein ectodomain. J Virol 88:11611–11616. https://doi.org/10.1128/JVI.01726-14

11. Piñana M, Vila J, Gimferrer L, et al (2017) Novel human metapneumovirus with a 180-nucleotide duplication in the G gene. Future Microbiol 12:565–571. https://doi.org/10.2217/fmb-2016-0211

12. Piñana M, Andrés C, Gimferrer L, et al (2017) A novel human metapneumovirus carrying a 111-nucleotide duplication within the G gene detected at a tertiary university hospital in Catalonia since the 2015-2016 season. In: 20th ESCV Annual Meeting. p 52

13. Saikusa M, Kawakami C, Nao N, et al (2017) 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

14. Saikusa M, Nao N, Kawakami C, et al (2017) 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

15. García-García ML, Calvo C, Rey C, et al (2017) Human metapneumovirus infections in hospitalized children and comparison with other respiratory viruses. 2005-2014 prospective study. PLoS One 12:1–12. https://doi.org/10.1371/journal.pone.0173504

16. Schuster JE, Khuri-Bulos N, Faouri S, et al (2015) Human Metapneumovirus Infection in Jordanian Children. Pediatr Infect Dis J 34:1335–1341. https://doi.org/10.1097/inf.0000000000000892

17. Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

18. Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, et al (2010) ALTER: Program-oriented conversion of DNA and protein alignments. Nucleic Acids Res 38:14–18. https://doi.org/10.1093/nar/gkq321

19. Kozlowski LP, Bujnicki JM (2012) MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics 13:1. https://doi.org/10.1186/1471-2105-13-111

20. Walsh I, Seno F, Tosatto SCE, Trovato A (2014) PASTA 2.0: An improved server for protein aggregation prediction. Nucleic Acids Res 42:301–307. https://doi.org/10.1093/nar/gku399

21. Gupta R, Jung E, Brunak S (2004) Prediction of N-glycosylation sites in human proteins. Prep

22. Steentoft C, Vakhrushev SY, Joshi HJ, et al (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J

442 32:1478–1488. https://doi.org/10.1038/emboj.2013.79

443 23. Estrada J, Bernadó P, Blackledge M, Sancho J (2009) ProtSA: A web
444 application for calculating sequence specific protein solvent accessibilities in the
445 unfolded ensemble. BMC Bioinformatics 10:1–8. https://doi.org/10.1186/1471-
446 2105-10-104

447 24. Bernado P, Blanchard L, Timmins P, et al (2005) A structural model for unfolded
448 proteins from residual dipolar couplings and small-angle x-ray scattering. Proc
449 Natl Acad Sci 102:17002–17007. https://doi.org/10.1073/pnas.0506202102

450 25. Bernadó P, Blackledge M, Sancho J (2006) Sequence-specific solvent
451 accessibilities of protein residues in unfolded protein ensembles. Biophys J
452 91:4536–4543. https://doi.org/10.1529/biophysj.106.087528

453 26. Reiche J, Jacobsen S, Neubauer K, et al (2014) Human metapneumovirus:
454 Insights from a ten-year molecular and epidemiological analysis in Germany.
455 PLoS One 9:. https://doi.org/10.1371/journal.pone.0088342

456 27. Zhang L, Liu W, Liu D, et al (2018) Epidemiological and clinical features of
457 human metapneumovirus in hospitalised paediatric patients with acute
458 respiratory illness: a cross- sectional study in Southern China, from 2013 to
459 2016. BMJ Open 8:6–12. https://doi.org/10.1136/bmjopen-2017-019308

460 28. Saikusa M, Nao N, Kawakami C, et al (2019) Predominant detection of the
461 subgroup A2b human metapneumovirus strain with 111-nucleotide duplication in
462 Yokohama City, Japan in 2018. Jpn J Infect Dis 350–352.
463 https://doi.org/10.7883/yoken.jjid.2019.124

464 29. Piñana M, Andrés C, Gimferrer L, et al (2018) Human metapneumovirus: are the
465 new duplications within the G gene responsible for doubling its prevalence? In:
466 21st congress of the European Society for Clinical Virology

467 30. Regev L, Meningher T, Hindiyeh M, et al (2012) Increase human

metapneumovirus mediated morbidity following pandemic influenza infection. PLoS One 7:2–7. https://doi.org/10.1371/journal.pone.0034750

31. Neemuchwala A, Duvvuri VR, Marchand-Austin A, et al (2015) Human metapneumovirus prevalence and molecular epidemiology in respiratory outbreaks in Ontario, Canada. J Med Virol 87:269–274. https://doi.org/10.1002/jmv.24024

TABLES

Table 1. Primers and PCR conditions.

Gene	Primers	CAN97-83 Initial position	CAN97-83 Final position	PCR conditions	Fragment length (bp)	Reference
G	GF: GAGAACATTCGRRCRATAGAYA	6,247	6,268	50ºC×30' – 95ºC×15' – 45x (95ºC×30" – 59ºC×30" – 72ºC×1') – 72ºc×10'	924	Ludewick HP et al., 2005
	GR: AGATAGACATTRACAGTGGATT	7,149	7,170			
F	FF: CAATGCWGGRATAACACCAGC	3,693	3,713	50ºC×30' – 95ºC×15' – 45x (95ºC×30" – 55ºC×30" – 72ºC×1') – 72ºC×10'	745	Designed for this study
	FR: ATTGAAYTGATCYTCAGGAAAC	4,416	4,437			

Table 2. Demographic data of total patients and patients with HMPV.

Season	Total patients			Patients with HMPV					
	Pediatric	Adult	Total	Pediatric		Adult		Total	
2014 - 2015	1,939	1,744	3,683	54	3%	36	2%	90	2%
2015	194	547	741	3	2%	2	<1%	5	1%
2015 - 2016	2,591	2,209	4,800	64	2%	87	4%	151	3%
2016	222	564	786	10	5%	0	0%	10	1%
2016 - 2017	2,832	1,927	4,759	90	3%	61	3%	151	3%
Overall	**7,778**	**6,991**	**14,769**	**221**	**3%**	**186**	**3%**	**407**	**3%**

Table 3. Distribution of genotypes and lineages throughout the study period.

Genetic group	Season										Total	
	2014-2015		2015		2015-2016		2016		2016-2017			
A1	0	0%	0	0%	0	0%	0	0%	0	0%	**0**	**0%**
A2a	9	10%	1	20%	0	0%	0	0%	1	1%	**11**	**3%**
A2b	12	14%	0	0%	22	16%	1	10%	2	1%	**37**	**10%**
A2c	13	15%	0	0%	65	46%	8	80%	67	47%	**153**	**40%**
$A2c_{wt}$	11	85%	0	0%	46	71%	6	75%	31	46%	**94**	**61%**
$A2c_{180dup}$	2	15%	0	0%	18	28%	1	12.5%	25	37%	**46**	**30%**
$A2c_{111dup}$	0	0%	0	0%	1	1%	1	12.5%	11	17%	**13**	**9%**
B1	37	42%	4	80%	22	16%	1	10%	42	29%	**106**	**27%**
B2	17	19%	0	0%	31	22%	0	0%	31	22%	**79**	20%
B2a	5	29%	0	0%	2	6%	0	0%	3	10%	**10**	**13%**
B2b	12	71%	0	0%	29	94%	0	0%	28	90%	**69**	**87%**
Total	88	100%	5	100%	140	100%	10	100%	143	100%	**386**	**100%**

$A2c_{wt}$: A2c strains without any duplication

$A2c_{180dup}$: A2c strains carrying the 180-nucleotide duplication

$A2c_{111dup}$: A2c strains carrying the 111-nucleotide duplication

Table 4. Clinical features of patients infected with HMPV.

Factor	Pediatric cases			Adult cases		
	URTI (n=42)	LRTI (n=161)	p	URTI (n=77)	LRTI (n=85)	p
Season*			0.908			0.020
2014-2015	10 (25.6 %)	42 (26.1 %)		14 (18.4 %)	19 (22.6 %)	
2015-2016	10 (25.6 %)	44 (27.3 %)		46 (60.5 %)	33 (39.3 %)	
2016	1 (2.56 %)	9 (5.59 %)		0 (0.00 %)	0 (0.00 %)	
2016-2017	18 (46.2 %)	66 (41.0 %)		16 (21.1 %)	32 (38.1 %)	
Age	2.18 [0.88;6.41]	1.17 [0.49;2.61]	0.011	66.8 [53.7;77.1]	74.7 [61.3;82.8]	0.001
Age group			0.001			0.031
0-2 years old	19 (45.2 %)	95 (59.0 %)		0 (0.00 %)	0 (0.00 %)	
2-5 years old	10 (23.8 %)	52 (32.3 %)		0 (0.00 %)	0 (0.00 %)	
5-15 years old	11 (26.2 %)	14 (8.70 %)		0 (0.00 %)	0 (0.00 %)	
15-64 years old	2 (4.76 %)	0 (0.00 %)		34 (43.6%)	23 (27.4%)	
>64 years old	0 (0.00 %)	0 (0.00 %)		44 (56.4%)	61 (72.6%)	
Sex			1.000			0.197
Female	18 (42.9 %)	67 (41.6 %)		45 (58.4 %)	40 (47.1 %)	
Male	24 (57.1 %)	94 (58.4%)		32 (41.6%)	45 (52.9 %)	
Subgenotype			0.763			0.052
A/B	0 (0.00 %)	0 (0.00 %)		0 (0.00 %)	1 (1.18 %)	
A2	24 (57.1 %)	83 (51.6 %)		37 (48.1 %)	46 (54.1 %)	
B1	11 (26.2 %)	44 (27.3 %)		17 (22.1 %)	26 (30.6 %)	
B2	7 (16.7 %)	34 (21.1 %)		23 (29.9 %)	12 (14.1 %)	
Sublineage			0.032			0.246
A2a	5 (11.9%)	2 (1.2%)		1 (1.3 %)	2 (2.4 %)	
A2b	6 (14.3%)	14 (8.7%)		7 (9.1 %)	7 (8.3 %)	
A2c	13 (31.0%)	67 (41.6 %)		29 (37.7 %)	37 (44.0 %)	
B1	11 (26.2)	44 (27.3%)		17 (22.1%)	26 (31.0%)	
B2a	0 (0.00 %)	5 (3.1%)		3 (3.9 %)	2 (2.4 %)	
B2b	7 (16.7%)	29 (18.0%)		20 (26.0 %)	10 (11.9%)	
Duplication			0.768			0.048
111	1 (2.4 %)	5 (3.1%)		1 (1.3 %)	6 (7.1 %)	
180	6 (14.3%)	16 (9.9 %)		7 (9.1 %)	15 (17.6 %)	
no	35 (83.3%)	140 (87.0%)		69 (89.6%)	64 (75.3%)	
Comorbidities			0.046			0.937
Yes	13 (31.0 %)	80 (49.7 %)		63 (81.8 %)	71 (83.5 %)	
Non	29 (69.0 %)	81 (50.3 %)		14 (18.2 %)	14 (16.5 %)	
Respiratory comorbidities			<0.001			0.799
Asthma	1 (2.38 %)	32 (19.9 %)		0 (0.00 %)	0 (0.00 %)	
Pneumopathy	0 (0.00 %)	20 (12.42%)		0 (0.00 %)	0 (0.00 %)	
EPOC	0 (0.00 %)	0 (0.00 %)		15 (19.5 %)	19 (22.4 %)	
Non	41 (97.6 %)	109 (67.7 %)		62 (80.5 %)	66 (77.6 %)	
Cardiopathy			0.532			<0.001
Yes	2 (4.76 %)	14 (8.70 %)		11 (14.3 %)	35 (41.2 %)	
Non	40 (95.2 %)	147 (91.3 %)		66 (85.7 %)	50 (58.8 %)	

Oncohematology			0.276			0.022
Yes	4 (9.52 %)	8 (4.97 %)		31 (40.3 %)	19 (22.4 %)	
Non	38 (90.5 %)	153 (95.0 %)		46 (59.7 %)	66 (77.6 %)	
Immunodepression			0.008			0.203
Immunodeficiency	1 (2.38 %)	2 (1.24 %)		23 (29.9 %)	17 (20.0 %)	
TPH	3 (7.14 %)	0 (0.00 %)		0 (0.00 %)	0 (0.00 %)	
Non	38 (90.5 %)	159 (98.8 %)		54 (70.1 %)	68 (80.0 %)	
Diabetes mellitus			.			0.114
Yes	0 (0.00 %)	0 (0.00 %)		12 (15.6 %)	23 (27.1 %)	
Non	42 (100 %)	161 (100 %)		65 (84.4 %)	62 (72.9 %)	
Prematurity			0.422			.
Yes	3 (7.14 %)	21 (13.0 %)		0 (0.00 %)	0 (0.00 %)	
Non	39 (92.9 %)	140 (87.0 %)		77 (100 %)	85 (100 %)	
Chronic kidney disease			.			0.221
Yes	0 (0.00 %)	0 (0.00 %)		9 (11.7 %)	17 (20.0 %)	
Non	42 (100 %)	161 (100 %)		68 (88.3 %)	68 (80.0 %)	
Bacteria co-infection			0.244			0.005
Yes	4 (9.52 %)	7 (4.35 %)		10 (13.0 %)	28 (32.9 %)	
Non	38 (90.5 %)	154 (95.7 %)		67 (87.0 %)	57 (67.1 %)	
Viral co-infection			0.384			1000
Yes	11 (26.2 %)	30 (18.6 %)		5 (6.49 %)	5 (5.88 %)	
Non	31 (73.8 %)	131 (81.4 %)		72 (93.5 %)	80 (94.1 %)	
Antibiotic			<0.001			<0.001
Yes	7 (16.7 %)	85 (52.8 %)		42 (54.5 %)	79 (92.9 %)	
Non	35 (83.3 %)	76 (47.2 %)		35 (45.5 %)	6 (7.06 %)	
Duplication *vs* the rest			0.743			0.018
A2c w/ duplication	7 (22.6 %)	21 (17.9 %)		8 (13.3 %)	21 (36.2 %)	
Other types	24 (77.4 %)	96 (82.1 %)		69 (86.7 %)	64 (63.8 %)	
A2c sublineage			0.202			0.034
A2c w/ duplication	7 (53.8 %)	21 (31.3 %)		8 (10.4 %)	21 (24.7 %)	
A2c w/o duplication	6 (46.2 %)	46 (68.7 %)		21 (89.6 %)	16 (75.3 %)	

487

488

Table 5. Severity of HMPV disease in pediatric patients. HFNC = high flux nasal cannula; CO = conventional oxygenotherapy (low flow nasal cannula or oxygen mask); MV = mechanical ventilation.

Pediatric patients										
Factor	**Subgenotypes**				**Duplications *vs* other strains**			**A2c sublineage**		
	A2	B1	B2	p	A2c w/ duplication	Other strains	p	A2c w/ duplication	A2c w/o duplication	p
Length hospital stay	4.50 [3.00;7.75]	5.00 [4.00;7.00]	5.50 [3.00;7.25]	0.880	5.00 [4.00;11.00]	5.00 [3.00;7.00]	0.199	5.00 [4.00;11.0]	4.00 [3.00;6.00]	0.166
Respiratory support				1.000			0.297			0.406
Yes	47 (81.0 %)	27 (79.4 %)	19 (79.2 %)		14 (93.3%)	79 (78.2%)		14 (93.3 %)	25 (78.1 %)	
Non	11 (19.0 %)	7 (20.6 %)	5 (20.8 %)		1 (6.67%)	22 (21.8%)		1 (6.67 %)	7 (21.9 %)	
Type of respiratory support				0.520			0.140			0.391
HFNC	13 (22.4 %)	7 (20.6 %)	7 (29.2 %)		3 (20.0%)	24 (23.8%)		3 (20.0 %)	7 (21.9 %)	
CO	28 (48.3 %)	20 (58.8 %)	10 (41.7 %)		8 (53.3%)	50 (49.5%)		8 (53.3 %)	16 (50.0 %)	
MV	6 (10.3 %)	0 (0.00 %)	2 (8.33 %)		3 (20.0%)	5 (5.00%)		3 (20.0 %)	2 (6.25 %)	
Non	11 (19.0 %)	7 (20.6 %)	5 (20.8 %)		1 (6.67%)	22 (21.8%)		1 (6.67 %)	7 (21.9 %)	
Length respiratory support	3.00 [2.00;5.75]	4.00 [2.00;5.75]	3.00 [2.00;6.00]	0.874	3.00 [3.00;9.00]	4.00 [2.00;5.00]	0.208	3.00 [3.00;9.00]	3.00 [2.00;4.25]	0.122
ICU admission				0.031			0.633			1.000
Yes	9 (15.5 %)	0 (0.00 %)	2 (8.33 %)		2 (13.3%)	9 (8.9%)		2 (13.3 %)	5 (15.6 %)	
Non	49 (84.5 %)	34 (100 %)	22 (91.7 %)		13 (86.7%)	92 (91.1%)		13 (86.7 %)	27 (84.4 %)	
Length ICU stay	6.00 [4.00;7.00]	. [.;.]	13.0 [9.00;17.0]	0.340	9.00 [7.50;10.50]	5.00 [4.00;7.00]	0.340	9.00 [7.50;10.5]	4.00 [4.00;6.00]	0.155
Exitus				.			.			.
Yes	0 (0 %)	0 (0 %)	0 (0 %)		0 (0 %)	0 (0 %)		0 (0%)	0 (0%)	
Non	58 (100 %)	34 (100 %)	24 (100 %)		15 (100%)	101 (100%)		15 (100 %)	32 (100 %)	

Table 6. Severity of HMPV disease in adult patients. HFNC = high flux nasal cannula; CO = conventional oxygenotherapy (low flow nasal cannula or oxygen mask); MV = mechanical ventilation

Adult patients										
Factor	Subgenotypes				Duplications *vs* other strains			A2c sublineage		
	A2	B1	B2	p	A2c w/ duplication	Other strains	p	A2c w/ duplication	A2c w/o duplication	p
Length hospital stay	4.00 [2.00;9.00]	9.00 [2.00;12.0]	4.00 [2.00;7.25]	0.327	4.00 [2.00;9.00]	6.00 [2.00;10.50]	0.442	4.00 [2.00;9.00]	2.50 [2.00;7.50]	0.787
Respiratory support				0.655			0.639			0.596
Yes	30 (90.9 %)	15 (88.2 %)	8 (80.0 %)		11 (84.6%)	42 (89.4%)		11 (84.6 %)	13 (92.9 %)	
Non	3 (9.09 %)	2 (11.8 %)	2 (20.0 %)		2 (15.4%)	5 (10.6%)		2 (15.4 %)	1 (7.14 %)	
Type of respiratory support				0.145			0.780			0.450
HFNC	4 (12.1 %)	1 (5.88 %)	1 (10.0 %)		2 (15.4%)	4 (8.5%)		2 (15.4 %)	1 (7.14 %)	
CO	25 (75.8 %)	9 (52.9 %)	6 (60.0 %)		8 (61.5%)	32 (68.1%)		8 (61.5 %)	12 (85.7 %)	
MV	1 (3.03 %)	5 (29.4 %)	1 (10.0 %)		1 (7.69%)	6 (12.8%)		1 (7.69 %)	0 (0.00 %)	
Non	3 (9.09 %)	2 (11.8 %)	2 (20.0 %)		2 (15.4%)	5 (10.6%)		2 (15.4 %)	1 (7.14 %)	
Length respiratory support	2.00 [1.00;5.00]	3.00 [1.00;10.0]	1.00 [1.00;5.50]	0.304	1.00 [1.00;4.00]	2.00 [1.00;6.00]	0.449	1.00 [1.00;4.00]	1.00 [1.00;4.25]	0.938
ICU admission				0.151			0.182			1.000
Yes	2 (6.06 %)	4 (23.5 %)	2 (20.0 %)		0 (0%)	8 (17%)		0 (0.00 %)	1 (7.14 %)	
Non	31 (93.9 %)	13 (76.5%)	8 (80.0 %)		13 (100%)	39 (83%)		13 (100 %)	13 (92.9 %)	
Length ICU stay	9.50 [6.75;12.2]	11.0 [5.00;21.0]	10.5 [8.75;12.2]	0.939	. [.;.]	10.5 [5.5;15.25]	.	. [.;.]	15.0 [15.0;15.0]	.
Exitus				0.096			1.000			1.000
Yes	2 (6.06 %)	1 (5.88 %)	3 (30.0 %)		1 (7.69%)	5 (10.6%)		1 (7.69 %)	1 (7.14 %)	
Non	31 (93.9 %)	16 (94.1 %)	7 (70.0 %)		12 (92.3%)	42 (89.4%)		12 (92.3 %)	13 (92.9 %)	

FIGURES

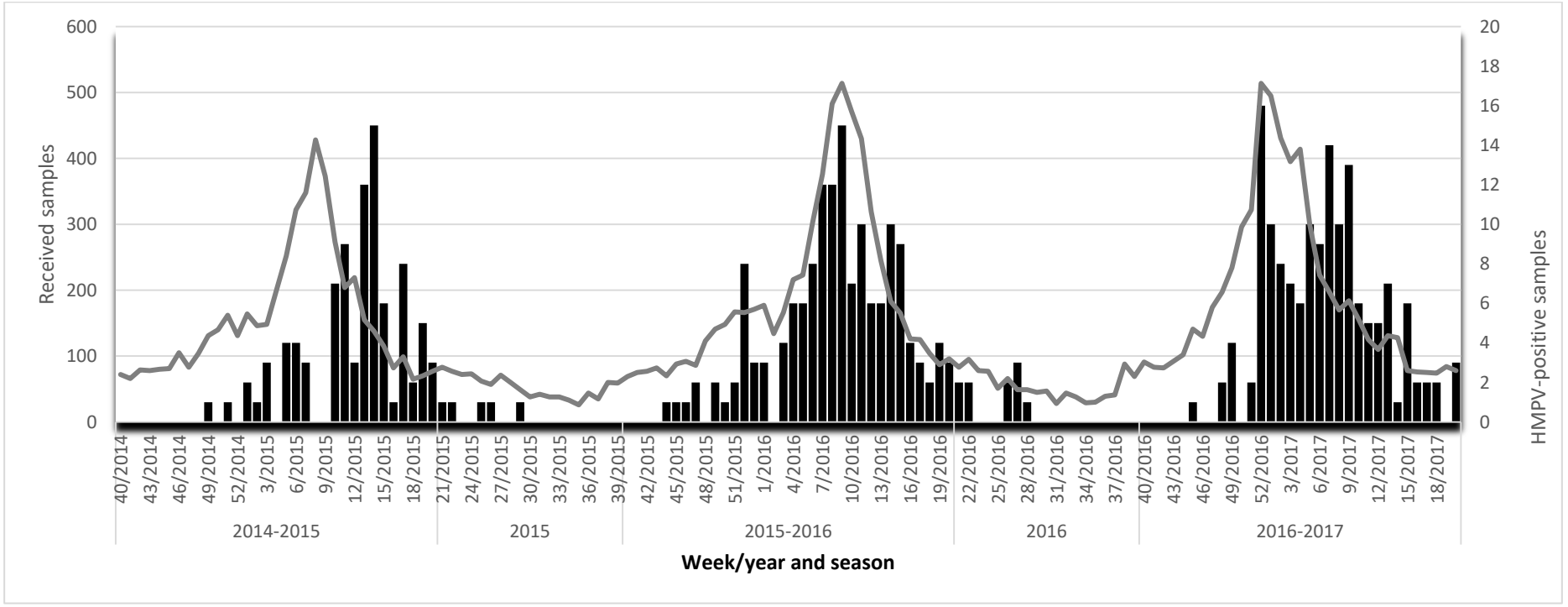

Figure 1. Seasonality of HMPV. Weekly distribution of laboratory-confirmed HMPV is represented in black bars and that of total received samples is in a grey line.

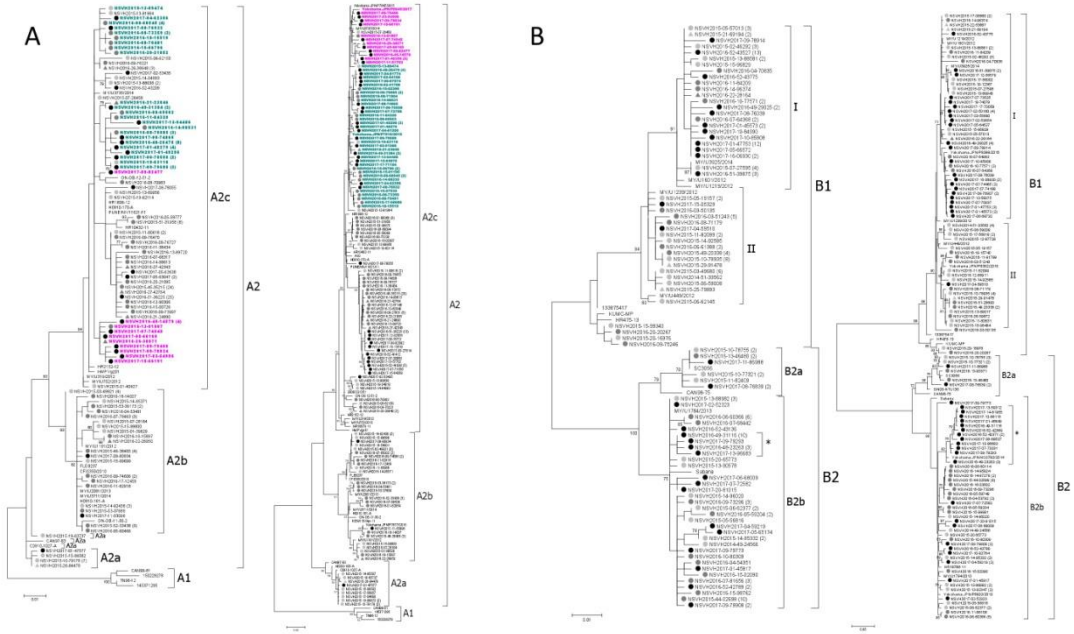

Figure 2. Phylogenetic trees. Phylogenetic trees of F and G sequences from (A) HMPV-A and (B) HMPV-B, respectively. The analyses were performed on G sequences from nucleotide position 6340–6891 in reference to CAN97-83 strain (accession number AY297749) for HMPV-A and 6319-6921 CAN98-75 (accession number AY297748) for HMPV-B. On F sequences, nucleotide positions were 3846-4287 in reference to CAN97-83 strain and 3843-4284 in reference to CAN98-75. All 4 phylogenetic trees were inferred by using the maximum likelihood method, based on the Kimura 2-parameter for the F gene and General Time Reversible for the G gene. Numbers at the tree branch nodes represent the measure of support calculated by the bootstrap method (1000 replicates); only those exceeding 70% are shown. Sequences are marked with solid circles or triangles depending on whether they were collected during the seasonal or the interseasonal period (2014-2015 and 2015 in light grey, 2015-2016 and 2016 in dark grey and 2016-2017 in black). A2c sequences with duplications in the G protein have their name in bold turquoise for the 180-nucleotide duplication and in bold pink for the 111-nucleotide duplications.

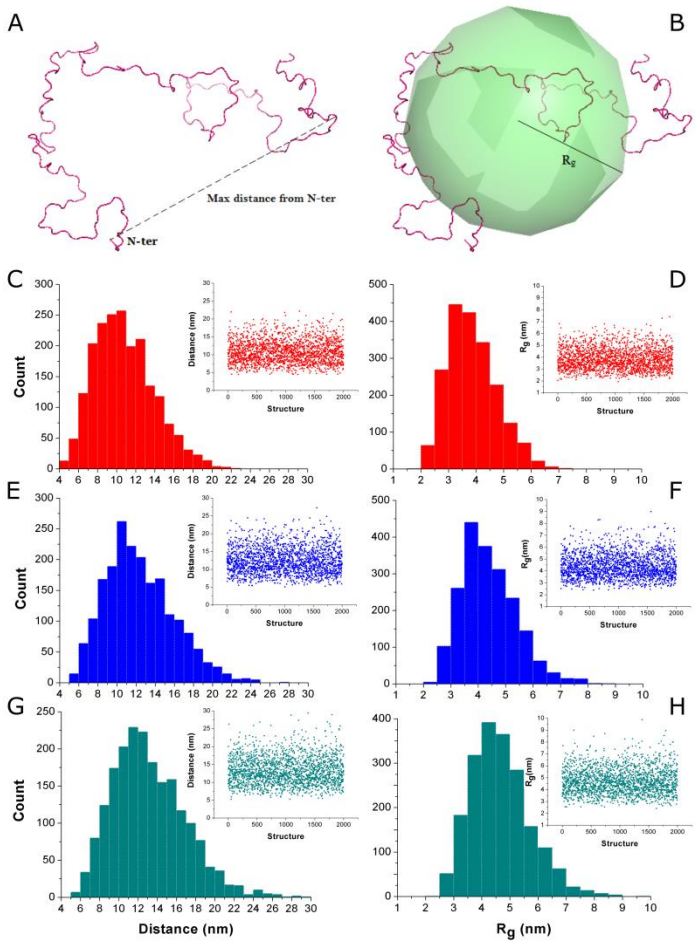

Figure 3. Maximum distance and radius of gyration analyses of the disordered ensembles of G protein. A) Representation of the maximum distance measured between the N atom of the first extracellular residue (Asn52) and the more distant atom to this. B) Depiction of the radius of gyration calculated for an individual conformation. C, E, G) Histograms of the maximum distances measured in the disordered ensemble for the wild type, 111- and 180-nucleotide duplications variants of the G protein, respectively. Insets at the right-hand part of each panel depict scatter-plots of maximum distances versus structure. D, F, H) Histograms of the radius of gyration of structures in the disordered ensemble for the wild type, 111- and 180-nucleotide duplications variants of the G protein, respectively. Similarly, insets at the right-hand part of each panel show scatter-plots of radius of gyration versus structure of each ensemble.

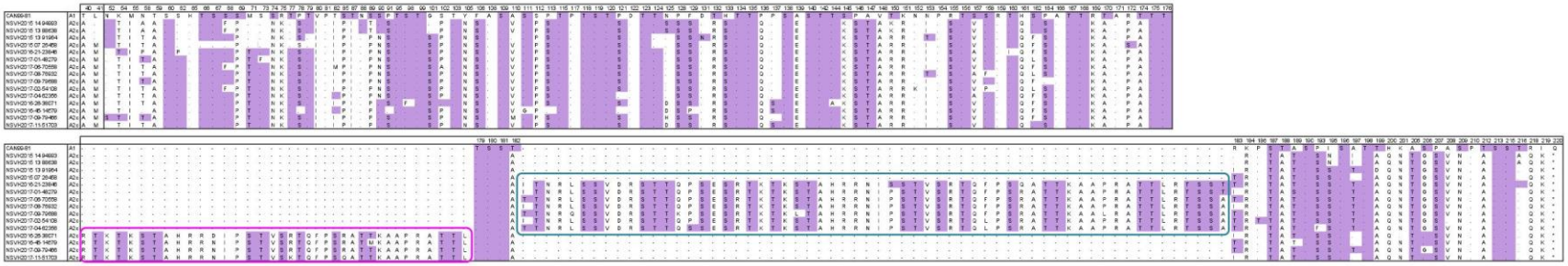

Figure 4. Glycosylation pattern of HMPV-A G protein. Multiple alignment of deduced G amino acid sequences of HMPV-A. Only positions reflecting an amino acid change or a putative N- or O-glycosylated site are shown, amino acids were numerated following the reference sequence CAN99-81 (accession number AY574224). Amino acid positions 40 and 41 correspond to the transmembrane domain, while the remaining positions correspond to the ectodomain. The 180-nucleotide duplications in the ectodomain are framed in a turquoise box, the 111-nucleotide duplications are in a pink box. N- and O-glycosylated sites are marked in purple.