Archives of Virology
https://doi.org/10.1007/s00705-018-3986-4

BRIEF REPORT

Comparative genomics of human rubulavirus 2

Maja Šantak[1] · Gordana Mlinarić-Galinović[2] · Tatjana Vilibić-Čavlek[2,3] · Irena Tabain[3]

Received: 5 April 2018 / Accepted: 27 June 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract
Although human rubulavirus 2 (HPIV2) is an important respiratory pathogen, little is known about its molecular epidemiology. We performed a comparative analysis of the full-length genomes of fourteen HPIV2 isolates belonging to different genotypes. Additionally, evolutionary analyses (phylogenetic reconstruction, sequence identity, detection of recombination and adaptive evolution) were conducted. Our study presents a systematic comparative genetic analysis that complements prior analyses and utilizes full-length HPIV2 genomes to provide a basis for future work on the clinical significance, molecular variation and conservation, and evolution of HPIV2.

Human rubulavirus 2 (human parainfluenza virus type 2, HPIV2) is a non-segmented, negative-strand RNA virus classifable within the family *Paramyxoviridae*, genus *Rubulavirus*. The genomic RNA is 15 654 nt long and contains six genes encoding for seven proteins in the following order: the *Nucleoprotein* gene (*N*), the *Phosphoprotein /V* protein gene (*P/V*), the *Matrix* protein gene (*M*), the *Fusion* protein gene (*F*), the *Haemagglutinin neuraminidase* gene (*HN*) and the *RNA dependent RNA polymerase* (or *Large* protein) gene (*L*). HPIV2 enters the cell by attaching to sialic acid as a receptor. Binding of the HN protein to the receptor initiates conformational change of the F protein, which ultimately leads to the fusion of the viral and cellular membranes [1, 2] and the entrance of the viral ribonucleoprotein complex into the cell. The ribonucleoprotein complex is composed of viral RNA tightly bound by the viral N protein. N protein binds two viral proteins: P protein and RNA-dependent RNA polymerase (RdRp) (L protein). Together, this complex is responsible for viral transcription and replication. To initiate short uncapped leader RNA transcription or full-length anti-genome replication, RdRp binds to the extreme 3′ terminus of the genome and recognizes the leader sequence (*le*), which is approximately 50 nt in length and located at the 3′ terminus of the genome [3]. The same complex is responsible for the synthesis of the complete genomic RNA as it recognizes the short trailer sequence (*tr*) of the anti-genome RNA.

Human parainfluenza viruses (HPIV1, HPIV2, HPIV3 and HPIV4) are the second major cause of hospitalization for respiratory tract illness in young children (only respiratory syncytial virus (RSV) precedes) [4]. HPIV2 and HPIV4 cause acute respiratory infection (ARI, croup and pneumonia) in infants and young children [5], although their prevalence rates are lower than those of HPIV3 and HPIV1. The relative contribution to the burden of disease in infants and young children is estimated at approximately 7:3.5:2:1 for HPIV3, HPIV1, HPIV2 and HPIV4, respectively [6–8]. Almost all children encounter HPIVs within the first few years after birth, but immunity is incomplete, and re-infections occur throughout life. However, after primary infection, protection against severe disease develops [4]. No licensed vaccines are available for any HPIV types due to incomplete knowledge regarding the immune response to these viruses.

Epidemiological and phylogenetic studies of HPIV2 infections are rare due to the low number of hospitalized patients with HPIV2 and the fact that most respiratory

Handling Editor: William G Dundon.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00705-018-3986-4) contains supplementary material, which is available to authorized users.

✉ Maja Šantak
msantak@unizg.hr

[1] Centre for Research and Knowledge Transfer in Biotechnology, University of Zagreb, Rockefellerova 10, 10000 Zagreb, Croatia

[2] School of Medicine, University of Zagreb, Šalata 3, 10000 Zagreb, Croatia

[3] Croatian National Institute of Public Health, Rockefellerova 12, 10000 Zagreb, Croatia

Published online: 10 August 2018

infections are treated symptomatically outside of the hospital setting; thus, HPIV2 infection is not typically laboratory confirmed. Therefore, it is difficult to estimate the exact number of HPIV2-infected patients and to follow the genetic variation and evolution of the virus. However, sometimes, especially in children, the elderly, or immunocompromised persons, hospitalization is needed due to the severity of the disease, which makes this virus a relevant pathogen from both clinical and public health perspectives. Virological surveillance is important to determine the virus circulation and unbiased significance of viral infection in a population. Previous phylogenetic studies of HPIV2 strains circulating in France [9], Saudi Arabia [10] and Croatia [11] identified four clusters (G1 to G4) with additional subclusters in clusters G1 and G4; however, it seems that when more isolates are added to the phylogenetic analysis (Figure 1), the G3 genotype could also be further divided into two subclusters, G3a and G3b. These studies were based on the coding region of the *HN* gene because the HN protein has been shown to be extensively antigenically variable in contrast to structural proteins, which are reported to be antigenically well conserved in a number of HPIV2 clinical isolates [11]. We have shown that both genes encoding surface glycoproteins, the *HN* gene and the *F* gene, individually are equally suitable for phylogenetic analyses, although longer sequences spanning both genes are advantageous when the analysed isolates are temporally and geographically highly related and the number of nucleotide differences is extremely low [12].

In this study, we performed a systematic comparative genetic analysis that complements prior analyses and utilizes full-length HPIV2 genomes to provide a basis for future work on the clinical significance, molecular variation and conservation, and evolution of HPIV2.

To date, there is not much data regarding the evolution of HPIV2, and a complete molecular analysis of the HPIV2 genome has not been done. The number of complete HPIV2 genomes in GenBank is very low. Presently, there are 35 complete HPIV2 genomes reported but only four (one genome from genotype G1a, one from genotype G3 and two from genotype G4b) actually have complete sequences of all 15 654 nucleotides, while the other 31 sequences lack up to 268 nucleotides at their terminal regions. Therefore, we determined the whole-genome sequences of an additional ten HPIV2 isolates (five belonging to genotype G1a and five belonging to genotype G3, Figure 1). Ten HPIV2 samples were isolated from nasopharyngeal aspirates or swabs from patients hospitalized in Zagreb, Croatia, during 2011, 2012, and 2014-2016 (the study was approved by the Ethics Committee of the University Hospital for Infectious Diseases "Dr. Fran Mihaljević", number 01-692-1-2014). Virus isolation was performed according to the method reported previously [12]. Viral RNA was isolated either from clinical samples or from Vero cell supernatants using previously reported methods [13]. cDNA synthesis, PCR amplification and sequencing conditions are reported in [12]. Primer sequences used for PCR and sequencing are shown in Supplementary Table 1. The sequences generated in this study were submitted to GenBank under accessiosn numbers MG836416 to MG836425. The accession numbers for sequences obtained from GenBank that were included in alignments are the following: DQ072586-DQ072589, JF912194, JF912195, NC_003443, AF533010-AF533012, D00865, AB176531, AB189948-AB189953. Multiple sequence alignments were made using the CLUSTAL W program integrated into MEGA6.0 software (www.megasoftware.net). A phylogenetic tree based on the *HN* gene was generated in MEGA6.0 using maximum likelihood (ML) analysis with the Kimura 2-parameter substitution model and 1,000 bootstrap replicates. The average similarity of HPIV2 genomes of three genotypes (G1a, G3 and G4b) was assessed using SimPlot v3.5.1. software [14]. The same software was used to conduct a bootscan analysis of the HPIV2 genomes. Sequences were aligned by either Clustal Omega (www.ebi.ac.uk) or Clone Manager v6.0 (Sci-Ed Software). A similarity plot of the complete genomic sequences of different HPIV2 genotypes, in which the Greer strain was used as the query sequence, clearly indicated that some regions in the genome are more variable than others (Figure 2). The most divergent regions are the non-coding regions of the genome with an overall similarity below 86% (Table 1). The exception is the 5' non-coding region of the *L* gene, which has an overall similarity of 90.7% (Table 1). This finding is not unexpected because these regions do not code for amino acids and therefore have higher flexibility in their nucleotide content. In spite of their variability, these regions are not eligible for phylogenetic analysis and evolutionary monitoring because they are short regions (between 43 and 807 nt) and because selective pressure does not alter the composition of these regions.

High degrees of similarity are seen within the coding regions, where the overall similarity is above 90% (Table 1), with the exception of the coding regions of the *F* and *HN* genes, which encode for surface glycoproteins and have similarities of 87.3% and 88.4%, respectively. Thus, we confirmed that the *F* and *HN* regions are the most variable regions of this virus (Table 1) and determine that they should continue to be used as phylogenetic markers. Usually, only one region (more often the *HN* gene) is used. Previously, we showed that closely related isolates, e.g., those from the same season, can be falsely proclaimed to be identical if only one of these two regions is analysed; therefore, we have suggested [12] that both regions should be used, although it may be a demanding process to analyse a 4-kb region.

Leader (*le*) and trailer (*tr*) promoters are also found to be highly similar between the fourteen HPIV2 strains (Supplementary Figure S1, Table 1). The trailer region

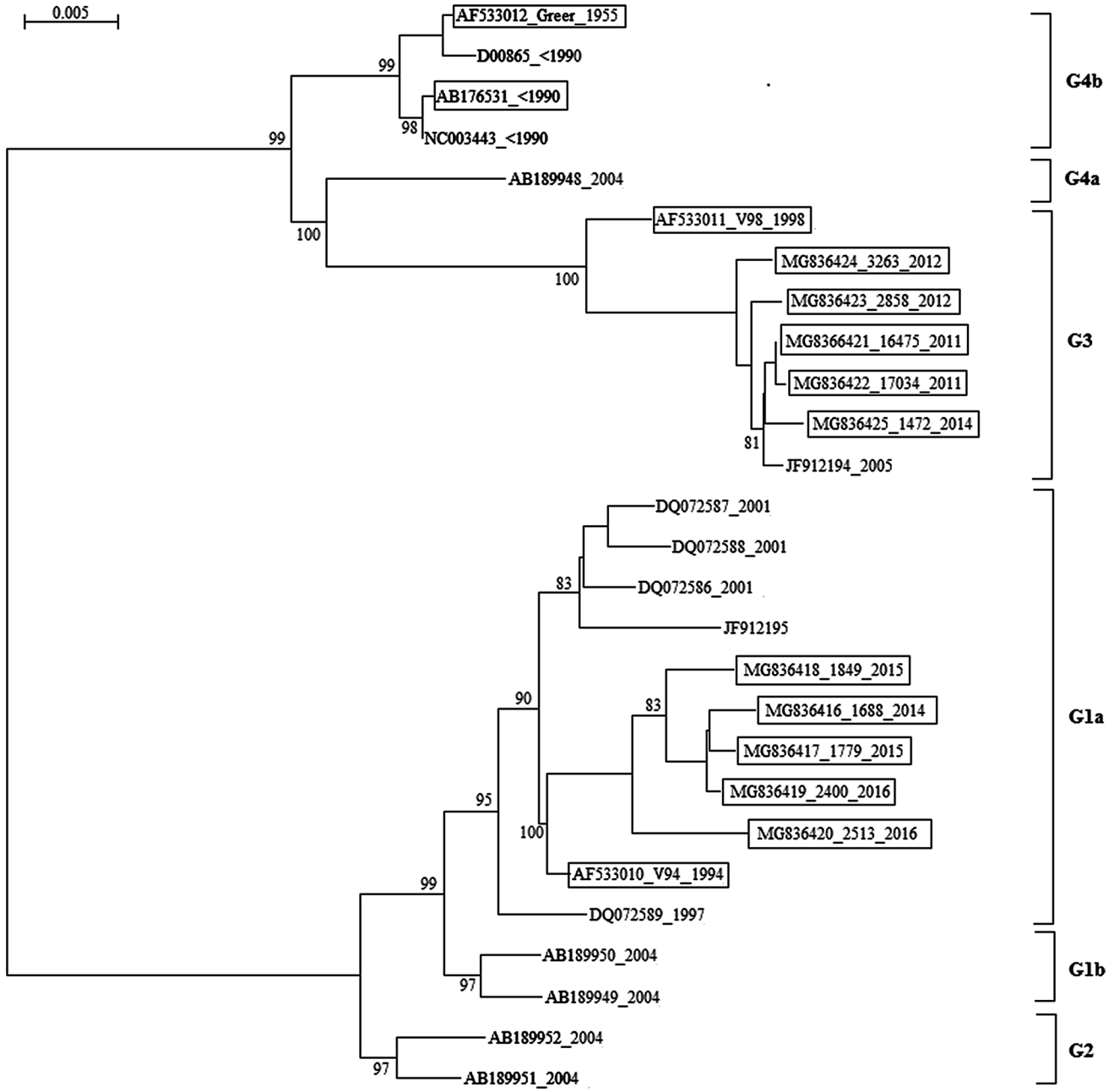

Fig. 1 Maximum likelihood phylogenetic tree of the *HN* gene ORF of HPIV2 strains constructed by MEGA6.0 with 1000 bootstrap replicates. Isolates are presented either with the accession number from GenBank or with the accession number from GenBank and their isolate numbers. The year of isolation is indicated for each sequence. Strains whose complete sequences were used for the analyses are squared. The genotypes are indicated by the brackets on the right side. Only bootstrap values over 80% are displayed at the branch nodes. The scale bar indicates 0.01 nt differences per site, over the indicated region

is identical in all strains, while the leader region shows 90.9% similarity. These regions are multifunctional entities (reviewed in [15]) which explains a high level of evolutionary stability (Supplementary Figure S1). A regulatory element in the *le* region between 25 and 30 nt in length that modulates RNA synthesis to gain optimal growth conditions for infected cells has been identified [16]. This region is 100% conserved in all analysed sequences.

HPIV2 and mumps rubulavirus are both human pathogens and taxonomically belong to the same genus, *Rubulavirus*. Since isolates of these two species are phylogenetically

Table 1 Total variability of the HPIV2 genome regions, for the fourteen HPIV2 complete genome sequences

Genome region			No. of overall nt differences/identity (%)
Region *	Start nt	Stop nt	
leader	1	55	5/90.9
NP 3' nc	56	157	15/85.1
NP cd	158	1 789	137/91.6
NP 5' nc/V 3' nc	1 790	1 996	56/72.8
V cd	1 997	2 674	56/91.7
P cd	1 997	3 184	111/90.7
V 5' nc/M 3' nc	2 675	3 482	117/85.5
M cd	3 483	4 616	98/91.3
M 5' nc/F 3' nc	4 617	4 792	34/80.6
F cd	4 793	6 448	210/87.3
F 5' nc/HN 3' nc	6 449	6 820	57/84.6
HN cd	6 821	8 536	198/88.4
HN 5' nc/L 3' nc	8 537	8797	42/83.8
L cd	8 798	15 589	438/93.5
L 5' nc	15 590	15 633	4/90.7
trailer	15 634	15 654	0/100

*nc – non-coding region; cd – coding region

closely related within the *Paramyxoviridae* family, an analysis of their similarity was performed on two viruses isolated in the 1950s: HPIV2 Greer virus from 1955 and mumps virus L-Zagreb from 1953. The average identity was 56.5% with 6.6% nt gaps. Mumps rubulavirus has an additional gene encoding the SH protein. A similarity plot (Supplementary Figure S2) shows that the coding regions of other genes are highly similar between the two viral species, especially in the N segment (except for the 5' end) and the complete L segment. The same is reflected at the protein level (Supplementary Figure S3), where the identity in the N protein is 49.4% and 59.2% in the L protein. Other proteins have a lower identity rate: 36.0% in the V protein, 36.5% in the P protein, 41.1% in the M protein, 38.1% in the F protein and 40.5% in the HN protein. The leader and trailer regions have high similarity as well, with 64.6% and 50.0% similarity, respectively. In contrast to the coding regions and the terminal sections of the genome, the non-coding regions have a much lower level of similarity (Supplementary Figure S3), and the majority of identified gaps are found in these regions.

The overall protein identity for each of the fourteen HPIV2 isolates is depicted in Table 2 and ranges from 88.1% to 98.6% depending on the protein. The average identity of the F protein is 90.4% (Table 2), indicating that this protein has a higher level of variability than proteins in the RdRp complex. However, more than half of the substitutions (31/53, or 58.5%) are found between 372-551 aa. In contrast, the 176-371 aa region contains only 5 substitutions, indicating a highly conserved region spanning almost 200 aa. HN protein is identified as the HPIV2 protein with the lowest identity level, showing an average identity of only 88.1% (Table 2). Although the HN protein is a protein with high variability, there is not an identifiable region in which substitutions are aggregated; rather, they are fairly uniformly distributed. In contrast to the fact that the HN coding region is the second most variable region in the HPIV2 genome, there are short regions of the HN protein that are highly evolutionary conserved even at the intragenus level. When the HN protein of HPIV2 was compared to the mumps rubulavirus HN protein (Supplementary Figure S3), regions 228-NRKSCS-233 and 393-GAEGRV/L-398 with neuraminidase and receptor binding activities [17, 18], respectively, were proven to be conserved in both HPIV2 and mumps rubulavirus (Supplementary Figure S3). A large conserved region spanning 151-186 aa in HPIV2 and 163-198 aa in mumps rubulavirus was identified with a 91.7% identity (33 out of 36 aa are identical). Additional analysis involving HPIV3 and HPIV5 (data not shown) revealed that this region is also conserved among all four viruses, indicating its functional relevance. The function of this region of the HN protein is so far unknown. However, *in silico* structural analysis (data not shown) indicates that the region 165-187 aa is located in close proximity to the neuraminidase region 228-NRKSCS-233 and forms part of the neuraminidase propeller. Furthermore, the region 148-209 aa was identified as the region of HN protein that promotes F-mediated cell fusion [19]. These conserved regions prove that these viruses, although having evolved from the same origin and developed numerous unique regions, have kept major functional regions intact or highly conserved. Located on the inner side of the viral envelope, M protein has a much higher level of identity than surface glycoproteins, with an average identity of 95.5% (Table 2). The majority of these substitutions (11/17, or 64.7%) are located in a region spanning less than 100 aa, between 129 and 225 aa. This region has an average identity of only 88.5%. In contrast, the highly conserved regions between 1 and 128 aa and between 236 and 377 aa contain only 2 and 3 substitutions, respectively. For the P protein, the average identity was calculated to be 92.7% (Table 2), but no particularly variable or conserved regions could be identified. Since the V protein shares two-thirds of its protein sequence with the P protein, the variability of the V protein is comparable to that of the P protein. It was found that the average identity was 93.3% (Table 2), with neither highly variable nor highly stable regions identified.

The average identity for the nucleoprotein is 93.4% (Table 2). Inspection of the N protein sequence alignment indicated that the region between 204 and 373 aa is a highly conserved region with 100% identity for all 14 HPIV2 samples. In contrast, the region between 422 and 543 aa (C-terminal segment of the protein) is highly variable, with

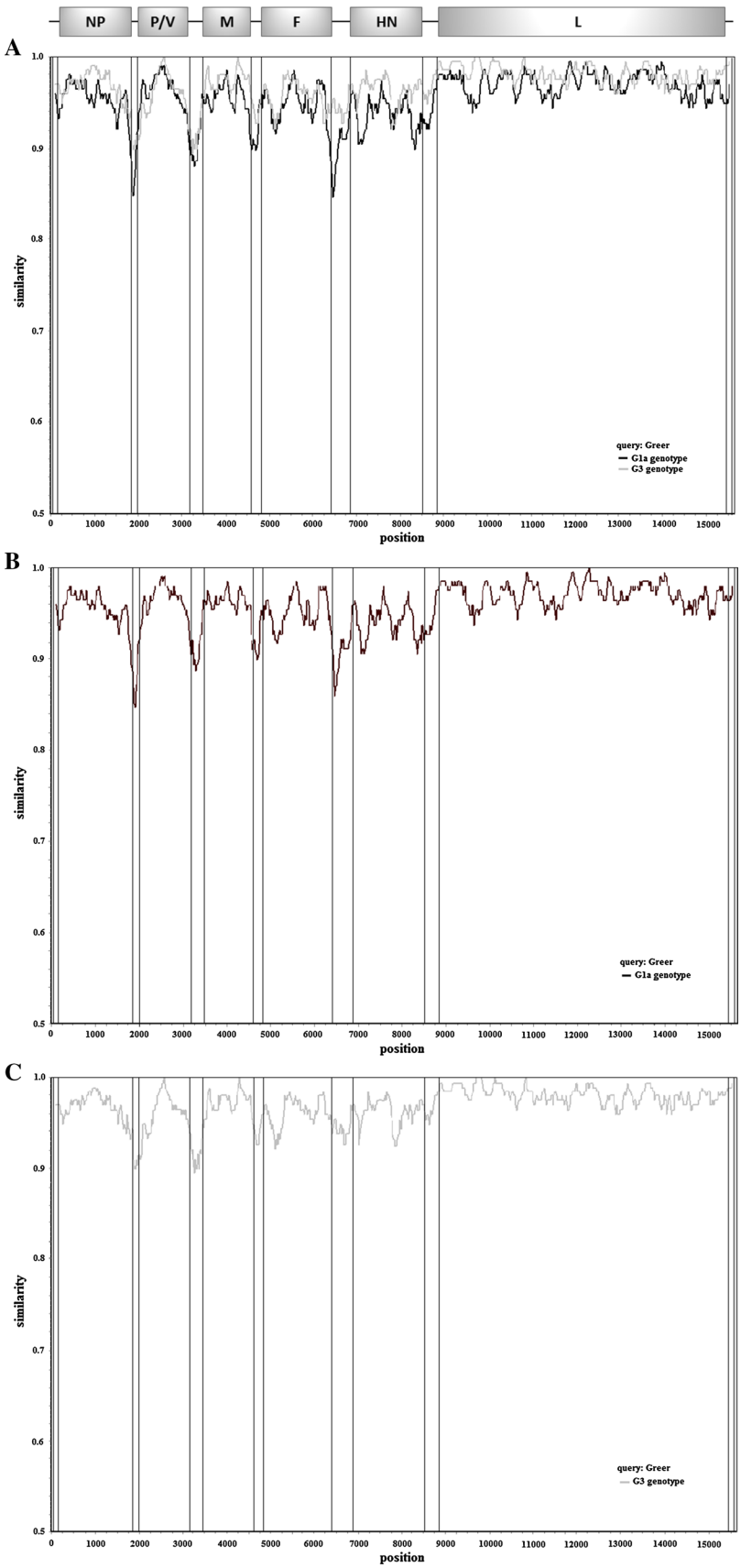

Fig. 2 Similarity Plot against the query sequence (Greer) for a panel of HPIV2 sequences belonging to genotypes G1a (A and B) or G3 (A and C). The analysis was conducted via SimPlot using a sliding window of 200 nucleotides moving in steps of 20 nucleotides with conditions: gapstrip on, Kimura (2-parameter) and T/t: 2.0. A schematic of the complete HPIV2 genome is depicted at the top

Table 2 Total variability of the HPIV2 protein sequences and synonymous versus non-synonymous substitutions rates for each protein gene, for the fourteen HPIV2 strains sequenced

Protein	Protein size (aa)	No. of total aa substitutions/identity (%)	No. of overall synonymous substitutions	No. of overall non-synonymous substitutions	ω value*
nucleoprotein (NP)	543	36/93.4	90	47	0.067-0.160
V protein	225	15/93.3	38	18	0.025-0.204
phosphoprotein (P)	395	29/92.7	78	33	0.048-0.178
matrix (M) protein	377	17/95.5	72	26	0.030-0.188
fusion (F) protein	551	53/90.4	142	68	0.035-0.128
hemagglutinin neuraminidase (HN)	571	68/88.1	108	90	0.120-0.280
large (L) protein	2 263	31/98.6	396	42	0.015-0.042

* $\omega = K_a/K_s$, K_a – synonymous substitution rate, K_s – non-synonymous substitution rate. $p < 0.0001$, the p value was computed by Fisher exact test

63.9% (23/36) of the substitutions found in this segment and an average identity of 81%. The L protein is proven to be a highly conserved protein with the average identity of 98.6% among the fourteen HPIV2 isolates (Table 2). Several conserved regions with 100% identity were identified: 330-484 aa, 797-897 aa, 1029-1178 aa and 1406-1593 aa. Additionally, the region between 1 and 229 aa contains only 2 substitutions. This protein does not seem to contain any highly variable regions. Thus, the N and L proteins show very low levels of variability with the exception of the C-terminal portion of the N protein. Similar result were obtained even at the intragenus level (Supplementary Figure S3), which indicates that there are high demands for the preservation of the primary structure for these proteins. The N-terminal portion (1-370 aa) of the nucleoprotein is highly conserved with an identity of 97.3% (10/370 aa positions are variable), while the C-terminal portion (371-543 aa) seems to be rather variable with only 83.8% identity (28/173 aa positions are variable). The conserved N-terminal region was identified as an essential region for interactions between monomers of the N protein, while the region between 295 and 494 aa was required for binding of P protein [20]. Later, a region spanning 403-494 aa on the N protein was identified as an essential region for binding to the L protein [21]. Therefore, it seems that there are not very strict demands for the primary structure of the N protein for binding the P protein and L protein, but the self-assembly process is highly sensitive to changes in the primary structure. In the same publication [21], the authors identified a region in the P protein from 278 to 353 aa as a region required for binding to the L protein. In contrast to the variable region on the N protein that binds L protein, a region of the P protein with the same function is more conserved (70/75 aa, 93.3% identity) indicating certain structural demands.

Selective pressure drives natural selection and confers beneficial adaptations to the population. Thus, it is a major mechanism of evolution. To detect the influence of selection, the ratio of the number of substitutions causing amino acid replacements (nsyn sites) per total possible nsyn sites (k*a*) was divided by the number of silent substitutions (syn sites) per possible syn sites (k*s*). The k*s,* k*a* and their ratio *ω* were calculated using KaKs_Calculator software [22] with the method of model averaging (MYN and GY). Statistical significance was set at $p < 0.0001$. The number of synonymous changes was higher than that of non-synonymous changes, indicating negative selection pressure (Table 2). This is in accordance with the fact that most protein-coding genes are considered to be under the effect of purifying selection [23]. The average acceptance rate (ω) is < 1, indicating that, on average, purifying selection dominates the evolution of all genes and does not affect the integrity of the encoded proteins. The K_a/K_s ratio is especially low for the polymerase region (ω value ranges from 0.015 to 0.042 for all models), while it is much higher for the HN region (ω value ranges from 0.120 to 0.280 for all models), indicating that negative selection is less pronounced for the HN gene, while it strongly affects the L gene.

To estimate the force of selective pressures (positive selections) across various regions of the viral genome, all coding regions were analysed using multiple methods from the Datamonkey server [24] (http://www.datamonkey.org/): single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), internal branch fixed-effects likelihood (IFEL), random effects likelihood (REL) and mixed effects model of evolution (MEME). Sites were considered to be positively selected if they met the cut-off criteria of p-value < 0.1 for SLAC, FEL, IFEL and MEME, or Bayes factors > 50 for REL. The codons which were identified using at least two methods were accepted as positively selected codons. In all coding regions, except for the V region, at least one positively selected codon was found. Coding regions for proteins that are not present at the viral surface, namely, the N, P and M proteins and RdRp, had only one identified positively selected codon: positions 487 (Pro and

Leu), 212 (Ala and Val), 309 (Asn and Thr) and 1400 (Thr and Leu), respectively. In contrast the F and HN proteins, which are viral surface glycoproteins, were found to contain four (positions 46 (Thr and Ser), 104 (Gln and Arg), 303 (Gln and Leu) and 545 (Gly, Glu and Lys)) and five (positions 65 (Pro and Ser), 152 (Asn and Pro), 345 (Gln and Arg), 479 (Pro and Leu) and 547 (Leu and Phe)) positively selected codons, respectively. Therefore, it seems that surface glycoproteins are under greater selective pressure than other proteins. Epitopes for either of the surface glycoproteins have not been identified thus far, and so it is not possible to comment on the position of positively selected codons. However, this is a strong indication that a mechanism to evade the host immune system is a major driving force for viral evolution.

Natural recombination events in paramyxoviruses have been reported in Newcastle disease virus (NDV) [25–27], in which recombination has occurred between different lineages. Since we assessed HPIV2 isolates that coincided spatially and temporally, we also analysed potential recombination events. A standard bootscan analysis (SimPlot) was used to analyse possible recombination events in the fourteen HPIV2 isolates sequenced. The analysis detected evidence of possible recombination events. Therefore, we conducted a more intense search for potential recombination breakpoints. Several methods from the Datamonkey server [24] (http://www.datamonkey.org/) and the Recombination Detection Program v.4.95 (RDP4) [28] were used. Various methods, such as RDP, GENECONV, BootScan, MaxChi, Chimaera and SiScan, all with default settings, were used by the RDP4 program. Possible breakpoints were detected at $p < 0.01$. The GARD method examined 50 349 models and found evidence of 5 breakpoints at positions 5825, 6527, 8357, 8680 and 11504. SBP analysis was performed using HKY85 and TrN93 models and yielded a potential breakpoint at position 5054. Finally, the RDP4 program detected possible breakpoints at positions 2261-2710, 3990-5979, 5955-6201, 6779-7198 and 7985-8573. None of the identified potential breakpoints overlapped, and sequence analysis of the identified potential breakpoints could not confirm the recombination. Thus, this finding does not seem to be convincing, especially because the predicted recombined segments were shorter than 700 nt. We conclude at this point that the false-positive detection of potential recombination sites is likely due to a high level of similarity between the analysed sequences and that given the evidence to date, homologous recombination seems to play little or no role in the evolution of HPIV2. However, since we analysed a limited population, analysis of a larger number of HPIV2 complete genome sequences might lead to a different conclusion.

In conclusion, thus far, there is not much known about HPIV2 in spite of the fact that it is a relevant human pathogen. Therefore, this study provides some insight into the genetic/phylogenetic variations and molecular characterizations of this virus.

Acknowledgements This work has been supported by the Croatian Science Foundation under the project number 6255 and by the grant "Strengthening the capacity of CerVirVac for research in virus immunology and vaccinology", KK.01.1.1.01.0006, awarded to the Scientific Centre of Excellence for Virus Immunology and Vaccines and co-financed by the European Regional Development Fund.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Lamb RA, Paterson RG, Jardetzky TS (2006) Paramyxovirus membrane fusion: lessons from the F and HN atomic structures. Virology 344:30–37
2. Bose S, Heath CM, Shah PA, Alayyoubi M, Jardetzky TS, Lamb RA (2013) Mutations in the parainfluenza virus 5 fusion protein reveal domains important for fusion triggering and metastability. J Virol 87:13520–13531
3. Whelan SP, Barr JN, Wertz GW (2004) Transcription and replication of nonsegmented negative-strand RNA viruses. Curr Top Microbiol Immunol 283:61–119
4. Hall CB (2001) Respiratory syncytial virus and parainfluenza virus. N Engl J Med 344:1917–1928
5. Chanock RM (1956) Association of a new type of cytopathogenic myxovirus with infantile croup. J Exp Med 104:555–576
6. Murphy BR (1988) Current approaches to the development of vaccines effective against parainfluenza viruses. Bull World Health Organ 66:391–397
7. Reed G, Jewett PH, Thompson J, Tollefson S, Wright PF (1997) Epidemiology and clinical impact of parainfluenza virus infections in otherwise healthy infants and young children 5 years old. J Infect Dis 175:807–813
8. Yano T, Fukuta M, Maeda C, Akachi S, Matsuno Y, Yamadera M, Kobayashi A, Nagai Y, Kusuhara H, Kobayashi T, Amano H, Nishinaka T, Ochiai H, Watanabe M, Nakamura H, Suga S, Ihara T (2014) Epidemiological investigation and seroprevalence of human parainfluenza virus in Mie Prefecture in Japan during 2009–2013. Jpn J Infect Dis 67:506–508
9. Terrier O, Cartet G, Ferraris O, Morfin F, Thouvenot D, Hong SS, Lina B (2008) Characterization of naturally occurring parainfluenza virus type 2 (hPIV-2) variants. J Clin Virol 43:86–92
10. Almajhdi FN, Alshaman MS, Amer HM (2012) Human parainfluenza virus type 2 hemagglutinin–neuramindase gene: sequence and phylogenetic analysis of the Saudi strain Riyadh 105/2009. Virol J 9:316
11. Tsurudome M, Nishio M, Komada H, Bando H, Ito Y (1989) Extensive antigenic diversity among human parainfluenza type 2 virus isolates and immunological relationships among paramyxoviruses revealed by monoclonal antibodies. Virology 171:38–48
12. Šantak M, Slović A, Ljubin-Sternak S, Mlinarić Galinović G, Forčić D (2016) Genetic diversity among human parainfluenza virus type 2 isolated in Croatia between 2011 and 2014. J Med Virol 88:1733–1741

13. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal Biochem 162:156–159
14. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73:152–160
15. Noton SL, Fearns R (2015) Initiation and regulation of paramyxovirus transcription and replication. Virology 479–480:545–554
16. Matsumoto Y, Ohta K, Goto H, Nishio M (2016) Parainfluenza virus chimeric mini-replicons indicate a novel regulatory element in the leader promoter. J Gen Virol 97:1520–1530
17. Jorgensen ED, Collins PL, Lomedico PT (1987) Cloning and nucleotide sequence of Newcastle disease virus hemagglutinin–neuraminidase mRNA: identification of a putative sialic acid binding site. Virology 156:12–24
18. Mirza AM, Deng R, Iorio RM (1994) Site-directed mutagenesis of a conserved hexapeptide in the paramyxovirus hemagglutinin–neuraminidase glycoprotein: effects on antigenic structure and function. J Virol 68:5093–5099
19. Tsurudome M, Kawano M, Yuasa T, Tabata N, Nishio M, Komada H, Ito Y (1995) Identification of regions on the hemagglutinin–neuraminidase protein of human parainfluenza virus type 2 important for promoting cell fusion. Virology 213:190–203
20. Nishio M, Tsurudome M, Ito M, Kawano M, Kusagawa S, Komada H, Ito Y (1999) Mapping of domains on the human parainfluenza virus type 2 nucleocapsid protein (NP) required for NP-phosphoprotein or NP–NP interaction. J Gen Virol 80:2017–2022
21. Nishio M, Tsurudome M, Ito M, Ito Y (2000) Mapping of domains on the human parainfluenza type 2 virus P and NP proteins that are involved in the interaction with the L protein. Virology 273:241–247
22. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J (2006) KaKs calculator: calculating Ka and Ks through model selection and model averaging. Genom Proteom Bioinform 4:259–263
23. Cristianini N, Hahn MW (2007) Introduction to computational genomics: a case studies approach. University Press, Cambridge
24. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457
25. Chong YL, Padhi A, Hudson PJ, Poss M (2010) The effect of vaccination on the evolution and population dynamics of avian paramyxovirus-1. PLoS Pathog 6:e1000872
26. Yin Y, Cortey M, Zhang Y, Cui S, Dolz R, Wang J, Gong Z (2011) Molecular characterization of Newcastle disease viruses in Ostriches (*Struthio camelus* L.): further evidences of recombination within avian paramyxovirus type 1. Vet Microbiol 149:324–329
27. Satharasinghe DA, Murulitharan K, Tan SW, Yeap SK, Munir M, Ideris A, Omar AR (2016) Detection of inter-lineage natural recombination in avian paramyxovirus serotype 1 using simplified deep sequencing platform. Front Microbiol 7:1907
28. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1(1):vev003 **(eCollection 2015)**