Genome-Wide Analysis of Human Metapneumovirus Evolution

Jin Il Kim[1], Sehee Park[1], Ilseob Lee[1], Kwang Sook Park[1], Eun Jung Kwak[1], Kwang Mee Moon[1], Chang Kyu Lee[2], Joon-Yong Bae[1], Man-Seong Park[1], Ki-Joon Song[1]*

1 Department of Microbiology, the Institute for Viral Diseases and Korea Bank for Pathogenic Viruses, College of Medicine, Korea University, Seoul, Republic of Korea, 2 Department of Laboratory Medicine, College of Medicine, Korea University, Seoul, Republic of Korea

* songmicr@korea.ac.kr

Abstract

Human metapneumovirus (HMPV) has been described as an important etiologic agent of upper and lower respiratory tract infections, especially in young children and the elderly. Most of school-aged children might be introduced to HMPVs, and exacerbation with other viral or bacterial super-infection is common. However, our understanding of the molecular evolution of HMPVs remains limited. To address the comprehensive evolutionary dynamics of HMPVs, we report a genome-wide analysis of the eight genes (N, P, M, F, M2, SH, G, and L) using 103 complete genome sequences. Phylogenetic reconstruction revealed that the eight genes from one HMPV strain grouped into the same genetic group among the five distinct lineages (A1, A2a, A2b, B1, and B2). A few exceptions of phylogenetic incongruence might suggest past recombination events, and we detected possible recombination breakpoints in the F, SH, and G coding regions. The five genetic lineages of HMPVs shared quite remote common ancestors ranging more than 220 to 470 years of age with the most recent origins for the A2b sublineage. Purifying selection was common, but most protein genes except the F and M2-2 coding regions also appeared to experience episodic diversifying selection. Taken together, these suggest that the five lineages of HMPVs maintain their individual evolutionary dynamics and that recombination and selection forces might work on shaping the genetic diversity of HMPVs.

Introduction

Human metapneumovirus (HMPV) is a single-stranded RNA virus that belongs to the family *Paramyxoviridae* [1, 2]. Mostly in young children and the elderly, HMPV causes respiratory distress similar to that associated with human respiratory syncytial virus (hRSV), ranging from upper respiratory disease to bronchiolitis and pneumonia [3–5]. Based on the previous reports, HMPV infection appears to be seasonal [6], and co-infection with other respiratory pathogens is common [7–10]. However, there are still no vaccines or antivirals that can be used to treat HMPV infection [11, 12].

Among the eight genes (N, P, M, F, M2, SH, G, and L) of HMPVs, the genetic property of the membrane glycoproteins F and G defines two major lineages, A and B, which can be further

classified into the four different lineages (A1, A2, B1, and B2) and two additional sublineages (A2a and A2b) [3, 13, 14]. Most of these lineage/sublineage strains have co-circulated, except for the A1 lineage that had been only detected until 2003 [15]. However, no specific lineages appear to have predominated among circulating HMPV strains [16, 17], and the dominant lineages vary year by year [18–21].

Previously, a close genetic relationship between HMPV and avian metapneumovirus (aMPV) subtype C was suggested by comparing sequence identity, genomic organization, and phylogenetic location [22–24]. A possible transmission hypothesis of the HMPV from avian species was then testified by analyzing the genomic sequences of N, F, and G coding regions and by determining serum antibody levels against the aMPV [25, 26]. Zoonotic cases of HMPV infection from humans to great apes were also reported [27–29]. Since then, various studies have assessed the genetic relationships between the HMPV lineages [18, 21, 30–32]. However, these all were based on the analyses of partial genes, and comprehensive understanding of evolutionary dynamics of HMPVs still remains limited.

As presented in the cases of RNA virus [10, 33–35], evolutionary forces, such as mutation and recombination, might have contributed to the genetic diversity of HMPVs. Through interactions or sequential actions, all of these forces might affect the accumulation of mutations arisen spontaneously or in response to the immune systems of infectees, which would eventually shape the evolutionary structure of HMPVs [36]. During this course of evolution, beneficial mutations are retained, but deleterious mutations are removed via purifying or negative selection [37]. For HMPV genes, negative selection force appeared to manage its effects on viral evolution, but only a limited number of residues has been detected as diversifying or positive selection sites, especially in the membrane glycoproteins G and F [18, 25]. Recombination is also considered an evolutionary force working on the genetic diversity of the *Pneumovirinae* viruses [38], and it was suspected in the aMPVs, based on the phylogenetic incongruence of the same viral genes [39]. However, there have been no confirmed cases of recombination reported in HMPVs.

A vaccine is the most effective prevention measure against infectious diseases [40]. Using computational approaches, antigenic clustering and transmission of some genetic signatures of other RNA viruses have been successfully demonstrated [41, 42]. Recently, a novel Baysian method of estimating phylogenetic signals against an uncertain prior condition has been suggested to investigate the genetic footprints of virulence and antigenic evolution of some viruses [43]. Given the ongoing diversifying progress of HMPVs, the computational methods may have crucial importance for developing effective medical arsenals against the circulation of divergent HMPV strains. When accompanied with the classical approaches of *in vitro* and *in vivo* virological testing, these may unveil scientific priorities much sooner for the development of vaccines and antivirals against HMPVs.

In this study, we analyzed 103 complete genome sequences of HMPVs using computational estimation methods of phylogenetic reconstruction, recombination, and natural selection. Based on the derived phylogenetic relationships and potential breakpoints of recombination, we suggested recombination as a possible mechanism underlying the evolutionary dynamics of HMPVs. We also provided the profiles of selection pressure analyses and discussed their evolutionary effects on the HMPV genetic diversity.

Materials and Methods

Sequence preparation

A total of 166 complete MPV sequences were downloaded from the Virus Pathogen Resource database (http://www.viprbrc.org/brc/home.spg?decorator=vipr). These were mainly

contributed by two HMPV sequencing projects (BioProject PRJNA73051 and 237298). Two prototype HMPV sequences, NL/00/01 (GenBank accession number AF371337) and NL/99/01 (AY525842), were not included in this study because no collection dates were available from them. Among 166 complete genomes, the sequences from unknown hosts (n = 39), no collection date information (n = 2: DQ843658 and DQ843659), avian origins (n = 10: HG934338, KC915036, DQ009484, NC007652, EF199771, EF199772, AB548428, FJ977568, HG934339, and JF424833), and gorilla sequence (n = 1: HM197719) were removed at first. In addition, 11 sequences having missing nucleotide information in the protein coding regions (KF530158, missing 5749–6141 and 6640–7212 nucleotide regions; KF530160, 1091–1423 and 6244–7220; KF530165, 6670–7236; KF530166, 6332–7098; KF530177, 8074–8142; KF530181, 6716–7234; KF530182, 5862–7477 and 7911–8183; KF530183, 4428–4898 and 7446–7783; KF530185, 6814–6887; KF686741, 3212–3575 and 6355–7095; and KF686743, 6416–7152) were also removed. Hence, 103 complete sequences of HMPVs including a Korean HPMV KUMC-MP strain (KF516922) were prepared for genomic investigation. After initial alignment using the MUSCLE method in MEGA5.2 [44], datasets of the eight genes were established by extracting corresponding coding region sequences. Each dataset was then re-aligned using the MAFFT program (v7.130b) [45] and trimmed at the N- and C-terminal regions to ensure that each dataset retained the open reading frame or common protein coding regions. The stop codon in the C-terminal region and sequence irregularities found at the C-termini of the SH and G coding regions were also removed. The resulting HMPV datasets were assigned to nucleoprotein N (nucleotide region: 31–1,182; nucleotide length: 1,152), phosphoprotein P (1–885; 885), matrix protein M (1–762; 762), fusion glycoprotein F (1–1,617; 1,617), matrix protein-2 M2 (1–724; 724; for M2-1 coding region, 1–561; and for M2-2, 512–724), small hydrophobic protein SH (1–531; 531), attachment glycoprotein G (1–444; 444), and RNA-dependent RNA polymerase L (1–6,015; 6,015). The complete genome sequence set was constituted by concatenating from the N through the L genes together (N-P-M-F-M2-SH-G-L in order). The aligned genomic sequences are provided as S1 Data.

Phylogenetic trees and evolutionary dynamics

Phylogenetic relationships, evolutionary rates (nucleotide, synonymous, and nonsynonymous substitutions/site/year), and the time (year) of the most recent common ancestor (tMRCA) were estimated using a time-framed Bayesian evolution analysis approach via a Markov Chain Monte Carlo (MCMC) inference method, implemented in the BEAST package (v1.8.1) [46]. The estimates were obtained based on the two tree models (Bayesian skygrid and Bayesian skyline) using the prior setting of the general time-reversal (GTR)+I+Γ substitution model and the lognormal relaxed molecular clock. The MCMC analysis ran for 200 million iterations, with sampling every 200 thousand iteration after 10% burn-in. Two independent runs for each dataset were combined and assessed to ensure their convergence in Tracer v1.6 [47]. The MCMC tree samples were used to infer the maximum clade credibility (MCC) trees for each dataset using TreeAnnotator v1.8.1. The MCC trees were then edited for better visualization using Fig-Tree v1.4.2. Relative genetic diversity ($N_e\tau$: N_e, the effective population size; τ, generation time) was also estimated by means of Bayesian skyline reconstruction to assess the population dynamics of each of the HMPV lineage sequences. The estimates were presented as mean values along with the lower and upper limits of the 95% highest probability density (HPD).

Recombination and natural selection

Two models of recombination analysis, the single breakpoint of recombination (SBP) and genetic algorithms for recombination detection (GARD) methods, were used to screen

potential breakpoints in the HMPV genomes in the condition of best-fit substitution models and beta-gamma site-to-site rate variation with four rate classes, using the web server of Data-monkey (http://www.datamonkey.org) together with the HyPhy package (v2.2) [48–50]. Based on the detected breakpoints, the nucleotide sequence sets were divided into pre- and post-breakpoint regions and used for the comparison of phylogenetic replacements using MEGA5.2 (maximum likelihood tree, GTR+I+Γ, 500 bootstrap replications) [44, 51]. Natural selection was also assessed using the same web server and HyPhy program. Site-specific pressures were analyzed by estimating the ratio of non-synonymous (dN) to synonymous substitutions (dS) at every codon using the single likelihood ancestor counting (SLAC) method employing the best-fit REV nucleotide substitution model (cutoff p-value, 0.05) [52]. The mixed effects model of evolution (MEME) method was also used to determine the sites of episodic and pervasive positive selection [53].

Isolation of a HMPV from clinical nasopharyngeal specimens

Clinical nasopharyngeal specimens have been sampled from the patients at Korea University Medical Center (KUMC) Guro Hospital to screen the presence of various respiratory pathogens in the specimens, which would be exempted from the approval of Institutional Review Board of KUMC Guro Hospital. A HPMV strain, named as KUMC-MP, was isolated from one of the specimens and deposited to Korea Bank for Pathogenic Viruses, Korea University College of Medicine. From this strain, we obtained its complete genome sequence (GenBank accession number KF516922) by reverse transcription-PCR and used it in this study. The sequences of the primer sets were presented in S1 Table.

Results

Phylogenetic relationships of HMPV genomes and potential recombination events

We first determined the phylogenetic relationships of the complete genome and individual N, P, M, F, M2, SH, G, and L genes from 103 complete HMPV sequences (S2 and S3 Tables). In the MCC trees, the sequences of the complete and individual genes constituted the five different lineages of A1, A2a, A2b, B1, and B2, and the eight genes from each virus were all assigned to the same lineage (Figs 1–3 and S1 Fig).

There were, however, two exceptions to the lineage consistency. The first exception was a TN94-49 virus isolated in the U.S. in 1994 (TN94, GenBank accession number JN184400). The N, P, and F genes of TN94 were located in the A2a sublineage (Fig 1A, 1B and 1C), whereas its SH, G, and L genes were in the A2b sublineage (Fig 2A, 2B and 2C) and its M and M2 genes were at the A2 parental locations (Fig 3A and 3B). The second example was more divergent. The N, P, M, M2, SH, and L genes of the HMPV/AUS/144834728/2003/A virus (AUS03, GenBank accession number KC562241) were all grouped with other A1 lineage sequences (Figs 1–3); however, the F gene was located independently between the A1 and A2 lineages (Fig 1C), and the G was assigned to the A2b sublineage (Fig 2B).

As the genes of HMPVs are consistently positioned in a linear genome [1], phylogenetic discordance between the genes of an individual HMPV might suggest past recombination events [39]. To test this hypothesis, we analyzed the sequence sets (nucleotide and amino acid) of nine protein-coding regions (after the M2 gene sequences were separated into two protein-coding regions M2-1 and M2-2) using the SBP and GARD methods, which could detect potential recombination breakpoints [49]. We screened the nucleotide sequence sets first, and then, checked the corresponding amino acid sequence sets of nine coding regions. The GARD
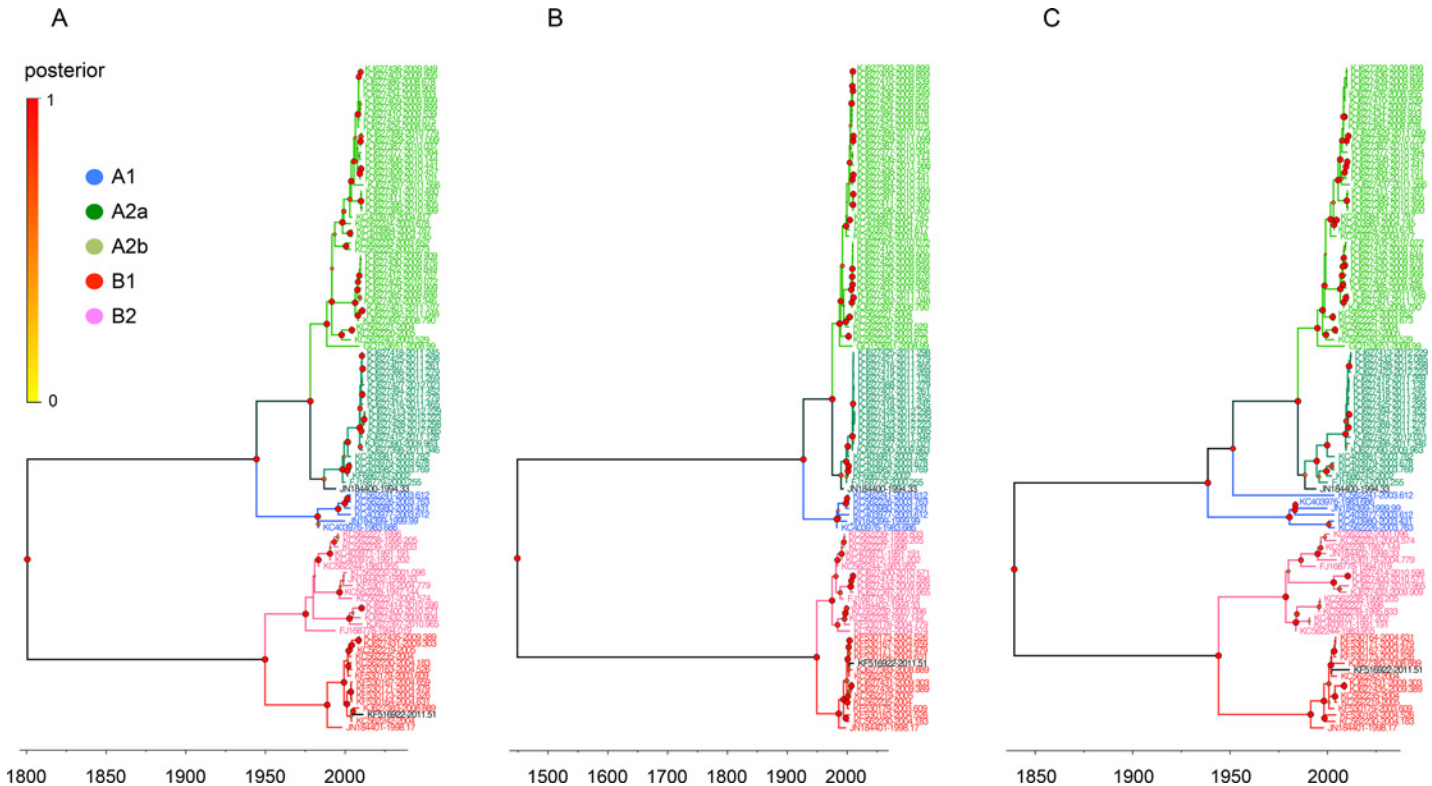
Fig 1. Phylogenetic relationships of HMPV N, P, and F gene sequences. The relative phylogenetic relationships of HMPV N (A), P (B), and F (C) gene sequences were defined in time-framed maximum clade credibility (MCC) trees. The five different colors represent each lineage (blue, A1; green, A2a; lime, A2b; red, B1; and pink, B2). The genes of JN184400 and KF516922 are colored deep green and black, respectively. The isolation time point of each sequence is expressed as the year fraction at the end of the sequence accession number. As the color of circles in the tree nodes, the size of circles in the node represents the posterior probability of their clustering (the bigger size demonstrates the higher probability).

detected no breakpoints of recombination. However, the SBP method located three potential breakpoints: 1,207 nucleotide residue in the F gene, 153 amino acid residue in the SH gene, and 104 amino acid residue in the G gene (Table 1). To confirm these, we divided each sequence set into pre- and post-breakpoint regions and checked the placements of the same viral sequences in the phylogenetic trees (Fig 4). For the F and G genes, we found that some strains showed different phylogenetic locations (S4 Table). The pre- and post-breakpoint sequences of the F gene of AUS03 virus was located in the A1 and A2b lineages (Fig 4A and 4B), respectively, while those of the G gene of HMPVgz01 (GenBank accession number GQ153651) were placed in the A2b and A2a, respectively. For the G gene of TN94 (GenBank accession number JN184400), we also observed its different locations in the pre- (parental of A2a and A2b sublineages) and post-breakpoint (A2a sublineage) trees (Fig 4E and 4F). A drastic change between the lineages was observed in the pre- and post-breakpoint trees of the SH gene. The pre-breakpoint region of the A1 SH sequences was placed in the same ancestral location in relation to those of the A2 lineage (Fig 4C), as seen in other HMPV trees (Figs 1–3). However, the post-breakpoint region of the A1 SH sequences appeared to have a more close evolutionary relationship to the A2b sequences by being located between the A2a and A2b sublineages (Fig 4D and S4 Table). When the detected recombination candidates were removed from the sequence sets, there was no more recombination breakpoint detected in the F gene, but the same residues were still detected in the SH and G genes which suggesting a possible remainder of intra-lineage recombination imprints (S5 Table). The complete genome sequences also resulted in a similar result
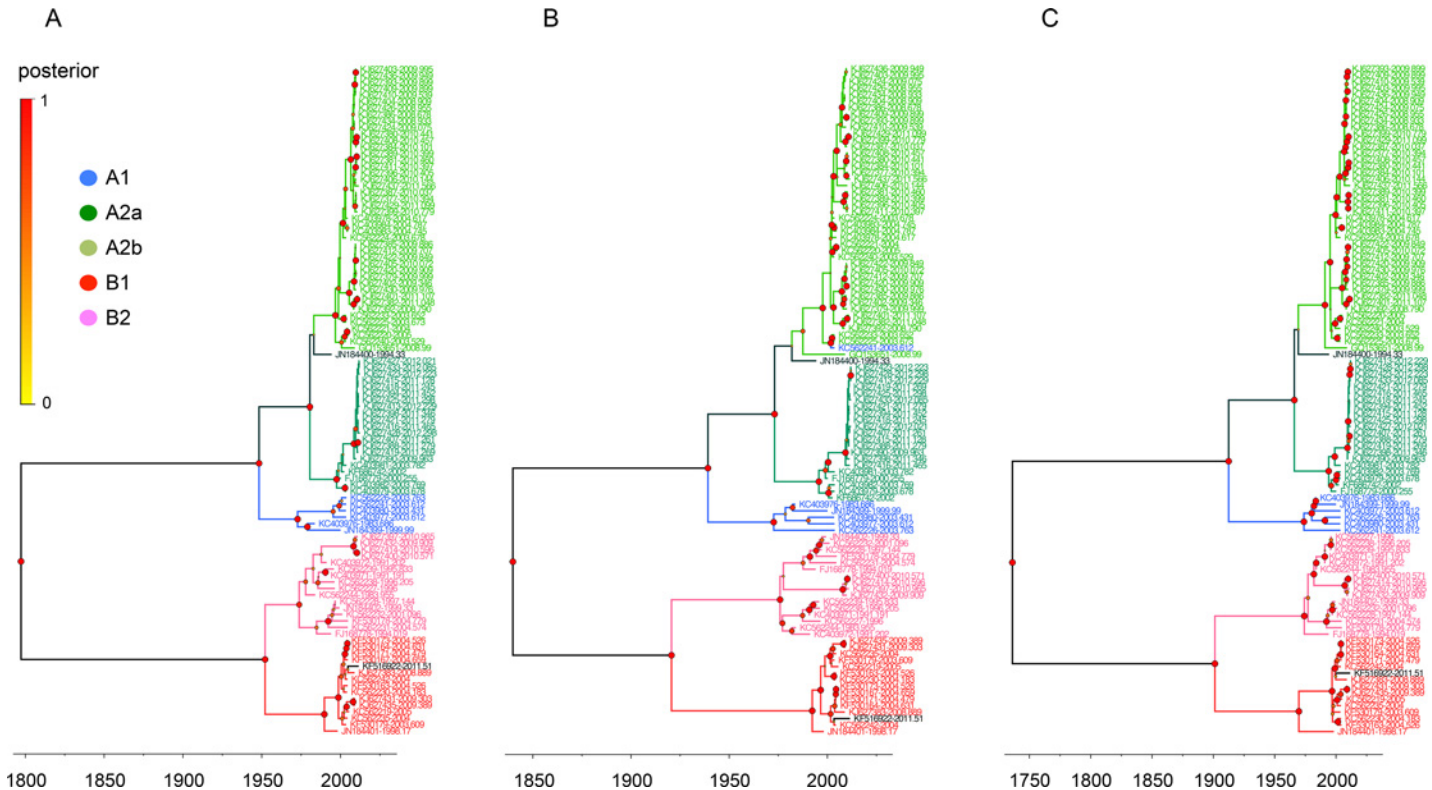
Fig 2. Phylogenetic relationships of HMPV SH, G, and L gene sequences. The relative phylogenetic relationships of HMPV SH (A), G (B), and L (C) gene sequences were defined in MCC trees. The five different colors represent each lineage (blue, A1; green, A2a; lime, A2b; red, B1; and pink, B2). The genes of JN184400 and KF516922 are colored deep green and black, respectively. Please see the Fig 1 legend for the information of isolation time points of sequences and posterior probability of tree branches.

doi:10.1371/journal.pone.0152962.g002

of recombination (Table 1 and S2 Fig). Considered together, these results indicate that genetic recombination within and between the HMPV lineages might have affected the evolutionary dynamics of HMPV genomes.

Selection profiles of HMPV genomes

Understanding the effects of selection pressures on the genetic diversity is of importance to interpret the evolutionary dynamics of HMPVs [54]. To address these, we analyzed overall dN/dS ratios of the complete and individual genes of HMPVs, and the SH and G genes exhibited higher dN/dS estimates than others. The three lowest dN/dS ratios were found in the N, M, and F genes. By the SLAC method, only the three codons (113, 127, and 139) in the G gene appeared to have been under positive selection whereas most others have been affected by negative selection to varying degrees (N, 13.28%; P, 13.56%; M, 11.41%; F, 12.43%; M2-1, 12.30%; M2-2, 2.82%; SH, 7.34%; G, 8.78%; and L, 9.61%) (Table 2). It is difficult to estimate positive selection using the sequence alignment of protein coding regions because most selection events might be episodic that happened only in some evolutionary lineages. To reduce this methodological limitation, the MEME method was suggested for the detection of both pervasive and episodic histories of positive selection [53], and it detected such sites in the N (1 site, 0.26% selection ratio), P (2, 0.68%), M (1, 0.39%), M2-1 (1, 0.53%), SH (3, 1.69%), G (5, 3.38%), and L (11, 0.54%) coding regions (Table 2 and S6 Table). However, the F and M2-2 coding regions exhibited no traces of positive selection, again. For the three surface glycoprotein genes (F, SH,
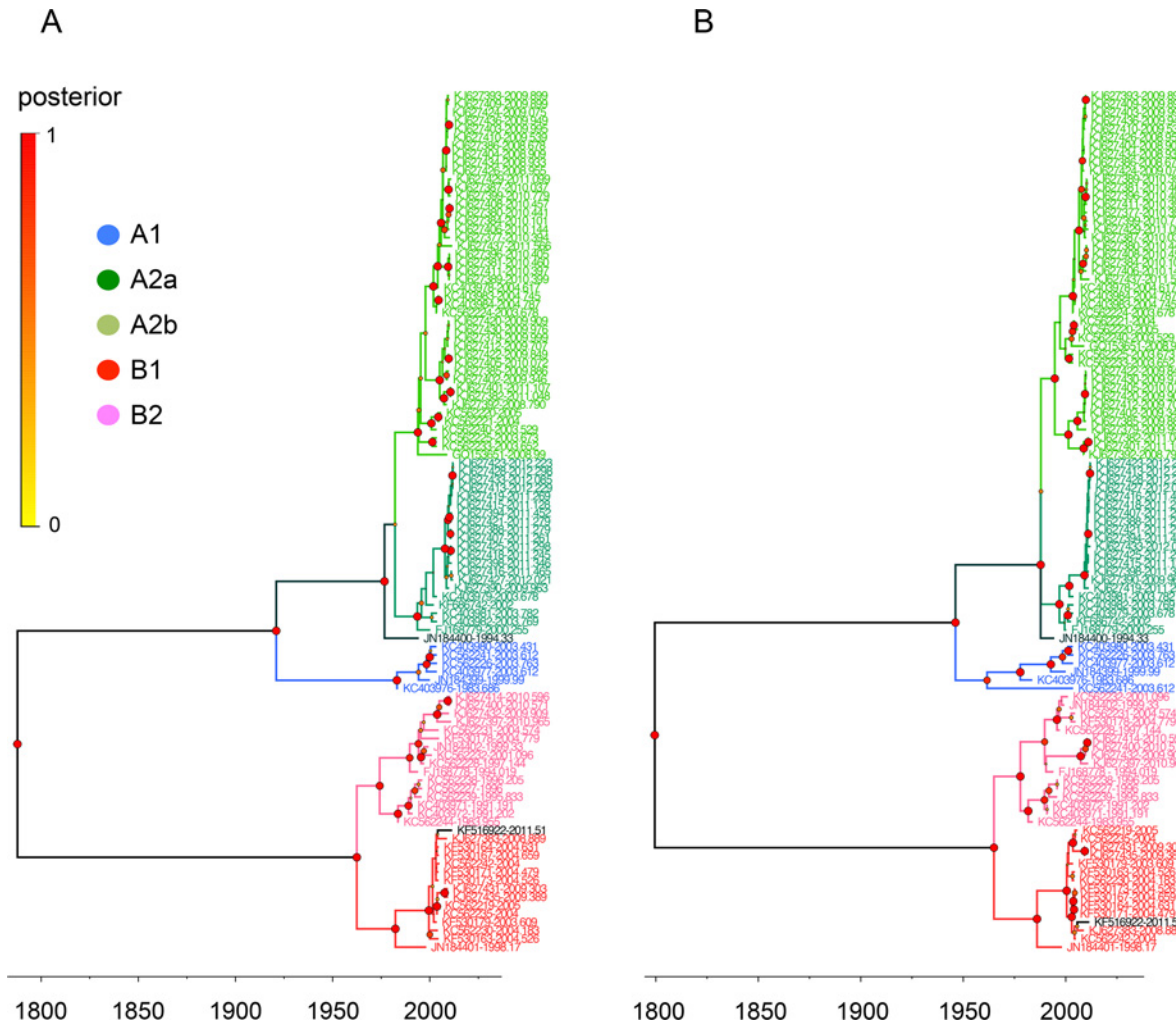
Fig 3. Phylogenetic relationships of HMPV M and M2 gene sequences. The relative phylogenetic relationships of HMPV M (A) and M2 (B) gene sequences were defined in MCC trees. The five different colors represent each lineage (blue, A1; green, A2a; lime, A2b; red, B1; and pink, B2). The genes of JN184400 and KF516922 are colored deep green and black, respectively. Please see the Fig 1 legend for the information of isolation time points of sequences and posterior probability of tree branches.

doi:10.1371/journal.pone.0152962.g003

and G), we further investigated the positive selection profiles of the five genetic lineages, and they mostly exhibited lineage-specific selection, compared with the results of overall gene sequences (S6 Table). The complete genome sequence set produced somewhat different

Table 1. Potential recombination breakpoints in the HMPV genomes.

Coding region	Potential breakpoint (BP)		Nucleotide region	
	Nucleotide	Amino acid	Pre-BP	Post-BP
Complete	4,308 (6773.75, 100%)[a]	n.d.[b]	1–4,305	4,309–12,180
F	1,207 (16.60, 99.99%)	n.d.	1–1,206	1,210–1,617
SH	n.d.	153 (2327.67, 100%)	1–456	460–531
G	n.d.	104 (4528.41, 100%)	1–309	313–444

[a] IC improvement score and % model-averaged support based on the Akaike information criterion (AIC).

[b] n.d., not detected.
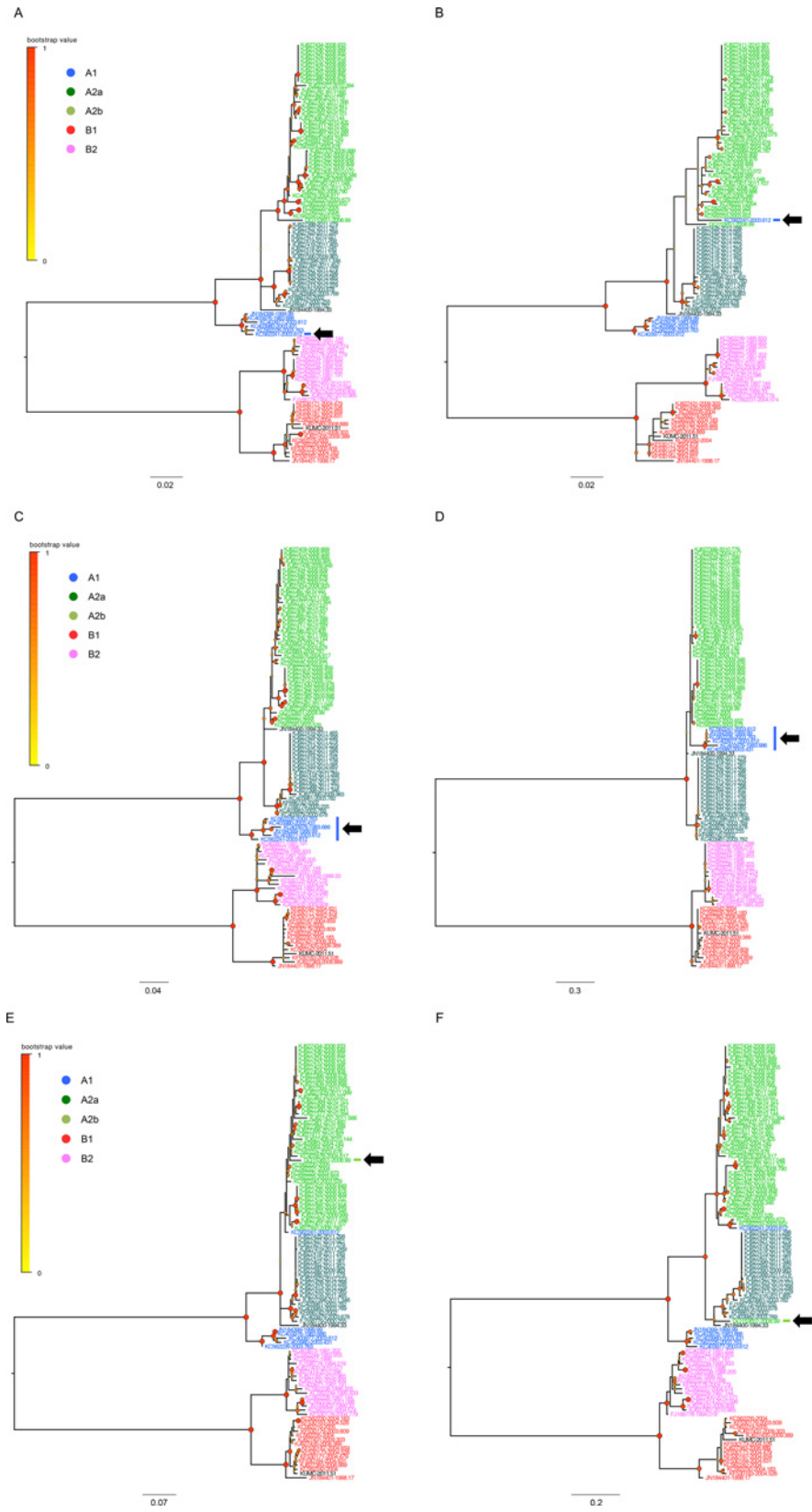
doi:10.1371/journal.pone.0152962.t001

Fig 4. Phylogenetic placements of the pre- and post-breakpoint region sets of F, SH, and G gene sequences. The phylogenetic placements of the same viral sequences were compared using the datasets of pre- (A, C, and E) and post-breakpoint (B, D, and F) regions divided by the detected recombination breakpoints in the F (A and B), SH (C and D), and G (E and F) genes. The trees were reconstructed by MEGA5.2 and edited in the FigTree (v1.4). The scales in the trees indicate the number of substitutions per site.

doi:10.1371/journal.pone.0152962.g004

positive selection codons, but for the G protein gene, it included all of the codons that were detected in the individual gene analysis (S6 and S7 Tables). These all suggest that not only negative selection but also pervasive and episodic events of positive selection interact to shape the genetic diversity of HMPVs.

Evolutionary profiles of HMPV genomes

The evolutionary rate of the complete genome of HMPVs was approximately 0.52×10^{-3} substitution/site/year (95% highest probability density (HPD) 0.27–0.76). The two membrane glycoprotein genes, SH and G, showed higher evolutionary rates (1.57 and 2.13×10^{-3} substitution/site/year, respectively) whereas those of the nucleocapsid and polymerase complex protein genes were found to be relatively low: N, 0.65×10^{-3} (0.36–0.92×10^{-3}); P, 0.85 (0.42–1.29); M, 0.76 (0.38–1.16); M2, 0.74 (0.42–1.06); and L, 0.43 (0.19–0.63) (Table 3). However, another membrane glycoprotein gene F exhibited only 0.60×10^{-3} (0.32–0.90) evolutionary rate, which was the second lowest among all of the tested genes. At the translated amino acid level, the F protein gene again showed the lowest evolutionary rate (non-synonymous substitution), along with other N, M, and L protein genes, at less than 1×10^{-5} substitution/site/year. The G protein gene exhibited the highest evolutionary rate of amino acid at 1.10×10^{-3} substitution/site/year, followed by SH, P, M2-2, and M2-1, with 0.61, 0.20, 0.19, and 0.11×10^{-3} substitution/site/year, respectively (Table 3).

We also estimated the circulation times of the common ancestors of the HMPV genes. The tMRCA of complete genome sequences of HMPVs was estimated as approximately 390.92 years of age, based on the most recent sequence age (2012.30 in year fraction), and that of the eight genes was also quite remote, ranging from 224.03 to 472.45 (N, 255.01; P, 288.46; M,

Table 2. Selection profiles estimated from the HMPV genomes.

Coding region	Codon #	Overall dN/dS	Number of selected codons		
			Positive		Negative
			SLAC	MEME	SLAC
Complete	4,060	0.098	6 (0.15)[a]	54 (1.33)	798 (19.66)
N	384	0.035	n.d.[b]	1 (0.26)	51 (13.28)
P	295	0.120	n.d.	2 (0.68)	40 (13.56)
M	254	0.026	n.d.	1 (0.39)	29 (11.41)
F	539	0.045	n.d.	n.d.	67 (12.43)
M2-1	187	0.073	n.d.	1 (0.53)	23 (12.30)
M2-2	71	0.160	n.d.	n.d.	2 (2.82)
SH	177	0.380	n.d.	3 (1.69)	13 (7.34)
G	148	0.517	3 (2.03)	5 (3.38)	13 (8.78)
L	2,005	0.053	n.d.	11 (0.54)	197 (9.61)

[a] The selection ratio (%) is provided in parenthesis.

[b] n.d., not detected.

doi:10.1371/journal.pone.0152962.t002

Table 3. Evolutionary rates estimated from the HMPV genomes.

Coding region	NT length analyzed	Evolutionary rate (10^{-3} substitutions/site/year)		
		Nucleotide	**Synonymous**	**Nonsynonymous**
Complete	12,133	0.52 (0.27–0.76) [a]	-	-
N	1,152	0.65 (0.36–0.92)	0.52	$< 10^{-5}$
P	885	0.85 (0.42–1.29)	0.50	0.20
M	762	0.76 (0.38–1.16)	0.55	$< 10^{-5}$
F	1,617	0.60 (0.32–0.90)	0.43	$< 10^{-5}$
M2	727	0.74 (0.42–1.06)	-	-
M2-1	564	-	0.62	0.11
M2-2	216	-	0.53	0.19
SH	531	1.57 (0.71–2.49)	0.57	0.61
G	444	2.13 (1.10–3.18)	0.72	1.10
L	6,015	0.43 (0.19–0.63)	0.32	$< 10^{-5}$

[a] Lower and upper limits of 95% HPD are provided in parenthesis.

doi:10.1371/journal.pone.0152962.t003

254.42; F, 289.30; M2, 224.03; SH, 255.70; G, 352.79; and L, 472.45) (S8 Table). However, the tMRCAs of each individual lineage were of recent origin (less than 50 years), except for the F, M2, and G genes in the A1 lineage, which were estimated to be 70.19 (95% HPD, 39.73–104.60), 54.56 (34.20–80.81), and 75.64 (38.83–121.64) years, respectively. Compared with other lineages, the A2b sequences appeared to have the most recent tMRCAs (15.88 to 31.45 years) (Fig 5). These suggest that the nucleocapsid and polymerase complex protein genes along with the surface glycoprotein F gene are less dynamic in molecular evolutionary rates
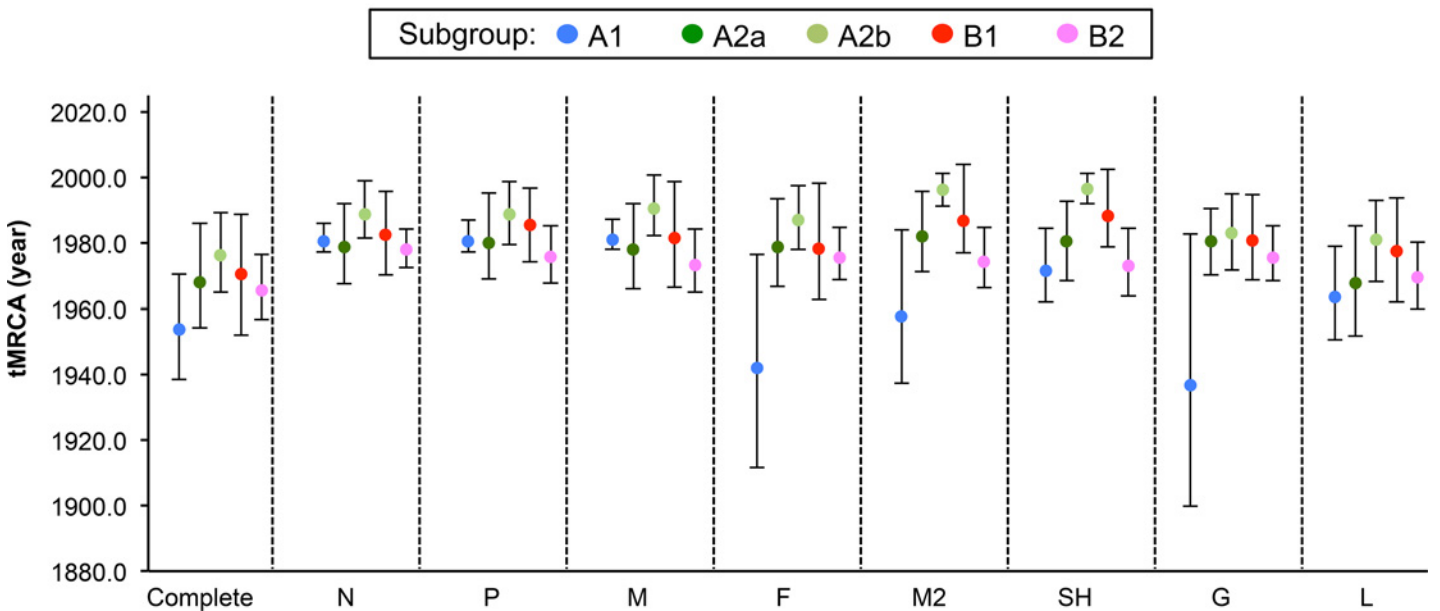

Fig 5. Time of the most recent common ancestor (tMRCA) of HMPV gene sequences. Based on a Bayesian analysis method implemented in BEAST v1.8.1, the estimated tMRCA of the eight HMPV genes is represented for each lineage. The lower and upper limits correspond to the 95% highest probability density (HPD).

doi:10.1371/journal.pone.0152962.g005

than two other surface glycoprotein SH and G genes and that the tMRCA estimates might be indicative of divergent genetic distances between the individual lineages of HMPVs.

Skyline-plot reconstruction and genetic diversity of HMPVs

The demographic legacy of HMPV genetic diversity was inferred using a Bayesian skyline-plot reconstruction method [25, 39]. It was revealed that the genetic diversity of the HMPV genes ranged from 4.38 to 215.79 (S7 Table). The A2a, A2b, B1, and B2 lineages of the N, P, M, and L genes showed the increases of genetic diversity between 2000 and 2005, whereas the A1 lineage of these genes displayed a similar peak in rather earlier times (approximately 1990–1995). Unlike the results of dN/dS ratio and evolutionary rates (Tables 2 and 3), the L gene showed the highest genetic diversity among all of the tested genes (Fig 6 and S9 Table). Among the three surface glycoprotein genes, the G gene exhibited the highest genetic diversity, and the A lineages of G gene exhibited more divergent demographic patterns than the B lineages whereas the two other glycoprotein genes, F and SH, showed similar patterns of diversity between the lineage lineages (Fig 6 and S9 Table). The genetic diversity of complete genome sequences appeared to be relatively higher, which might be mostly affected by that of the L gene due to its largest genome size among the eight genes of HMPVs (S9 Table).

Phylogenetic location and genetic similarity of KUMC-MP virus genes

A KUMC-MP virus (GenBank accession number KF516922) was isolated from a nasopharyngeal aspirate specimen collected from an infant (eight months old, male) who had recovered from a lower respiratory infection and meningitis. In the MCC trees, eight genes of the KUMC-MP virus were all classified with the other B1 lineage sequences according to their phylogenetic relationships (Figs 1–3). More than 97.7% genetic similarity was found with the viruses from Australia and the U.S. isolated around 2004. Against the A lineage sequences, however, the SH and G genes of the KUMC-MP virus exhibited only 46.8% and 17% similarity, respectively (SH, KJ627401, isolated in 2011; and G, KJ627399, isolated in 2010), and the six other genes of KUMC-MP showed 72.4–81.8% similarity (Table 4).

Discussion

Most previous studies evaluating the molecular evolution of HMPVs utilized a few genes instead of using a complete genome set. By great help of two recent genomic sequencing projects of HMPVs (BioProject PRJNA73051 and 237298), we could, for the first time, investigate the evolutionary footprints of the eight, complete genes of 103 HMPVs. However, most sequences were obtained from Peru (n = 57), Australia (n = 23), and USA (n = 18) regions (S2 Table). This might bias the results we observed even though the five genetic lineages of HMPVs were all present in the phylogenetic trees (Figs 1–3) and no certain lineages dominated in these geographical regions.

   Given the individual tMRCAs, each lineage of HMPVs appeared to be relatively recent origins (Fig 5 and S6 Table). However, when all of the lineage sequences were considered together, the overall tMRCAs extremely increased to the ranges of 224.03 to 472.45 years (S8 Table). Although a previous report also mentioned remote tMRCAs of some genes of HMPVs [25], the tMRCAs presented in this study provided comprehensive information of the eight genes with regard to the times of their undetected circulation and their relation to the inter-lineage diversity of HMPVs. In fact, the higher tMRCAs of HMPV genes were consistently matched with the higher mean values of genetic diversity, except for the N and SH genes (S8 and S9 Tables). These might suggest that the lineages of HMPVs experience ongoing evolution of genetic divergence and propose a question of what forces affect the evolutionary dynamics of
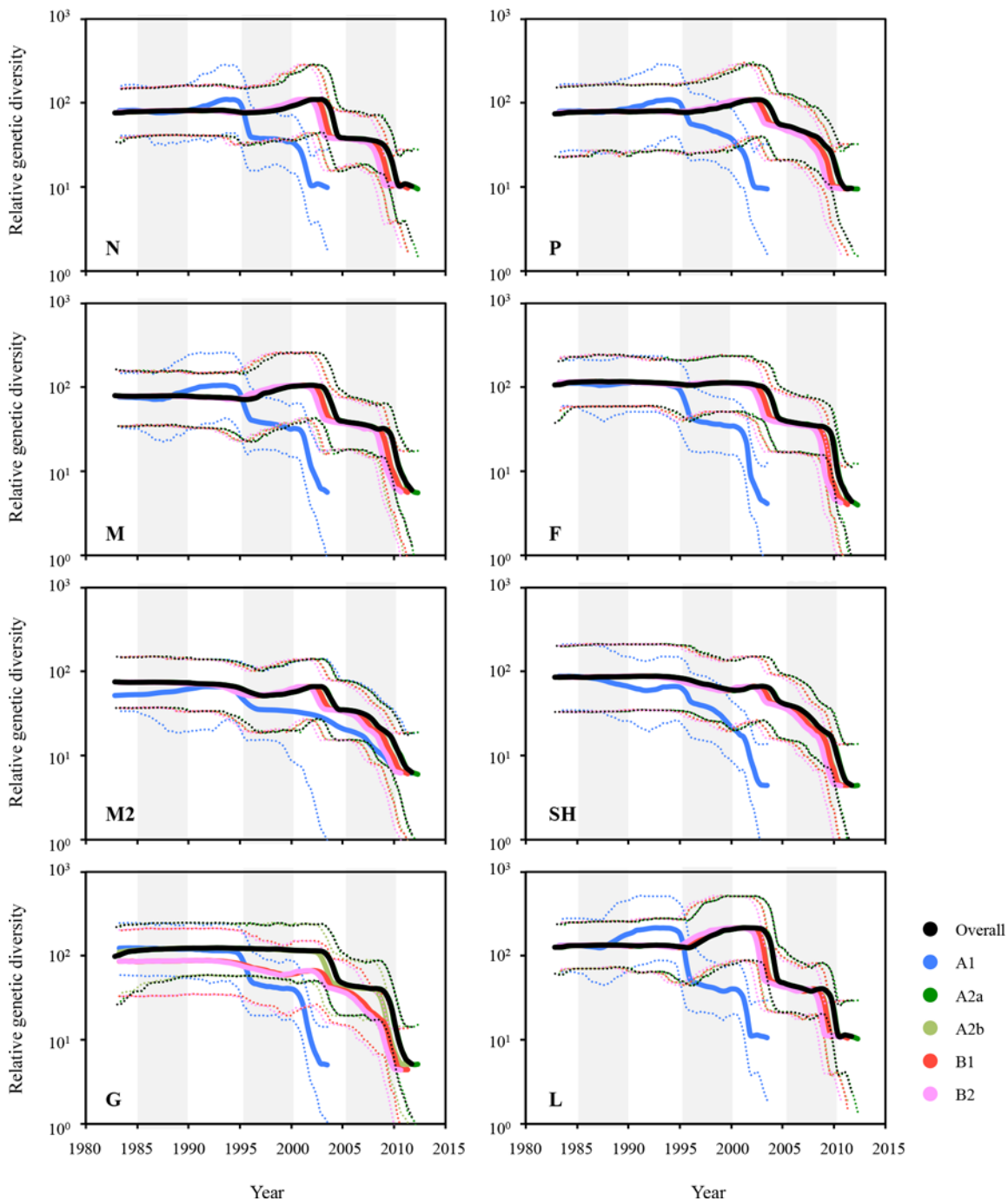
Fig 6. Population dynamics of HMPV sequences in skyline reconstruction plots. The demographic inferences of HMPV genetic diversity were depicted using a Bayesian skyline reconstruction plot method. The plots for the eight genes were then reconstituted into the five lineages. Solid lines represent mean diversity, and dotted lines represent the lower and upper limits of the 95% HPD.

HMPVs. We suspected genetic recombination first. In the MCC trees, we observed that the genes of TN94 and AUS03 viruses exhibited phylogenetic discordance (Figs 1–3). Given that HMPVs have a linear, single-stranded RNA genome, phylogenetic congruence is expected among the eight genes of the same virus. However, the TN94 virus harbored A2/A2a/A2b-lineage genes, and the AUS03 virus carried A1/A2/A2b-lineage genes. We speculated that past

Table 4. Sequence identity of KUMC-MP genes compared with other HMPV sequences.

Classification	% similarity to KUMC-MP gene							
	N	P	M	F	M2	SH	G	L
Overall	90.7	87.1	90.4	89.7	90.7	75.5	61.3	90
Maximum	99.3	98.5	99.9	99.3	98.7	97.9	97.7	99.1
	KC562242	KC562242	KF530164	KF530164	KC562242	KC562230	KF530164	KF530164
		KF530164	KF530167	KF530167	KF530163	KC562242		KF530171
		KF530171	KF530171	KF530171		KF530171		
			KF530173	KF530173				
	(B1/USA)	(B1/AUSa, USA)	(B1/AUS)	(B1/AUS)	(B1/AUS, USA)	(B1/AUS, USA)	(B1/AUS)	(B1/AUS)
Minimum	81.8	72.4	80.6	79.1	80.3	46.8	17	79.6
	KJ627405	JN184400	KC403981	KJ627381	KC403977	KJ627401	KJ627399	JN184399
				KJ627389				KC403980
				KJ627396				KC562226
				KJ627411				KC562241
				KJ627437				
	(A2b/Peru)	(A2a/USA)	(A2a/AUS)	(A2b/Peru)	(A1/AUS)	(A2b/Peru)	(A2b/Peru)	(A1/AUS, USA)

a AUS, Australia.

recombination events might have caused these results, as was suspected in a previous study [39]. To test this hypothesis, we examined the complete and individual sequence sets using the two different methods, SBP and GARD. When evaluating the nucleotide sequence sets by the SBP method, only the F gene appeared to retain one potential breakpoint of recombination. For the corresponding amino acids, the SH and G were also detected by the same method (Table 1). Different sensitivity to recombination analysis between nucleotide and amino acid sequences cannot be fully understood here, but it might be resulted from the various magnitudes of signal-to-noise ratios retained in our sequence sets because nucleotide sequence sets might have more obscure genetic relationships that should be taken into account for phylogenetic reconstruction, compared with amino acid sequence sets [55]. Genetic diversity retained in the sequence sets might be another reason for different sensitivity to recombination analysis. However, it might be still speculative because none was evidenced by the second method GARD (Table 1). These might be then spurious, so we tried to confirm the results by checking the lineage placements of the same viral gene sequences in the trees that were reconstructed using the sequence sets of pre- and post-breakpoint regions. One or two different viral sequences per gene exhibited a change of lineages in the phylogenetic trees of F and G genes (Fig 4A, 4B, 4E and 4F). For the SH gene, the pre- and post-breakpoint regions of the A1 lineage sequences all appeared to have different phylogenetic relationships to those of the A2a and A2b lineage sequences (Fig 4C and 4D). The different relationship of the same viral sequences in the phylogenetic trees indirectly indicated that previous recombination events were imprinted in the HMPV genomes and demonstrated that the inter-lineage interaction due to recombination might affect the genetic diversity of HMPVs. Also, the intra-lineage imprints of recombination could be another factor that worked on the HMPV genomes because the SH and G genes still retained the same recombination imprints even after the removal of the sequences of phylogenetic incongruences between the pre- and post-breakpoints trees (S5 Table). When analyzed with the complete genome sequences, a similar result was also obtained, and it was confirmed again in the different phylogenetic placements between the pre- and post-breakpoint sequence sets (S2 Fig and S4 Table). Especially for the KC562241 strain,

the two post-breakpoint sets of the complete genome sequences failed to demonstrate its potential as a recombinant strain (S2B Fig). However, it was confirmed when removing the L gene region from the post-breakpoint sequence set because of its largest gene size (S2C Fig). The possibility of false negative in the GARD should be also counted because the genetic diversity reflected in the aligned sequences could be below the limit required for the detection of recombination, or an unequal distribution of time intervals across the intra- and inter-lineages might have distorted the results. Artifacts caused by some technical errors, such as plasmid contamination during sequence reading procedures or co-infection possibility of two different HMPVs in a targeted clinical specimen, should be also considered.

Recombination is considered to contribute to the genetic diversity of viruses [56, 57]. Through this genomic interaction, a virus may evade host immunity, expand its host range, or generate a totally new virus strain [58]. Given that the F genes of HMPVs appear to be relatively conserved across different lineages [59] and that the *Pneumovirinae* share a common fusion mechanism [1], recombination among the F genes might contribute to their genomic preservation. As observed in an example of the removal of deleterious mutations [60], the HMPV F protein might maintain its fusion activity through recombination process. In this case, natural selection might be another evolutionary force acting on the bottleneck clones [61]. Negative selection has important implications for RNA-virus evolution because a large number of rare nonsynonymous mutations resulting from viral replication are often purified from the population [62]. It was previously provided that the HMPV genes were affected by negative selection pressure [32]. As presented in Table 2, we also observed that most of the genes were affected by negative selection. Along with purifying, negative selection pressure, episodic diversifying selection could be another evolutionary force working on the HMPV genomes. Even though only three sites in G protein genes were positively selected by the SLAC method, the MEME method located the imprints of both pervasive and episodic diversifying selections in the seven protein coding regions (Table 2), and this was demonstrated again in the results of the complete and separate lineage sequences (S6 and S7 Tables). Especially, the residue 146 in the G gene was suggested as one of the positive selection codons among the A2a sublineage sequences, and it was also detected in the results of the de Graaf et al. [25] and Papenburg et al. [18] studies. This might indicate the importance of this residue for the adaptive evolution of HMPVs. However, these natural selection profiles should be interpreted with care due to the suspected past recombination events. Or, these all may be tested using mutant viruses. By using the reverse genetically-rescued viruses, the implications of recombination and natural selection acting on some target residues can be further elucidated as demonstrated in other RNA viruses [63, 64].

The demographic distribution of the effective population size is indicative of genetic diversity [58]. As reported previously [25], most of the HMPV genes showed relatively low genetic diversity, except for the L gene (Fig 6 and S7 Table). The L gene sequences showed the lowest evolutionary rates in both the nucleotide and amino acid estimates (Table 3). In the selection profiles, the L gene also exhibited a low dN/dS ratio and 9.61% negative selection (Table 2). Considering that the L gene encodes a viral RNA polymerase and that a larger genome is usually associated with a lower mutation rate, the low rates of substitution and dN/dS values found for the HMPV L gene are understandable. However, the L gene showed the most increased diversity during 2000–2005, compared with the other genes (Fig 6 and S9 Table). This difference should be further investigated to examine the effects of L gene diversity on HMPV evolution. The G gene also exhibited interesting demographic pattern. As seen in Fig 6, the five lineages of HMPVs shared very similar demographic distributions even though the A1 lineage always showed its demography at an earlier time period. However, the A lineages of G gene exhibited a higher genetic diversity than the B lineages. This lineage-dependent genetic

diversity should be also addressed in the near future to better understand the molecular dynamics of HMPV evolution. In addition to the relatively conserved F gene, which considered the best candidate vaccine antigen against HMPV infection [59, 65], the genetically divergent characteristics of G genes across different genetic lineages may provide information that can be applied into the development of lineage-specific or multivalent HMPV vaccines and antivirals [11].

We isolated the KUMC-MP virus from a clinical sample collected in 2011. Based on the analysis of genetic similarity, we suspected this virus might be genetically related with the B1 lineage viruses circulating in Australia or the U.S. around 2004 (Table 4). Given the sustained circulation of HMPVs in Korea [66, 67], more genomic data are needed to investigate the molecular epidemiology and evolutionary dynamics of HMPVs in Korea.

In conclusion, we investigated the evolutionary dynamics of HMPVs by analyzing the complete genome sequences. Based on our results, we suggest that recombination is one of the evolutionary forces to have affected the genetic diversity within and between the five genetic lineages of HMPVs and that not only negative selection but also episodic positive selection works on shaping the evolutionary structure of HMPVs.

Supporting Information

S1 Data. Data set of 103 HMPV genome sequences used in this study.
(ZIP)

S1 Fig. Phylogenetic relationship of the complete genome sequences of HMPVs. The relative phylogenetic relationship of complete genome sequences of HMPVs was defined in time-framed maximum clade credibility (MCC) trees. The five different colors represent each lineage (blue, A1; green, A2a; lime, A2b; red, B1; and pink, B2). The genes of JN184400 and KF516922 are colored deep green and black, respectively. The isolation time point of each sequence is expressed as the year fraction at the end of the sequence accession number. As the color of circles in the tree nodes, the size of circles in the node represents the posterior probability of their clustering (the bigger size demonstrates the higher probability).
(TIF)

S2 Fig. Comparison of phylogenetic placements between the pre- and post-breakpoint region sets of the complete genome sequences of HMPVs. The phylogenetic placements of the same viral sequences were compared using the datasets of pre- (A, using 1–4,305 nucleotide region) and post-breakpoint (B, using 4,309–12,180 nucleotide region and C, using 4,309–6,165 nucleotide region) regions divided by the detected recombination breakpoint (4,305 in nucleotide) in Table 1. The trees were reconstructed by MEGA5.2 and edited in the FigTree (v1.4). The scale in the trees indicate the number of substitutions per site.
(TIF)

S1 Table. Primers used for reverse transcription-PCR and DNA sequencing.
(DOCX)

S2 Table. Subtype and geographic information of complete HMPV sequences.
(DOCX)

S3 Table. Information of lineage, strain name, isolation region and year of HMPVs.
(DOCX)

S4 Table. Putative recombination strains based on the phylogenetic placements between the pre- and post-breakpoint sequence sets in Fig 4.
(DOCX)

S5 Table. Recombination breakpoints detected after removing putative recombination strains.
(DOCX)

S6 Table. Positive selection codons estimated from the HMPV genomes.
(DOCX)

S7 Table. Positive selection codons estimated from the complete genome sequences of HMPVs.
(DOCX)

S8 Table. Estimated time of most recent common ancestors (tMRCA) of the HMPV genomes.
(DOCX)

S9 Table. Index of relative genetic diversity of the HMPV genomes.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: K-JS. Performed the experiments: EJK KMM. Analyzed the data: JIK SP IL KSP J-YB MS-P K-JS. Contributed reagents/materials/analysis tools: CKL. Wrote the paper: JIK KSP MS-P K-JS.

References

1. Lamb RA, Parks GD. Fields Virology, Chapter 33. *Paramyxoviridae*. 6th ed. Philiadelphia, USA: Lippincott Williams & Wilkins; 2013.

2. van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA, et al. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nature medicine. 2001; 7(6):719–24. doi: 10.1038/89098 PMID: 11385510.

3. van den Hoogen BG, van Doornum GJ, Fockens JC, Cornelissen JJ, Beyer WE, de Groot R, et al. Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients. The Journal of infectious diseases. 2003; 188(10):1571–7. doi: 10.1086/379200 PMID: 14624384.

4. Papenburg J, Boivin G. The distinguishing features of human metapneumovirus and respiratory syncytial virus. Reviews in medical virology. 2010; 20(4):245–60. doi: 10.1002/rmv.651 PMID: 20586081.

5. Schuster JE, Williams JV. Human metapneumovirus. Pediatrics in review / American Academy of Pediatrics. 2013; 34(12):558–65. doi: 10.1542/pir.34-12-558 PMID: 24295817.

6. Edwards KM, Zhu Y, Griffin MR, Weinberg GA, Hall CB, Szilagyi PG, et al. Burden of human metapneumovirus infection in young children. The New England journal of medicine. 2013; 368(7):633–43. doi: 10.1056/NEJMoa1204630 PMID: 23406028; PubMed Central PMCID: PMC3662802.

7. Jain S, Williams DJ, Arnold SR, Ampofo K, Bramley AM, Reed C, et al. Community-acquired pneumonia requiring hospitalization among U.S. children. The New England journal of medicine. 2015; 372 (9):835–45. doi: 10.1056/NEJMoa1405870 PMID: 25714161.

8. Mackay IM, Bialasiewicz S, Jacob KC, McQueen E, Arden KE, Nissen MD, et al. Genetic diversity of human metapneumovirus over 4 consecutive years in Australia. The Journal of infectious diseases. 2006; 193(12):1630–3. doi: 10.1086/504260 PMID: 16703505.

9. Bosis S, Esposito S, Niesters HG, Crovari P, Osterhaus AD, Principi N. Impact of human metapneumovirus in childhood: comparison with respiratory syncytial virus and influenza viruses. Journal of medical virology. 2005; 75(1):101–4. doi: 10.1002/jmv.20243 PMID: 15543589.

10. Zappa A, Canuti M, Frati E, Pariani E, Perin S, Ruzza ML, et al. Co-circulation of genetically distinct human metapneumovirus and human bocavirus strains in young children with respiratory tract infections in Italy. Journal of medical virology. 2011; 83(1):156–64. doi: 10.1002/jmv.21940 PMID: 21108354.

11. Ren J, Phan T, Bao X. Recent vaccine development for human metapneumovirus. The Journal of general virology. 2015. doi: 10.1099/vir.0.000083 PMID: 25667325.

12. Feuillet F, Lina B, Rosa-Calatrava M, Boivin G. Ten years of human metapneumovirus research. Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology. 2012; 53 (2):97–105. doi: 10.1016/j.jcv.2011.10.002 PMID: 22074934.

13. Huck B, Egger M, Bertz H, Peyerl-Hoffman G, Kern WV, Neumann-Haefelin D, et al. Human metapneumovirus infection in a hematopoietic stem cell transplant recipient with relapsed multiple myeloma and rapidly progressing lung cancer. Journal of clinical microbiology. 2006; 44(6):2300–3. doi: 10.1128/JCM.00152-06 PMID: 16757646; PubMed Central PMCID: PMC1489415.

14. Buchholz UJ, Collins PL, Fouchier RA, van den Hoogen BG, Williams JV. Naming convention for human metapneumovirus strains. ICTV 2012. Available: http://talk.ictvonline.org/files/ictv_documents/m/gen_info/4329.aspx. Accessed 25 Jan 2016.

15. Schildgen V, van den Hoogen B, Fouchier R, Tripp RA, Alvarez R, Manoha C, et al. Human Metapneumovirus: lessons learned over the first decade. Clin Microbiol Rev. 2011; 24(4):734–54. doi: 10.1128/CMR.00015-11 PMID: 21976607; PubMed Central PMCID: PMCPMC3194831.

16. Carneiro BM, Yokosawa J, Arbiza J, Costa LF, Mirazo S, Nepomuceno LL, et al. Detection of all four human metapneumovirus subtypes in nasopharyngeal specimens from children with respiratory disease in Uberlandia, Brazil. Journal of medical virology. 2009; 81(10):1814–8. doi: 10.1002/jmv.21555 PMID: 19697417.

17. Chung JY, Han TH, Kim SW, Hwang ES. Genotype variability of human metapneumovirus, South Korea. Journal of medical virology. 2008; 80(5):902–5. doi: 10.1002/jmv.21129 PMID: 18360903.

18. Papenburg J, Carbonneau J, Isabel S, Bergeron MG, Williams JV, De Serres G, et al. Genetic diversity and molecular evolution of the major human metapneumovirus surface glycoproteins over a decade. Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology. 2013; 58(3):541–7. doi: 10.1016/j.jcv.2013.08.029 PMID: 24041471.

19. Neemuchwala A, Duvvuri VR, Marchand-Austin A, Li A, Gubbay JB. Human metapneumovirus prevalence and molecular epidemiology in respiratory outbreaks in Ontario, Canada. Journal of medical virology. 2015; 87(2):269–74. doi: 10.1002/jmv.24024 PMID: 25091145.

20. Li J, Ren L, Guo L, Xiang Z, Paranhos-Baccala G, Vernet G, et al. Evolutionary dynamics analysis of human metapneumovirus subtype A2: genetic evidence for its dominant epidemic. PLoS One. 2012; 7 (3):e34544. doi: 10.1371/journal.pone.0034544 PMID: 22479641; PubMed Central PMCID: PMC3316673.

21. Velez Rueda AJ, Mistchenko AS, Viegas M. Phylogenetic and phylodynamic analyses of human metapneumovirus in Buenos Aires (Argentina) for a three-year period (2009–2011). PloS one. 2013; 8 (4):e63070. doi: 10.1371/journal.pone.0063070 PMID: 23646177; PubMed Central PMCID: PMC3639999.

22. van den Hoogen BG, Bestebroer TM, Osterhaus AD, Fouchier RA. Analysis of the genomic sequence of a human metapneumovirus. Virology. 2002; 295(1):119–32. doi: 10.1006/viro.2001.1355 PMID: 12033771.

23. Govindarajan D, Samal SK. Sequence analysis of the large polymerase (L) protein of the US strain of avian metapneumovirus indicates a close resemblance to that of the human metapneumovirus. Virus research. 2004; 105(1):59–66. doi: 10.1016/j.virusres.2004.04.014 PMID: 15325081.

24. Lwamba HC, Alvarez R, Wise MG, Yu Q, Halvorson D, Njenga MK, et al. Comparison of the full-length genome sequence of avian metapneumovirus subtype C with other paramyxoviruses. Virus research. 2005; 107(1):83–92. doi: 10.1016/j.virusres.2004.07.002 PMID: 15567037.

25. de Graaf M, Osterhaus AD, Fouchier RA, Holmes EC. Evolutionary dynamics of human and avian metapneumoviruses. The Journal of general virology. 2008; 89(Pt 12):2933–42. doi: 10.1099/vir.0.2008/006957-0 PMID: 19008378.

26. Kayali G, Ortiz EJ, Chorazy ML, Nagaraja KV, DeBeauchamp J, Webby RJ, et al. Serologic evidence of avian metapneumovirus infection among adults occupationally exposed to Turkeys. Vector borne and zoonotic diseases. 2011; 11(11):1453–8. doi: 10.1089/vbz.2011.0637 PMID: 21736487.

27. Slater OM, Terio KA, Zhang Y, Erdman DD, Schneider E, Kuypers JM, et al. Human metapneumovirus infection in chimpanzees, United States. Emerging infectious diseases. 2014; 20(12):2115–8. doi: 10.3201/eid2012.140408 PMID: 25417845; PubMed Central PMCID: PMC4257807.

28. Palacios G, Lowenstine LJ, Cranfield MR, Gilardi KV, Spelman L, Lukasik-Braum M, et al. Human metapneumovirus infection in wild mountain gorillas, Rwanda. Emerging infectious diseases. 2011; 17 (4):711–3. doi: 10.3201/eid1704.100883 PMID: 21470468; PubMed Central PMCID: PMC3377396.

29. Buitendijk H, Fagrouch Z, Niphuis H, Bogers WM, Warren KS, Verschoor EJ. Retrospective serology study of respiratory virus infections in captive great apes. Viruses. 2014; 6(3):1442–53. doi: 10.3390/v6031442 PMID: 24662675; PubMed Central PMCID: PMC3970160.

30. Escobar C, Luchsinger V, de Oliveira DB, Durigon E, Chnaiderman J, Avendano LF. Genetic variability of human metapneumovirus isolated from Chilean children, 2003–2004. Journal of medical virology. 2009; 81(2):340–4. doi: 10.1002/jmv.21399 PMID: 19107972.

31. Pizzorno A, Masner M, Medici C, Sarachaga MJ, Rubio I, Mirazo S, et al. Molecular detection and genetic variability of human metapneumovirus in Uruguay. Journal of medical virology. 2010; 82 (5):861–5. doi: 10.1002/jmv.21752 PMID: 20336730.

32. Lo Presti A, Cammarota R, Apostoli P, Cella E, Fiorentini S, Babakir-Mina M, et al. Genetic variability and circulation pattern of human metapneumovirus isolated in Italy over five epidemic seasons. The new microbiologica. 2011; 34(4):337–44. PMID: 22143806.

33. Heeney JL, Dalgleish AG, Weiss RA. Origins of HIV and the evolution of resistance to AIDS. Science. 2006; 313(5786):462–6. doi: 10.1126/science.1123016 PMID: 16873637.

34. Lukashev AN. Recombination among picornaviruses. Rev Med Virol. 2010; 20(5):327–37. doi: 10.1002/rmv.660 PMID: 20632373.

35. Gaunt E, Harvala H, Osterback R, Sreenu VB, Thomson E, Waris M, et al. Genetic characterization of human coxsackievirus A6 variants associated with atypical hand, foot and mouth disease: a potential role of recombination in emergence and pathogenicity. J Gen Virol. 2015; 96(Pt 5):1067–79. doi: 10.1099/vir.0.000062 PMID: 25614593; PubMed Central PMCID: PMC4631059.

36. Salemi M, Vandamme A-M, Lemey P. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 2nd ed. Cambridge, UK; New York: Cambridge University Press; 2009. xxvi, 723 p. p.

37. Domingo E, Escarmis C, Sevilla N, Baranowski E. Population dynamics in the evolution of RNA viruses. Advances in experimental medicine and biology. 1998; 440:721–7. PMID: 9782350.

38. Collins PL, Karron RA. Fields Virology, Chapter 38. *Respiratory Syncytial Virus and Metapneumovirus*. 6th ed. Philidelphia, USA: Lippincott Williams & Wilkins; 2013.

39. Padhi A, Poss M. Population dynamics and rates of molecular evolution of a recently emerged paramyxovirus, avian metapneumovirus subtype C. Journal of virology. 2009; 83(4):2015–9. doi: 10.1128/JVI.02047-08 PMID: 19052092; PubMed Central PMCID: PMC2643776.

40. Rappuoli R, Pizza M, Del Giudice G, De Gregorio E. Vaccines, new opportunities for a new society. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111 (34):12288–93. doi: 10.1073/pnas.1402981111 PMID: 25136130; PubMed Central PMCID: PMC4151714.

41. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. Mapping the antigenic and genetic evolution of influenza virus. Science. 2004; 305(5682):371–6. doi: 10.1126/science.1097211 PMID: 15218094.

42. Rhee SY, Blanco JL, Jordan MR, Taylor J, Lemey P, Varghese V, et al. Geographic and Temporal Trends in the Molecular Epidemiology and Genetic Mechanisms of Transmitted HIV-1 Drug Resistance: An Individual-Patient- and Sequence-Level Meta-Analysis. PLoS medicine. 2015; 12(4): e1001810. doi: 10.1371/journal.pmed.1001810 PMID: 25849352.

43. Vrancken B, Lemey P, Rambaut A, Bedford T, Longdon B, Gunthard HF, et al. Simultaneously estimating evolutionary history and repeated traits phylogenetic signal: applications to viral and host phenotypic evolution. Methods in ecology and evolution / British Ecological Society. 2015; 6(1):67–82. doi: 10.1111/2041-210X.12293 PMID: 25780554; PubMed Central PMCID: PMC4358766.

44. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution. 2011; 28(10):2731–9. doi: 10.1093/molbev/msr121 PMID: 21546353; PubMed Central PMCID: PMC3203626.

45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 2013; 30(4):772–80. doi: 10.1093/molbev/mst010 PMID: 23329690; PubMed Central PMCID: PMC3603318.

46. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012; 29(8):1969–73. doi: 10.1093/molbev/mss075 PMID: 22367748; PubMed Central PMCID: PMC3408070.

47. Rambaut A, Suchard MA, Drummond AJ. Tracer v1.6. Available: http://tree.bio..ed.ac.uk/software/tracer/ 2009.

48. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010; 26(19):2455–7. doi: 10.1093/bioinformatics/btq429 PMID: 20671151; PubMed Central PMCID: PMC2944195.

49. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. Bioinformatics. 2006; 22(24):3096–8. doi: 10.1093/bioinformatics/btl474 PMID: 17110367.

50. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21 (5):676–9. doi: 10.1093/bioinformatics/bti079 PMID: 15509596.

51. Nei M, Kumar S. Molecular Evolution and Phylogenetics. New York: Oxford University Press; 2000.

52. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution. 2005; 22(5):1208–22. doi: 10.1093/molbev/msi105 PMID: 15703242.

53. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS genetics. 2012; 8(7):e1002764. doi: 10.1371/journal.pgen.1002764 PMID: 22807683; PubMed Central PMCID: PMC3395634.

54. Pond SL, Poon AFY, Frost SD. Ch. 14: Estimating selection pressures on alignmebts of coding sequences. In: Lemey P, Salemi M, Vandamme AM, editors. The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing. 2nd ed. Cambridge, U.K.: Cambridge University Press; 2009.

55. Opperdoes FR. Chapter 9: Phylogenetic analysis using protein sequences. In: Lemey P, Salemi M, Vandamme AM, editors. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and hypothesis Testing. 2nd ed. Cambridge, U.K.: Cambridge University Press; 2009.

56. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nature reviews Microbiology. 2011; 9 (8):617–26. doi: 10.1038/nrmicro2614 PMID: 21725337; PubMed Central PMCID: PMC3324781.

57. Perez-Losada M, Arenas M, Galan JC, Palero F, Gonzalez-Candelas F. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2015; 30:296–307. doi: 10.1016/j.meegid.2014.12.022 PMID: 25541518.

58. Holmes EC. Fields Virology, Chapter 11. Virus Evolution. 6th ed. Philidelphia, USA: Lippincott Williams & Wilkins; 2013.

59. Melero JA, Mas V. The Pneumovirinae fusion (F) protein: A common target for vaccines and antivirals. Virus research. 2015. doi: 10.1016/j.virusres.2015.02.024 PMID: 25738581.

60. Keightley PD, Eyre-Walker A. Deleterious mutations and the evolution of sex. Science. 2000; 290 (5490):331–3. PMID: 11030650.

61. Duarte E, Clarke D, Moya A, Domingo E, Holland J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(13):6015–9. PMID: 1321432; PubMed Central PMCID: PMC402129.

62. Hughes AL, Hughes MA. More effective purifying selection on RNA viruses than in DNA viruses. Gene. 2007; 404(1–2):117–25. doi: 10.1016/j.gene.2007.09.013 PMID: 17928171; PubMed Central PMCID: PMC2756238.

63. Kim JI, Lee I, Park S, Lee S, Hwang MW, Bae JY, et al. Effects of a hemagglutinin D222G substitution on the pathogenicity of 2009 influenza A (H1N1) virus in mice. Archives of virology. 2014; 159 (10):2559–65. doi: 10.1007/s00705-014-2104-5 PMID: 24824345.

64. Meng J, Lee S, Hotard AL, Moore ML. Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. mBio. 2014; 5(5):e01704–14. doi: 10.1128/mBio.01704-14 PMID: 25249281; PubMed Central PMCID: PMC4173764.

65. Cox RG, Williams JV. Breaking in: human metapneumovirus fusion and entry. Viruses. 2013; 5(1):192–210. doi: 10.3390/v5010192 PMID: 23325326; PubMed Central PMCID: PMC3564117.

66. Kim HR, Cho AR, Lee MK, Yun SW, Kim TH. Genotype variability and clinical features of human metapneumovirus isolated from Korean children, 2007 to 2010. The Journal of molecular diagnostics: JMD. 2012; 14(1):61–4. doi: 10.1016/j.jmoldx.2011.09.004 PMID: 22056628.

67. Noh JY, Song JY, Cheong HJ, Choi WS, Lee J, Lee JS, et al. Laboratory surveillance of influenza-like illness in seven teaching hospitals, South Korea: 2011–2012 season. PLoS one. 2013; 8(5):e64295. doi: 10.1371/journal.pone.0064295 PMID: 23717587; PubMed Central PMCID: PMC3661466.