Virology Journal

Genetic diversity and evolution of human metapneumovirus fusion protein over twenty years

Chin-Fen Yang[†4], Chiaoyin K Wang[†4], Sharon J Tollefson[1], Rohith Piyaratna[1], Linda D Lintao[4], Marla Chu[4], Alexis Liem[4], Mary Mark[4], Richard R Spaete[4], James E Crowe Jr[1,2,3] and John V Williams*[1,2,3]

Address: [1]Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, USA, [2]Monroe Carell Jr Children's Hospital at Vanderbilt, Nashville, TN, USA, [3]Department of Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN, USA and [4]MedImmune Vaccines, Inc, Mountain View, CA, USA

Email: Chin-Fen Yang - yangc@medimmune.com; Chiaoyin K Wang - wangk@medimmune.com; Sharon J Tollefson - sharon.tollefson@vanderbilt.edu; Rohith Piyaratna - rohith.piyaratna@vanderbilt.edu; Linda D Lintao - lintaol@medimmune.com; Marla Chu - chum@medimmune.com; Alexis Liem - liema@medimmune.com; Mary Mark - markm@medimmune.com; Richard R Spaete - delta_gee@prodigy.net; James E Crowe - james.crowe@vanderbilt.edu; John V Williams* - john.williams@vanderbilt.edu

* Corresponding author   †Equal contributors

This article is available from: http://www.virologyj.com/content/6/1/138

Abstract

Background: Human metapneumovirus (HMPV) is an important cause of acute respiratory illness in children. We examined the diversity and molecular evolution of HMPV using 85 full-length F (fusion) gene sequences collected over a 20-year period.

Results: The F gene sequences fell into two major groups, each with two subgroups, which exhibited a mean of 96% identity by predicted amino acid sequences. Amino acid identity within and between subgroups was higher than nucleotide identity, suggesting structural or functional constraints on F protein diversity. There was minimal progressive drift over time, and the genetic lineages were stable over the 20-year period. Several canonical amino acid differences discriminated between major subgroups, and polymorphic variations tended to cluster in discrete regions. The estimated rate of mutation was 7.12×10^{-4} substitutions/site/year and the estimated time to most recent common HMPV ancestor was 97 years (95% likelihood range 66-194 years). Analysis suggested that HMPV diverged from avian metapneumovirus type C (AMPV-C) 269 years ago (95% likelihood range 106-382 years).

Conclusion: HMPV F protein remains conserved over decades. HMPV appears to have diverged from AMPV-C fairly recently.

Background

Human metapneumovirus (HMPV) is a recently described respiratory virus in the order *Mononegavirales*, family *Paramyxoviridae*, subfamily *Pneumovirinae*, genus *Metapneumovirus* [1]. HMPV is a leading cause of lower respiratory infection (LRI) in infants and children worldwide [2-13]. HMPV is also associated with severe disease in immunocompromised hosts or persons with underlying conditions [14-20]. Most reports of HMPV molecular epidemiology have included only a few seasons, and the

genetic variability of HMPV over decades has not been determined. Candidate vaccines for HMPV are under development [21-25], and the fusion (F) protein is the major antigenic determinant of protection [22,24,26-28] Therefore, it is critical to understand the potential for immune escape through virus evolution over time, and the likelihood that immunity against a particular F protein included in a vaccine candidate will be broadly protective.

The virus most closely related genetically to HMPV is avian metapneumovirus type C (AMPV-C) [1]. AMPV is an emerging pathogen of poultry that was identified in 1979. Subtypes AMPV-A and AMPV-B circulate in Europe and Africa, while AMPV-C was discovered in Minnesota and has been detected in the US and Korea [29,30]. Productive experimental infection of poultry with HMPV has not been successful, and serological studies have failed to detect evidence of human infection by AMPV [1]. Recent data suggest that F protein is responsible for this species restriction [31]. Thus, HMPV infection of humans may arise from a relatively recent trans-species transmission from AMPV-C.

We analyzed full-length F gene sequences from 68 isolates of HMPV collected over a 20-year period from otherwise healthy children with respiratory disease and 17 published full-length F gene sequences from other regions of the world. Our data show that HMPV F is highly conserved geographically over several decades. Distinct amino acid changes were present between different genetic lineages, but these amino acids were conserved within lineages. Variations that were present clustered in discrete regions, suggesting antigenic sites possibly driven by selective immune pressure. However, HMPV F gene sequences did not display progressive drift over time, unlike influenza viruses. The mutation rate of HMPV was similar to that of other RNA viruses, and the time to most recent common ancestor suggested recent divergence from AMPV-C.

Results

Comparison of sequence identity between subgroups

Full-length F gene sequences were obtained for 68 Tennessee strains of HMPV and assigned to one of the four proposed lineages (A1, A2, B1, or B2) based on phylogenetic analysis, discussed further below [32]. Of the 68 strains sequenced, 34 (50%) were of the B2 lineage, 18 (26%) A2, 7 (10%) B1 and 9 (13%) A1 lineage. Sequences obtained in this study were compared to 17 published full-length HMPV F gene sequences. The overall mean nucleotide identity between all 85 isolates was 89%, with a minimum identity of 83.7% (Table 1). The identity within major groups was higher, mean 96% (minimum 93.9%) between A1 and A2, and mean 97% (minimum 93.5%) between B1 and B2. The B2 lineage diverged more from the A lineages than the B1 lineage. B2 mean identity with A1 and A2 was 86.7% and 89.7%, respectively, while B1 identity with A1 and A2 was 91.3% and 94.7%, respectively. Mean nucleotide identity was >97% within all minor lineages, although the minimum identity for the B2 isolates was the lowest at 93.5%, showing more diversity within this lineage.

Amino acid identity was more conserved than nucleotide identity between and within all groups, with overall minimum identity of 93.7% and mean identity 96.3%. Amino acid identity within major groups was 98.7% for A1 and A2, and 99.3% for B1 and B2. The minimum amino acid identity between all lineages was approximately 94%; the greater divergence of the B2 lineage at the nucleotide level was not represented in the amino acid sequence.

Table 1: Comparison of nucleotide and amino acid identity of full-length human metapneumovirus F genes within or between subgroups.

Group	Number of sequences	Minimum % nt identity	Mean % nt identity	Minimum % aa identity	Mean % aa identity
A1	13	97.5	98.2	99.3	99.6
A2	23	97.2	98.7	98.9	99.6
All A1+A2	36	93.9	96	98	98.7
B1	11	97.6	98.5	98.7	99.3
B2	38	93.5	97.5	99.4	99.9
All B1+B2	49	93.5	97	98.7	99.3
All A1+B1	24	84	91.3	93.7	97
All A1+B2	51	83.7	86.7	94.2	95.7
All A2+B1	34	84	94.7	93.9	98.1
All A2+B2	61	84.1	89.7	94.6	96.7
All	85	83.7	89	93.7	96.3

nt = nucleotide; aa = amino acid.

Distinct and conserved amino acid changes between lineages

There were a number of amino acid residues distinct to each group or subgroup (Table 2). The greatest number of divergent and subgroup-specific residues was identified in the F1 domain, between the two heptad repeat (HR) regions. At several positions all subgroups had either arginine or lysine but maintained a basic residue: 82, 348, 450, 479 and 518; only position 82 has been shown to be cleaved during infection [33,34]. Many subgroup-specific residues were similar biochemically between groups. Some variations, however, were unexpected, such as the presence of a proline at position 404 only in B subgroup viruses. Fourteen cysteine residues were conserved among all isolates except one Japanese sequence (JPS03.178) with a reported C292W variation [35]. Three potential N-glycosylation sites were conserved in all sequences: N58, N172 and N350 (Figure 1).

There were a number of single amino acid variations present in only one or a few sequences; these amino acids are listed in Table 3 and shown graphically in Figure 1. Many of these variant amino acids were biochemically quite dissimilar, though the biological significance of this finding is not clear. Interestingly, the highest variability was in the region between amino acids 260 to 300, analogous to the major antibody antigenic site A of the related RSV F protein [36] (Figure 1). Some of the variations in this region, such as E294G, were present in viruses of both the A1 and A2 subgroups. Viruses of the A2 lineage had the greatest number of such variations in the region between amino acids 230 to 300, but not elsewhere in the protein.

Phylogenetic diversity and evolution over time

We performed phylogenetic and evolutionary analysis of the aligned full-length F sequences with six different models using the BEAST program suite [37]. The phylogenetic

Table 2: Comparison of distinct amino acid variations in the indicated functional domains of F protein between groups or subgroups of unique human metapneumovirus strains.

Functional domain	AA residues in domain	No. of AA	AA Position	A (n = 36)	B (n = 49)	A1 (n = 13)	A2 (n = 23)	B1 (n = 11)	B2 (n = 38)
Signal peptide	1-22	2	6	V	M/V	V	V	M/V	M/V
			9	F	I				
F2 subunit	23-102	2	61	A/S	T/(S*)	A	A/S	T	T/(S*)
			82			R	K	K	K
Fusion peptide	103-125	1	122	V	I				
Heptad repeat A	131-172	4	135	T	N				
			139	N	G				
			143			K	K/T	Q/(K*)	K/T
			167	D	E				
F1 subunit	173-453	10	175	R	S				
			185	A/D	A	A	D	A	A
			233	N	Y				
			286	V	I				
			296	K/R	N/D	K	K/R	N/(D*)	D
			312	Q	K				
			348	K	R				
			404	N/S	P	N	N/S	P	P
			449	V/I	I	V	V/I	I	I
			450	K	R/K				
Heptad repeat B	454-486	3	466	S/N	S	S	N/S	S	S
			479	R	K/(R*)	R	R	K/(R*)	K
			482	S/(N*)	N/(S*)	S	S/(N*)	N/(S*)	N
Transmembrane	490-514	6	498	I	I/V	I	I	V	I
			503	S	L				
			504	T/S	T/A	T	S	T	T/A
			507	L	S				
			510	V/I	I	V	V/I	I	I
			511	F	I				
Cytoplasmic tail	515-539	4	518	K/(R*)	R/(K*)	K	K/(R*)	R/(K*)	R
			528	S	N				
			533	N	G				
			539			N/(S*)	S/N	S	S

Amino acids (AA) in bold type did not vary within the group or subgroup.
*Amino acids found in only one isolate within the group or subgroup.
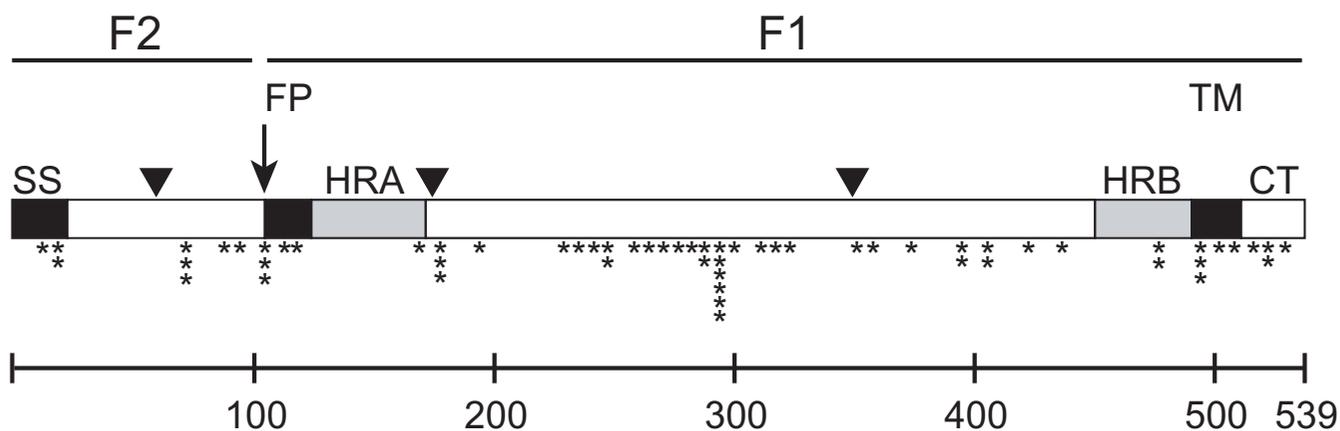
Figure 1
Schematic representation of putative structure and mutation map of human metapneumovirus F protein. SS = signal sequence; FP = fusion peptide; HRA = heptad repeat A; HRB = heptad repeat B; TM = transmembrane domain; and CT = cytoplasmic tail. Arrow indicates cleavage site; arrowheads indicate putative N-glycosylation sites. Amino acid variations are indicated by asterisks, with the number of asterisks representing the number of distinct strains in which the variation was found.

Table 3: Distinct amino acid variations detected in the indicated domains of F protein in human metapneumovirus strains.

Domain	AA positions of domain	Subgroup			
		A1	A2	B1	B2
Signal peptide	1-22	K20Q		S21N	S21N
F2 subunit	23-102		D72E (3)		
		E93V	E96G		
Cleavage site	102-103	S101P (2)*		S101P	
Fusion peptide	103-125		T114A		A115T
Heptad repeat A	131-172		K172R		
F1 subunit	173-453		K179Q	R179K (2)	
			S232P	F196Y	
		I248F	G239E		
		L249P (2)	G261E		
			M270T		
			V271I	D280G	
			C292G	I285T	
			C292W		
		E294G (4)	E294G (2)		
			K296R (3)		
			N298S		
			Y310N		
			A314T		E323K
			I352V		N358K
			H368N	R396W (2)	
			N404S (2)		
			T419I	K438R	
Heptad repeat B	454-486			D475E (2)	
Transmembrane	490-514	I492T	I492V	I492V	
		I514T	L507P		
Cytoplasmic tail	515-539	K519R	P520Q	P525L	P520T

* Numbers in parentheses indicate the number of distinct strains with the variation.

tree representing the sequence relationships by nucleotide substitutions identified four genetic subgroups (Figure 2), consistent with previous analyses [32]. The four distinct subgroups remained stable over time, and viruses within these lineages were closely related genetically, despite being isolated at time points separated by as many as twenty years. Thus, the clustering did not correlate closely with chronological origin of the sequences. For example, one subcluster within the B2 lineage contained nearly identical sequences from Tennessee in 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1998, 1999 and 2001, as well as Netherlands in 1994 and Canada in 1998 and 2000 (Figure 2). Similar clustering of chronologically and geographically disparate sequences was present within each subgroup. In the A1 subgroup, Tennessee sequences from 1994, 1996, and 2003 were closely related to Canadian sequences from 1999 and 2000 and a Japanese sequence from 2003. To examine further the evolution of HMPV F gene sequences over time, we aligned sequences within each subgroup in chronological order (see Additional files 1, 2, 3 and 4). A few nucleotide changes persisted in later chronological viruses and thus represented progressive evolution at those sites. However, the majority of the nucleotide changes from year to year were not preserved and often reverted in subsequent isolates, showing a lack of major drift over time.

Analysis of multiple sequences collected over time allowed a molecular clock calculation of viral nucleotide changes. The mutation rate of HMPV F was 7.12×10^{-4} substitutions/site/year (95% HPD 4.23×10^{-4}, 1.01×10^{-3}). The estimated time to most recent common ancestor (tMRCA) of all HMPV strains was 97 years (95% HPD 66-194) (Figure 2). The estimated time of divergence of the A subgroup into A1 and A2 was 51 years (95% HPD 38-92) and between B1 and B2 subgroups 40 years (95% HPD 38-97). Similar analysis using the limited number of available AMPV full-length F sequences (n = 24, including 16 AMPV-C F sequences collected between 1998-2007) suggested a tMRCA between AMPV-C and HMPV of 269 years (95% HPD 106-382) (Figure 2). However, very few full-length AMPV type C F sequences were available, and most were obtained within the last few years. The effect of these limitations is reflected in the wide 95% HPD intervals and thus the estimates for divergence of HMPV from AMPV-C must be considered with some caution.

Discussion
We analyzed 85 full-length HMPV F gene sequences obtained over a twenty-year period from Tennessee, Canada, Japan, and the Netherlands. Our data confirm that there are four distinct genetic lineages of HMPV, provisionally designated as A1, A2, B1 and B2 [32]. These data further show that these genetic subgroups are stable over time in circulating viruses in a population of children with

respiratory illnesses. Thus, HMPV does not appear to exhibit progressive genetic evolution, unlike influenza virus that exhibits rapid genetic drift associated with antigenic variation resulting in immune escape. In this respect, HMPV appears to be similar to other paramyxoviruses. RNA viruses mutate frequently due to the infidelity and lack of proofreading ability of RNA-dependent RNA polymerases [38]. Our data confirm that the HMPV polymerase also allows frequent errors resulting in the circulation of field strains with nucleotide variations at a similarly high rate. The rate of mutations we identified in HMPV F (7.12×10^{-4} substitutions/site/year) was intermediate between the lower rate of measles virus H gene mutation (9×10^{-5} substitutions/site/year) and the higher rate of influenza A virus HA (1.8×10^{-3} substitutions/site/year)[39]. Nonetheless, while paramyxoviruses including RSV and measles exhibit mutations and genotype variation over time [40,41], these nucleotide mutations do not result in progressive antigenic "drift" over time with loss of neutralizing epitopes [36,42,43]. This finding is in contrast to the data from studies of the influenza virus hemagglutinin protein, which progressively evolves both genetically and antigenically, necessitating annual vaccine updates [44-47]. The reason for the lack of directional antigenic drift in paramyxoviruses is not clear. There could be functional constraints on paramyxovirus fusion proteins to prevent such drastic amino acid changes. The fact that nucleotide diversity is greater than amino acid diversity among HMPV F sequences supports this hypothesis. In contrast to paramyxoviruses, the analogous influenza virus hemagglutinin and human immunodeficiency virus gp120 fusion proteins are capable of substantial mutation to escape neutralizing antibodies without loss of function. Alternatively, the nature of immune pressure on fusion protein sequences by human antibodies could differ between paramyxoviruses and orthomyxoviruses. Experimental live wild-type virus challenge of previously infected adults with a single lot of virus can achieve productive infection in a repetitive fashion within months of previous infection with the same virus [48]. The mechanism of the functional constraints on paramyxovirus fusion protein diversity warrants further investigation.

The finding that HMPV F gene sequences do not evolve rapidly in a progressive fashion is important for the development of monoclonal antibodies (mAbs) and vaccines. The HMPV F protein is the major determinant of protection in animal models [21,22,24,26,28]. Studies with a limited number of virus strains in these models suggest a degree of cross-protective efficacy mediated by prior infection with viruses of differing subgroups [26,49]. The high degree of conservation of F protein over time suggests that interventions such as mAbs or vaccines likely will not need to be continuously updated.
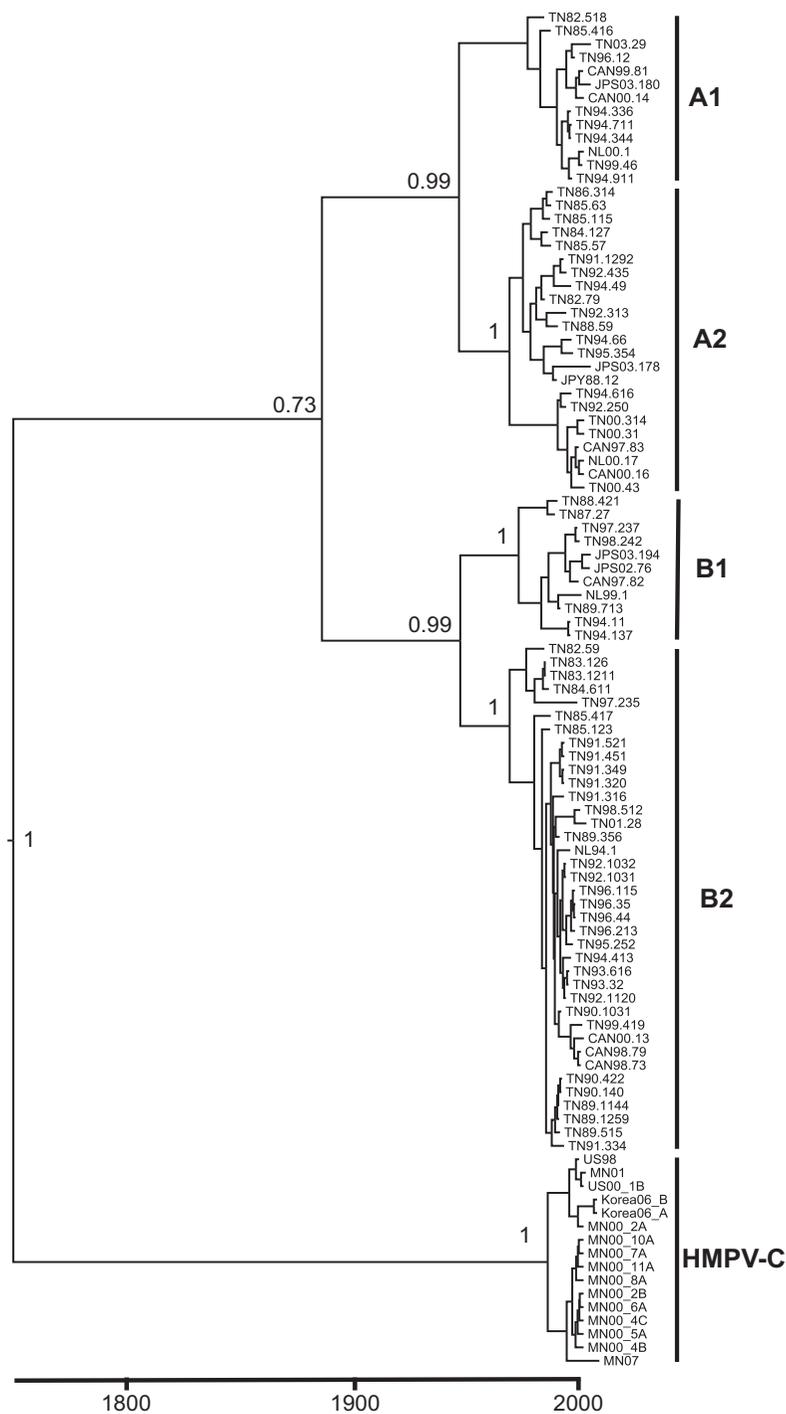
Figure 2
Maximum clade credibility tree of HMPV and AMPV F nucleotide diversity by tMRCA. Phylogenetic analysis of 85 full-length HMPV F nucleotide sequences from Canada (CAN), Japan (JPS or JPY), Tennessee (TN), or the Netherlands (NL) and 16 AMPV F sequences. The first two digits of the HMPV sequence names indicate the year of the isolate. The names of the AMPV sequences indicate geographic origin (US = United States; UK = United Kingdom; MN = Minnesota) and year. The posterior probability of divergence is indicated at each node. Mean TMRCA nodes on the MCC tree differ slightly from those reported in text, although all are contained with the same 95% HPD values. Scale bar represents time in years. Tree was constructed as described in Methods.

We identified a number of group and subgroup-specific amino acid residues, some in putative functional domains. The biological importance of these variations is not clear, since definitive evidence of pathogenic differences between HMPV strains has not been described. A previous analysis of 84 partial HMPV F sequences did not identify subgroup-specific amino acid differences between A1 and A2 viruses; however, a 441-nt gene segment was analyzed and most of the viruses were of recent derivation [32]. The subgroup-specific amino acid changes in F genes also were conserved over time, raising the question of whether these residues possess critical biological features for virus infection or transmission. Some of these variant amino acids were found in regions presumed to be essential, such as the heptad repeat (HR) regions. Synthetic HR peptides mediate potent *in vitro* inhibition of HMPV infection [50,51] and the HR are predicted to form a six-membered helical bundle [50], suggesting that HMPV F is a Class I viral fusion protein. We have cultivated multiple strains of all four subgroups that exhibit similar growth kinetics and syncytial formation *in vitro*, and similar levels of replication *in vivo* in rodents (data not shown), suggesting that the fusion function of all these strains is intact despite amino acid variations.

The HMPV F protein, like other Class I viral fusion proteins, requires cleavage for activation and most strains require exogenous trypsin for in vitro growth. Schickli et al described a cleavage site mutation S101P that arose in two strains of HMPV during cell passage and was associated with trypsin-independent viral growth *in vitro* [34]. The variant viruses did not differ from wild type in replication in Syrian hamsters [34]. We identified an S101P variation in three distinct viruses in this study from 1989, 1994, and 1999. The F gene sequences in the current study were amplified directly from specimens collected from children with URI, and thus these viruses are natural variants. One of these had an associated E93V variation that also was observed by Schickli et al. None of these three viruses in our study was associated with more severe clinical disease (data not shown).

Human and rodent F-specific mAbs have been described with neutralizing activity *in vitro* and protective effects *in vivo*, and several overlapping antigenic sites have been identified using these mAbs [52,53]. However, the precise location of these epitopes on the protein has not been defined. We find it intriguing that the greatest concentration of amino acid variations among these 85 field isolates lies in a region found between residues 260 to 300, which is roughly analogous to the major antibody antigenic site A in the human RSV F protein [36,42]. The precise definition of neutralizing epitopes, especially conserved epitopes for broadly neutralizing antibodies, is critical for the development of prophylactic mAbs.

Phylogenetic and evolutionary analysis of multiple full-length HMPV and AMPV F sequences obtained over twenty years showed that HMPV may have diverged from AMPV-C nearly 300 years ago, and the divergence of the four HMPV genotypes likely occurred within the last hundred years. de Graaf et al recently reported estimated tMRCA values of ~120 years for the four HMPV genotypes and 200 years for HMPV divergence from AMPV-C [54]. These estimates were based on analysis of 76 HMPV G sequences, 107 partial HMPV F sequences, 12 partial AMPV-C F sequences, 21 HMPV N sequences, and 15 AMPV-C N sequences from isolates collected over approximately 12 years. Thus, the number of genes included was greater, but the spread in years was less and most sequences were from recent isolates. Despite these differences, we estimated remarkably similar rates of divergence for both major and minor subgroups. Padhi et al analyzed published HMPV G sequences and estimated a tMRCA of only 25-50 years; however, the majority of viruses in that study were isolated between 2001 and 2003 [55]. Analysis of complete genome sequences from HMPV strains obtained over many years would provide the most robust estimates of genetic diversity and evolution.

Our phylogenetic and evolutionary analysis suggest that HMPV may have diverged fairly recently from AMPV, although the power of this analysis was limited by the small number of available AMPV F gene sequences. Successful productive infection of chickens and turkeys with HMPV has not been reported [1], although inflammation, HMPV RNA and antigen could be detected in turkey poults inoculated with a large inoculum of HMPV [56]. HMPV and AMPV contain analogous open reading frames in the same order that are distinct from those of the *Pneumovirus* genus, and metapneumoviruses lack the NS1 and NS2 genes of pneumoviruses [57]. This finding suggests that HMPV diverged from AMPV-C. Other viruses including influenza and HIV are thought to have originated in animal reservoirs but are now established primary human pathogens; HMPV may have arisen as a human pathogen by similar zoonotic transfer.

Methods

HMPV isolates

Virus sequences were derived from specimens collected over a twenty-year period from 1982-2002 in the Vanderbilt Vaccine Clinic, as previously described [2,3]. Nasal wash specimens were collected from children <5 years of age with acute respiratory tract illness. We extracted RNA from these samples and used quantitative real-time RT-PCR to test for HMPV by detection of nucleoprotein gene sequences [2]. Specimens that tested positive for HMPV were subjected to nested RT-PCR for the F gene as described below. Viral nomenclature used in this study uses a letter code representing the geographic site of isola-

tion (*e.g.*, "TN" represents Tennessee) followed by the year of isolation, month in which the virus was isolated and isolate number.

RNA extraction, RT-PCR and sequencing of F genes

RNA was extracted from 220 µl of nasal wash sample on a Qiagen BioRobot 9604 Workstation using the QIAamp Viral RNA kit (Qiagen), as described [2]. Amplification of the entire F open reading frame (ORF) was carried out by RT-PCR followed by nested PCR. The primers used to amplify the F regions were FF1 (5'-ATGTCTGTACTTC-CCAAA-3') and FR (5'-CCCGYACTTCATATTTGCA-3') for RT-PCR, and FF2 (5'-AATATGCAAGACTTGGAGCC-3' and 5'-AGGATCTGCAAGAGCTGGAG-3') and FR (5'-CCCGYACTTCATATTTGCA-3') for nested PCR. The Thermoscript/Platinum Taq Polymerase Kit (Invitrogen) was used in a 50 µL RT-PCR reaction with 10 µL of diluted RNA as template. The RT-PCR was carried out at 50°C for 50 min and 95°C for 3 min, followed by 5 cycles of 94°C for 30 sec, 50°C for 1 min, and 68°C for 3 min, and additional 30 cycles of 94°C for 30 sec, 55°C for 1 min, and 68°C for 3 min. For nested PCR, 2 µL of RT-PCR product was added to a 50 µL reaction using Platinum PCR Supermix (Invitrogen). The reaction was incubated at 95°C for 3 min followed by 5 cycles of 94°C for 30 sec, 50°C for 30 sec, and 68°C for 2 min, and additional 30 cycles of 94°C for 30 sec, 55°C for 30 sec, and 68°C for 2 min. For all reactions a final extension at 68°C for 7 min was included. The resulting products were about 1.9 kb for the F ORF and flanking sequences. The majority of PCR products generated after RT-PCR and nested PCR were specific and migrated as a single band of the expected size (data not shown). Agarose gel purification of the desired PCR products was performed when multiple products were generated.

Sequencing reactions were carried out using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Eight sequencing primers were used for each fragment to ensure a two-fold coverage of the open reading frame. Sequencing primers are available upon request. The products were processed by capillary electrophoresis using ABI 3730 DNA Analyzer (Applied Biosystems), and analyzed using DNA Sequencing Analysis (Applied Biosystems) and Sequencher (Gene Codes Corp.).

Sequence alignment and phylogenetic analysis

Final sequences were edited and aligned using the ClustalW algorithm in MacVector version 10.0 (Accelrys) and MEGA version 3.1 [58]. Published AMPV and HMPV F sequences were obtained from GenBank (Accession numbers AY145287-AY145301, AY304360-AY304362, AYAY622381, EF051124, EF081369, EF199771-EF199772, EF589610, AF176593, AF187153-AF187154,

AF298642-AF298650, AF368170, AF085228, AJ400728, AJ400730, DQ175630-DQ175634, DQ207607, D00850, EU658938, Y14290-Y14294). Sequences identified in this study have been submitted to GenBank under accession numbers EU857542-EU857610. Pairwise sequence alignment, multiple sequence alignment, and percent nucleotide identity calculations were performed using MacVector version 9.0. Inference of phylogeny and overall rates of evolutionary change (nucleotide substitutions per site per year) and the time to most recent common ancestor (tMRCA) were estimated using the Bayesian Markov chain Monte Carlo (MCMC) approach available in the BEAST package http://beast.bio.ed.ac.uk/[37]. Because the sequences analyzed were very closely related and exhibited few multiple substitutions at single nucleotide sites, we used the simple HKY85 model of nucleotide substitution in each case, as more complex models sometimes failed to converge (data not shown). Data sets were analyzed under demographic models of constant population size, exponential population growth, and expansion population growth, using strict or relaxed (uncorrelated logarithmic) molecular clocks. Comparison of the output of each model showed that the relaxed clock, exponential population growth model gave the best estimation based on 95% highest posterior density (HPD)(not shown). All runs were visually examined to ensure convergence and Estimated Sample Size of >200. MCMC chains were run for 30 million steps with a burn-in rate of 10%, and two separate runs were combined using the Log Combiner program http://beast.bio.ed.ac.uk/[37], with uncertainty in parameter estimates reported as the 95% HPD. Output sets of trees were combined using LogCombiner and analyzed with the TreeAnnotator program to produce a Maximum Clade Credibility tree with a posterior probability limit of >50%. Final tree was produced with FigTree [37].

Competing interests

Authors' contributions

CFY, CKW, LDL, MC, AL, MM, RRS, and RP performed RT-PCR, cloning and sequencing of HMPV isolates. SJT cultivated HMPV isolates and performed RT-PCR, cloning and sequencing. JVW and JEC conceived the study, participated in its design and coordination, and helped to draft the manuscript. JVW performed sequence alignment and phylogenetic analysis. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplemental Figure 1. Nucleotide sequence alignment of full-length F genes from subgroup A1 HMPV isolates, listed in chronological order.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1743-422X-6-138-S1.pdf]

Additional file 2

Supplemental Figure 2. Nucleotide sequence alignment of full-length F genes from subgroup A2 HMPV isolates, listed in chronological order.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1743-422X-6-138-S2.pdf]

Additional file 3

Supplemental Figure 3. Nucleotide sequence alignment of full-length F genes from subgroup B1 HMPV isolates, listed in chronological order.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1743-422X-6-138-S3.pdf]

Additional file 4

Supplemental Figure 4. Nucleotide sequence alignment of full-length F genes from subgroup B2 HMPV isolates, listed in chronological order.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1743-422X-6-138-S4.pdf]

References

1. Hoogen BG van den, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA, Osterhaus AD: **A newly discovered human pneumovirus isolated from young children with respiratory tract disease.** *Nat Med* 2001, **7**:719-724.
2. Williams JV, Wang CK, Yang CF, Tollefson SJ, House FS, Heck JM, Chu M, Brown JB, Lintao LD, Quinto JD, *et al.*: **The role of human metapneumovirus in upper respiratory tract infections in children: a 20-year experience.** *J Infect Dis* 2006, **193**:387-395.
3. Williams JV, Harris PA, Tollefson SJ, Halburnt-Rush LL, Pingsterhaus JM, Edwards KM, Wright PF, Crowe JE Jr: **Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children.** *N Engl J Med* 2004, **350**:443-450.
4. Hoogen BG van den, van Doornum GJ, Fockens JC, Cornelissen JJ, Beyer WE, de Groot R, Osterhaus AD, Fouchier RA: **Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients.** *J Infect Dis* 2003, **188**:1571-1577.
5. Peiris JS, Tang WH, Chan KH, Khong PL, Guan Y, Lau YL, Chiu SS: **Children with respiratory disease associated with metapneumovirus in Hong Kong.** *Emerg Infect Dis* 2003, **9**:628-633.
6. Mullins JA, Erdman DD, Weinberg GA, Edwards K, Hall CB, Walker FJ, Iwane M, Anderson LJ: **Human metapneumovirus infection among children hospitalized with acute respiratory illness.** *Emerg Infect Dis* 2004, **10**:700-705.
7. McAdam AJ, Hasenbein ME, Feldman HA, Cole SE, Offermann JT, Riley AM, Lieu TA: **Human metapneumovirus in children tested at a tertiary-care hospital.** *J Infect Dis* 2004, **190**:20-26.
8. Mackay IM, Bialasiewicz S, Jacob KC, McQueen E, Arden KE, Nissen MD, Sloots TP: **Genetic diversity of human metapneumovirus over 4 consecutive years in Australia.** *J Infect Dis* 2006, **193**:1630-1633.
9. Foulongne V, Guyon G, Rodiere M, Segondy M: **Human metapneumovirus infection in young children hospitalized with respiratory tract disease.** *Pediatr Infect Dis J* 2006, **25**:354-359.
10. Esper F, Martinello RA, Boucher D, Weibel C, Ferguson D, Landry ML, Kahn JS: **A 1-year experience with human metapneumovirus in children aged <5 years.** *J Infect Dis* 2004, **189**:1388-1396.
11. Ebihara T, Endo R, Kikuta H, Ishiguro N, Ishiko H, Hara M, Takahashi Y, Kobayashi K: **Human metapneumovirus infection in Japanese children.** *J Clin Microbiol* 2004, **42**:126-132.
12. Dollner H, Risnes K, Radtke A, Nordbo SA: **Outbreak of human metapneumovirus infection in norwegian children.** *Pediatr Infect Dis J* 2004, **23**:436-440.
13. Boivin G, De Serres G, Cote S, Gilca R, Abed Y, Rochette L, Bergeron MG, Dery P: **Human metapneumovirus infections in hospitalized children.** *Emerg Infect Dis* 2003, **9**:634-640.
14. Williams JV, Martino R, Rabella N, Otegui M, Parody R, Heck JM, Crowe JE Jr: **A prospective study comparing human metapneumovirus with other respiratory viruses in adults with hematologic malignancies and respiratory tract infections.** *J Infect Dis* 2005, **192**:1061-1065.
15. Williams JV, Crowe JE Jr, Enriquez R, Minton P, Peebles RS Jr, Hamilton RG, Higgins S, Griffin M, Hartert TV: **Human metapneumovirus infection plays an etiologic role in acute asthma exacerbations requiring hospitalization in adults.** *J Infect Dis* 2005, **192**:1149-1153.
16. Vicente D, Montes M, Cilla G, Perez-Trallero E: **Human metapneumovirus and chronic obstructive pulmonary disease.** *Emerg Infect Dis* 2004, **10**:1338-1339.
17. Pelletier G, Dery P, Abed Y, Boivin G: **Respiratory tract reinfections by the new human Metapneumovirus in an immunocompromised child.** *Emerg Infect Dis* 2002, **8**:976-978.
18. Madhi SA, Ludewick H, Abed Y, Klugman KP, Boivin G: **Human metapneumovirus-associated lower respiratory tract infections among hospitalized human immunodeficiency virus type 1 (HIV-1)-infected and HIV-1-uninfected African infants.** *Clin Infect Dis* 2003, **37**:1705-1710.
19. Larcher C, Geltner C, Fischer H, Nachbaur D, Muller LC, Huemer HP: **Human metapneumovirus infection in lung transplant recipients: clinical presentation and epidemiology.** *J Heart Lung Transplant* 2005, **24**:1891-1901.
20. Englund JA, Boeckh M, Kuypers J, Nichols WG, Hackman RC, Morrow RA, Fredricks DN, Corey L: **Brief communication: fatal human metapneumovirus infection in stem-cell transplant recipients.** *Ann Intern Med* 2006, **144**:344-349.
21. Buchholz UJ, Nagashima K, Murphy BR, Collins PL: **Live vaccines for human metapneumovirus designed by reverse genetics.** *Expert Rev Vaccines* 2006, **5**:695-706.
22. Cseke G, Wright DW, Tollefson SJ, Johnson JE, Crowe JE Jr, Williams JV: **Human metapneumovirus fusion protein vaccines that are immunogenic and protective in cotton rats.** *J Virol* 2007, **81**:698-707.
23. Herfst S, de Graaf M, Schickli JH, Tang RS, Kaur J, Yang CF, Spaete RR, Haller AA, Hoogen BG van den, Osterhaus AD, Fouchier RA: **Recovery of human metapneumovirus genetic lineages a and B from cloned cDNA.** *J Virol* 2004, **78**:8264-8270.
24. Tang RS, Mahmood K, Macphail M, Guzzetta JM, Haller AA, Liu H, Kaur J, Lawlor HA, Stillman EA, Schickli JH, *et al.*: **A host-range restricted parainfluenza virus type 3 (PIV3) expressing the human metapneumovirus (hMPV) fusion protein elicits protective immunity in African green monkeys.** *Vaccine* 2005, **23**:1657-1667.
25. Mok H, Tollefson SJ, Podsiad AB, Shepherd BE, Polosukhin VV, Johnston RE, Williams JV, Crowe JE Jr: **An alphavirus replicon-based human metapneumovirus vaccine is immunogenic and protective in mice and cotton rats.** *J Virol* 2008, **82**:11410-11418.
26. Herfst S, de Graaf M, Schrauwen EJ, Ulbrandt ND, Barnes AS, Senthil K, Osterhaus AD, Fouchier RA, Hoogen BG van den: **Immunization**

of Syrian golden hamsters with F subunit vaccine of human metapneumovirus induces protection against challenge with homologous or heterologous strains. *J Gen Virol* 2007, **88:**2702-2709.

27. Skiadopoulos MH, Biacchesi S, Buchholz UJ, Amaro-Carambot E, Surman SR, Collins PL, Murphy BR: **Individual contributions of the human metapneumovirus F, G, and SH surface glycoproteins to the induction of neutralizing antibodies and protective immunity.** *Virology* 2006, **345:**492-501.

28. Mok H, Tollefson SJ, Podsiad AB, Shepherd BE, Polosukhin VV, Johnston RE, Williams JV, Crowe JE Jr: **An Alphavirus Replicon-Based Human Metapneumovirus Vaccine is Immunogenic and Protective in Mice and Cotton Rats.** *J Virol* 2008, **82(22):**11410-8.

29. Njenga MK, Lwamba HM, Seal BS: **Metapneumoviruses in birds and humans.** *Virus Res* 2003, **91:**163-169.

30. Toquin D, de Boisseson C, Beven V, Senne DA, Eterradossi N: **Subgroup C avian metapneumovirus (MPV) and the recently isolated human MPV exhibit a common organization but have extensive sequence divergence in their putative SH and G genes.** *J Gen Virol* 2003, **84:**2169-2178.

31. de Graaf M, Schrauwen EJ, Herfst S, van Amerongen G, Osterhaus AD, Fouchier RA: **Fusion protein is the main determinant of metapneumovirus host tropism.** *J Gen Virol* 2009, **90:**1408-1416.

32. Hoogen BG van den, Herfst S, Sprong L, Cane PA, Forleo-Neto E, de Swart RL, Osterhaus AD, Fouchier RA: **Antigenic and genetic variability of human metapneumoviruses.** *Emerg Infect Dis* 2004, **10:**658-666.

33. Schowalter RM, Smith SE, Dutch RE: **Characterization of human metapneumovirus F protein-promoted membrane fusion: critical roles for proteolytic processing and low pH.** *J Virol* 2006, **80:**10931-10941.

34. Schickli JH, Kaur J, Ulbrandt N, Spaete RR, Tang RS: **An S101P substitution in the putative cleavage motif of the human metapneumovirus fusion protein is a major determinant for trypsin-independent growth in vero cells and does not alter tissue tropism in hamsters.** *J Virol* 2005, **79:**10678-10689.

35. Ishiguro N, Ebihara T, Endo R, Ma X, Kikuta H, Ishiko H, Kobayashi K: **High genetic diversity of the attachment (G) protein of human metapneumovirus.** *J Clin Microbiol* 2004, **42:**3406-3414.

36. Beeler JA, van Wyke Coelingh K: **Neutralization epitopes of the F glycoprotein of respiratory syncytial virus: effect of mutation upon fusion function.** *J Virol* 1989, **63:**2941-2950.

37. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007, **7:**214.

38. Drake JW: **Rates of spontaneous mutation among RNA viruses.** *Proc Natl Acad Sci USA* 1993, **90:**4171-4175.

39. Jenkins GM, Rambaut A, Pybus OG, Holmes EC: **Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis.** *J Mol Evol* 2002, **54:**156-165.

40. Rima BK, Earle JA, Baczko K, ter Meulen V, Liebert UG, Carstens C, Carabana J, Caballero M, Celma ML, Fernandez-Munoz R: **Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture.** *J Gen Virol* 1997, **78(Pt 1):**97-106.

41. Garcia O, Martin M, Dopazo J, Arbiza J, Frabasile S, Russi J, Hortal M, Perez-Brena P, Martinez I, Garcia-Barreno B, *et al.*: **Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein.** *J Virol* 1994, **68:**5448-5459.

42. Crowe JE, Firestone CY, Crim R, Beeler JA, Coelingh KL, Barbas CF, Burton DR, Chanock RM, Murphy BR: **Monoclonal antibody-resistant mutants selected with a respiratory syncytial virus-neutralizing human antibody fab fragment (Fab 19) define a unique epitope on the fusion (F) glycoprotein.** *Virology* 1998, **252:**373-375.

43. Tamin A, Rota PA, Wang ZD, Heath JL, Anderson LJ, Bellini WJ: **Antigenic analysis of current wild type and vaccine strains of measles virus.** *J Infect Dis* 1994, **170:**795-801.

44. Cox NJ, Black RA, Kendal AP: **Pathways of evolution of influenza A (H1N1) viruses from 1977 to 1986 as determined by oligonucleotide mapping and sequencing studies.** *J Gen Virol* 1989, **70(Pt 2):**299-313.

45. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, *et al.*: **The global circulation of seasonal influenza A (H3N2) viruses.** *Science* 2008, **320:**340-346.

46. Six HR, Webster RG, Kendal AP, Glezen WP, Griffis C, Couch RB: **Antigenic analysis of H1N1 viruses isolated in the Houston metropolitan area during four successive seasons.** *Infect Immun* 1983, **42:**453-458.

47. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA: **Mapping the antigenic and genetic evolution of influenza virus.** *Science* 2004, **305:**371-376.

48. Hall CB, Walsh EE, Long CE, Schnabel KC: **Immunity to and frequency of reinfection with respiratory syncytial virus.** *J Infect Dis* 1991, **163:**693-698.

49. Skiadopoulos MH, Biacchesi S, Buchholz UJ, Riggs JM, Surman SR, Amaro-Carambot E, McAuliffe JM, Elkins WR, St Claire M, Collins PL, Murphy BR: **The two major human metapneumovirus genetic lineages are highly related antigenically, and the fusion (F) protein is a major contributor to this antigenic relatedness.** *J Virol* 2004, **78:**6927-6937.

50. Miller SA, Tollefson S, Crowe JE Jr, Williams JV, Wright DW: **Examination of a fusogenic hexameric core from human metapneumovirus and identification of a potent synthetic peptide inhibitor from the heptad repeat 1 region.** *J Virol* 2007, **81:**141-149.

51. Deffrasnes C, Hamelin ME, Prince GA, Boivin G: **Identification and evaluation of a highly effective fusion inhibitor for human metapneumovirus.** *Antimicrob Agents Chemother* 2008, **52:**279-287.

52. Williams JV, Chen Z, Cseke G, Wright DW, Keefer CJ, Tollefson SJ, Hessell A, Podsiad A, Shepherd BE, Sanna PP, *et al.*: **A recombinant human monoclonal antibody to human metapneumovirus fusion protein that neutralizes virus in vitro and is effective therapeutically in vivo.** *J Virol* 2007, **81:**8315-8324.

53. Ulbrandt ND, Ji H, Patel NK, Riggs JM, Brewah YA, Ready S, Donacki NE, Folliot K, Barnes AS, Senthil K, *et al.*: **Isolation and characterization of monoclonal antibodies which neutralize human metapneumovirus in vitro and in vivo.** *J Virol* 2006, **80:**7799-7806.

54. de Graaf M, Osterhaus AD, Fouchier RA, Holmes EC: **Evolutionary dynamics of human and avian metapneumoviruses.** *J Gen Virol* 2008, **89:**2933-2942.

55. Padhi A, Verghese B: **Positive natural selection in the evolution of human metapneumovirus attachment glycoprotein.** *Virus Res* 2008, **131:**121-131.

56. Velayudhan BT, Nagaraja KV, Thachil AJ, Shaw DP, Gray GC, Halvorson DA: **Human metapneumovirus in turkey poults.** *Emerg Infect Dis* 2006, **12:**1853-1859.

57. Hoogen BG van den, Bestebroer TM, Osterhaus AD, Fouchier RA: **Analysis of the genomic sequence of a human metapneumovirus.** *Virology* 2002, **295:**119-132.

58. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5:**150-163.