Evolutionary dynamics of human and avian metapneumoviruses

Miranda de Graaf,1 Albert D. M. E. Osterhaus,1 Ron A. M. Fouchier1 and Edward C. Holmes2,3

1Department of Virology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands
2Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA
3Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

Correspondence
Ron A. M. Fouchier
r.fouchier@erasusmc.nl

Received 29 August 2008
Accepted 19 September 2008

Human (HMPV) and avian (AMPV) metapneumoviruses are closely related viruses that cause respiratory tract illnesses in humans and birds, respectively. Although HMPV was first discovered in 2001, retrospective studies have shown that HMPV has been circulating in humans for at least 50 years. AMPV was first isolated in the 1970s, and can be classified into four subgroups, A–D. AMPV subgroup C is more closely related to HMPV than to any other AMPV subgroup, suggesting that HMPV has emerged from AMPV-C upon zoonosis. Presently, at least four genetic lineages of HMPV circulate in human populations – A1, A2, B1 and B2 – of which lineages A and B are antigenically distinct. We used a Bayesian Markov Chain Monte Carlo (MCMC) framework to determine the evolutionary and epidemiological dynamics of HMPV and AMPV-C. The rates of nucleotide substitution, relative genetic diversity and time to the most recent common ancestor (TMRCA) were estimated using large sets of sequences of the nucleoprotein, the fusion protein and attachment protein genes. The sampled genetic diversity of HMPV was found to have arisen within the past 119–133 years, with consistent results across all three genes, while the TMRCA for HMPV and AMPV-C was estimated to have existed around 200 years ago. The relative genetic diversity observed in the four HMPV lineages was low, most likely reflecting continual population bottlenecks, with only limited evidence for positive selection.

INTRODUCTION

Human metapneumovirus (HMPV) is a respiratory pathogen that was first described in 2001 (van den Hoogen et al., 2001). HMPV causes respiratory illnesses ranging from mild upper respiratory symptoms to severe lower respiratory tract disease (van den Hoogen et al., 2001, 2003; Williams et al., 2004). HMPV is an enveloped, non-segmented, negative-strand RNA virus and was classified as a member of the family Paramyxoviridae, the subfamily Pneumovirinae, the genus Metapneumovirus. HPMV seroprevalence in the human population reaches 100% by the age of five (van den Hoogen et al., 2001). Experimental HMPV infections in macaques have been shown to induce transient protection (van den Hoogen et al., 2007), which is in agreement with relatively frequent occurrences of reinfection in humans (Ebihara et al., 2004). HMPV can be divided in two main genetic lineages (A and B), each consisting of at least two sublineages – A1, A2, B1 and B2 (van den Hoogen et al., 2004). Annual circulation of all four lineages has been observed worldwide, with reports of a predominating strain changing on a yearly basis in some studies, but not in others (Galiano et al., 2006; Sloots et al., 2006; van den Hoogen et al., 2004).

The only other member of the genus Metapneumovirus is avian metapneumovirus (AMPV). AMPV has been found to infect domestic poultry worldwide, causing acute respiratory infections (Cook, 2000b). AMPVs have been classified into four subgroups, A through to D (Bayon-Auboyer et al., 1999; Eterradossi et al., 1995; Juhasz & Easton, 1994; Seal, 1998). AMPV subgroups A and B were first detected in South Africa in 1978 and were subsequently found in Europe, Israel and Asia (Cook, 2000a). AMPV subgroup D was first detected in France (Bayon-Auboyer et al., 1999). Finally, AMPV subgroup C was first detected in the USA in 1996 and is more closely related to HMPV than to any other AMPV subgroup (Govindarajan & Samal, 2004, 2005; Govindarajan et al., 2004; Toquin et al., 2003; van den Hoogen et al., 2002; Yunus et al., 2003).
The genome organizations of HMPV and AMPV are similar: from the 3' to 5' ends the genomes encode nucleoprotein (N), phosphoprotein (P), matrix protein (M), fusion protein (F), M2-1 and M2-2, small hydrophobic protein (SH), attachment protein (G) and the large polymerase protein (L). Overall, HMPV and AMPV-C share 80% amino acid identity, with N being most conserved protein and SH and G the most variable proteins (Biacchini et al., 2003; van den Hoogen et al., 2001, 2002). The viral RNA is encapsidated by the N protein, which, together with the P and L proteins, forms the ribonucleoprotein complex required for genome replication. With analogy to respiratory syncytial virus (RSV), M2-1 is a transcriptional elongation factor that enhances synthesis of readthrough mRNAs (Collins et al., 1996; Fears & Collins, 1999; Hardy & Wertz, 1998), although for HMPV its function is not completely understood as M2-1 is not essential for virus viability (Buchholz et al., 2005; Herfst et al., 2004). HMPV has three surface glycoproteins: SH, G and F. The F protein is synthesized as an inactive precursor (F₂) that is cleaved by host proteases into the functional subunits F₁ and F₂. The F protein of HMPV is highly immunogenic and protective, whereas G and SH are not (Herfst et al., 2007; Skiadopoulos et al., 2006).

The close genetic relationship between HMPV and AMPV-C is not reflected in host range. To date, no AMPV-C infections have been reported in humans and it has been shown that HMPV does not replicate in turkeys (van den Hoogen et al., 2001). However, using both chimeric viruses and minireplicon systems it has been shown that the ribonucleoprotein complex proteins of HMPV and AMPV-C are functionally interchangeable (de Graaf et al., 2008; Govindarajan et al., 2006; Pham et al., 2005). Spill-over events of pathogens from birds to humans are not uncommon (Reed et al., 2003). Since HMPV is most closely related to AMPV-C, it is plausible that an AMPV-C-like virus was the ancestor of HMPV. Retrospective analysis of serum samples collected from humans in 1958 revealed that HMPV has been widespread in the human population for at least 50 years. Thus, any cross-species transmission event must have taken place before 1958 (van den Hoogen et al., 2001). Herein, we test this hypothesis using sequence data from the N, G and F genes of HMPV and AMPV-C collected over a 25-year-sampling period. In addition, to understand the evolutionary dynamics of HMPV in more detail we estimated rates of nucleotide substitution, times to common ancestry and relative genetic diversity of all HMPV lineages using a Bayesian coalescent approach, as well as gene- and site-specific selection pressures.

### METHODS

**Data preparation.** All gene sequence datasets were compiled from GenBank. Two additional nearly complete HMPV genome sequences were submitted to GenBank (accession nos FJ168778 and FJ168779). The data for the G, F and N genes were divided into subsets comprising the individual sublineages of HMPV, AMPV-C or combinations thereof. For the G gene, sequence alignments were made for HMPV-A1 (ORF nt position 1–616, n=19), HMPV-A2 (nt 1–616, n=29), HMPV-B1 (nt 1–634, n=14), HMPV-B2 (nt 1–619, n=14) and for all HMPV lineages (nt 1–442, n=76). For the F gene, sequence alignments were compiled for HMPV-A1 (n=22), HMPV-A2 (n=30), HMPV-B1 (n=25), HMPV-B2 (n=30) and AMPV-C (n=12) using ORF nt positions 781–1221. For the N gene, sequence alignments were made for HMPV (complete ORF, n=21) and AMPV-C (complete ORF, n=15). All sequence alignments were created using both MUSCLE (Edgar, 2004) and CLUSTAL W software running within the BioEdit (version 7.0.1) (Hall, 1999) program, and adjusted manually where necessary. Sequence alignments are available from the authors on request.

**Estimating evolutionary dynamics.** Overall rates of evolutionary change (nucleotide substitutions per site, per year), the time to the most recent common ancestor (TM RCA) in years and relative genetic diversity (expressed as Nₑ, where Nₑ is the effective population size and t the generation), were estimated using the BEAST program (http://beast.bio.ed.ac.uk/). This program employs a Bayesian Markov Chain Monte Carlo (MCMC) approach, utilizing the number and temporal distribution of genetic differences among viruses sampled at different times (Drummond et al., 2005, 2006; Drummond & Rambaut, 2007). To estimate the substitution rates with as much accuracy as possible, we compared a variety of models of demographical history, namely constant population size, exponential population growth, and Bayesian skyline reconstruction, and with both strict (constant) and relaxed (variable) molecular clocks, and using a variety of prior values for the substitution rate. As similar rate/ date estimates were obtained under all demographical models, we only report those from the Bayesian skyline model as this employs the least constrained coalescent prior. Similarly, a preliminary analysis of a subset of the data collected here (data not shown) revealed that the uncorrelated lognormal relaxed clock model gave the best estimation, although rate/date estimates were again highly congruent across all clock models, suggesting that they are robust. Because the sequences analysed here were very closely related, sampled over a relatively narrow time-span and exhibited little multiple substitution 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). All MCMC chains were run for a sufficient length of time to ensure convergence (allowing a 10% burn-in) and uncertainty in parameter estimates is reported as values of the 95% highest probability density (HPD). Finally, the MCMC samples were summarized to infer the maximum clade credibility (MCC) trees of each dataset using TreeAnnotator v1.4.7 (Drummond & Rambaut, 2007), with posterior probability values shown for the major branches. The MCC trees were depicted with estimated branch lengths in time (years).

**Analysis of selection pressures.** We estimated overall (i.e. genome-specific) and site-specific pressures in all datasets using the Datamonkey web interface of the HY-PHY package (www.datamonkey.org) (Pond & Frost, 2005). This method quantifies selection pressures as the ratio of non-synonymous (dN) to synonymous substitutions (dS) per site, using the random effects likelihood (REL) method for datasets smaller than 40 sequences, the fixed effects likelihood (FEL) method for datasets smaller up to 100 sequences and the single likelihood ancestor counting (SLAC) for the datasets containing more than 100 sequences and the mean dS/dN (ω) values. These were run with three different nucleotide substitution models – HKY85, TrN93 and REV – although for most alignments the TrN93 model was the best-fitting model of nucleotide substitution.

For an additional analysis of selection pressures we used the CODEML program (Yang et al., 2000) to estimate the maximum-likelihood ω value for all branches of each phylogeny (the ‘one-ratio’ model), and compared this to the ω values estimated separately for the external (ω₆)
and internal \((o_e)\) branches using the ‘two-ratio’ model. Input phylogenetic trees for this analysis were estimated under the general time reversible (GTR) + I + G4 substitution model, using PAUP* (Swofford, 2003), and employing tree bisection–reconnection (TBR) branch-swapping.

**RESULTS**

**Evolutionary dynamics of metapneumoviruses**

Rates of nucleotide substitution and TMRCAs for the G, N and F genes of HMPV and AMPV were estimated using the BEAST program. For the G gene, separate datasets for all four sublineages (HMPV-A1, -A2, -B1 and -B2) were analysed both individually and as part of combined datasets containing all sequences of HMPV-A, HMPV-B or HMPV. No dataset containing both HMPV and AMPV-C sequences was made, since it was not possible to ensure proper alignment in this case. For the F gene, similar datasets were compiled with an additional dataset containing all HMPV and AMPV-C sequences. Because of limited sequence availability, no N gene dataset was made for the individual HMPV lineages. MCC trees were also inferred for each metapneumovirus dataset and plotted in (mean) evolutionary time (Fig. 1). The mean rate of nucleotide substitution for the G gene of HMPV was \(3.5 \times 10^{-5}\) (HPD, 2.3–4.8 \times 10^{-5}) nucleotide substitutions per site, per year. Somewhat lower rates were found for the F and N genes, at \(9.0 \times 10^{-4}\) (HPD, 6.6–11.5 \times 10^{-4}) and \(8.5 \times 10^{-4}\) (HPD, 2.0–15.0 \times 10^{-4}) substitutions per site, per year, respectively. Because of the small size of some of the individual sublineage datasets, and hence the large sampling error, we fixed their substitution rates – using the values obtained from the complete datasets determined above – to obtain a better estimate of TMRCAs. Accordingly, the mean TM RCA for HMPV-A1, -A2, -B1 and -B2 were estimated to be 14 (HPD, 11–20), 26 (HPD, 25–28), 13 (HPD, 10–16) and 18 (HPD, 12–24) years, respectively, based on the G gene, and 12 (HPD, 6–20), 28 (HPD, 24–34), 13 (HPD, 8–20) and 18 (HPD, 12–24) years based on the F gene datasets. The estimated TM RCA of the two main lineages HMPV-A and HMPV-B are strikingly similar, at 41 (HPD, 24–61) and 34 (HPD, 20–51) years for the F gene, and 41 (HPD, 32–51) and 37 (HPD, 20–26) years for the G gene. The TM RCA for HMPV as a whole (i.e. all lineages) was estimated to be 122 (HPD, 54–204) years for the G, 119 (HPD, 79–166) for the F, and 133 (HPD, 40–264) for the N genes. Thus, the genetic diversity between ~1996 and ~2001 (although the 95% HPD interval is wide), a pattern that was seen for all lineages and in both the G and F genes.

**Standing genetic diversity**

The demographical history of the HMPV dataset was inferred using the Bayesian skyline plot (Fig. 2). Notably, the relative genetic diversity \((N_e\tau)\) in all HMPV lineages was very low, ranging from only 3.3 to 96.8. In addition, the Bayesian skyline plot also showed that there was an increase and subsequent decrease in genetic diversity between ~1996 and ~2001 (although the 95% HPD interval is wide), a pattern that was seen for all lineages and in both the G and F genes.

**Selection pressures**

Our analysis of selection pressures through computation of the ratio of non-synonymous \((d_{NS})\) to synonymous \((d_k)\) substitutions per site \(d_{NS}/d_k (\omega)\) revealed limited positive selection and abundant negative selection for the G (mean \(\omega=0.524\)) and F proteins (mean \(\omega=0.075\)) (Table 1). It should be noted that for the HMPV set the G sequences were shortened to ensure proper alignment. Neither positive nor negative selection could be detected for the N protein (mean \(\omega=0.038\)). Overall, this analysis revealed that the mean \(\omega\) is roughly 10-fold higher for the G than for the F and N genes, compatible with the localized action of immune selection on the former. Positively selected sites in the F protein were only found for the A1 (aa 34) and A2 (aa 34 and 50) sets. Both sites are located in the extracellular domain of F. Positively selected sites for the G protein were found for the A1 (aa 82, 125, 156 and 158), A2 (aa 146 and 151), B2 (aa 85) and HMPV (aa, 85, 86, 93, 111, 112, 113, 140 and 142) sets and were all located in the extracellular domain. By using the NetOglyc (version 3.1) program it was predicted that all but one (aa 86) of the positively selected sites found in the G protein corresponded to potential O-linked glycosylated sites in one or more strains of that dataset (Hansen et al., 1998).

To obtain a more precise picture of selection pressures we also estimated \(\omega\) separately for the internal \((o_i)\) and external \((o_e)\) branches of each phylogeny using the CODEML program (Table 2). In case of the F and N genes the \(\omega\) values were higher on the tips compared with the internal branches of the trees, resulting in an increased \(o_d/o_e\) value, and strongly suggesting that most of the non-synonymous mutations in these proteins are transiently deleterious (so that they are purged by purifying selection before they can reach high frequencies). However, values of both \(o_i\) and \(o_e\) values were high in the case of the G gene, resulting in a low \(o_d/o_e\) value, and again suggestive of immune selection.

**DISCUSSION**

In the present study we explored the evolutionary dynamics of HMPV and AMPV through analyses of a large set of HMPV and AMPV-C G, F and N gene sequences. For HMPV, our results were highly consistent across all three datasets, indicating that the admittedly sparse sampling did
Fig. 1. MCC trees from the Bayesian analysis of the HMPV G (a), the HMPV and AMPV-C N (b) and F (c) sequences. The posterior probabilities of the branches are depicted in each case. Note that the mean TMRCAs on the MCC trees can differ from those reported in the text, although all are contained within the same 95% HPD values. Scale bar represents time in years.
not introduce major biases. In particular, both the highly variable G sequences and the less variable F and N sequences gave similar demographical signatures and times to common ancestry, although the nucleotide substitution rates were higher for the G gene. The results for the G gene were comparable to those described previously (Padhi & Verghese, 2006). Datasets containing HMPV-A1 (a), HMPV-A2 (b), HMPV-B1 (c), HMPV-B2 (d), HMPV-A1 and -A2 (e), HMPV-B1 and -B2 (f) and HMPV (g) sequences from the F gene (left panels) and G (right panels) were used. The thick line represents the median estimate and the dashed lines show the 95% HPD limits.

**Fig. 2.** Bayesian skyline plots of the changing levels of relative genetic diversity (N_e(t)) of HMPV shown on a log scale against time backwards since the most recent (t_0=2006). Datasets containing HMPV-A1 (a), HMPV-A2 (b), HMPV-B1 (c), HMPV-B2 (d), HMPV-A1 and -A2 (e), HMPV-B1 and -B2 (f) and HMPV (g) sequences from the F gene (left panels) and G (right panels) were used. The thick line represents the median estimate and the dashed lines show the 95% HPD limits.
Overall, evolutionary rates were high and of the level expected for RNA viruses (Hanada et al., 2004; Jenkins et al., 2002). In addition, the genetic diversity found within the different sublineages of HMPV was of a very recent origin, and HMPV as a whole was characterized by low levels of standing genetic variation (as reflected in estimates of \( N_e \)). Such limited genetic diversity, which is also characteristic of human influenza A virus (Rambaut et al., 2008), as well as other paramyxoviruses (Pomeroy et al., 2008) is most likely due to repeated (annual) population bottlenecks, which periodically purge genetic variation.

Overall, the Bayesian skyline plots for HMPV suggest that the population dynamics of this virus are rather complex, with both increases and decreases in population size, and again expected for a virus where there are distinct peaks and troughs in prevalence. The general trend for all four lineages was an increase and subsequent decrease in population size between ~1996 and ~2001. Unfortunately the resolution of these Bayesian skyline plots was not sufficient to allow detailed analyses of the population dynamics in each viral lineage, or their precise cause. However, for the future it will be interesting to determine whether the increases and decreases in genetic diversity are in-phase or out-of-phase, as demonstrated in some other RNA viruses (Adams et al., 2006). At present it is not possible to compare these demographics to epidemiological studies published, as none of the studies cover a comparable period of time and include data of viruses isolated from patients and healthy individuals, including children as well as adults. Our analysis also provides new insights into the timescale of HMPV evolution. Although HMPV had already been shown to be widely spread in the human population in 1958, our coalescent analysis reveals that the TMRCA of HMPV existed approximately 70 years before this, so that HMPV was circulating in human populations long before its detection. Since HMPV and AMPV-C are highly similar in terms of protein composition and sequence identity it has been speculated that HMPV is the result of a cross-species transmission event from birds to humans (van den Hoogen et al., 2001). Further, since there are a wider variety of AMPV subgroups compared with HMPV subtypes it seems most likely that the direction of transmission is from birds to humans. As the TMRCA for HMPV and AMPV-C was calculated to be around 200 years ago, we suggest that the cross-species transmission event from birds to humans occurred at about this time.

### Table 1. Comparison of mean values of \( \omega (d_{NS}/d_S) \) of the G, F and N genes between the different HMPV lineages

| Lineage | Gene | \( \omega \) | Positively selected sites* | Negatively selected sites* | Method |
|---------|------|-------------|--------------------------|---------------------------|-------|
| HMPV-A1 | G    | 0.544       | 4 (aa 82, 125, 156, 158) | 2                         | REL   |
| HMPV-A1 | F    | 0.041       | 1 (aa 294)               | 1                         | REL   |
| HMPV-A2 | G    | 0.518       | 2 (aa 146, 151)          | 3                         | REL   |
| HMPV-A2 | F    | 0.166       | 2 (aa 294, 310)          | 1                         | REL   |
| HMPV-B1 | G    | 0.495       | 0                        | 37                        | REL   |
| HMPV-B1 | F    | 0.108       | 0                        | 0                         | REL   |
| HMPV-B2 | G    | 0.693       | 1 (aa 85)                | 0                         | REL   |
| HMPV-B2 | F    | 0.100       | 0                        | 27                        | REL   |
| HMPV   | G    | 0.524       | 8 (aa 85, 86, 93, 111, 112, 113, 140, 142) | 28                        | REL   |
| HMPV   | F    | 0.075       | 0                        | 30                        | SLAC  |
| HMPV   | N    | 0.038       | 0                        | 0                         | REL   |

*P<0.1.

### Table 2. The mean ratio \( \omega (d_{NS}/d_S) \) for all branches in the tree, and the separate \( \omega \) values estimated for the external (\( \omega_e \)) and internal (\( \omega_i \)) branches of the phylogenies of the HMPV G, F and N genes

| Virus (gene) | No. sequences | Sequence length (codons) | \( \omega \) (all branches) | \( \omega_i \) (internal branches) | \( \omega_e \) (external branches) |
|--------------|---------------|--------------------------|-----------------------------|-----------------------------------|-----------------------------------|
| HMPV N       | 20            | 394                      | 0.027                       | 0.021                             | 0.060                             |
| HMPV F       | 107           | 147                      | 0.042                       | 0.015                             | 0.151                             |
| HMPV G       | 76            | 147                      | 0.422                       | 0.422                             | 0.421                             |
| HMPV A G     | 48            | 147                      | 0.399                       | 0.433                             | 0.358                             |
| HMPV B G     | 28            | 147                      | 0.497                       | 0.476                             | 0.523                             |
Despite its capacity to cause human disease, we found few positively selected sites for the F and G genes and none for the N gene of HMPV. Indeed, negatively selected sites were abundant for both the F and G genes. The positively selected sites found for the F protein were located in the extracellular domain of the F1 chain in the cysteine-rich region. Although the location of the antigenic sites of the F protein of HMPV have yet to be identified, some neutralizing monoclonal antibodies specific for the F protein have been described previously (Ulbrandt et al., 2006). The epitopes of these monoclonal antibodies do not correspond to the location of the positively selected sites described here. For RSV, most antigenic sites are located in the cysteine-rich region (Lopez et al., 1998). However, alignment of HMPV and RSV F sequences showed that the positively selected sites did not correspond to any of these antigenic sites. All positively selected sites found for the G protein were located in the extracellular domain, and all but one corresponded to potential O-linked glycosylated sites in one or more HMPV strain. For RSV, there is a strong association between the positively selected sites and neutralizing epitopes (Zlateva et al., 2004). However, the G protein of HMPV is only weakly immunogenic and protective in animal models in contrast to RSV (Olmsted et al., 1986; Skiadopoulos et al., 2006). The mean ω was approximately 10-fold higher for the G gene than for the N and F genes. In addition, separate estimates of the ω ratio on internal versus external branches revealed a major difference between the G, F and N genes, with a far greater ωF/ωN in the former protein. Hence, the N and F genes are evidently more constrained than the G gene, so that mutations in these genes are more likely to be deleterious and result in an elevation in the number of (young) non-synonymous mutations towards the tips of the trees. In contrast, a greater proportion of the mutations occurring in the G gene are likely to be either neutral or positively selected (although this latter process is difficult to detect using the analytical methods employed here), resulting in a higher level of non-synonymous diversity and lower ωF/ωN values. Similar results have been observed in the G gene of bovine respiratory syncytial virus (Pybus et al., 2007; Zwickl, 2006).

In conclusion, we were able to elucidate the evolutionary history and dynamics of metapneumoviruses by using dated genomic AMPV and HMPV sequences. Metapneumoviruses have high substitution rates like most RNA viruses due to high mutation rates, short generation times and large population sizes. Although such high mutation rates are not necessarily adaptive, and most often result in deleterious mutations, as reflected in the purifying selection on the F and N genes, they may occasionally facilitate viral emergence in novel hosts.

ACKNOWLEDGEMENTS

We thank Byron Martina for stimulating discussions, Roel van Eijk and Theo Bestbroer for excellent technical assistance, and Nancy Ulbrandt for sharing unpublished results.

REFERENCES

Adams, B., Holmes, E. C., Zhang, C., Mammen, M. P., Jr, Nimmannitya, S., Kalayanarooj, S. & Boots, M. (2006). Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. Proc Natl Acad Sci U S A 103, 14234–14239.

Bayon-Auboyer, M. H., Jestin, V., Toquin, D., Cherbonnel, M. & Eterradossi, N. (1999). Comparison of F-, G- and N-based RT-PCR protocols with conventional virological procedures for the detection and typing of turkey rhinotracheitis virus. Arch Virol 144, 1091–1109.

Biacchesi, S., Skiadopoulos, M. H., Boivin, G., Hanson, C. T., Murphy, B. R., Collins, P. L. & Buchholz, U. J. (2003). Genetic diversity between human metapneumovirus subgroups. Virology 315, 1–9.

Buchholz, U. J., Biacchesi, S., Pham, Q. N., Tran, K. C., Yang, L., Luongo, C. L., Skiadopoulos, M. H., Murphy, B. R. & Collins, P. L. (2005). Deletion of M2 gene open reading frames 1 and 2 of human metapneumovirus: effects on RNA synthesis, attenuation, and immunogenicity. J Virol 79, 6588–6597.

Collins, P. L., Hill, M. G., Cristina, J. & Grosfeld, H. (1996). Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. Proc Natl Acad Sci U S A 93, 81–85.

Cook, J. K. (2000a). Avian pneumovirus infections of turkeys and chickens. Vet J 160, 118–125.

Cook, J. K. (2000b). Avian rhinotracheitis. Rev Sci Tech 19, 602–613.

de Graaf, M. & Samal, S. K. (2004). Sequence analysis of the large polymerase (L) protein of the US strain of avian metapneumovirus subgroup C indicates that it possesses the longest genome among metapneumoviruses. Virus Genes 30, 331–333.
Evolutionary dynamics of metapneumoviruses

Govindarajan, D., Yunus, A. S. & Samal, S. K. (2004). Complete sequence of the G glycoprotein gene of avian metapneumovirus subgroup C and identification of a divergent domain in the predicted protein. *J Gen Virol* **85**, 3671–3675.

Govindarajan, D., Buchholz, U. J. & Samal, S. K. (2006). Recovery of avian metapneumovirus subgroup C from cDNA: cross-recognition of avian and human metapneumovirus support proteins. *J Virol* **80**, 5790–5797.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Hanada, K., Suzuki, Y. & Gojobori, T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol* **21**, 1074–1080.

Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. & Brunak, S. (1998). NetGlyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* **15**, 115–130.

Hardy, R. W. & Wertz, G. W. (1998). The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J Virol* **72**, 520–526.

Herfst, S., de Graaf, M., Schickli, J. H., Tang, R. S., Kaur, J., Yang, C. F., Spaeet, R. R., Haller, A. A., van den Hoogen, B. G. & other authors (2004). Recovery of human metapneumovirus genetic lineages A and B from cloned cDNA. *J Virol* **78**, 8264–8270.

Herfst, S., de Graaf, M., Schrauwen, E. J., Ulbrandt, N. D., Barnes, A. S., Senthil, K., Osterhaus, A. D., Fouchier, R. A. & van den Hoogen, B. G. (2007). Immunization of Syrian golden hamsters with F subunit vaccine from cloned cDNA. *J Virol* **78**, 2972–2979.

Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* **54**, 156–165.

Juhasz, K. & Easton, A. J. (1994). Extensive sequence variation in the attachment (G) protein gene of avian pneumovirus: evidence for two distinct subgroups. *J Gen Virol* **75**, 2873–2880.

Lopez, J. A., Bustos, R., Orvell, C., Berois, M., Arbiza, J., Garcia-Barreno, B. & Melero, J. A. (1998). Antigenic structure of human respiratory syncytial virus fusion glycoprotein. *J Virol* **72**, 6922–6928.

Olmsted, R. A., Elango, N., Prince, G. A., Murphy, B. R., Johnson, P. R., Barreno, B. & Melero, J. A. (1998). Antigenic structure of human respiratory syncytial virus fusion glycoprotein. *J Virol* **72**, 6922–6928.

Pybus, O. G., Rambaut, A., Belshaw, R., Freczekton, R. P., Drummond, A. J. & Holmes, E. C. (2007). Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* **24**, 845–852.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K. & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619.

Reed, K. D., Meece, J. K., Henkel, J. S. & Shulka, S. K. (2003). Birds, migration and emerging zoonoses: West Nile virus, lytic disease, influenza A and enteropathogens. *Clin Med Res* **1**, 5–12.

Seal, B. S. (1998). Matrix protein gene nucleotide and predicted amino acid sequence demonstrate that the first US avian pneumovirus isolate is distinct from European strains. *Virus Res* **58**, 45–52.

Skiaadopulos, M. H., Biacchesi, S., Buchholz, U. J., Amarocco, E., Surman, S. R., Collins, P. L. & Murphy, B. R. (2006). Individual contributions of the human metapneumovirus F, G, and SH surface glycoproteins to the induction of neutralizing antibodies and protective immunity. *Virology* **345**, 492–501.

Sloots, T. P., Mackay, I. M., Bialasiewicz, S., Jacob, K. C., McQueen, E., Harnett, G. B., Siebert, D. J., Masters, B. I., Young, P. R. & Nissen, M. D. (2006). Human metapneumovirus, Australia, 2001–2004. *Emerg Infect Dis* **12**, 1263–1266.

Swofford, D. L. (2003). *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.

van Hoogen, B. G., Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A. & Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med* **7**, 719–724.

van Hoogen, B. G., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. (2002). Analysis of the genomic sequence of a human metapneumovirus. *Virology* **295**, 119–132.

van Hoogen, B. G., van Doornum, G. J. J., Fockens, J. C., Cornelissen, J. J., Beyer, W. E. P., de Groot, R., Osterhaus, A. D. M. E. & Fouchier, R. A. M. (2003). Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients. *J Infect Dis* **188**, 1571–1577.

van Hoogen, B. G., Herfst, S., Spong, L., Cane, P. A., Forleo, E., de Swart, R. L., Osterhaus, A. D. M. E. & Fouchier, R. A. M. (2004). Antigenic and genetic variability of human metapneumoviruses. *Emerg Infect Dis* **10**, 658–666.

van Hoogen, B. G., Herfst, S., de Graaf, M., Spong, L., van Lavieren, R., van Amerongen, G., Yuksel, S., Fouchier, R. A., Osterhaus, A. D. & de Swart, R. L. (2007). Experimental infection of macaques with human metapneumovirus induces transient protective immunity. *J Gen Virol* **88**, 1251–1259.

Williams, J. V., Harris, P. A., Tollefson, S. J., Halburnt-Rush, L. L., Pingsterhaus, J. M., Edwards, K. M., Wright, K. D., Crowe, J. E., Jr, Williams, J. V., Harris, P. A., Tollefson, S. J., Halburnt-Rush, L. L., Pingsterhaus, J. M., Edwards, K. M., Wright, K. D., Crowe, J. E., Jr (2004). Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N Engl J Med* **350**, 443–450.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. C. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.

Yunus, A. S., Govindarajan, D., Huang, Z. & Samal, S. K. (2003). Deduced amino acid sequence of the small hydrophobic protein of...
US avian pneumovirus has greater identity with that of human metapneumovirus than those of non-US avian pneumoviruses. *Virus Res* 93, 91–97.

**Zlateva, K. T., Lemey, P., Vandamme, A. M. & Van Ranst, M.** (2004). Molecular evolution and circulation patterns of human respiratory syncytial virus subgroup A: positively selected sites in the attachment G glycoprotein. *J Virol* 78, 4675–4683.

**Zwickl, D. J.** (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD thesis, The University of Texas at Austin.