Molecular Comparisons of Full Length Metapneumovirus (MPV) Genomes, Including Newly Determined French AMPV-C and –D Isolates, Further Supports Possible Subclassification within the MPV Genus

Paul A. Brown*, Evelyne Lemaître, François-Xavier Briand, Céline Courtillon, Olivier Guionie, Chantal Allié, Didier Toquin, Marie-Hélène Bayon-Auboyer*, Véronique Jestin, Nicolas Eterradossi

French Agency for Food, Environmental and Occupational Health Safety (ANSES), Avian and Rabbit Virology Immunology and Parasitology Unit (VIPAC), Université Européenne de Bretagne, Ploufragan/Plouzané laboratory, Ploufragan, France

Abstract

Four avian metapneumovirus (AMPV) subgroups (A–D) have been reported previously based on genetic and antigenic differences. However, until now full length sequences of the only known isolates of European subgroup C and subgroup D viruses (duck and turkey origin, respectively) have been unavailable. These full length sequences were determined and compared with other full length AMPV and human metapneumoviruses (HMPV) sequences reported previously, using phylogenetics, comparisons of nucleic and amino acid sequences and study of codon usage bias. Results confirmed that subgroup C viruses were more closely related to HMPV than they were to the other AMPV subgroups in the study. This was consistent with previous findings using partial genome sequences. Closer relationships between AMPV-A, B and D were also evident throughout the majority of results. Three metapneumovirus “clusters” HMPV, AMPV-C and AMPV-A, B and D were further supported by codon bias and phylogenetics. The data presented here together with those of previous studies describing antigenic relationships also between AMPV-A, B and D and between AMPV-C and HMPV may call for a subclassification of metapneumoviruses similar to that used for avian paramyxoviruses, grouping AMPV-A, B and D as type I metapneumoviruses and AMPV-C and HMPV as type II.

Introduction

The Genus *Metapneumovirus* (MPV), in the family *Paramyxoviridae*, subfamily *Pneumovirinae*, includes globally important viruses in avian and human health. Avian MPV (AMPV) [1–5] cause respiratory and genital disorders in poultry having a severe economic impact on the industry [6]. Human MPV (HMPV), is responsible for bronchiolitis in infants [7,8] and severe infections in the elderly or immunocompromised adults [7,9–11]. AMPV and HMPV are now classified into the genus *Metapneumovirus* [12], to acknowledge the difference in genome order [13] and the absence of the non structural protein genes NS1 and NS2 as evidenced throughout the majority of results. Three metapneumovirus “clusters” HMPV, AMPV-C and AMPV-A, B and D were further supported by codon bias and phylogenetics. The authors have declared that no competing interests exist.
and most recently from Muscovy duck in China (Acc N’\nKC915036 and KF364615). Partial sequence (M gene) is available
for a Chinese chicken isolate of AMPV-C (acc N’\nJX422020).

In combination with sequences described previously[20,24,34,35], this study completes the sequences of European AMPV-C (French duck isolate, Fr-AMPV-C) and D (French
Turkey isolate, Fr-AMPV-D). Comparisons were made to full
length sequences of European AMPV-A, B and US-C and of all
HMPV-A and B sublineages.

Materials and Methods

Acquisition of Full length Fr-AMPV-C and D sequences

Fr-AMPV-C Muscovy duck/France/1999/99178 and Fr-
AMPV-D Turkey/France/1985/85035 (the latter previously
identified Fr/85/1 in refs [35,36]) were propagated in Vero cells
as described previously[20,36]. The 99178 and 85035 viruses
were shown experimentally to be pathogenic for SPF Muscovy
ducklings or turkeys, respectively [37]. Fr-AMPV-C and Fr-
AMPV-D virus stocks had a titer of log104.20 and log105.00
TCID50/ml, respectively [38]. Viral RNA was extracted using
QIaamp Viral RNA mini kit (Qiagen, France) according to
the manufacturer’s instructions. Primers were designed from previ-
ously published partial sequences (Table S1) and from the full
manufacturer's instructions. dsDNA was amplified from cDNA in overlap-
ning segments using Expand high fidelity enzyme (EHF, Roche,
France) according to manufacturer’s recommendations. PCR
products were purified using an Utraclean Gelspin kit (Mobio),
and sequenced using Big Dye Terminator v3.1 cycle sequencing
kit according to the manufacturer’s instructions. Each genome region
was amplified three times and PCR products were sequenced in
both directions.

Genetic comparisons

Full length Fr-AMPV-C and Fr-AMPV-D sequences were
assembled using vectorNTIv11 software, then aligned using
MEGA 5.2 [39] against available full genome sequences of
MPV downloaded from Genbank (four HMPV and 13 AMPV
genomes, see Acc No in Table 1). Open reading frames (ORFs)
were predicted and then compared with those reported previously
using MEGA 5.2. Program “getorf” from EMBOSS (emboss-
sourceforge.net) was used to detect potential ORFs which had
been defined as a region of at least 150 nucleotides between two
STOP codons. The amino acid [aa] sequences of these ORFs were
compiled in a database file. ORFs identified from Fr-AMPV-C
and Fr-AMPV-D were compared to all other MPV ORFs by local
BLAST [40] and were submitted to global BLAST search online.

Codon usage

The extent of codon bias was evaluated among the studied
MPV. To measure the general non-uniformity of the synonymous
codon usage, the effective number of codons (Nc) [41] was
calculated based on the longest MPV gene (L). Nc values range
from 20 when only one of the possible synonymous codons is used
for each amino acid, to 61 when all synonymous codons are used
equally. The closer the Nc value is to 20, the stronger the bias in
codon usage and the more non random codon usage is. It is
generally admitted that genes have a significant codon bias when
the Nc value is less than or equal to 35 [42].

In the Nc value calculation formula: Nc = 2+9/F2+1/F3+5/
F4+3/F6, F2 corresponds to the probability that two randomly
chosen codons for an amino acid, possibly encoded by two distinct
codons, are identical. F3 is the probability that three randomly
chosen codons for an amino acid with three synonymous codons
are identical and so on for F4 and F6. The Nc value was
determined using CodonW 1.4.4 (http://codonw.sourceforge.net)
and was correlated to the percentage of G+C at the third position
(GC3) as it has been shown previously to be a major factor
influencing the synonymous codon usage pattern in the HMPV
genome [43].

Phylogenetics. All available AMPV full-length genome
sequences and one representative of each of the four HMPV
sublineages were aligned using Clustal W. Alignments were also
checked manually for a good correspondence of the common
coding regions. Phylogenetic analysis was performed using MEGA
5.2 with the Neighbor-Joining method (1000 bootstrap replicates)
and the Kimura-2-parameter substitution model.

Results and Discussion

Sequence overview

The full length consensus sequences for Fr-AMPV-C and Fr-
AMPV-D were 14152bp- and 13415bp-long, respectively. Table
S1 presents the previously released sequences for these two viruses.
The present report provides newly determined sequences equal to
73 and 78% of the total genome sequence for these viruses,
respectively. The full length genomes were consistent in the order
(3’-leader-P-N-M-F-M2-SH-G-L-trailer-5’) and in the size of
known ORFs for MPV genomes (Table 1). Both sequences have
been submitted to EMBL (Accession numbers HG934338 and
HG934339 respectively). The Fr-AMPV-C and Fr-AMPV-D
genomes, like several other AMPVs, were found not to conform
to the “rule of six” [44], a feature that separates pneumovirinae
from paramyxovirinae [45]. In general, genome lengths were
conserved amongst AMPV subgroups A, B and D and amongst
HMPV sublineages, however clear differences could be seen in the
genome lengths of AMPV subgroup C viruses, mostly resulting
from the different lengths of their G genes (Table 1).

Phylogenetics

Three significant clusters were observed, one grouping all
HMPVs, a second grouping AMPV-Cs and a third grouping the
AMPV-A, B and D subgroups (Fig. 1).

Within the AMPV-C cluster, viruses isolated from Muscovy
ducks (SO1, GDY and 99178) formed a separate sub-lineage from
the others. This separation is potentially related to species rather
than geographic origin as the Asian SO1, GDY, PL-1 and PL-2
isolates were split into different clusters (SO1 and GDY with the
European 99178 isolate and PL-1 and 2 with the US isolates,
Fig. 1), although these geographical relationships could be blurred
if AMPV were shed by migratory birds, in the overlap between the
East Asian-Australasian flyway with both the East Atlantic and/
Pacific Americas flyway in the Northern hemisphere [46].

Nucleoprotein (N) ORF

The first ORF in the AMPV genome encodes the N protein,
which is a component of the polymerase complex and important
for the formation of the nucleocapsid helical structure [43,47].
The N ORF of Fr-AMPV-C was more closely related to that of
other AMPV-Cs and HMPVs in terms of length and aa
conservation than it was to AMPV-A, B or D.
| Origin | Accession number | Strain | Species | Host | Country | Start | End | S.1 | S.2 | S.3 | Genotype |
|--------|-----------------|--------|---------|------|---------|-------|-----|-----|-----|-----|----------|
| Duck   | KF654338        | France | Muscovy | Duck | France | 54    | 59  | 111 | 113 | 118 | 11412   |
| China  | KF350836        | China  | Muscovy | Duck | China  | 53    | 58  | 111 | 113 | 118 | 11412   |
| Duck   | KF354151        | China  | Muscovy | Duck | China  | <13   | <18 | 54  | 55  | 118 | >11414b |
| Duck   | DQ004884        | USA    | Goose   | Duck | USA    | 13    | 18  | 54  | 55  | 118 | >11407   |
| Mallard| DQ850753        | USA    | Goose   | Duck | USA    | 18    | 23  | 54  | 55  | 118 | >11407   |
| Duck   | AY536038        | Canada | Cackleduck | USA | Canada | 53    | 58  | 111 | 113 | 118 | 11415b |
| Turkey | AY272780        | USA    | Cackleduck | Turkey | USA | 53    | 58  | 111 | 113 | 118 | 11415b |
| Duck   | AY219777        | Korea  | Precinct | Duck | Korea | 53    | 58  | 111 | 113 | 118 | 11415b |
| Turkey | AY098317        | United Kingdom | Turkey | Turkey | United Kingdom | 55    | 59  | 111 | 113 | 118 | 11415b |
| Turkey | AY406145        | Italy   | Turkey | Turkey | Italy | 55    | 59  | 111 | 113 | 118 | 11415b |
| Turkey | B0162080        | France  | Turkey | Turkey | France | 62    | 67  | <13 | <18 | <18 | >13414a |
| Turkey | AY522489        | France  | Turkey | Turkey | France | 62    | 67  | <13 | <18 | <18 | >13414a |
| Turkey | AY522491        | France  | Turkey | Turkey | France | <13   | <18 | <13 | <18 | <18 | >13414a |

Species 2 and 3 refer to the second and third nucleotide changes, respectively, in the genome. The Genotype column indicates the full genome length, including the leader and trailer sequences respectively.

\( \text{doi:10.1371/journal.pone.0102740.t001} \)
Fr-AMPV-C, N ORF (394aa) was identical to previously reported AMPV-C N ORFs regardless of host species and to those of all HMPV sublineages (Table 1). High aa identity was observed with all other AMPV-C sequences (99%) and with all HMPV sublineages (89–90%) however, the identities with AMPV-A, B and D were lower (70, 71 and 73%, respectively). Two aa positions (44 and 137) were found to be specific of Fr-AMPV-C N compared to all other AMPV-Cs, notably these were the same amino acids as found in HMPV subgroups at this position (Fig. 2). Neither of these amino acids was within the three conserved regions identified previously among pneumoviruses (Barr et al 1991) and represented as boxes A–C in Fig. 2 the two latter (B and C) being merged in MPV forming a larger conserved domain B/C (aa 241–327 Fig. 2). Four separate regions (see grey shaded boxes in Fig. 2) were also highly conserved in all MPVs.

Subgroups A, B and D viruses were more closely related in terms of length and aa identity than they were to AMPV-Cs or HMPVs. Indeed, the length of the Fr-AMPV-D N protein (391aa) was identical to that of both AMPV-A and B and three amino acids shorter than AMPV-C and HMPV N proteins. Amino acid identities were high with A and B (89–90%) but lower with subgroup C and HMPV N proteins (71–74%). Localization of aa differences can be seen in Fig. 2, which also supports a relationship between Fr-AMPV-D and subgroups A and B.

Phosphoprotein (P) ORF

The second main ORF in the AMPV genome encodes the P protein, which is also part of the polymerase complex. Consistent with N protein comparisons, the P ORFs of AMPV-Cs and HMPVs were more closely related than they were to AMPV-A, B and D. In the same respect, subgroups A, B and D also demonstrated closer relationships.

The length of the Fr-AMPV-C, P ORF (294aa) was identical to other previously reported subgroup C P ORFs regardless of host species, and to those of all HMPV subgroups (Table 1). The P ORFs of AMPV-A, B and D were 16–17aa shorter (Table 1). The full length Fr-AMPV-C P sequence demonstrated a high aa conservation of 96–97% with all AMPV-C sequences, 67–68% with HMPV subgroups and 56%, 54% and 53% with AMPV-A, B and D respectively. Sequence conservation in the carboxy terminal half of the P protein (aa160–294) was notably higher for all the studied MPVs than it was in the amino terminal half (aa11–159). The carboxy terminal half has been reported to support most of the interactions with the N protein and polymerase complex, as reviewed by Easton et al., 2004 [45]. The high conservation of the P interaction domain between AMPV-C and HMPV is consistent with the finding that a recombinant chimeric HMPV with the P gene derived from AMPV-C was able to replicate in Vero cells [48,49].

In common with all subgroup C and HMPV P protein sequences analyzed previously [50], Fr-AMPV-C P lacked cysteine residues and maintained high conservation within the region (aa185–240) proposed to play a role in maintaining the structural integrity of the nucleocapsid complex [51]. More recently, this region in the HMPV P sequence has been shown to contain a short molecular recognition element (aa198–211) and a small domain (aa171–193) responsible for P tetramerization [52] (Fig. 3). In the later domain subgroup C viruses were fully conserved and
only differed at two amino acid positions from all HMPV sequences. Sequences of subgroups A, B and D in this domain were fully conserved but they differed at three and five amino acid positions from HMPVs and AMPV-Cs respectively (Fig. 3). Interestingly the first 14aa of the molecular recognition element were 100% conserved in all MPVs with the exception of strain SO1 which contained just one aa difference (Fig. 3 grey shaded box).

The length of the Fr-AMPV-D P protein (278aa) was identical to subgroup A and just one aa shorter than the P protein of subgroup B. Amino acid identities were 71–72% conserved with subgroups A and B. In contrast to subgroup C and HMPV P proteins, subgroups A, B and D contained cysteine residues (Fig. 3). Cysteine 56 was conserved in subgroups A, B and D and cysteine 64 was conserved between subgroups A and B. Subgroups A, B and D all differed extensively at the extreme C terminus of the P protein (Fig. 3).

Matrix protein (M) ORF

The third main ORF in the AMPV genome encodes the M protein which orchestrates the assembly of viral components at the plasma membrane, through interactions with the viral glycoproteins and nucleocapsid [53,54]. The length of the M ORF (254aa) was identical in all MPVs and its aa sequence was extremely conserved amongst all AMPV-Cs (99%) and between AMPV-A, B and D (90 to 94%). High aa conservation was also seen between AMPV-C and HMPV (87–88%) but to a lesser extent between AMPV-C and AMPV-A, B and D (78–79%). The hexapeptide (aa14–19) with no known function but conserved across all pneumoviruses [55] was also highly conserved in all MPVs, with the exception of Fr-AMPV-C that contained one conservative aa change (V→A) at position 18. Three cysteine residues (aa110, 147 and 239) were also conserved in all MPVs.

Fusion protein (F) ORF

The fourth main ORF in the AMPV genome encodes the highly antigenic, type I membrane fusion protein F. In paramyxoviruses, F is synthesized as an inactive single precursor F0, which is directed to the endoplasmic reticulum by its N-terminal signal peptide. F0 is then cleaved at an arginine-rich cleavage site, mostly by host endoproteases such as furins, into functional F1–F2 subunits held together by disulfide bonds. The F1 subunit remains inserted into the virus membrane by its carboxy-terminal transmembrane domain [56]. The F2 subunit of both HMPV and human and bovine RSV (HRSV, BRSV) has been reported to determine cellular host range [57,58]. The length of the Fr-AMPV-C, F ORF (537aa) was identical to previously reported AMPV-C F ORFs, one aa shorter than the F of AMPV-A, B and D, and two aa shorter than HMPV F (Table 1). The aa sequence of Fr-AMPV-C was extremely conserved (98–99%) with all other AMPV-Cs, highly conserved (91–92%) with all HMPV sublineages and slightly less conserved (71–73%) with AMPV-A, B and D. The cleavage site was located in the same position (aa 99–102) in all MPVs (grey box Clv in Fig. 4). The cleavage sequence (RRKR) conserved in all AMPV-Cs was not consistent with the typical furin cleavage (R-X-X-R/K-R) site found in AMPV-A, B and D (RRRR, RKKR and RQKR respectively), however a less typical recognition site (R-X-X-R) has also been shown to be functional [59].

The sequence of the signal peptide (aa 1–18 Fig. 4) at the N terminal end of F2 was extremely subgroup specific in the avian viruses, a rather surprising finding considering its function, with at best 39% identity between subgroups D and C however, a higher identity (56–61%) was seen between subgroup C and HMPV. Two cysteine residues (aa 28 and 60) (Fig. 4) remained conserved in all MPVs, including Fr-AMPV-C and D (with the exception of position 28 in the AY579780 APV/C0 sequence), which further supports their already suggested possible structural role [50].

In the F1 subunit, the fusion related domain (103–125) [60] was 100% conserved in all MPVs, with the exception of one aa change in AMPV-B. Other interesting conserved features in all MPV sequences included i) the position of the 12 extracellular cysteine residues (Fig. 4), a finding that is consistent with their possible involvement in protein secondary structure through the formation of disulfide bonds (Van den Hoogen et al., 2002), and ii) a proposed N-Linked glycosylation site (aa353–355, Fig. 4) [50]. Other features appeared subgroup or strain specific. For example, all AMPV-C F1 sequences contained a glycine residue (G) at amino acid 294 (Fig. 4), a position previously reported in HMPV to be influential in low pH-triggered fusion and syncitial phenotype [61,62], and in AMPV-C to contribute to the increased protective capacity of a genetically modified virus [63]. An integrin binding domain 329RGD331 (grey box Ibd in Fig. 4) has been identified in the F1 subdomain of the HMPV F protein, and changes to either of its first two residues have been shown to be detrimental for fusion activity [64]. No such typical RGD domain exists in the AMPV F protein; in contrast, all subgroup C sequences contained a motif 329RSD331 and subgroup A, B and D contained a motif 329RDD331. The subgroup specific modifications in this biologically significant domain also support the closer relationship between subgroups A, B and D.

The F1 cytoplasmic tail also exhibited inter subgroup variation (Fig. 4). Indeed, intra subgroup identities in this part of F1 were extremely high (subgroup C sequences 96–100%, subgroup A 100%, and HMPV 88–100%), whereas an extremely low conservation between subgroups was observed, with at best 56% between AMPV-B and D and as low as 0–4% between AMPV-A and HMPV. In spite of this low conservation, a TTG motif was conserved in the F1 cytoplasmic tails of AMPV-A, B and D (Fig. 4). The cytoplasmic tails of several paramyxovirus fusion proteins have been shown to be important in virus assembly [65].

Finally, several regions in the F1 subunits of pneumoviruses and MPVs are important in the production of neutralizing antibodies [66–69]. Brown et al 2009 [66] demonstrated that two regions (211–310 and 336–479) of the AMPV-A F protein were recognized by neutralizing antibodies to both subgroup A and B but not subgroup C virus. These regions appeared highly conserved (mean 95%) between AMPV-A, B and D but much less so with AMPV-C and HMPV (71–84%). Such identities are consistent with the lack of cross neutralization of AMPV-A, B and D with subgroup C viruses, and further suggest that neutralizing epitopes within regions 211–310 or 336–479 of AMPV-A and B are also likely to exist in subgroup D. These genetic data thus correlate with the previously reported antigenic cross-reactivity between AMPV subgroups [66,70].
Figure 3. Amino acid comparisons of the phosphoprotein ORF of AMPV-A, B, C and D and HMPV subgroups A and B. A region responsible for the tetramerization of the P protein is shown together with a molecular recognition element (MoRE). Grey box represents a highly conserved region in MPVs in the MoRE. End of protein.

doi:10.1371/journal.pone.0102740.g003
The M2 protein ORF

The M2 gene contains two overlapping ORFs (M2.1 and M2.2). M2.1 is involved in virus synthesis and enhances the processivity of the viral polymerase whilst M2.2 has been suggested to alter the balance between transcription and replication [45]. M2.2 has also been shown to be important for adaptation to Vero cells [71]. Fr-AMPV-C M2.1 was identical in length (184aa) to all other subgroup C sequences, however two, four and three aa shorter than AMPV-A and B, AMPV-D and HMPV sequences, respectively. Fr-AMPV-C amino acid identities were again highly conserved with all other subgroup C sequences (98%) and with all HMPV sublineages (84–85%), which is consistent with the finding that the polymerase complex proteins (M2.1, N, P and L) of either virus are biologically active in heterologous rescues [49,72]. Similarly, high identities (87–90%) were also observed between AMPV-A, B and D, however identities between subgroup C with AMPV-A, B and D (71–74%) were moderately lower. The three cysteine residues found in all pneumoviruses [50] within the first 30 aa of M2.1 remained conserved in both Fr-AMPV-C and D. M2.1 is intra cellular and

Figure 4. Amino acid comparisons of the fusion ORF of AMPV-A, B, C and D and HMPV subgroups A and B. Previously identified biologically important domains are labeled and underlined. Diamonds indicate conserved cysteine residues. Open boxes highlight other domains discussed in the relevant paragraph. Cleavage site (Clv) and integrin binding domain (Ibd). *end of protein.

doi:10.1371/journal.pone.0102740.g003

Figure 5. Amino acid comparisons of the previously reported conserved domain III (Poch et al., 1990) in the L ORF of AMPV-A, B, C and D and HMPV subgroups A and B. Four core motifs (A, B, C and D) described by Poch et al 1990 are underlined. A pentapeptide conserved in mononegavirales is highlighted in grey within motif C. Open boxes highlight regions in MPVs with 100% conservation. *end of protein.

doi:10.1371/journal.pone.0102740.g005
conservation of cysteines has been shown in RSV to be important for the formation of structural metal binding motifs [73,74]. In M2.2, a similar level of conservation was seen between subgroup C sequences (93–99%) but conservation was notably lower with the HMPV sublineages (54–58%) and with AMPV-A, B and D (20–24%) which was consistent with previous literature [50]. The highest inter subgroup identity was seen between AMPV-A, B and D (64–72%).

Three cysteine residues were conserved across all MPVs (aa7, 16 and 56) and a further two conserved between AMPV-A, B and D (aa22 and 59). Cysteines 7, 16 and 22 fell within a region (aa0–25) identified in HMPV as critical to promote viral gene transcription [75].

Small hydrophobic protein ORF (SH)

SH is a small type II membrane glycoprotein protein localized in the endoplasmic reticulum, golgi and cell surface [76]. SH has been shown to be non-essential for virus attachment, infectivity or virion assembly [29,77,78]. However, a SH deletion in AMPV-A contributed to an altered syncytial phenotype and a reduced immunogenicity [79].

SH length (175aa) was conserved across all AMPVs with the exception of subgroup B (180aa). This was in contrast to the varying lengths seen in HMPV SH (177–183aa). A range of 83–100% aa identity was seen between all subgroup C SH sequences. SH conservation between AMPV-A, B and D was considerably lower (42–49% aa identity), and even lower between HMPVs and all APMVs (14–31%). In the SH transmembrane domain, subgroup-C sequences demonstrated a closer relationship with HMPVs (39–50%) than they did with subgroups A, B or D (19–30%). Subgroups A, B and D were more closely related in their transmembrane domains (70–86%).

Further relationships between AMPV-C/HMPV or AMPV-A/B/D were evident in the conservation of cysteine residues. AMPV-A, B and D had fourteen (3 in the intracellular and 11 in the extra cellular domain) and AMPV-Cs and HMPVs had nine in the extracellular domain. Seven cysteines were conserved across all MPVs in the extracellular domain.

These features make the SH protein the second most variable protein (after the G gene) in the MPV genome, with respect to inter-subgroup aa identity. Interestingly inter-subgroup differences did not prevent the restoration of a typical phenotype when SHB was introduced into a SH-deleted AMPV-A genome background. A similar result could not be achieved using SHC [79].

Glycoprotein ORF (G)

G is a heavily glycosylated type II membrane protein, involved in, but not essential for virus attachment [29,80,81]. Most recently it is emerging as an inhibitor of the cellular host immune response to viral infection [82–84].

We have reported previously genetic analysis of the large G ORF in Fr-AMPV-C [24] and D [24,35]. Both studies showed that G exhibited the most extensive divergence between subgroups in terms of length and sequence identity. Differences in the length of the G protein ectodomain amongst AMPV-C isolates have been also reported [18,85–87]. In the present study, the length of both Chinese AMPV-C G sequences were identical to that of Fr-AMPV-C (585aa), whilst both Korean AMPV-C G sequences were shorter (264aa) and more closely resembled AMPV/CO (table 1). Intra subgroup C identities including the two Chinese and two Korean AMPV-C G sequences were within the range reported previously (75–83%) [24]. The two pairs of Asian AMPV-C G sequences were highly conserved (intra pair identity = 97 and 99.6%, respectively) and the four viruses exhibited the conserved intracellular and trans membrane domains and the ten extracellular cysteine residues previously reported to be conserved in all AMPV-Cs [24]). Remarkably, 19 out of 22 aa differences between the two Chinese sequences were confined to a short domain (aa269–299) immediately at the N terminal end of the previously identified, variable part of G ectodomain.
The polymerase protein ORF (L)

The final ORF of metapneumovirus genomes encode the large RNA-dependent RNA polymerase protein L. It is a major part of the polymerase complex responsible for most of the enzymatic processes involved in transcription and replication [88]. It is also responsible for viral messenger RNA capping, polyadenylation, methylation and phosphorylation processes [89].

The length of Fr-AMPV-C L (2005aa) was identical to all other AMPV-C and HMPV sequences, one aa shorter than that of AMPV-A and B and two shorter than that of Fr-AMPV-D. Extremely high aa conservation was observed amongst subgroup-C viruses (98–100%), and amongst HMPV sublineages (94–99%). Closer relationships were observed between subgroup-C viruses and HMPVs (80–81%) than between these viruses and subgroups A, B, D (63–64%). Subgroup D again demonstrated a closer relationship with subgroups A and B (84–86%).

Six functional domains (I–VI) have been identified in the L proteins of non segmented negative strand viruses [88], with domain III including four highly conserved core polymerase motifs (A–D) [50] The newly identified sequences were consistent with these findings in motifs A, B and C, however some variation was seen in motif D (Fig. 5) and motifs A and C appeared to be larger in MPVs (Fig. 5). Two additional regions were observed where all MPVs were completely conserved (Fig. 5). The QGDNQ pentapeptide found in motif C within domain III was replaced in all MPVs by NGDNQ. AMPV-A, B and D shared four or five amino acids in motif D that were not represented in the subgroup C viruses or HMPVs (Fig. 5). Further conservation was observed between all MPVs in the ATP-binding motif (aa 1677–1721) identified previously [50] and in five previously unidentified regions scattered through the L ORF were all MPVs were 100% conserved over 15 or more aa (aa15–29, aa549–573, aa656–670.

### Table 2. AMPV transcription start and stop sequences.

| gene | AMPV subgroup | transcription start sequence | transcription stop sequence |
|------|---------------|-----------------------------|----------------------------|
| N    | A             | GGGACAAGT                   | AGTTA(Xn)Poly A            |
|      | B             | ...A.(Xn)Poly A              |                            |
|      | C             | ...A.(Xn)Poly A              |                            |
|      | D             | ...A.(Xn)Poly A              |                            |
| P    | A             | ...Xn)Poly A                |                            |
|      | B             | ...Xn)Poly A                |                            |
|      | C             | ...Xn)Poly A                |                            |
|      | D             | ...Xn)Poly A                |                            |
| M    | A             | ...Xn)Poly A                |                            |
|      | B             | ...Xn)Poly A                |                            |
|      | C             | ...Xn)Poly A                |                            |
|      | D             | ...Xn)Poly A                |                            |
| F    | A             | ...Xn)Poly A                |                            |
|      | B             | ...G...Xn)Poly A            |                            |
|      | C             | ...G...Xn)Poly A            |                            |
|      | D             | ...Xn)Poly A                |                            |
| M2   | A             | ...Xn)Poly A                |                            |
|      | B             | ...Xn)Poly A                |                            |
|      | C             | ...Xn)Poly A                |                            |
|      | D             | ...Xn)Poly A                |                            |
| SH   | A             | ...Xn)Poly A                |                            |
|      | B             | ...G...Xn)Poly A            |                            |
|      | C             | ...G...Xn)Poly A            |                            |
|      | D             | ...Xn)Poly A                |                            |
| G    | A             | ...Xn)Poly A                |                            |
|      | B             | ...C.(Xn)Poly A             |                            |
|      | C             | ...Xn)Poly A                |                            |
|      | D             | ...Xn)Poly A                |                            |
| L    | A             | A...C.A.                    | ...Xn)Poly A              |
|      | B             | G...C.A.                    | ...Xn)Poly A              |
|      | C             | A...C.AG                    | ...Xn)Poly A              |
|      | D             | ...C...                    | ...Xn)Poly A              |

Sequences in bold represent consensus sequence.

doi:10.1371/journal.pone.0102740.t002

Metapneumovirus Molecular Comparisons and Subclassification
aa1250–1265 and aa1297–1319). Finally two regions of subgroup specific sequences were observed towards the N terminal end of the L protein (302–320, and 431–446).

Codon usage in the L gene

Different groups were revealed by the codon bias analysis i) HMPVs (Nc = 41.9 to 43.2) ii) AMPV-Cs (47.1 and 47.3) and iii) AMPV-A, B, D (51.4 to 52.7) (Fig. 6). Data points were close to the curve (expected value of Nc if the bias was solely due to the GrC content at the third position) demonstrating that the biases were mostly due to the GC content. Interestingly, AMPV-C and HMPV demonstrated a different codon bias profile, indicating that many of their proteins shared high aa similarity, a feature that most probably reflects their adaptation to a specific host. Another striking aspect of the codon bias study was that AMPV-A, B and D had a very similar codon bias, although the genetic distances were important between these viruses (Fig. 1). It is not known whether the bias picture would change if all protein sequences in the full length genome were used, however, this has been performed for HMPV and resulted in a very similar bias [43].

Non coding regions, intergenic regions and leader and trailer sequences

The numbers of nt between two consecutive ORFs (thus a sequence encompassing the 3‘NCR of previous gene, intergenic region and 5‘NCR of subsequent gene) in Fr-AMPV-C were consistent with other subgroup Cs (Table 1). The smallest (21 nt) was between M2-2/SH and the largest (187 nt) between SH/G. The only notable differences in lengths occurred between G/L of AMPV-C strains 2a/97, PL1 and PL-2 (Table 1), where these three strains exhibited 28 nts, 16 nts less than all other AMPV-Cs. The numbers of nt between consecutive ORFs in subgroups A, B and D were more similar than they were to subgroup Cs or HMPVs (Table 1). Although the typical AMPV gene start signal GGGACAAGT and gene stop signal AGTTA(Xn)Poly A [13,90–92] were mostly conserved amongst AMPV subgroups, some differences were observed (Table 2).

We have previously described the 3 ‘ and 5 ‘ sequence extremities of Fr-AMPV-C [34] discussing. Here complete leader and trailer sequences showed varying levels of conservation amongst all MPVs (67–97.5%). The highest level of conservation was seen between the leader and trailer sequences of subgroup C viruses and HMPVs (79–85%). This was consistent with the heterologous rescue of AMPV-C and HMPV minigenomes using different polymerase complexes [49,72]. Remarkably, subgroup D had a leader sequence of 62 nt which was 7 nt longer than any MPV leader sequence reported to date.

Conclusion

This study provides the full length genome sequences for two new AMPV strains including the first full length sequence for AMPV subgroup D. Results supported previous reports that AMPV-C viruses are indeed more closely related with HMPVs than they are with other AMPV subgroups, and further demonstrate that AMPV-D is more closely related with the AMPV-A and B subgroups. Ideally, this study might be extended by sequencing more AMPV-D isolates. Unfortunately only two such isolates are currently available worldwide, both isolated in France, on the same date and within close proximity, consequently efforts to obtain new AMPV-D isolates should be continued. The three MPV “clusters” HMPV, AMPV-C and AMPV-ABD were also further supported based on phylogenetics, sequence comparisons and codon bias studies.

These data combined with those of previous reports indicating antigenic relationships between subgroups A, B and D [20,70] and between subgroup C and HMPV [93] may call for a sub classification of MPVs comparable to that implemented for avian paramyxovirus, where viruses are first grouped into serotypes (type number) then separated into genotypes [94,95]. Transposing a similar approach into the MPV genus would result in grouping AMPV-A, B and D as type I MPVs and AMPV-Cs and HPMVs as type II.

Supporting Information

Table S1 Partial nucleotide sequences previously released for Fr-AMPV-C and Fr-AMPV-D. (DOCX)

Acknowledgments

The authors gratefully acknowledge the Conseil Régional de Bretagne and of Département des Côtes d’Armor.

Author Contributions

Conceived and designed the experiments: PB NE. Performed the experiments: EL CC CA PB DT MHBA. Analyzed the data: PB FXB EL CC OG NE. Contributed to the writing of the manuscript: PB NE VJ.

References

1. Buys SB, Du Prees JH (1980) A preliminary report on the isolation of a virus causing sinusitis in turkeys in South Africa and attempts to attenuate the virus. Turkeys 28: 56.
2. Giraud P, Bennejean G, Guittet M, Toquin D (1986) Turkey rhinotracheitis in France: preliminary investigations on a ciliostatic virus. Vet Rec 119: 606–607.
3. McDougall JS, Cook JK (1986) Turkey rhinotracheitis: preliminary investigations. Vet Rec 118: 206–207.
4. Wilding GP, Baxter-Jones C, Grant M (1986) Ciliostatic agent found in rhinotracheitis. Vet Rec 118: 755.
5. Wood VT, Gough RE, Chettle N, Eddy R (1986) Preliminary observations on a virus associated with turkey rhinotracheitis. Vet Rec 119: 139.
6. Cook JK (2000) Avian pneumovirus infections of turkeys and chickens. Vet J 160: 118–125.
7. van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, et al. (2001) A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nat Med 7: 719–724.
8. de Graaf M, Ostertaus AD, Fouchier RA, Holmes EC (2000) Evolutionary dynamics of human and avian metapneumoviruses. J Gen Virol 89: 2933–2942.
9. Crowe JE, Jr. (2004) Human metapneumovirus as a major cause of human respiratory tract disease. Pediatr Infect Dis J 23: S215–221.
10. Faisley AR, Erdman D, Andreason LJ, Welsh EE (2003) Human metapneumovirus infections in young and elderly adults. J Infect Dis 187: 783–790.
11. Williams JV, Wang CK, Yang CF, Tollefon SJ, House FS, et al. (2006) The role of human metapneumovirus in upper respiratory tract infections in children: a 20-year experience. J Infect Dis 193: 387–393.
12. Pringle CR (1998) Virus taxonomy-San Diego 1998. Arch Virol 143: 1449–1455.
13. Ling R, Easton AJ, Pringle CR (1992) Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene order different from that of other pneumoviruses. J Gen Virol 73 (Pt 7): 1709–1715.
14. Randhawa JS, Marriott AC, Pringle CR, Easton AJ (1997) Rescue of synthetic mimivirus establishes the absence of the NS1 and NS2 genes from avian pneumovirus. J Virol 71: 9049–9054.
15. Govindarajan D, Samal SK (2004) 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 Res 105: 59–66.
16. Govindarajan D, Samal SK (2005) Analysis of the complete genome sequence of avian metapneumovirus subgroup C indicates that it possesses the longest genome among metapneumoviruses. Virus Genes 30: 331–333.
43. Zhong Q, Xu W, Wu Y, Xu H (2012) Patterns of synonymous codon usage on
41. Wright F (1990) The ‘effective number of codons’ used in a gene. Gene 87: 23–
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local
39. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5:
37. Toquin D, Guionie O, Allee C, Morin Y, Le CoQ L, et al. (2006) Compared
35. Bayon-Auboyer MH, Arnauld C, Toquin D, Eterradossi N (2000) Nucleotide
29. Naylor CJ, Brown PA, Edworthy N, Ling R, Jones RC, et al. (2004)
25. Turpin EA, Stallknecht DE, Slemons RD, Zsak L, Swayne DE (2008) Evidence
23. Shin HJ, Ngjmi MK, McCombs B, Halvorson DA, Nagaraja KV (2000) Avian
19. Yunus AS, Govindarajan D, Huang Z, Samal SK (2003) Deduced amino acid
18. de Graaf M, Herfst S, Verelst J, Osterhaus AD, et al. (2008) Respiratory syncytial virus (RSV) fusion subunit F2, not attachment protein G, determines the specificity of RSV infection. J Virol 77: 4609–4616.
17. Bayon-Auboyer MH, Arnould C, Toquin D, Eterradossi N (2000) Nucleotide sequences of the F, L and G protein genes of two non-A/non-B avian pneumoviruses (AVP). J Gen Virol 81: 2723–2733.
16. Bayon-Auboyer MH, 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.
15. Bayon-Auboyer MH, Eterradossi N, Njenga MK, McComb B, Halvorson DA, Nagaraja KV (2000) Avian pneumovirus (AVP) RNA from wild and sentinel birds in the United States has genetic homology with RNA from AVP isolates from domestic turkeys. J Clin Microbiol 38: 4282–4284.
14. Toquin D, Guionie O, Jestin V, Zwingelstein F, Aller C, et al. (2006) European and American subgroup C isolates of avian metapneumovirus belong to different genetic lineages. Virus Genes 32: 97–103.
13. Turpin EA, Stahlknecht DE, Slemons RD, Zsak L, Swayne DE (2008) Evidence of avian metapneumovirus spread beyond Minnesota among wild and domestic birds in central North America. Avian Dis 52: 902–908.
12. Cook JK, Huggins MB, Orbell SJ, Scne DA (1999) Preliminary antigenic characterization of an avian metapneumovirus isolated from commercial turkeys in Colorado, USA. Avian Pathology 28: 607–617.
11. Shi H, Njenga MK, McCombs B, Halvorson DA, Nagaraja KV (2000) Avian pneumovirus (AVP) RNA from wild and sentinel birds in the United States has genetic homology with RNA from AVP isolates from domestic turkeys. J Clin Microbiol 38: 4282–4284.
10. Toquin D, Bayon-Auboyer MH, Eterradossi N, Jestin V (1999) Isolation of a pneumovirus from a Muscovy duck. Vet Rec: 13, 451, 608.
9. Bennett RS, Neovziorski J, Velayuddhan BT, Nagaraja KV, Zeman DH, et al. (2004) Evidence of avian metapneumovirus spread beyond Minnesota among wild and domestic birds in central North America. Avian Dis 52: 902–908.
8. Weiner SL, Yoon S, Wang J, Zhang C, et al. (2013) Avian metapneumovirus subgroup C infection in chickens, China. Emerg Infect Dis 19: 1092–1094.
7. Catelli E, Cecchinato M, Savage CE, Jones RC, Naylor CJ (2006) Demonstration of comparable rates of attenuation of the field and vaccine avian metapneumovirus vaccine. Vaccine 24: 6476–6482.
6. Lupini C, Cecchinato M, Riczzi M, Naylor CJ, Catelli E (2011) A turkey rhinotracheitis outbreak caused by the environmental spread of a vaccine-derived avian metapneumovirus. Avian Pathol 40: 325–330.
5. Naylor CJ, Brown PA, Edworthy N, Ling R, Jones RC, et al. (2004) Development of a reverse-genetics system for Avian pneumovirus demonstrates that the small hydrophobic (SH) and attachment (G) genes are not essential for virus viability. J Virol 83: 3219–3237.
4. Sugiyama M, Ito H, Hata Y, Ono E, Ito T (2010) Complete nucleotide sequences of avian metapneumovirus subgroup B genome. Virus Genes 41: 389–395.
3. Liedberg HC, Alvarez R, Winge MG, Yu Q, Halvorson D, et al. (2005) Comparison of the full-length genome sequence of avian metapneumovirus subgroup C with other paramyxoviruses. Virus Res 107: 83–92.
2. Bennett RS, LaRue R, Shaw D, Yu Q, Nagaraja KV, et al. (2005) A wild goose metapneumovirus containing a large attachment glycoprotein is avirulent but immunoprotective in domestic turkeys. J Virol 79: 14034–14042.
1. Lee E, Song MS, Shin JY, Lee YM, Kim CJ, et al. (2007) Genetic characterization of avian metapneumovirus subgroup C isolated from pheasants and a quail-origin bird market. Virus Res 128: 195–205.
Brown PA, Briend FX, Guionie O, Lemaire E, Courtillot C, et al. (2013) An alternative method to determine the 5′ extremes of non-segmented, negative sense RNA viral genomes using positive replication intermediate 5′ tailing: Application to two members of the Paramyxoviridae family. J Virol Methods 193: 121–127.
71. Chubbie J, Naylor CJ (2011) Avian metapneumovirus M2-2 protein inhibits replication in Vero cells: modification facilitates live vaccine development. Vaccine 29: 9493–9498.
72. Govindarajan D, Buchholz UJ, Samal SK (2006) Recovery of avian metapneumovirus subgroup C from cDNA; cross-recognition of avian and human metapneumovirus support proteins. J Virol 80: 5790–5797.
73. Esperante SA, Noval MG, Alhieri TA, de Oliveira GA, Silva JL, et al. (2013) Fine modulation of the respiratory syncytial virus M2-1 protein quaternary structure by reversible zinc removal from its Cys(3)-His(1) motif. Biochemistry 52: 6779–6789.
74. Tanner SJ, Ariza A, Richard CA, Kyle HF, Dods RL, et al. (2014) Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. Proc Natl Acad Sci U S A 111: 1580–1585.
75. Ren J, Wang Q, Kolli D, Prusak DJ, Tseng CT, et al. (2012) Human metapneumovirus M2-2 protein inhibits innate cellular signaling by targeting MAVS. J Virol 86: 13049–13061.
76. Deng Q, Weng Y, Lu W, Demers A, Song M, et al. (2011) Topology and cellular localization of the small hydrophobic protein of avian metapneumovirus. Virus Res 160: 102–107.
77. de Graaf M, Herfst S, Aarhion J, Burgers PC, Zaaaraoui-Boutahar F, et al. (2013) Small hydrophobic protein of human metapneumovirus does not affect virus replication and host gene expression in vitro. PLoS One 8: e38572.
78. Karron RA, Bumaaguir DA, Georgiou AE, Whitehead SS, Adams J, et al. (1997) Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vivo: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. Proc Natl Acad Sci U S A 94: 13961–13966.
79. Li J, Ling R, Randhawa JS, Shaw K, Davis PJ, et al. (1996) Sequence of the nucleosidase protein gene of subgroup A and B avian pneumoviruses. Virus Res 41: 185–191.
80. Yu Q, Davis PJ, Brown TD, Cavanagh D, Velayudhan B, et al. (1992) Sequence and in vitro expression of the M2 gene of turkey rhinotracheitis virus. J Virol 73 (Pt 6): 1355–1363.
81. Yu Q, Davis PJ, Li J, Cavanagh D (1992) Cloning and sequencing of the matrix protein gene of turkey rhinotracheitis virus reveal a gene order different from that of respiratory syncytial virus. Virology 186: 426–434.
82. Alvarez R, Jones LP, Seal BS, Kapczynski DR, Tripp RA (2004) Serological cross-reactivity of members of the Metapneumovirus genus. Virus Res 105: 67–73.
83. de Almeida RS, Hammoumi S, Gil P, Briand FX, Molia S, et al. (2013) New avian paramyxoviruses type I strains identified in Africa provide new outcomes for phylogeny reconstruction and genotype classification. PLoS One 8: e76413.
84. Alonso S, Collins PL, et al. (2005) Infection of nonhuman primates with recombinant human metapneumovirus lacking the SH, G, or M2-2 protein catagorizes each as a nonessential accessory protein and identifies vaccine candidates. J Virol 79: 12609–12613.
85. Cox RG, Livesay SB, Johnson M, Ohi MD, Williams JV (2012) The human metapneumovirus fusion protein mediates entry via an interaction with RGD-binding integrins. J Virol 86: 12148–12160.
86. Tao X, Kolli D, Ren J, Liu T, Garofalo RP, et al. (2013) Human metapneumovirus glycoprotein G disrupts mitochondrial signaling in airway epithelial cells. PLoS One 8: e62568.