There are amendments to this paper

OPEN

Molecular evolution and genetic variations of V and W proteins derived by RNA editing in Avian Paramyxoviruses

Pachineella Lakshmana Rao, Ravi Kumar Gandham & Madhuri Subbiah ✉

The newly assigned subfamily *Avulavirinae* in the family *Paramyxoviridae* includes avian paramyxoviruses (APMVs) isolated from a wide variety of avian species across the globe. Till date, 21 species of APMVs are reported and their complete genome sequences are available in GenBank. The APMV genome comprises of a single stranded, negative sense, non-segmented RNA comprising six transcriptional units (except APMV-6 with seven units) each coding for a structural protein. Additionally, by co-transcriptional RNA editing of phosphoprotein (*P*) gene, two mRNAs coding for accessory viral proteins, V and W, are generated along with unedited P mRNA. However, in APMV-11, the unedited mRNA codes for V protein while +2 edited mRNA translates to P protein, similar to members of subfamily *Rubulavirinae* in the same family. Such RNA editing in paramyxoviruses enables maximizing the coding capacity of their smaller genome. The three proteins of *P* gene: P, V and W, share identical N terminal but varied C terminal sequences that contribute to their unique functions. Here, we analyzed the *P* gene editing site, V and W sequences of all 21 APMV species known so far (55 viruses) by using bioinformatics and report their genetic variations and molecular evolution. The variations observed in the sequence and hexamer phase positions of the *P* gene editing sites is likely to influence the levels and relative proportions of P, V and W proteins' expressions which could explain the differences in the pathogenicity of APMVs. The V protein sequences of APMVs had conserved motifs similar to V proteins of other paramyxoviruses including the seven cysteine residues involved in MDA5 interference, STAT1 degradation and interferon antagonism. Conversely, W protein sequences of APMVs were distinct. High sequence homology was observed in both V and W proteins between strains of the same species than between species except in APMV-3 which was the most divergent APMV species. The estimates of synonymous and non-synonymous substitution rates suggested negative selection pressure on the V and W proteins within species indicating their low evolution rate. The molecular clock analysis revealed higher conservation of V protein sequence compared to W protein indicating the important role played by V protein in viral replication, pathogenesis and immune evasion. However, we speculate the genetic diversity of W proteins could impact the degree of pathogenesis, variable interferon antagonistic activity and the wide host range exhibited by APMV species. Phylogenetically, V proteins of APMVs clustered into three groups similar to the recent classification of APMVs into three new genera while no such pattern could be deciphered in the analysis of W proteins except that strains of same species grouped together. This is the first comprehensive study describing in detail the genetic variations and the molecular evolution of *P* gene edited, accessory viral proteins of Avian paramyxoviruses.

Avian paramyxoviruses (APMVs) are a group of paramyxoviruses known to infect a variety of bird species across the globe. Till date, 21 species of APMVs have been identified and the list is expected to grow with increase in viral surveillance in wild and domesticated birds. With recent ICTV 2019 classification, these viruses now belong to three genera, *Metaavulavirus* (APMV-2, -5, -6, -7, -8, -10, -11, -14, -15 and -20), *Orthoavulavirus* (APMV-1, -9, -12, -13, -16, APV-A, APV-B and APV-C) and *Paraavulavirus* (APMV-3 and -4) within the new subfamily

National Institute of Animal Biotechnology, Hyderabad, 500032, Telangana, India. ✉e-mail: madhuri@niab.org.in

Avulavirinae under family *Paramyxoviridae* in the order *Mononegavirales*[1]. APMV-1 to -9 were isolated before 1980, APMV-10 to -13 were identified by 2015 and all the other APMVs were reported in the recent years[2].

Avian Paramyxoviruses are enveloped with a single stranded, non-segmented, negative sense RNA genome of size 13 to 17 kb[2]. The prototype virus, APMV-1 of genus *Orthoavulavirus*, also well known as Newcastle disease virus or NDV, is the most extensively studied virus in this group. NDV causes severe economically important disease in poultry. There are five pathotypes of NDV based on the clinical signs exhibited by infected chickens: (a) viscerotropic velogenic or highly virulent, pantropic NDV causing severe mortality (b) neurotropic velogenic or highly virulent NDV, specifically causing neurological illnesses and high mortality (c) mesogenic or moderately virulent NDV, with mortalities as high as 50% and reducing egg production (d) lentogenic–either respiratory or enteric type NDV, low virulence and causing low reduction in egg production and (e) asymptomatic or avirulent NDV. However, the pathotype classification is not always clear-cut[3]. There are more than thousand strains of NDV which have been isolated, sequenced and found to exhibit wide spectrum of virulence. The viral RNA genome encodes for six genes arranged in tandem, each coding for six structural proteins, 3'-N-P-M-F-HN-L-5'[4,5]. N is nucleocapsid protein, each N protomer is known to bind exactly 6 nucleotides of genomic and antigenomic RNA of most paramyxoviruses thus imposing a hexamer phase on the entire RNA genome. In nature, the genomic length of paramyxoviruses is polyhexameric ($6n + 0$) which is found to be necessary for efficient replication and this is called the 'rule of six'[6,7]. N together with P, phosphoprotein and L, large polymerase protein, forms viral RNA dependent RNA polymerase complex essential for viral genome transcription and replication; M, matrix protein, is seen within the envelope, aids in virus assembly and budding; two viral glycoproteins, F, fusion protein and HN, hemagglutinin-neuraminidase protein, are studded on the envelope and assist with fusion of virus with host membrane and receptor binding, respectively[4]. NDV F protein is known as the virulence determinant; the virulent strains have unique multiple basic amino acids, at least three arginine (R) or lysine (K) residues, at fusion protein cleavage site starting at amino acid position 113, and a phenylalanine residue at position 117[3]. APMV-6 is also known to express an additional small hydrophobic (SH) protein from *SH* gene located between *F* and *HN* genes[8]. Further, by co-transcriptional RNA editing of *P* gene, two mRNAs, V and W are expressed[4,9–12]. Also, in certain paramyxoviruses, by a process of alternative transcription initiation in *P* gene (+1 reading frame), accessory C proteins are generated[13,14]. Thus by these mechanisms, paramyxoviruses are able to efficiently utilize over 95% of their small RNA genome for expression of viral proteins[12].

The *P* gene carries a slippery sequence, a stretch of adenosine (A) nucleotides and guanosine (G) nucleotides called the 'editing site' where insertions of 1 G or 2 G nucleotides occur during transcription of *P* gene by the stuttering viral polymerase that reiteratively reads the template base[12,15]. A single G nucleotide addition leads to +1 frameshift in the ORF, generating V mRNA with a frequency of 25 to 35% and two G nucleotides addition leads to +2 frameshift in the ORF generating W mRNA with a frequency of 2 to 8.5% and the unedited mRNA (60-70%) codes for P protein in NDV[12,16]. Among the paramyxoviruses, members of the subfamily *Rubulavirinae* and APMV-11 of genus *Metaavulavirus*, encode V protein from their unedited transcript, while P protein is coded by +2 frameshift and W protein is expressed by +1 frameshift[17,18]. The resulting three mRNAs from *P* gene (P, V and W mRNAs) share common N terminal sequences and differ both in length and amino acid composition in their C terminal region. Their specific functions are dictated by their unique C terminal sequences. Studies on V protein of APMV-1 and other paramyxoviruses have revealed that V protein is multifunctional, targets STAT1 degradation, interferes with MDA5, is interferon antagonist[19–22], inhibits apoptosis[23,24], assist in viral replication[20,25] and plays important roles in tissue tropism, virulence determination[11,20,26] and host range restriction[21]. On the other hand, there is very limited information about the function of W protein. The W protein of Nipah virus has been shown to impact viral pathogenesis and support the virus to evade the host immunity[27–29]. In APMV-1, the nuclear localization of W protein and its incorporation into the virion has been recently reported[16,30].

The complete genome sequences of all 21 species of APMV have been described individually[17,31–62]. A comprehensive comparative analysis of complete genome and structural genes of 20 species of Avulaviruses has been recently published[2]. Three clades of viruses were concluded based on phylogenomic analysis of 20 APMVs: Clade I included APMV-2, -5, -6, -7, -8, -10, -11, -14, -15 and -20 (currently classified under new genus *Metaavulavirus*), clade II comprised of APMV -1, -9, -12, -13, -16, APV-A, -B and -C (currently assigned under genus *Orthoavulavirus*) and clade III included APMV-3 and -4 (currently under new genus *Paraavulavirus*)[1,2]. One of the viruses, previously classified as APMV-17 (South Korean "avian paramyxovirus 17") has now been proposed as a separate species (APMV-21) based on phylogenies of complete genomes, complete *F* and *L* genes, PASC and STD analysis[2,63]. Nevertheless, very little is known about the *P* gene edited accessory viral proteins of APMVs. We have examined and analyzed 55 viruses belonging to all 21 APMV species identified till date and discuss here the genetic diversity and molecular evolution of *P* gene edited proteins, V and W.

Materials and methods

Sequence information. The full length sequences of *P* gene available for all the 21 species of APMVs were obtained from National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Additionally, publications reporting the complete genome sequence of these viruses were also referred for identification of *P* gene editing site, for prediction of sequences of V and W proteins. A total of 55 viruses belonging to 21 species of APMVs were analyzed in this study which included (their GenBank accession numbers are provided in Table 1) four strains each of avirulent APMV-1, moderately virulent APMV-1 and highly virulent APMV-1; eight isolates of APMV-6; five isolates each of APMV-2 and APMV-8; four isolates of APMV-10; three isolates of APMV-13; two isolates each of APMV-3, APMV-4, APMV-5; one isolate each of APMV-7, APMV-9, APMV-11, APMV-12, APMV-14, APMV-15, APMV-16, APMV-20, APMV-21 and APV-A, APV-B and APV-C. Detailed information of these isolates along with metadata such as hosts, year and location of isolation in addition to their sequence information are provided in Table 1. The complete sequences of *P* gene ORF, V proteins and W proteins

S.No	APMV species		Genbank Id	Year of Isolation	Place of isolation	Isolated from	P ORF (Nt)/P Protein length (aa)	P gene 'RNA editing site' sequence (numbers indicate the position of nucleotide in the genome, editing site is bolded and hexamer phase of the start of the short C run in edit site is underlined)	Subunit hexamer phasing at the P gene mRNA editing site	V ORF (Nt)/V Protein length (aa)	W ORF (Nt)/W Protein length (aa)	No. of unique amino acids in the C terminal region of W protein
1	APMV-1 (NDV)	Velo-genic	KJ808820.1	2012	China	Pigeon	1188 /395	2283CGAU **UUUUCC** CCGG	5	720/239	684/227	94
2			FJ386394.2	2009	China	Chicken	1188 /395	2277CGAU **UUUUCC** CGGG	5	720/239	666/221	88
3			GU187941.1	1987	India	Chicken	1188/395	2277CGAU **UUUUCC** CGGG	5	720/239	531/196	62
4			JQ015296.1	2011	China	Chicken	1188/395	2283CGAU **UUUUCC** CGGG	5	720/239	534/177	43
5		Meso-genic	KX761866.1	2010	China	Duck	1188/395	2277CGAU **UUUUCC** CGGG	5	720/239	684/227	94
6			KT445901.1	1945	Palestine	Duck	1188/395	2277CGAU **UUUUCC** CGGG	5	720/239	666/221	88
7			AY562986.1	1993	USA	Anhinga	1188/395	2283UGAU **UUUUCC** CGGG	5	720/239	540/179	45
8			KJ736742.1	1998	Germany	Pigeon	1188/395	2283GAU **UUUUCC** CCGG	5	720/239	414/137	3
9		Lento-genic	KM885162.1	2013	China	Duck	1200/399	2277CGAU **UUUUCC** CGGG	5	738/245	552/183	49
10			JN688862.1	2009	China	Chicken	1200/399	2277CGAU **UUUUCC** CGGG	5	738/245	552/183	49
11			JX193078.1	2009	China	Duck	1188/395	2277CGAU **UUUUCC** CGGG	5	720/239	540/179	45
12			AB524405.1	1991	Alaska	Goose	1200/399	2277CGAU **UUUUCC** CGGG	5	738/245	552/183	49
13	APMV - 2		HM159993.1	2006	England	Chicken	1200/399	2089CAAUUU **UUCCCC** UU	3	699/232	624/207	67
14			HQ896023.1	1999	China	Gallus gallus	1200/399	2089CAAUUU **UUCCCC** UU	3	699/232	624/207	67
15			EU338414.1	1956	USA	Chicken	1200/399	2089CAAUUU **UUCCCC** UU	3	699/232	624/207	67
16			HM159994.1	1980	Kenya	Gadwell	1200/399	2089CAAUUU **UUCCCC** UU	3	699/232	624/207	67
17			HQ896024.1	2001	China	Gallus gallus	1200/399	2089CAAUUU **UUCCCC** UU	3	699/232	624/207	67
18	APMV - 3		EU403085.1	1975	Netherland	PKT	1161/386	2078AAAAA **UUUCCC** CCG	4	747/248	378/125	14
19			EU782025.1	1968	USA	Turkey	1173/390	2083AAAAAU **UUCCCC** CG	3	759/252	384/127	12
20	APMV - 4		FJ177514.1	1975	China	Duck	1182/393	2052G AAAAUU **UCCCCC** C	2	675/224	414/137	2
21			JX133079.1	2010	South Africa	Alopochen aegyptiacus	1182/393	2052G AAAAUU **UCCCCC** C	2	675/224	414/137	2
22	APMV - 5		GU206351.1	1974	Japan	Budgerigar	1341/446	2359AAUUUU **UUCCCG** GG	3	834/277	564/187	5
23			LC168750.1	1975	Japan	Melopsitt-acus undulatus	1341/446	2357GU AAUUUU **UUCCGG** G	3	834/277	564/187	5
24	APMV - 6		JX522537.1	2011	China	Anas poecilorhyn-cha	1293/430	2146AGU **UUUUUC** CCCUU	6	825/274	489/162	4
25			KP762799.1	2013	Kazakhstan	Red-crested pochard	1293/430	2146AGU **UUUUUC** CCCUU	6	825/274	489/162	5
26			EU622637.2	1977	Hong Kong	Duck	1293/430	2146AGU **UUUUUC** CCCUU	6	807/268	534/177	5
27			AY029299.1	1998	Taiwan	Duck	1293/430	2146AGU **UUUUUC** CCCUU	6	807/268	534/177	20
28			KF267717.1	2011	China	Anas poecilorhyncha	1293/430	2146AGU **UUUUUC** CCCUU	6	825/274	489/162	20
29			KT962980.1	2009	Russia	Anas crecca	1293/430	2139AGUU **UUUUCC** CCUU	5	825/274	594/197	5
30			EF569970.1	2003	Russia	Domestic goose	1293/430	2146AGU **UUUUUC** CCCUU	6	807/268	534/177	40
31			GQ406232.1	2007	Italy	Duck	1293/430	2134AGU **UUUUUC** CCGCU	6	816/271	474/157	4
32	APMV - 7		FJ231524.1	1975	USA	Dove	1185/394	2150AAUUU **UUUCCC** AAU	4	753/250	378/125	1
33	APMV - 8		FJ215863.2	1976	USA	Canada goose	1218/405	2095UAAUUU **UUCCCG** GG	3	717/238	612/203	65
34			MF448515.1	2013	Kazakhstan	Little stint	1218/405	2095UAAUUU **UUCCCG** GG	3	717/238	519/172	34
35			JX901129.1	1978	Japan	Pintail duck	1218/405	2095UAAUUU **UUCCCG** GG	3	717/238	519/172	34
36			FJ619036.1	1976	USA	Canada goose	1218/405	2095UAAUUU **UUCCCG** GG	3	717/238	519/172	34
37			MF448514.1	2013	Kazakhstan	Whooper swan	1218/405	2095UAAUUU **UUCCCG** GG	3	717/238	519/172	34
38	APMV - 9		EU910942.1	1978	USA	Domestic duck	1260/419	2319GAUU **UUUUCC** CGAG	5	792/263	585/194	56
39	APMV - 10		HM755886.2	2007	United Kingdom	Penguin	1218/405	2094A GAAUUU **UUCCCG** UG	3	741/246	519/172	34
40			HM147142.3	2007	Falkland Islands	Penguin	1218/405	2094A GAAUUU **UUCCCG** UG	3	741/246	519/172	34
41			HM755887.2	2007	United Kingdom	penguin	1218/405	2094A GAAUUU **UUCCCG** UG	3	741/246	519/172	34
42			HM755888.2	2007	United Kingdom	penguin	1218/405	2094A GAAUUU **UUCCCG** UG	3	741/246	519/172	34
43	APMV - 11		JQ886184.1	2010	France	Common snipe	1344 /447(2 G insertion)	2333AA AAUUCU **UCCCCC** A	2	834/277 (V mRNA is the unedited transcript)	546/181(1 G insertion)	15

Continued

44	APMV - 12	KC333050.1	2005	Italy	Wigeon	1218/405	^{2324}UGA**UU UUUUCC CU**AG	5	762/253	564/187	31
45		KU646513.1	2013	Kazakhstan	Wild goose	1194/397	^{2318}UGA**UU UUUUCC CG**UC	5	726/241	453/150	2
46	APMV - 13	KX119151.2	2011	Ukraine	White-fronted goose	1194/397	^{2318}UGA**UU UUUUCC CG**UC	5	726/241	453/150	2
47		LC041132.1	2000	Japan	Anahersp	1194/397	^{2318}UGA**UU UUUUCC CG**UC	5	726/241	453/150	2
48	APMV - 14	KX258200.1	2011	Japan	Duck	1227/408	^{2094}A AAA**UU UCCCCG GC**	2	759/252	471/156	12
49	APMV – 15	KX932454.2	2012	Brazil	Calidrisfuscicollis	1248/415	^{2039}AA **UUUUUC CCCCCU G**	6	759/252	438/145	2
50	APMV - 16	KY511044.1	2014	South Korea	Wild duck	1200/399	^{2289}UGA**UU UUUUUC CC**AGU	6	738/245	423/140	5
51	APV-A (APMV – 17)	KY452442.1	2014	Antarctica	Pygoscelispapua	1209/402	^{2105}AA**UUUUUUCCCC**AGU	Hexamer phasing position not determined as their genome lengths did not conform to 'rule of six'	771/256	492/163	12
52	APV-B (APMV – 18)	KY452443.1	2014	Antarctica	Pygoscelispapua	1179/392	^{2116}AA**UUUUUUCCCC**AGU		711/236	531/176	35
53	APV-C (APMV – 19)	KY452444.1	2014	Antarctica	Pygoscelispapua	1134/377	^{2153}GAA**UUUUUUCCC**AGU		666/221	393/130	4
54	APMV - 20	MF033136.1	2014	Kazakhstan	Gull	1296/431	^{2171}AA AA**UUUU CCCCCU C**	1	792/263	498/165	17
55	APMV - 21	MF594598.1	2015	South Korea	Wild birds	1320/439	^{2270}CGA**UU UUUUCC CG**UC	5	915/304	432/143	3

Table 1. Detailed information on V and W proteins of APMV species and strains analyzed in this study. A summary on the viruses analyzed in this study with specifications of their P, V and W protein sequences. Nt: nucleotides, aa: amino acids.

used in this study that were either directly collected from NCBI or derived by prediction using DNASTAR software suite are provided in Supplementary Files S1, S2 and S3.

Sequence alignment, comparison and prediction of conserved motifs/domains. Multiple sequence alignments of V and W proteins were performed using the TCOFFEE multiple alignment algorithm, mode 'expresso' and the sequence similarities were colored through ESPript[64–66]. All residues/amino acid positions mentioned in the results and discussion correspond to APMV-1 strain KJ808820.1, the strain that appears first in the alignment file. Individual sequences were also analyzed in NCBI's interface, conserved domain (CD)-search[67] and aligned sequences were run in DREME version 5.0.5 software[68] to identify conserved motifs/domains. The intraclade amino acid percentage identity was estimated using Megalign software from DNASTAR.

Prediction of Nuclear Localization Signal (NLS) and Nuclear Export Signal (NES) in V and W proteins. The nuclear localization signal (NLS) in V and W proteins of APMV species were identified using online tool, cNLS mapper with a cut-off score of 5.0 that predicted NLS specific to the importin αβ pathway[69]. The presence of nuclear export signal (NES) in V and W proteins of APMV species was predicted using online tool, NetNES 1.1 server that predicted leucine-rich NES using a combination of neural networks and hidden Markov models[70] and using LocNES that predicted the classical NESs in CRM1 cargoes[71].

Phylogenetic analysis and evolutionary divergence. Phylogenetic analysis was performed using MEGA7 software. For drawing the phylogenetic trees, evolutionary history was inferred by using the Maximum Likelihood method with JTT matrix-based model[72] for V proteins and Dayhoff matrix based model for W proteins[73]. For drawing the phylogenetic tree of V proteins, bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed[74]. Branches corresponding to partitions, reproduced in less than 80% bootstrap replicates, were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. Initial trees for heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then topology with superior log likelihood value was selected. A discrete gamma distribution was used for V proteins tree to model evolutionary rate differences among sites (16 categories (+G, parameter = 2.4693)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 5.67% sites). For drawing the phylogenetic tree of W proteins, a discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1488)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.70% sites). Analysis of both the trees involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 141 and 71 positions in the final dataset for drawing V and W proteins' phylogenetic trees, respectively.

The estimates of evolutionary divergence over sequence pairs between groups were analyzed for V and W proteins of all 21 APMV species using MEGA7[75]. Briefly, based on the maximum likelihood fits of 56 different amino acid substitution models, the final analyses were conducted in MEGA7 using JTT matrix-based model[72]

for V proteins and Dayhoff matrix based model[73] for W proteins. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis included 55 amino acid sequences. All positions containing gaps and missing data were eliminated.

Selection pressure analysis. The number of nonsynonymous substitutions per nonsynonymous site (dN), the number of synonymous substitutions per synonymous site (dS), and the dN/dS ratios for the nucleotide sequences of V and W proteins of all 21 species were analyzed for the entire sequence and also their shared N terminal and unique C terminal regions. The shared portion in the N-terminus of all the three proteins was considered up to the RNA editing site (KKG motif). The C terminal regions of V and W proteins of all 21 species were considered after RNA editing site (KKG motif). The dN/dS ratio of 21 species of APMV nucleotide sequences were estimated by DnaSP v6.12.03 software[76]. The protein was considered under positive selection or diversifying when the dN/dS ratio is >1 and negative or purifying selection when dN/dS ratio <1.

Evolutionary rate analysis. To estimate evolutionary rates of different APMV species in V and W nucleotide sequences, the substitution rate analysis was performed by BEAST v 1.10.4 software[77]. The substitution model GTR and site heterogeneity model G + I was found to be the best by MEGA7 and was used here to study the substitution rate of V and W sequences. The tree prior coalescent, constant size was used for individual and all the species. The uncorrelated relaxed clock with lognormal was implemented. The MCMC chain 4×10^8 cycles was used to reach the ESS value more than 200 to converge the data except for W proteins of APMV-8 strains where MCMC chain length of 2×10^8 cycles was used. The final data analysis was performed using tracer v 1.7.1 software.

Results

RNA editing site and prediction of V and W protein sequences.
Previous reports suggested identical *P* gene RNA editing sites for APMV-1, -2, -4, -5, -6, -7, -8, -9, -10, -12, -13, -15, -20, APV -A, -B and –C and varied *P* gene editing site sequences for APMV-3, -11, -14 and -16. We observed the following conserved pattern in the RNA editing sequences among APMVs: U_3C_6 for APMV-4; U_4C_4 for APMV-14; U_4C_5 for APMV-20; U_5C_3 for APMV -1 (all strains except KJ736742.1 and KJ808820.1), APMV-8 and APMV-10; U_5C_4 for APMV -1 (strains KJ736742.1 and KJ808820.1) and APMV-2; U_5C_6 for APMV-15; U_6C_3 for APMV -5 (strain GU206351.1), APMV -7, -9, -12, -13, -16, -19, -21; U_6C_2 for APMV-5 (strain LC168750.1); U_6C_4 for APMV -6, -17, -18 and $UUCUUC_5$ for APMV-11. Further, variations were observed in the cis-acting sequence at the editing site: the sequences immediately upstream of the editing site were $^{3'}$AA in APMV -2, -3, -4, -5, -7, -8, -10, -11, -14, -15, -17, -18, -19, -20 and $^{3'}$GA in APMV -1, -9, -12, -13, -16 and -21 (all orthoavulaviruses) while $^{3'}$AG was conserved among APMV-6 strains.

The hexamer phase of the start of the template C run in the *P* gene editing site (Table 1) revealed that this position was conserved within each species except in APMV-3 and for one strain of APMV-6 (KT962980.1). The hexamer phasing positions for APV -A, -B, -C were not determined as their genome lengths did not conform to 'rule of six'. The start of the C run was at hexamer position 1 for APMV-20; at hexamer position 2 for APMV -4, -11 and -14; at hexamer position 3 for APMV -2, -3 (EU782025), -5, -8 and -10; at hexamer position 4 for APMV-3 (EU403085) and APMV-7; at hexamer position 5 for APMV-1, one strain of APMV-6 (KT962980.1), APMV -9, -12, -13, and -21 and at hexamer position 6 for APMV -15, -16 and all strains of APMV-6 except one strain (KT962980.1).

All APMVs except APMV-11 expressed P protein from unedited mRNA. The editing site of APMV-11 resembled that of other paramyxoviruses that insert 2G for generating P mRNA or the 'genomic V' viruses[12]. The P protein of APMV-11 is derived from 2G nucleotides insertion, W protein from single G nucleotide addition and unedited mRNA expresses the V protein[17]. The V and W protein sequences of the other 20 APMV species were predicted by insertion of single G and two G nucleotides at the *P* gene RNA editing site, respectively.

Amino acid sequence analysis: percentage identity and conservations.
The V and W protein sequences of all 21 APMV species shared common N terminal region with P protein. The variations in the amino acid sequences were minimum within the first 60 amino acids in N terminal region. The N-terminal portions of P, V and W proteins of metaavulaviruses and orthoavulaviruses showed closer identity than paraavulaviruses. In ortho and paraavulaviruses, the N- and C- terminal regions showing high homology up to 93%. Metaavulaviruses showed 100% identity in all their C- and N- terminal regions. The C terminal region of W-protein sequences showed 0.0–100% identity, as both the amino acid composition and length variations at C-terminal portion for all the species were higher (Table 2). As described previously, the soyuz1 and soyuz2 motifs were observed within N terminal region in all APMVs except in APMV-3 strains[78]. Additionally, conserved domains (CD) were predicted in APMV-1 mesogenic strain Komarov (CD between aa 25 to 167) for large tegument protein UL36 (superfamily member PHA03247), in APMV-14 (CD between aa 31 and 144) for gene regulated by oestrogen in breast cancer-GREB1 (superfamily member, pfam15782) and in APMV-21 (CD between aa 53 and 120) for Tumor necrosis factor receptor superfamily member cd13415.

Comparison of V protein sequences of APMV species.
The V protein of APV-C was the shortest (221 aa, MW: 23.73 kDa) and that of APMV-21 was the longest (304 aa, MW: 31.98 kDa) among the 21 APMV species. The length of V protein (in aa) conserved within species was as follows: APMV-2 strains (232 aa), APMV-4 strains (224 aa), APMV-5 strains (277 aa), APMV-8 strains (238 aa), APMV-10 strains (246 aa) and APMV-13 strains (241 aa). However, in APMV-1, -3 and -6, variation in the length of V protein was observed between strains within the same species. Between species, the similarity in the V protein length was observed as follows:

V protein length of 252 aa was observed in APMV-14, -15 and one strain of APMV-3; APMV-11 and -5 showed 277 aa long V protein; APMV -9 and -20 had 263 aa long V protein while the V protein length of APMV-16 and two lentogenic strains of APMV-1 were 245 aa. The lowest amino acid identity (10.4%) was observed between V proteins of APMV-3 strain Wisconsin and APV-A. The lowest amino acid divergence was noticed between APV-B and APV-C (70.4) and both showed identity of 53.7% at amino acid level which was the highest between APMV species (Supplementary File S4).

The multiple sequence alignment of V protein sequences of APMV species revealed higher amino acid conservations in both N (majorly in the first 60 amino acids) and C terminal regions (Fig. 1). All viruses in this study had the following conserved motifs similar to V proteins of other known paramyxoviruses (i) *KKG* motif in the N terminal region, which is the coding sequences at the *P* gene mRNA editing site (residues 132-134, corresponding to APMV-1 strain KJ808820.1, the first strain in the alignment file) except in APMV-3, APMV-4, APMV-12, APMV-13, APMV-14 and APMV-20 (ii) *HRRE* motif (residues 177 – 180, corresponding to APMV-1 strain KJ808820.1), (iii) *WCNP* motif (residues 195-198, corresponding to APMV-1 strain KJ808820.1) and (iv) conserved seven cysteine-rich domain. Interestingly, the following amino acids were also conserved in the C terminal region in majority of APMV species with few exceptions: Proline residues at five positions: (a) position 175 in ten species (except in APMV -3,-4,-6, -7, -8, -9, -10, -11, -15, -20 and -21), (b) position 198 in all APMVs, (c) position 202 (except in APMV-4), (d) position 207 in all APMVs and (e) position 218 in all APMVs; Glycine residues at three positions: (a) position 176 (except in APMV-4 and -6), (b) position 188 (except in certain strains of APMV-1 and APMV-7, -9, -11, -12, -13, -14, -15, -16, APV -A and -B), (c) position 215 (except in a single lentogenic strain, KM885162, of APMV-1, all strains of APMV-3, all strains of APMV-10 and interestingly in all these viruses, the glycine residue was replaced with arginine residue); Serine residues at two positions: (a) position 182 (except in APMV -4, -11, -20) and (b) position 194 (except in APMV -3, -4, -5, -6, -12, -14); Arginine residue at position 208 (except in APMV -3, -4, -5, -7, -8, -11, -13, -16, APV-A, -B and –C); Leucine residue at position 223 (except in APMV-5, -6, -7, -8 and -11) and Aspartic acid residue at position 227 (except in APMV-2, -3, -7, -9, -11 and -21). The percentage amino acid conservation in the C terminal region of V proteins of all 21 APMV species was between 30% (in the longest V protein, that of APMV-21) and 48% (in the shortest V protein, that of APV-C). The NLS were predicted in V proteins of APMV-5 and APMV-20 by cNLS mapper with a cut off score of 5.0. The NES were identified only in APMV-5 strains (Table 3b).

Comparison of W protein sequences of APMV species.

The length of W protein varied between 125 and 227 amino acids (aa) with calculated molecular weights between 13.30 kDa and 24.38 kDa. APMV-3 strain Netherland and APMV-7 had the shortest W protein (125 aa) and two strains of APMV-1, mesogenic strain KX761866.1 and velogenic strain KJ808820.1 (227 aa) had the longest W protein.

The length (in aa) of W protein was conserved among all the strains of APMV-2 (207 aa), all the strains of APMV-4 (137 aa), all the strains of APMV-5 (187 aa), all the strains of APMV-10 (172 aa) and all the strains of APMV-13 (150 aa). Also, similarity in the W protein length was noticed between the following species: W protein length of 172 aa was observed in all strains of APMV-10 and all strains of APMV-8 except strain FJ215863.2; W protein length of 177 aa was deduced in one strain of APMV-1 (JQ015296.1) and 3 strains of APMV-6 (EU622637.2, AY029299.1, EF569970.1), while the W protein length of APMV-2 and APMV-5 were 187 aa. Variations in the W protein length between strains within the same species were observed in APMV-1 (227, 221, 196, 183, 179, 177, 137 aa), APMV-3 (125, 127 aa), APMV-6 (157, 162, 177, 197 aa) and APMV-8 (172, 203 aa). APMV-1 strains analyzed in this study, had the longest unique C terminal region when compared to other APMV species (Table 1 and Fig. 2).

The lowest amino acid percentage identity (2.4) was observed between W proteins of APMV-12 and APMV-3 strain Netherland. Incidentally, APMV-3 strain Netherland also showed the lowest amino acid identity with W proteins of other APMV species. The lowest amino acid divergence (83.7) was noticed between APMV-1 isolate HN1007 (KX761866.1) and APMV-16, their W protein amino acid identity was 48.6% which was the highest homology observed between APMV species (Supplementary File S5).

The NLS were identified in five out of the twelve strains of APMV-1, in one of the eight strains of APMV-6 (GQ406232.1) and in APMV-9 in their C terminal region while NLS in W protein of APMV-20 was observed in the N terminal region (shared with P and V proteins). The presence of NES were predicted in these viruses except in one strain of APMV-1 (JQ015296.1) and APMV-20 (Table 3a).

Phylogenetic tree and evolutionary distance analysis.

Based on their V protein sequences, phylogenetically the APMV species formed three distinct groups: group 1 consisted of APMV -3 strains, group 2 consisted of APMV -1, -9, -12, -13, -16, -21, APV -A, -B and -C (all orthoavulaviruses) and group 3 consisted of APMV-2, -5, -6, -7, -8, -10, -11, -14, -15, -20 and -4 (all metaavulaviruses and one paraavulavirus) (Fig. 3). The highest evolutionary divergence of 2.20 was observed between APMV-3 & APMV-4 and APMV-3 & APMV-12 followed by a divergence value of 1.91 between APMV-3 & APMV-5 and APMV-3 & APMV-11. The lowest divergence was between APV-A and APV-B (0.47) followed by APMV-9 & APMV-21 (0.50). The distance between the strains of the same species was noticed more in APMV-3 (0.4) followed by APMV-1 (0.297), which was further reiterated by their lower percentage of amino acid homology (Table 4).

The phylogenetic tree obtained from W protein sequences analysis showed clustering of strains of the same species (Fig. 4). The evolutionary distance analyses of W proteins of APMV species revealed that APMV-3 species is more divergent than other APMV species. The highest evolutionary divergence was noticed between APMV-3 & APV-C species (10.322) followed by APMV-3 & APMV-13 (8.594) and APMV-3 & APMV-12 (8.270). The lowest divergence was observed between APMV-9 & APMV-21 (0.400) followed by APV-A & APV-B (0.407). The distance between the strains of the same species was more in APMV-3 (0.619) followed by APMV-1 (0.256) which was also apparent from their lower percentage of amino acid homology (Table 5).

	Genus *Orthoavulavirus*	Genus *Metaavulavirus*	Genus *Paraavulavirus*
P complete	31.8–99.5	22.6–100	24.6–91.9
V complete	13.5–99.2	26.1–100	18.6–92.9
V- N terminal	31.0–98.0	22.5–100	19.5–93.3
V- C terminal	30.3–98.9	26.3–100	27.0–92.1
W complete	26.3–99.5	19.6–100	14.8–92.7
W- N terminal	31.0–99.2	22.4–100	16.7–93.3
W- C terminal	0.0–100	0.0–100	50.0–100

Table 2. Intraclade percentage amino acid sequence identity for *P*-gene products in APMV species**.** P complete, V complete and W complete refer to the entire protein sequences. N terminal refers to N terminal region of respective proteins shared with P protein and C terminal refers to the unique C terminal region of respective proteins. The amino acid sequences of a particular protein of all virus strains, within the particular genus, were analyzed and the percentage amino acid identity was calculated using Lasergene DNASTAR MegAlign software.

Selection pressure analysis. The dN/dS ratio was used to determine the natural selection pressure acting on the *P* gene edited products. The dN/dS ratio was estimated by DnaSP v6.12.03 for APMV species that comprised of more than one strain. The dN/dS values were significantly less than 1 for both V and W sequences (complete, N- and C-terminal regions) of most species explaining that they are under negative selection pressure. Only the C terminal region of V proteins of APMV-3 strains showed positive selection with dN/dS> 1 (Table 6).

Evolutionary rate analysis. The substitution rate of the V and W nucleotide sequences of APMV species that comprised of more than two strains were estimated by uncorrelated relaxed clock with lognormal using BEAST software. APMV-10 comprised of four strains, which were 98.65% to 100% identical to each other and hence the substitution rate could not be determined. The substitution rate was highest in APMV-13 followed by APMV-2, APMV-6 for both V and W proteins. The overall substitution rate was 7.37×10^{-5} for V protein and 8.07×10^{-5} for W protein (Table 7).

Discussion

Avian paramyxoviruses are known to infect a variety of bird species across the globe. Currently 21 species (previously called as serotypes) of APMVs are characterized and more viruses could be identified in future with improved viral surveillance programs. Paramyxoviruses with their small genome have a unique strategy of maximizing their genomic information by expressing viral proteins through co-transcriptional RNA editing. This helps to avoid error catastrophe caused by higher mutation rates often associated with larger genomes. Additionally, these viruses follow the 'rule of six' for efficient replication. Though, detailed studies on APMV structural genes and their complete genomes are available, a comparative knowledge of their accessory proteins expressed through RNA editing is lacking. In this study, using bioinformatics approach, we analyzed the *P* gene editing site, predicted and studied the protein sequences of edited products- V and W, of all 21 APMV species (55 viruses) known till date.

The hexamer phasing at the *P* gene editing site within each virus group is conserved[7]. We observed conserved hexamer phasing between certain APMV species and also, within each APMVs except in APMV-3 and for one strain of APMV-6. The hexamer phase is known to regulate the mRNA editing pattern, though subtle, it is important; for example, in human and bovine parainfluenza virus type 3 (PIV-3) in which the hexamer positions at *P* gene editing site are 2 to 5, higher mRNA editing frequency (~70%) and more number of G insertions (1 to 6 at equal frequencies) are observed while least mRNA editing ~30% with only 1 to 3 G insertions occur in Sendai virus wherein hexamer phase position is 1 at the *P* gene editing site [7,79,80]. Based on the hexamer phasing position, it is anticipated that, in all APMVs except in APMV -20, the editing frequency could be extensive with possibilities of more number of G insertions. However, the cis-acting sequence of *P* gene editing site in APMV-20 ($^{3'}$AA), suggests higher mRNA editing frequency and increased number of G insertions as reported in human and bovine PIV-3[81,82]. Thus APMVs seem to follow PIV-3 RNA editing phenotype.

Another interesting observation is the unique editing site sequence of APMV-11 ($^{3'}A_4UUCUU C_5$), in which the unedited mRNA translates to V protein and it has been suggested that 2G insertions in mRNA translates to P protein[17]. In rubulaviruses with *P* gene editing site sequence of $^{3'}A_3UUCUC_4$, realignment of the nascent mRNA/template hybrid during 1G insertion would mean non permissible A:C base pairing hence the minimum insertion expected is 2G[83]. The base pairing between the nascent chain and the template genome to form a hybrid is important to prevent transcriptional slippage by the polymerase[80]. Similarly, in APMV-11, 1G and 2G insertions would lead to unstable A:C base pairing, hence 3G insertion (V protein) could be the minimum number of insertions expected, also, while a 4G insertion would translate to W protein, a 5G insertion would lead to P protein synthesis. It needs to be explored if APMV-11 expresses more V protein (from both unedited mRNA and 3G insertions) than other paramyxoviruses. Furthermore, it will be interesting to study if deletions in addition to G insertions could happen in APMV-11 and other APMVs with longer C runs at the editing site as described previously with recombinant Sendai virus and PIV-3 minigenomes[83].

Three factors, the editing site (sequence and the length of C runs), the type of sequences immediately upstream of the editing site (cis-acting sequence) and the hexamer phase positions are known to decide the editing phenotype (i.e. number of G insertions, deletions and frequency of mRNA editing) which further can influence the virus pathogenicity[81,82,84,85]. Among APMVs, variations are observed in (i) *P* gene editing site, (ii) hexamer phase
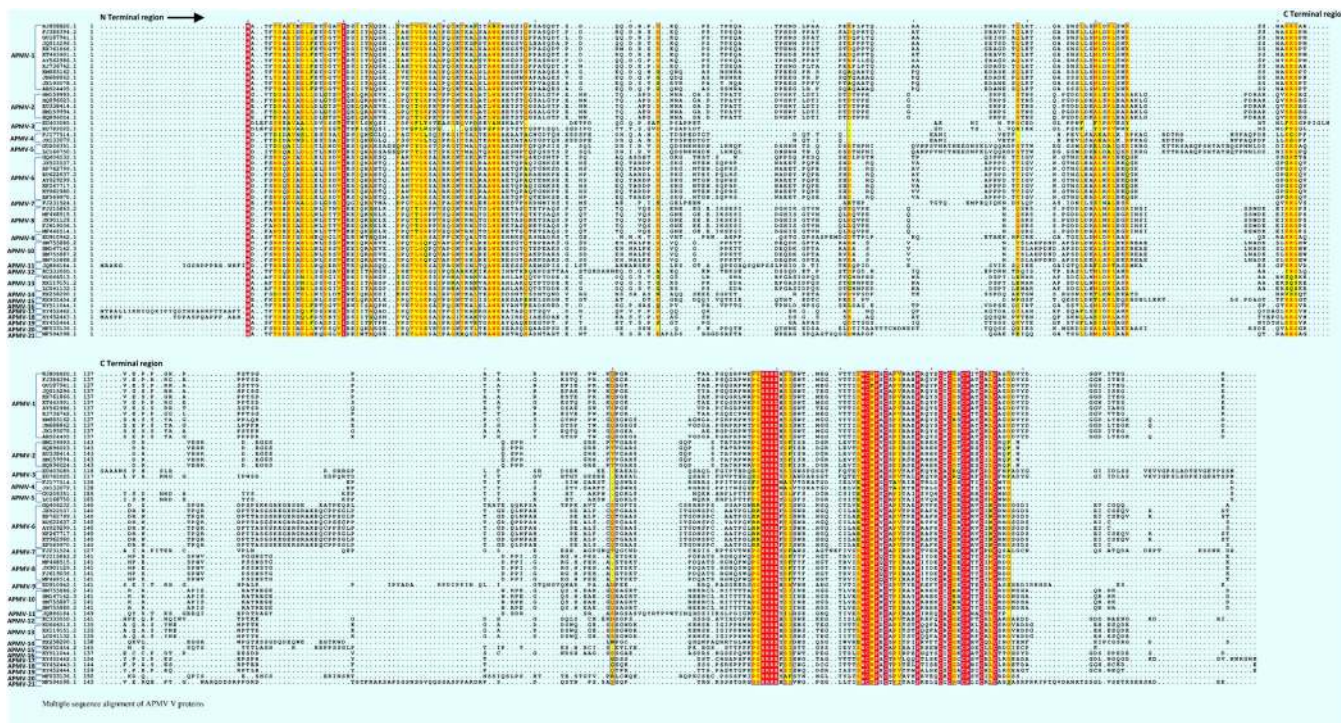
Figure 1. Multiple sequence alignment of V proteins of 21 APMV species. The conserved motifs and conserved amino acids are highlighted. The dots represent the gaps in the alignment. Highly conserved motifs and amino acids are in red and highlighted. The C terminal region is considered beyond the conserved KKG motif.

APMV species	Genbank ID	Predicted Nuclear localization signal sequence and position in W proteins	Predicted Nuclear export signal sequence and positionin W proteins
3a: NLS and NES within W proteins			
APMV -1	FJ386394.2	138RAPKRGTTNVRLNSREVNPAAETVRKDRRTK168	210IPLYL214 (predicted by NetNES 1.1 Server); 105VDTQLRTGASNSLLL119,202SDQGRAKTIPLYLRI216, 203DQGRAKTIPLYLRIM217 (predicted by LocNES)
	JQ015296.1	106DTQLKTGASNSLLSMLDKLSNKSSNAKKGAH136	None predicted
	KX761866.1	106DTQLKTGASNSLLSMLDKLSNKSSNAKKGAH136	210ILLHLWIMSSYLSTL224 (predicted by LocNES)
	KT445901.1	138RAPKRGTTNVRLNSREVNPAAETVRKDRRTK168	210IPLYL214 (predicted by NetNES 1.1 Server); 105VDTQLRTGASNSLLL119,202SDQGRAKTIPLYLRI216, 203DQGRAKTIPLYLRIM217 (predicted by LocNES)
	AY562986.1	106DAQLKTGASNSLLSMLDKLSNKSPNTKKGAH136	13LEI15 (predicted by NetNES 1.1 Server); 1MATFTDAEIDDILEIn (predicted by LocNES)
APMV-6	GQ406232.1	125RKLEKQGKSSGKTTADSSPGRDPSPSKKGARn	13LELSSDVI20(predicted by NetNES 1.1 Server)
APMV-9	EU910942.1	157RRRRKTRHPADQPTHRDTAWSPERPPSCQSR186	12LMEQSGLVI20 (predicted by NetNES 1.1 Server 125KEGATGGLLDMLDRI139 (predicted by LocNES)
APMV-20	MF033136.1	117KSNMGKDLDSALAKLEKKAASIKSDK142 (within N terminal region)	None predicted
3b: NLS and NES within V proteins			
APMV species	**Genbank ID**	**Predicted Nuclear localization signal sequence and position in V proteins**	**Predicted Nuclear export signal sequence and positionin V proteins**
APMV-5	GU206351.1	197PTTKHTAKPNQDRLSNQKRHR217	4TDDQAILDLLTL-N
	LC168750.1	nPTTKHTAKSNQDRLSNQKRHR217	4TDDQAILDLLTL-N
APMV-20		117KSNMGKDLDSALAKLEKKAASIKSDK142 (within N terminal region)	None predicted

Table 3. Predicted nuclear localization (NLS) and nuclear export signals (NES) in W (3a) and V (3b) proteins of 21 APMV species.

at the editing site and also (iii) the sequences immediately upstream of the editing site, all of which will determine the expression levels and relative proportions of P, V and W proteins in the APMVs which in turn could explain their differences in replication, pathogenicity and virus-host interactions.

S.No.	Species	GenBank ID	C terminal region of W proteins of APMV species
1	APMV-1	AY562986.1	(45)-AHGoVLKKDIINLRpsNmGIGRvmGVTRRspssMPGPPLEAGAQT
2	APMV-1	KX761866.1	(94)GAHGRvpRKGTTNLRpssKGvNRAAETvRRDRKTRsRPPLETRAQTRTQHIMDNGRsHnHQLVQSLMLSGQGRaKTILiHLWIMSSYLSTLCRR
3	APMV-1	JQ015296.1	(43)-AHGRvpKKGITNLRpNNTGTsRAMEATREDHsTRpRPSLETGA
4	APMV-1	GU187941.1	(62)-AHGRIpKKGIfNHRsNsKGaNRAAGTIRKDRRTRpRPPLETGAQTRTQHILDNGRsHNYQLV
5	APMV-1	FJ386394.2	(88)GAHGRApKRGTTNvRLNSRevNPAAETvRKDRRTKsRPPLETRAQTRTQHIMDNGRsHNYQLVQPLMLSDQGRaKTIPLYLRIMSSHL
6	APMV-1	KT445901.1	(88)GAHGRapKRGTTNvRLNSRevNPAAETvRKDRRTKsRPPLETRAQTRTQHIMDNGRsHNYQLVQPLMLSDQGRaKTIPLYLRIMSSHL
7	APMV-1	JX193078.1	(45)-AHLRvLHssALYsKgsscsGTNPTmGPIRGGsGTGfRPfLDPRAPA
8	APMV-1	JN688862.1	(49)-AQLRALYsRPPTPKGvPQRNKPNMEsIRGGLsGRRGTRsPQALDsRAPA
9	APMV-1	AB524405.1	(49)-AHLRALHsRPPTPKGaPQRNKPNvGPIRGGLsGRRGTRsPQALDsRAPA
10	APMV-1	KM885162.1	(49)-AHLRvLHsRPsTPKGaLQwNKPTMGPIRGGLsGKRGTRfPQALDPRAPA
11	APMV-1	KJ808820.1	(94)GGQGRapRKATINLWPsNTGIsRAvETTRKDRsTRpKPPLEAGAQTRAQHIMDNGRsHNHQLVQSLMCSSQGRaKTVLLYLWIMSNYLSTLCRR
12	APMV-1	KJ736742.1	(3) -GQG
13	APMV-2	HM159993.1	(67)GRRsGRaQGRGRQPVTTWKGaDsRsQERAAEHsHKAMATGTQEGVLIHLsRWETGsHKLVQPSLLSH
14	APMV-2	EU338414.1	(67)GRRsGRaQGRGRQPVTTWKGaDsRsQERAAEHsHKAMATGTQEGVLIHLsRWETGsHKLVQPSLLSH
15	APMV-2	HM159994.1	(67)GRRsGRaQGRGRQPVTTWKGaDsRsQERAAEHsHKAMATGTQEGVLIHLsRWETGsHKLVQPSLLSH
16	APMV-2	HQ896023.1	(67)GRRsGRaQGRGRQPATTWKGaDsRsQERAAEHsHKAMATGTQEGVLIHLsRWETGsHKLVQPSLLSH
17	APMV-2	HQ896024.1	(67)GRRsGRaQGRGRQPATTWKGaDsRsQERAAEHsHKAMATGTQEGVLIHLsRWETGsHKLVQPSLLSH
18	APMV-3	EU782025.1	(12)APPCsQRRNPLV
19	APMV-3	EU403085.1	(14)APNwTQRsGRLVtK
20	APMV-4	JX133079.1	(2) AP
21	APMV-4	FJ177514.1	(2) AS
22	APMV-5	LC168750.1	(5) APDLK
23	APMV-5	GU206351.1	(5) GPDLR
24	APMV-6	EF569970.1	(20)GKGAsQGAMPSIRvTHRGPT
25	APMV-6	KT962980.1	(40)GKGAsQGAMPSIRvTHRGPTQsCPLRsPPMPDWRsKHHRR
26	APMV-6	AY029299.1	(20)GKGAsQGAMPSIRvTHRGPT
27	APMV-6	JX522537.1	(5) GKGAs
28	APMV-6	KP762799.1	(5) GKGAs
29	APMV-6	EU622637.2	(20)GKGAsQGTMPFIRATHRGPT
30	APMV-6	KF267717.1	(5) GKGAs
31	APMV-6	GQ406232.1	(4) ARGT
32	APMV-7	FJ231524.1	(1) G
33	APMV-8	MF448514.1	(34)ALASRISLGPKQQLNRRPTHWQGTPRTCKHGQNP
34	APMV-8	FJ619036.1	(34)ALASRISLGPKQQLNRRPTHWQGTPRTCRHGQNP
35	APMV-8	JX901129.1	(34)ALASRISLGPKQQLNRRPTHWQGTPRTCKHGQNP
36	APMV-8	MF448515.1	(34)ALASRISLGPKQQLNRRPTHWQGTPRTCKHGQNP
37	APMV-8	FJ215863.2	(65)ALASRISLGPRRQLNRRPTHWQGTPRTCKHRQKPwRsHiKRQsILATRTQERvFIYHLPwHHQsH
38	APMV-9	EU910942.1	(56)GSiRDHKtRAPCTQANsLRRRKtRHPADQPTHRDTAWsPERPPsCQsRGETTARiR
39	APMV-10	HM755888.2	(34)ALEESSDIKsNQRRGKPPRGPKsRsGRPIsRPHQ
40	APMV-10	HM755887.2	(34)ALEESSDIKsNQRRGKPPRGPKsRsGRPIsRPHQ
41	APMV-10	HM147142.3	(34)ALEESSDIKsNQRRGKPPRGPKsRsGRPIsRPHQ
42	APMV-10	HM755886.2	(34)ALEESSDIKsNQRRGKPPRGPKsRsGRPIsRPHQ
43	APMV-11	JQ886184.1	(15)GRSVLNGIRRRADRT
44	APMV-12	KC333050.1	(31)GSWRATQLQRKKQwTKLFIRGcHPGRvQAWS
45	APMV-13	KU646513.1	(2) GR
46	APMV-13	KX119151.2	(2) GR
47	APMV-13	LC041132.1	(2) GR
48	APMV-14	KX258200.1	(12)PVARDPLLRRPR
49	APMV-15	KX932454.2	(2) DK
50	APMV-16	KY511044.1	(5) QRMLP
51	APV-A	KY452442.1	(12)GQLPLPRIKTLT
52	APV-B	KY452443.1	(35)GQLPCFRvKTNRTvHRvPEILFPLKETSLQARPPA
53	APV-C	KY452444.1	(4) GQIP
54	APMV-20	MF033136.1	(17)EKRSTANLGIPLFREDQ
55	APMV-21	MF594598.1	(3) GRc

Figure 2. Comparison of C terminal region of W proteins of APMV species. The C terminal region is considered beyond the conserved KKG motif (please refer Fig. 1). Number of aminoacids in the unique C terminal is mentioned within parenthesis. The basic amino acids (R, K and H) are bolded and enlarged. Amino acid residues that are conserved between strains of the same species are bolded.

With respect to the length and amino acid composition of V and W proteins, there were huge variations between species than within species, which was also reiterated by their dN/dS estimates. The V proteins were more conserved than W proteins. Higher sequence identity for V proteins was observed between the strains of the same species (exception was APMV-3) more often than between species. Phylogenetically, V protein analysis of APMV species grouped viruses similar to the individual gene-based phylogeny[2], all the members of genus *Orthoavulavirus* clustered into one group and all members of genus *Metaavulvirus* along with one of the avian paraavulaviruses (APMV-4) formed the second group while the avian paraavulavirus, APMV-3, formed an out-group. This was further affirmed by evolutionary distance analysis. The phylogenetic analysis and the evolutionary distance data of both V and W proteins clearly showed that APMV-3 strains are the most divergent.

The N terminal region of V and W proteins which is shared with P protein, showed highest conservation among APMVs. In their C terminal region, the V proteins of paramyxoviruses carry conserved arginine and isoleucine residues upstream of highly conserved seven cysteine residues (zinc binding domain), known to play important roles in MDA5 interference, STAT1 degradation and blocking interferon signaling to evade host immunity[86–88]. The V proteins of all 21 APMV species analyzed in this study had the seven cysteine residues and remarkably, many other amino acids were also conserved in their C terminal region. Though similar observations have been made earlier in other paramyxovirus V proteins, their functional importance is unknown yet[86].

The *V* and *W* genes were found to be under negative selection pressure with dN/dS <1 in all the species. This shows the conserved nature of the non-structural viral proteins within the species and probably indicates their functional importance, which is yet to be completely explored. Furthermore, the substitution rates of APMV species determined by molecular clock was varying across the species and was higher between species than within species. The substitution rates for W proteins was higher than V proteins except for APMV-1, where more strains were
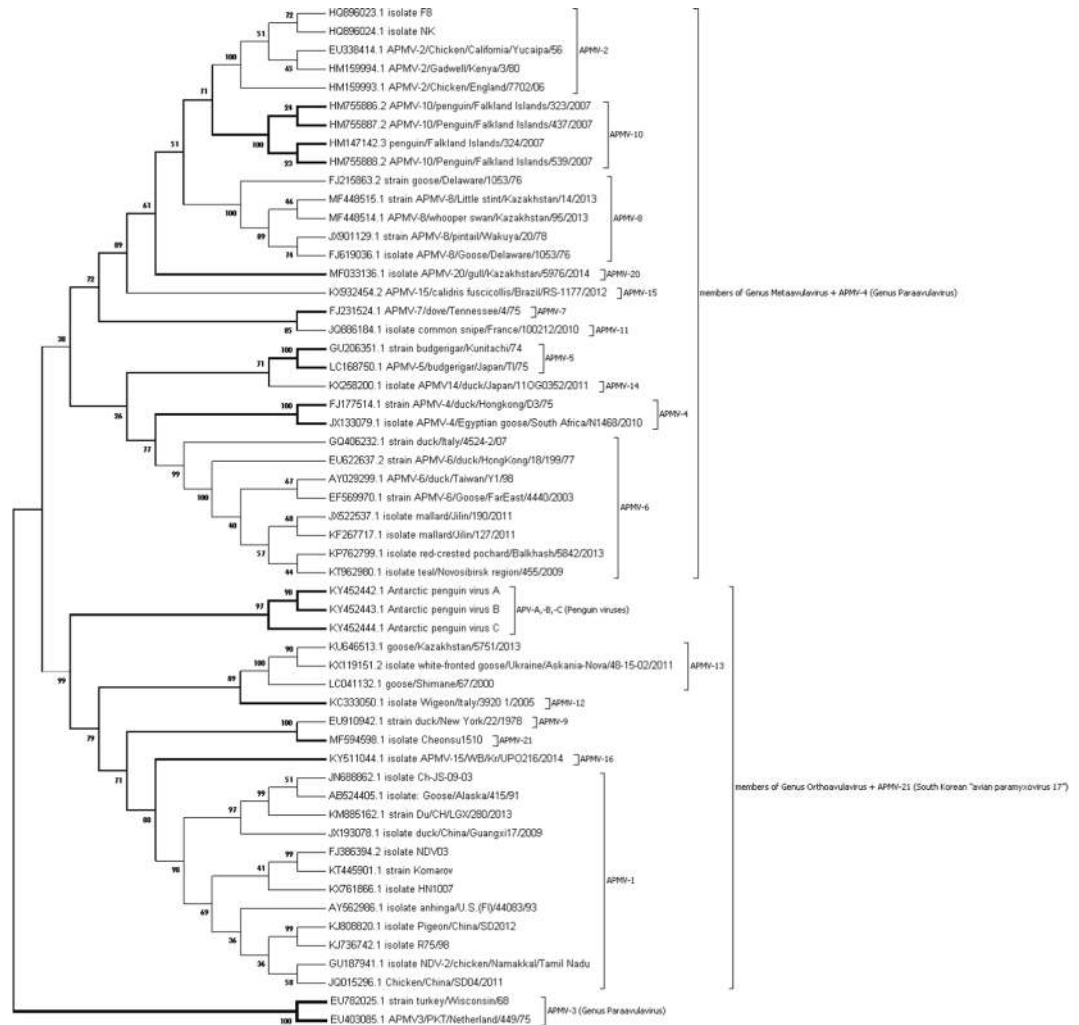
Figure 3. Phylogenetic tree derived from analysis of V proteins of APMV species by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Evolutionary analyses were conducted in MEGA7. All the orthoavulaviruses clustered together, all metaavulaviruses grouped along with APMV-4 (Paraavulavirus) while APMV-3 strains (Paraavulavirus) formed a separate branch.

APMV species	APMV-1	APMV-2	APMV-3	APMV-4	APMV-5	APMV-6	APMV-7	APMV-8	APMV-9	APMV-10	APMV-11	APMV-12	APMV-13	APMV-14	APMV-15	APMV-16	APV-A	APV-B	APV-C	APMV-20	APMV-21	Distance	Std. Err	Amino Acid Homology (%)
APMV-1		0.16	0.23	0.22	0.20	0.19	0.23	0.19	0.10	0.18	0.21	0.10	0.11	0.16	0.17	0.09	0.11	0.10	0.13	0.18	0.11	0.297361434	0.044035614	59.0-99.6
APMV-2	1.39		0.21	0.18	0.16	0.16	0.14	0.11	0.17	0.10	0.16	0.19	0.18	0.22	0.18	0.11	0.17	0.21	0.18	0.20	0.11	0.010158312	0.006115461	98.7-100
APMV-3	1.84	1.82		0.27	0.23	0.22	0.23	0.24	0.22	0.18	0.22	0.27	0.21	0.21	0.20	0.22	0.26	0.23	0.24	0.25	0.23	0.404530687	0.082313283	62.1
APMV-4	1.62	1.54	2.20		0.25	0.15	0.25	0.19	0.24	0.19	0.21	0.25	0.27	0.25	0.17	0.19	0.20	0.22	0.20	0.19		0.057117638	0.018962361	95.9
APMV-5	1.44	1.28	1.91	1.68		0.15	0.18	0.18	0.22	0.17	0.20	0.20	0.22	0.15	0.17	0.18	0.19	0.18	0.19	0.15	0.22	0.028150548	0.013900268	94.2
APMV-6	1.40	1.19	1.75	1.12	1.11		0.19	0.17	0.20	0.13	0.16	0.19	0.21	0.15	0.15	0.18	0.20	0.19	0.19	0.17	0.20	0.117040525	0.022292276	58.2-98.9
APMV-7	1.72	1.07	1.79	1.85	1.43	1.36		0.16	0.26	0.16	0.12	0.23	0.21	0.22	0.16	0.23	0.21	0.21	0.22	0.16	0.24	n/c	n/c	-
APMV-8	1.48	0.77	1.82	1.57	1.37	1.27	1.33		0.22	0.10	0.16	0.20	0.19	0.18	0.12	0.17	0.19	0.20	0.18	0.10	0.20	0.021611621	0.008070679	93.7-99.6
APMV-9	0.74	1.42	1.82	1.63	1.60	1.48	1.84	1.60		0.20	0.20	0.12	0.13	0.17	0.18	0.13	0.15	0.16	0.13	0.17	0.08	n/c	n/c	-
APMV-10	1.46	0.63	1.50	1.47	1.30	1.03	1.24	0.68	1.58		0.15	0.20	0.22	0.15	0.11	0.19	0.21	0.19	0.23	0.13	0.18	0.000000	n/c	98.8-100
APMV-11	1.63	1.26	1.91	1.56	1.44	1.17	0.96	1.16	1.51	1.12		0.21	0.22	0.22	0.14	0.21	0.22	0.22	0.20	0.16	0.20	n/c	n/c	-
APMV-12	0.75	1.56	2.02	1.81	1.44	1.41	1.81	1.51	0.92	1.50	1.54		0.10	0.18	0.21	0.11	0.13	0.13	0.13	0.18	0.11	n/c	n/c	-
APMV-13	0.84	1.66	1.79	1.92	1.53	1.50	1.62	1.39	0.96	1.70	1.50	0.72		0.18	0.21	0.12	0.13	0.14	0.14	0.23	0.13	0.018485243	0.009260324	96.3-98.3
APMV-14	1.25	1.32	1.72	1.73	1.06	1.11	1.49	1.35	1.33	1.11	1.54	1.33	1.38		0.16	0.18	0.18	0.16	0.18	0.15	0.18	n/c	n/c	-
APMV-15	1.29	0.79	1.56	1.35	1.25	1.21	1.18	0.82	1.43	0.74	1.02	1.46	1.48	1.20		0.17	0.20	0.18	0.18	0.14	0.19	n/c	n/c	-
APMV-16	0.58	1.43	1.69	1.48	1.29	1.39	1.70	1.37	0.90	1.45	1.63	0.80	0.93	1.24	1.28		0.11	0.13	0.13	0.17	0.12	n/c	n/c	-
APV-A	0.84	1.52	1.95	1.43	1.40	1.38	1.59	1.51	1.15	1.54	1.63	1.00	0.94	1.35	1.36	0.86		0.07	0.10	0.17	0.14	n/c	n/c	-
APV-B	0.81	1.44	1.83	1.52	1.32	1.43	1.53	1.55	1.14	1.48	1.65	1.00	1.03	1.25	1.42	0.91	0.47		0.09	0.16	0.14	n/c	n/c	-
APV-C	1.01	1.55	1.93	1.52	1.46	1.34	1.60	1.44	0.95	1.62	1.52	1.02	1.08	1.32	1.37	0.98	0.72	0.67		0.19	0.14	n/c	n/c	-
APMV-20	1.42	0.82	1.77	1.58	1.29	1.46	1.26	0.75	1.39	0.87	1.18	1.38	1.62	1.17	0.99	1.39	1.39	1.41	1.41		0.17	n/c	n/c	-
APMV-21	0.78	1.31	1.87	1.49	1.58	1.46	1.82	1.47	0.50	1.40	1.38	0.82	1.00	1.43	1.37	0.88	1.02	0.97	1.09	1.42		n/c	n/c	-

Table 4. Estimates of Evolutionary Divergence over Sequence Pairs between Groups, analyzed for V proteins of 21 APMV species. The number of amino acid substitutions per site from averaging over all sequence pairs between groups is shown. Standard error estimate(s) are shown above the diagonal. Analyses were conducted using the JTT matrix-based model. Evolutionary analyses were conducted in MEGA7.
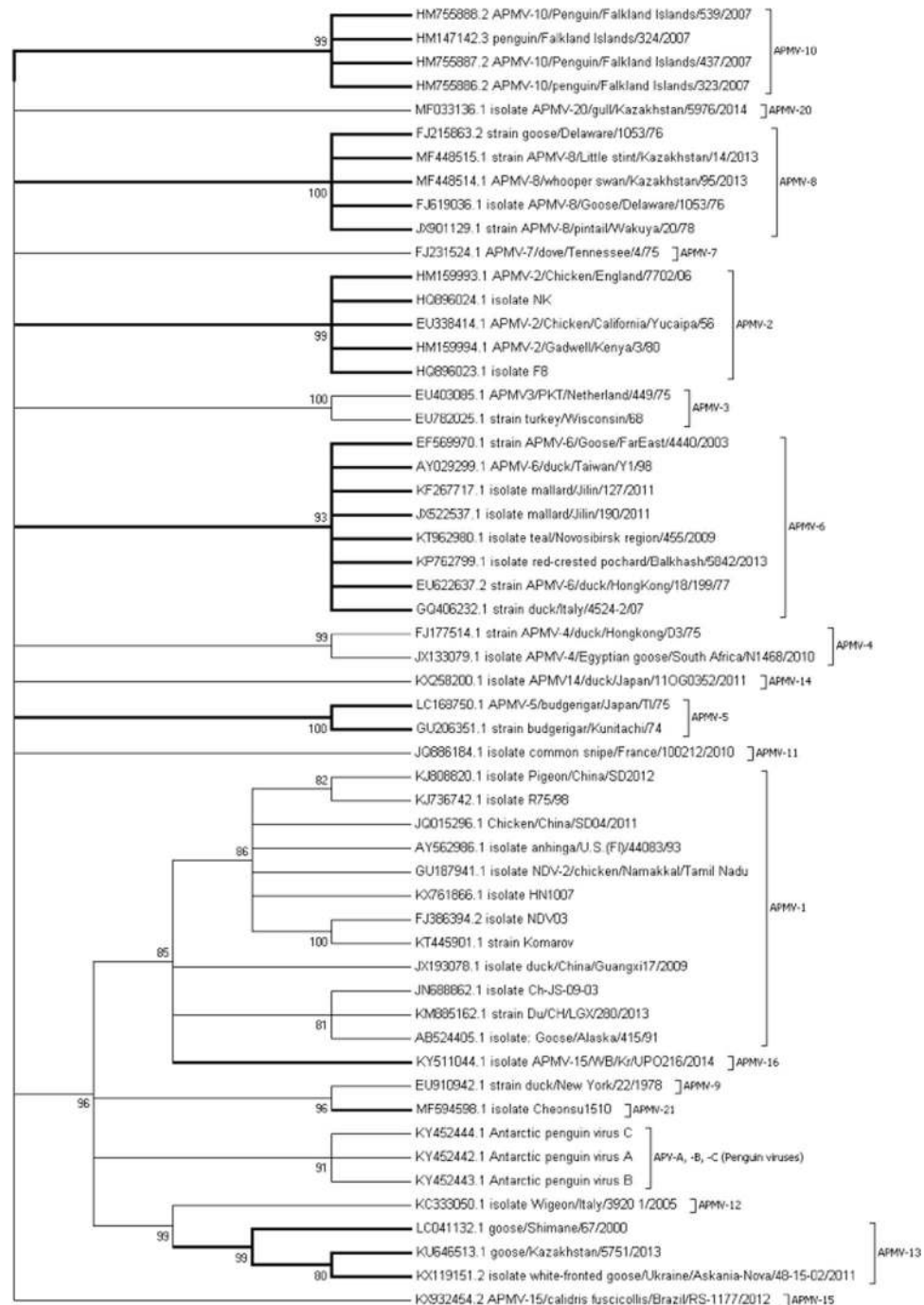
Figure 4. Phylogenetic tree derived from analysis of W proteins of APMV species by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the Dayhoff matrix based model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Evolutionary analyses were conducted in MEGA7. The tree shows clustering together of strains of the same species.

available for comparison. Though the substitution rate was slightly higher for V nucleotide sequence (5.15×10^{-4}) of APMV-1 compared to W nucleotide sequence (1.0×10^{-4}), it did not lead to changes in amino acid sequence as evident from dN/dS ratio estimates suggesting negative selection pressure. The higher conservation of V protein sequence implies its significant role in virus biology such as replication, pathogenesis and immune evasion.

In contrast to V proteins, the APMV W proteins were highly disordered, showed little sequence conservation when compared to V proteins and their divergence values were higher when compared to V proteins. The evolutionary data analysis of W proteins suggested higher sequence identity among strains of same species and higher variability between species. There were no conserved sequences or motifs in the C terminal region of W proteins except that most them carried large number of basic amino acids suggesting W protein to be highly basic as

APMV species	APMV-1	APMV-2	APMV-3	APMV-4	APMV-5	APMV-6	APMV-7	APMV-8	APMV-9	APMV-10	APMV-11	APMV-12	APMV-13	APMV-14	APMV-15	APMV-16	APV-A	APV-B	APV-C	APMV-20	APMV-21	Distance	Std. Err	Amino Acid Homology (%)
APMV-1		0.674	3.606	0.816	1.246	0.673	1.210	0.759	0.208	0.770	0.751	0.258	0.260	1.051	0.765	0.142	0.229	0.208	0.315	0.601	0.199	0.256688229	0.055048729	50.3-99.6
APMV-2	2.517		1.711	0.552	0.752	0.410	0.344	0.195	0.827	0.219	0.690	0.925	0.914	0.672	0.267	0.878	0.904	1.178	1.381	0.244	0.778	0.012357206	0.010920436	98.6-100
APMV-3	6.726	4.668		3.651	5.943	2.565	2.582	2.078	2.748	1.603	3.081	6.446	6.195	1.524	1.526	3.781	2.503	2.638	8.386	1.266	2.705	0.619433959	0.165791657	53.9
APMV-4	2.971	2.358	6.611		0.732	0.438	0.887	0.711	0.985	0.610	0.582	1.135	1.216	0.906	0.692	0.932	0.916	0.710	0.983	0.546	1.161	0.012344072	0.011612714	93.4
APMV-5	3.662	2.559	7.915	2.530		0.449	0.839	0.711	1.483	0.870	0.790	1.387	1.806	0.403	0.742	1.634	1.185	1.079	1.067	0.579	1.642	0.012103903	0.011637576	93
APMV-6	2.428	1.666	5.020	1.740	1.601		0.692	0.324	1.015	0.296	0.409	0.783	1.016	0.428	0.314	0.752	0.788	0.772	0.987	0.263	0.927	0.085996401	0.025678574	60.4-100
APMV-7	3.613	1.457	5.515	3.196	2.888	2.451		0.549	1.192	0.377	0.826	1.840	1.555	0.655	0.550	1.316	1.414	1.179	1.697	0.498	0.967	n/c	n/c	-
APMV-8	2.549	0.900	4.808	2.798	2.615	1.298	2.194		0.615	0.191	0.464	0.849	1.102	0.704	0.236	0.743	0.940	0.738	0.784	0.185	0.698	0.015515661	0.009522058	92.4-99.4
APMV-9	0.902	2.745	5.777	3.244	4.013	3.211	3.533	2.198		0.749	0.574	0.399	0.340	1.118	0.918	0.272	0.367	0.273	0.368	0.584	0.107	n/c	n/c	-
APMV-10	2.576	0.924	4.096	2.431	2.921	1.072	1.619	0.794	2.506		0.458	1.044	1.537	0.613	0.266	0.837	0.959	1.131	1.319	0.192	0.744	0.000001	n/c	98.3-100
APMV-11	2.636	2.490	6.187	2.321	2.567	1.624	2.794	1.728	1.932	1.716		1.092	1.001	0.624	0.530	0.650	1.027	0.974	0.870	0.432	0.625	n/c	n/c	-
APMV-12	1.108	2.912	8.270	3.461	3.962	2.561	4.829	2.625	1.605	3.079	3.380		0.190	1.542	0.957	0.328	0.299	0.331	0.420	0.864	0.351	n/c	n/c	-
APMV-13	1.218	2.984	8.594	3.770	4.431	3.108	4.374	3.089	1.420	3.750	3.007	0.756		2.673	0.898	0.343	0.340	0.329	0.369	1.005	0.310	0.033555163	0.016972415	95.3-98.0
APMV-14	3.165	2.378	4.481	3.164	1.527	1.678	2.322	2.449	3.137	2.036	2.199	4.134	5.562		0.762	1.326	1.541	1.678	1.257	0.480	1.257	n/c	n/c	-
APMV-15	2.635	1.098	3.990	2.744	2.539	1.267	2.188	0.990	2.927	1.025	1.927	2.922	2.655	2.511		0.920	0.795	0.765	0.939	0.324	0.580	n/c	n/c	-
APMV-16	0.581	2.831	6.387	3.043	4.016	2.401	3.613	2.381	1.088	2.494	2.270	1.289	1.477	3.463	2.832		0.283	0.233	0.343	0.653	0.248	n/c	n/c	-
APV-A	0.950	2.983	4.860	3.016	3.430	2.658	4.062	3.063	1.415	2.968	3.351	1.318	1.441	3.849	2.639	1.074		0.110	0.232	0.822	0.393	n/c	n/c	-
APV-B	0.909	3.558	5.515	2.400	3.247	2.687	3.437	2.451	1.073	3.429	3.145	1.455	1.449	4.084	2.652	0.934	0.407		0.167	0.689	0.231	n/c	n/c	-
APV-C	1.401	3.758	10.322	2.947	2.985	2.960	4.231	2.594	1.425	3.592	2.690	1.706	1.559	4.044	2.964	1.370	0.908	0.668		0.846	0.394	n/c	n/c	-
APMV-20	2.266	1.061	3.545	2.071	2.164	1.090	1.927	0.782	2.083	0.719	1.637	2.720	3.097	1.722	1.359	2.225	2.742	2.468	2.855		0.629	n/c	n/c	-
APMV-21	0.856	2.584	5.744	3.799	4.218	3.031	3.006	2.343	0.400	2.465	2.028	1.341	1.242	3.349	2.033	1.070	1.618	1.021	1.615	2.190		n/c	n/c	-

Table 5. Estimates of Evolutionary Divergence over Sequence Pairs between Groups, analyzed for W proteins of 21 APMV species. The number of amino acid substitutions per site from averaging over all sequence pairs between groups is shown. Standard error estimate(s) are shown above the diagonal. Analyses were conducted using the Dayhoff matrix based model. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). Evolutionary analyses were conducted in MEGA7.

Species	V -complete dN/dS ratio	W -complete dN/dS ratio	P-V-W shared N terminal region dN/dS ratio	V -C terminal region dN/dS ratio	W -C terminal region dN/dS ratio
APMV-1	0.435	0.583	0.604	0.660	0.893
APMV-2	0.637	0.115	0.459	0.000	—
APMV-3	0.489	0.651	0.212	1.092	0.437
APMV-4	0.327	0.776	0.636	0.987	—
APMV-5	0.450	0.646	0.534	0.631	0.297
APMV-6	0.513	0.600	0.582	0.935	0.000
APMV-8	0.439	0.548	0.529	—	—
APMV-10	0.388	0.684	0.534	0.00	0.00
APMV-13	0.398	0.320	0.502	0.143	0.00

Table 6. Selection Pressure Analysis of V and W proteins of 21 APMV species. Species wise calculation of dN/dS ratios. Only those species comprising of more than one strain were included for the substitution calculation. The dN/dS ratio was estimated by DnaSP v6.12.03[74]. The dN/dS value was zero in those where the number of non-synonymous substitutions was zero and the dN/dS value could not be determined (−) in those where the number of synonymous substitutions were zero.

Species	No. of strains	V protein substitution rate	W protein substitution rate
APMV-1	12	5.15×10^{-4}	1.00×10^{-4}
APMV-2	5	3.38×10^{-2}	5.51×10^{-2}
APMV-6	8	2.20×10^{-2}	2.22×10^{-2}
APMV-8	5	7.89×10^{-3}	8.21×10^{-3}
APMV-13	3	2.45×10^{-2}	9.64×10^{-2}
Total	**55 (21 APMV species)**	$\mathbf{7.37 \times 10^{-5}}$	$\mathbf{8.07 \times 10^{-5}}$

Table 7. Evolutionary Rate analysis by Molecular Clock- Estimated nucleotide substitution rates for V and W nucleotide sequences of all 21 APMV species. The best clock determined was uncorrelated relaxed clock with lognormal. The BEAST analysis could not done in species with less than 3 strains (APMV-3,-4,-5,-7,-9,-11,-12,-14,-15,-16,-17,-18,-19,-20 and -21) and data for APMV-10 could not be determined because APMV-10 strains were 98.65% to 100% identical to each other. The final data analysis was performed using tracer v 1.7.1 software.

described previously[12]. The exceptions were one strain of APMV-1 (KJ736742.1), APMV-4, APMV-7 and APV-C, and all of them had shorter W protein length. This genetic diversity seen in the W proteins may determine the degree of pathogenesis, variable interferon antagonistic activity and the wide host range exhibited by the APMV species.

The likelihood of W mRNA occurrence could be less than that of V mRNA, because of two unstable base pairing created by the two mismatches (2G) during polymerase stuttering. This skepticism becomes more compelling and doubts rise to whether W protein is expressed at all in those APMV species whose predicted W protein sequences have only fewer amino acid residues in their C terminal region- single (in APMV-7) or two (in APMV-4, APMV-13 & APMV-15) or three (in APMV-1 isolate R75/98) or four (in APMV-6, JX522537.1 & APV-C) or five (in APMV-5 & APMV-16) amino acids. However, equal or higher frequencies of insertions of

1G and 2G during RNA editing have been reported in certain paramyxoviruses such as Nipah virus and Bovine Parainfluenza virus type 3[89,90].

The presence of W mRNA of APMV-1 was first accounted in 1993 with a frequency of about 10%[12], and the W protein expression from APMV-1 lentogenic strain Clone 30 and from APMV-1 lentogenic strain La Sota and velogenic strain SG10 was recently confirmed[16,91]. We had earlier shown that W protein of APMV-1 mesogenic strain Komarov compartmentalized in the nucleus using plasmid system[30], while the same has also been documented during virus infection in cells in the above two studies. Here, we report NLS and NES of W proteins predicted only in certain APMV species, also, we could identify NLS only in five out of twelve strains of APMV-1 implying that the not all the W proteins of APMV-1 strains localize in the nucleus. The W protein sequence analysis of nearly 1000 strains of APMV-1 in our lab show variations in the W protein length between strains (unpublished data),which was also reported recently in an analysis of 286 strains of NDV[91], furthermore, W proteins of only about 50% of the strains analyzed by us are predicted to localize into the nucleus (data not shown) leading us to speculate that these differences in W proteins can attribute to the wide spectrum of pathogenicity and virulence observed in Newcastle disease.

Among paramyxoviruses, the W protein of Nipah and Hendra viruses, are the most well characterized. The nuclear localization of W protein of Nipah virus was found to modify p53 expression and activity[92], sequester inactive STAT1 within nucleus[93], prevent IRF3 phosphorylation, inhibit IFN signaling mediated both by the virus and TLR-3[29], modulate host immunity, influence the disease course and viral pathogenesis specifically neurovirulence[27,28,94]. Intriguingly, neither the lack of W protein nor its cytoplasmic localization in APMV-1 strain clone 30 had any effect on viral replication in cell culture[16]. Though no conserved motifs could be identified between the W proteins of APMVs and Nipah virus, it will be interesting to study if similar roles are executed by W proteins of any or all the APMV species. To our knowledge, this is the first comprehensive and comparative evolutionary study of the *P* gene edited accessory viral proteins of APMVs. The information obtained by this study will enable designing future studies to understand the specific functions of conserved motifs/amino acids of V and W proteins and decipher their evolutionary significance on the virus and as well as on the host.

References

1. Amarasinghe, G. K. *et al*. Taxonomy of the order Mononegavirales: update 2019. *Archives of Virology* **164**, 1967–1980 (2019).
2. Aziz-ul-Rahman, M. M. & Shabbir, M. Z. Molecular Phylogenetics and Evolution Comparative evolutionary and phylogenomic analysis of Avian avulaviruses. *Molecular Phylogenetics and Evolution* **127**, 931–951 (2018).
3. OIE. Infection with Newcastle disease virus. In OIE - Terrestrial Animal Health Code. Preprint at, http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/current/chapitre_nd.pdf (2019).
4. Lamb, R. A. & Parks, G. D. Paramyxoviridae: the viruses and their replication (Fields vir; & P. M. H. (Eds. B. N. Fields, D. N. Knipe, ed.). Lippincott, Williams, and Wilkins (2007).
5. Millar, N. S. & Emmerson, P. T. Molecular Cloning and Nucleotide Sequencing of Newcastle Disease. *Virus. Newcastle Disease* **8**, 79–97 (1988).
6. Calain, P. & Roux, L. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *Journal of General Virology* **67**, 4822–4830 (1993).
7. Kolakofsky, D. *et al*. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *Journal of Virology* **72**, 891–899 (1998).
8. Chang, P. C. *et al*. Complete nucleotide sequence of avian paramyxovirus type 6 isolated from ducks. *Journal of General Virology* **82**, 2157–2168 (2001).
9. Chambers, P. & Samson, A. C. Non-structural proteins in Newcastle disease virus-infected cells. *Journal of General Virology* **58**(Pt 1), 1–12 (1982).
10. Locke, D. P. *et al*. Newcastle disease virus phosphoprotein gene analysis and transcriptional editing in avian cells. *Virus Research* **69**, 55–68 (2000).
11. Mebatsion, T., Verstegen, S., De Vaan, L. T. C., Römer-Oberdörfer, A. & Schrier, C. C. A recombinant Newcastle Disease Virus with low-level V protein expression is immunogenic and lacks pathogenicity for chicken embryos. *Journal of Virology* **75**, 420–428 (2001).
12. Steward, M., Vipond, I. B., Millar, N. S. & Emmerson, P. T. RNA editing in Newcastle disease virus. *Journal of General Virology* **74**, 2539–2547 (1993).
13. Giorgi, C., Blumberg, B. M. & Kolakofsky, D. Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35**, 829–836 (1983).
14. Kolakofsky, D., Vidal, S. & Curran, J. Paramyxovirus RNA Synthesis and P Gene Expression. In K. D.W. (Ed.), The Paramyxoviruses,The Viruses. Springer, Boston, MA (1991).
15. Vidal, S., Curran, J. & Kolakofsky, D. A stuttering model for paramyxovirus P mRNA editing. *The EMBO Journal* **9**, 2017–2022 (1990).
16. Karsunke, J. *et al*. W protein expression by Newcastle disease virus. *Virus Research* **263**, 207–216 (2019).
17. Briand, F. X., Henry, A., Massin, P. & Jestin, V. Complete genome sequence of a novel avian paramyxovirus. *Journal of Virology* **86**, 7710–7710 (2012).
18. Paterson, R. G. & Lamb, R. A. RNA editing by G-nucleotide insertion in Mumps virus P-gene mRNA transcripts. *Journal of Virology* **64**, 4137–4145 (1990).
19. Childs, K. S., Andrejeva, J., Randall, R. E. & Goodbourn, S. Mechanism of mda-5 inhibition by paramyxovirus V proteins. *Journal of Virology* **83**, 1465–73 (2009).
20. Huang, Z., Krishnamurthy, S., Panda, A. & Samal, S. K. Newcastle disease virus V protein is associated with viral pathogenesis and functions as an alpha interferon Antagonist. *Journal of Virology* **77**, 8676–8685 (2003).
21. Park, M. S., García-Sastre, A., Cros, J. F., Basler, C. F. & Palese, P. Newcastle Disease Virus V protein is a determinant of host range restriction. *Journal of Virology* **77**, 9522–9532 (2003).
22. Park, M. S. *et al*. Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. *Journal of Virology* **77**, 1501–1511 (2003).
23. Chu, Z. *et al*. Newcastle Disease Virus V protein inhibits cell apoptosis and promotes viral replication by targeting CacyBP/SIP. *Frontiers in Cellular and Infection Microbiology* **8**, 304 (2018).

24. Wang, C. *et al.* Newcastle disease virus V protein inhibits apoptosis in DF-1 cells by downregulating TXNL1. *Veterinary Research* **49**, 102 (2018).

25. Chu, Z. *et al.* Newcastle disease virus V protein promotes viral replication in HeLa cells through the activation of MEK/ERK signaling. *Viruses* **10**, 489 (2018).

26. Alamares, J. G., Elankumaran, S., Samal, S. K. & Iorio, R. M. The interferon antagonistic activities of the V proteins from two strains of Newcastle disease virus correlate with their known virulence properties. *Virus Research* **147**, 153–157 (2010).

27. Satterfield, B. A. *et al.* The immunomodulating V and W proteins of Nipah virus determine disease course. *Nature Communications* **6**, 7483 (2015).

28. Satterfield, B. A., Geisbert, T. W. & Mire, C. E. Inhibition of the host antiviral response by Nipah virus: current understanding and future perspectives. *Future Virology* **11**, 331–344 (2016).

29. Shaw, M. L., Cardenas, W. B., Zamarin, D., Palese, P. & Basler, C. F. Nuclear localization of the Nipah virus W protein allows for inhibition of both virus-and toll-like receptor 3-triggered signaling pathways. *Journal of Virology* **79**, 6078–6088 (2005).

30. Vaidyanathan, S. P. & Gawai, S. & Subbiah, M. Poster Presentation, ICID 2016: Elucidation of the role of non-structural viral protein (W) of Newcastle disease virus. *International Journal of Infectious Diseases* **45**, 337–338 (2016).

31. Abolnik, C., De Castro, M. & Rees, J. Full genomic sequence of an African Avian Paramyxovirus Type 4 strain isolated from a wild duck. *Virus Genes* **45**, 537–541 (2012).

32. Fereidouni, S. *et al.* Next-generation sequencing of five new avian paramyxoviruses 8 isolates from Kazakhstan indicates a low genetic evolution rate over four decades. *Archives of Virology* **163**, 331–336 (2018).

33. Heiden, S., Grund, C., Höper, D., Mettenleiter, T. C. & Römer-Oberdörfer, A. Pigeon paramyxovirus type 1 variants with polybasic F protein cleavage site but strikingly different pathogenicity. *Virus Genes* **49**, 502–506 (2014).

34. Hiono, T., Matsuno, K., Tuchiya, K., Lin, Z., Okamatsu, M. & Sakoda, Y. Complete genome sequence of the avian paramyxovirus serotype 5 strain APMV-5/budgerigar/Japan/TI/75. *Genome Announcement* **4**, pii: e01005-16 (2016).

35. Jeong, J. *et al.* Complete genome sequence of a novel avian paramyxovirus isolated from wild birds in South Korea. *Archives of Virology* **163**, 223–227 (2018).

36. Karamendin, K. *et al.* Novel avian paramyxovirus isolated from gulls in Caspian seashore in Kazakhstan. *PLoS One* **12**, 1–15 (2017).

37. Karamendin, K. *et al.* Complete genome sequence of avian paramyxovirus strain APMV-6/red-crested pochard/Balkhash/5842/2013 from Kazakhstan. *Genome Announcements* **3**, e00158–15 (2015).

38. Kim, S. H., Subbiah, M., Samuel, A. S., Collins, P. L. & Samal, S. K. Roles of the fusion and hemagglutinin-neuraminidase proteins in replication, tropism, and pathogenicity of avian paramyxoviruses. *Journal of Virology* **85**, 8582–8596 (2011).

39. Kumar, S., Nayak, B., Collins, P. L. & Samal, S. K. Complete genome sequence of avian paramyxovirus type 3 reveals an unusually long trailer region. *Virus Research* **137**, 189–197 (2008).

40. Kumar, S., Nayak, B., Samuel, A. S., Xiao, S., Collins, P. L. & Samal, S. K. Complete genome sequence of avian paramyxovirus-3 strain Wisconsin: Evidence for the existence of subgroups within the serotype. *Virus Research* **149**, 78–85 (2010).

41. Kydyrmanov, A. I. *et al.* Novel avian paramyxovirus isolated from gulls in Caspian seashore in Kazakhstan. *PLoS One. 28* **12**(12), e0190339 (2017).

42. Lee, H. J. *et al.* A novel avian paramyxovirus (Putative Serotype 15) isolated from wild birds. *Frontiers in Microbiology* **8**, 786 (2017).

43. Li, X., Zhang, S., Wang, H., Zhao, J. & Zhang, G. Genomic characterization of two avian paramyxovirus type 2 isolates from chickens in China. *Virus Genes* **43**, 55–59 (2011).

44. Miller, P. J. *et al.* Evidence for a new avian paramyxovirus serotype 10 detected in Rockhopper Penguins from the Falkland Islands. *Journal of Virology* **84**, 11496–11504 (2010).

45. Nayak, B., Kumar, S., Collins, P. L. & Samal, S. K. Molecular characterization and complete genome sequence of avian paramyxovirus type 4 prototype strain duck/Hong Kong/D3/75. *Virology Journal* **5**, 124 (2008).

46. Neira, V. *et al.* Novel avulaviruses in penguins, Antarctica. *Emerging Infectious Diseases* **23**, 1212–1214 (2017).

47. Paldurai, A., Subbiah, M., Kumar, S., Collins, P. L. & Samal, S. K. Complete genome sequences of avian paramyxovirus type 8 strains goose/Delaware/1053/76 and pintail/Wakuya/20/78. *Virus Research* **142**, 144–153 (2009).

48. Tirumurugaan, K. G. *et al.* Genotypic and pathotypic characterization of Newcastle disease viruses from India. *PLoS One* **6**, e28414 (2011).

49. Samuel, A. S., Kumar, S., Madhuri, S., Collins, P. L. & Samal, S. K. Complete sequence of the genome of avian paramyxovirus type 9 and comparison with other paramyxoviruses. *Virus Research* **142**, 10–18 (2009).

50. Samuel, A. S., Paldurai, A., Kumar, S., Collins, P. L. & Samal, S. K. Complete genome sequence of avian paramyxovirus (APMV) serotype 5 completes the analysis of nine APMV serotypes and reveals the longest APMV genome. *PLoS One* **5**, 1–13 (2010).

51. Subbiah, M., Nayak, S., Collins, P. L. & Samal, S. K. Complete genome sequences of avian paramyxovirus serotype 2 (APMV-2) strains Bangor, England and Kenya: Evidence for the existence of subgroups within serotype 2. *Virus Research* **152**, 85–95 (2010).

52. Subbiah, M., Xiao, S., Collins, P. L. & Samal, S. K. Complete sequence of the genome of avian paramyxovirus type 2 (strain Yucaipa) and comparison with other paramyxoviruses. *Virus Research* **137**, 40–48 (2008).

53. Subbiah, M. *et al.* Pathogenesis of two strains of avian paramyxovirus serotype 2, Yucaipa and Bangor, in chickens and turkeys. *Avian Diseases* **54**, 1050–1057 (2010).

54. Terregino, C. *et al.* Antigenic and genetic analyses of isolate APMV/wigeon/Italy/3920-1/2005 indicate that it represents a new avian paramyxovirus (APMV-12). *Archives of Virology* **158**, 2233–2243 (2013).

55. Thampaisarn, R. *et al.* Characterization of avian paramyxovirus serotype 14, a novel serotype, isolated from a duck fecal sample in Japan. *Virus Research* **228**, 46–57 (2017).

56. Thomazelli, L. M. *et al.* Novel avian paramyxovirus (APMV-15) isolated from a migratory bird in South America. *PLoS One* **12**, 1–7 (2017).

57. Tian, Z. *et al.* Complete nucleotide sequence of Avian Paramyxovirus type 6 strain JL isolated from mallard ducks in China. *Journal of Virology* **86**, 13112 (2012).

58. Tsunekuni, R., Ito, H., Otsuki, K., Kida, H. & Ito, T. Genetic comparisons between lentogenic Newcastle disease virus isolated from waterfowl and velogenic variants. *Virus Genes* **40**, 252–255 (2010).

59. Wu, W. *et al.* Molecular and antigenic characteristics of Newcastle disease virus isolates from domestic ducks in China. *Infection, Genetics and Evolution* **32**, 34–43 (2015).

60. Xiao, S. *et al.* Complete genome sequence of avian paramyxovirus type 7 (strain Tennessee) and comparison with other paramyxoviruses. *Virus Research* **145**, 80–91 (2009).

61. Xiao, S. *et al.* Complete genome sequences of avian paramyxovirus serotype 6 prototype strain Hong Kong and a recent novel strain from Italy: Evidence for the existence of subgroups within the serotype. *Virus Research* **150**, 61–72 (2010).

62. Yamamoto, E., Ito, H., Tomioka, Y. & Ito, T. Characterization of novel avian paramyxovirus strain APMV/Shimane67 isolated from migratory wild geese in Japan. *The Journal of Veterinary Medical Science* **77**, 1079–1085 (2015).

63. Aziz-ul-Rahman & M. Z. Shabbir. One (1) new species in the genus Avulavirus (Mononegavirales: Paramyxoviridae). Taxonomic proposal document submitted to *ICTV*, code assigned 2019.014M (2019).

64. Armougom, F. *et al.* Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research* **34**, W604–W608 (2006).

65. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *Journal of Molecular Biology* **302**, 205–217 (2000).

66. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).
67. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**, D222–D226 (2015).
68. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011).
69. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. In *Proceedings of the National Academy of Sciences* **106**, 10171-10176 (2009).
70. la Cour, T. *et al.* Analysis and prediction of leucine-rich nuclear export signals. *Protein Engineering, Design and Selection* **17**, 527–536 (2004).
71. Xu, D. *et al.* LocNES: a computational tool for locating classical NESs in CRM1 cargo proteins. *Bioinformatics* **31**, 1357–1365 (2015).
72. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices. *Computer Applications in the Biosciences* **8**, 275–282 (1992).
73. Schwartz, R. M. & Dayhoff, M. O. Matrices for detecting distant relationships. *In Atlas of protein sequence and structure* **5**, 353–358 (1979).
74. Felsenstein, J. Confidence Limits On Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**, 783–791 (1985).
75. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).
76. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution* **34**, 3299–3302 (2017).
77. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* **4**, vey016 (2018).
78. Karlin, D. & Belshaw, R. Detecting remote sequence homology in disordered proteins: Discovery of conserved motifs in the N-termini of mononegavirales phosphoproteins. *PLoS One* **7**, e31719 (2012).
79. Kolakofsky, D. Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: A review. *Virology* **498**, 94–98 (2016).
80. Iseni, F. *et al.* Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA* **8**, 1056–1067 (2002).
81. Hausmann, S., Garcin, D., Morel, A. S. & Kolakofsky, D. Two nucleotides immediately upstream of the essential A6G3 slippery sequence modulate the pattern of G insertions during Sendai virus mRNA editing. *Journal of Virology* **73**, 343–351 (1999).
82. Hausmann, S., Garcin, D., Delenda, C. & Kolakofsky, D. The versatility of paramyxovirus RNA polymerase stuttering. *Journal of Virology* **73**, 5568–5576 (1999).
83. Jacques, J. P., Hausmann, S. & Kolakofsky, D. Paramyxovirus mRNA editing leads to G deletions as well as insertions. *The EMBO Journal* **13**, 5496–503 (1994).
84. Mebatsion, T., Verstegen, S., De Vaan, L. T., Römer-Oberdörfer, A. & Schrier, C. C. A recombinant Newcastle disease virus with low-level V protein expression is immunogenic and lacks pathogenicity for chicken embryos. *Journal of Virology* **75**, 420–428 (2001).
85. Mebatsion, T., de Vaan, L. T., de Haas, N., Römer-Oberdörfer, A. & Braber, M. Identification of a mutation in editing of defective Newcastle disease virus recombinants that modulates P-gene mRNA editing and restores virus replication and pathogenicity in chicken embryos. *Journal of Virology* **77**, 9259–9265 (2003).
86. Ramachandran, A. & Horvath, C. M. Dissociation of paramyxovirus interferon evasion activities: universal and virus-specific requirements for conserved V protein amino acids in MDA5 interference. *Journal of Virology* **84**, 11152–11163 (2010).
87. Steward, M. & Samson, A. C. & Errington W, E. P. The Newcastle disease virus V protein binds zinc. *Archives of Virology* **140**, 1321–1328 (1995).
88. Qiu, X. *et al.* Newcastle Disease Virus V Protein Targets Phosphorylated STAT1 to Block IFN-I Signaling. *PLoS One* **11**, e0148560–e0148560 (2016).
89. Kulkarni, S. *et al.* Nipah Virus edits its P Gene at High Frequency to Express the V and W Proteins. *Journal of Virology* **83**, 3982–3987 (2009).
90. Pelet, T., Curran, J. & Kolakofsky, D. The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *The EMBO Journal* **10**, 443–448 (1991).
91. Yang, Y. *et al.* Appropriate amount of W protein of avian avulavirus 1 benefits viral replication and W shows strain-dependent subcellular localization. *Virology* **538**, 71–85 (2019).
92. Martinez-Gil, L., Vera-Velasco, N. M. & Mingarro, I. Exploring the Human-Nipah Virus Protein-Protein Interactome. *Journal of Virology* **91**, pii: e01461-17. (2017).
93. Ciancanelli, M. J., Volchkova, V. A., Shaw, M. L., Volchkov, V. E. & Basler, C. F. Nipah Virus Sequesters Inactive STAT1 in the Nucleus via a P Gene-Encoded Mechanism. *Journal of Virology* **83**, 7828–7841 (2009).
94. Yoneda, M. *et al.* The nonstructural proteins of Nipah virus play a key role in pathogenicity in experimentally infected animals. *PLoS One* **5**, e12709–e12709 (2010).

Acknowledgements

Author contributions

M.S. conceived and designed the study, P.L.R. collected the data, M.S., P.L.R. and R.G. analyzed the data, M.S., P.L.R., R.G. drafted the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-66252-x.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.