Accepted Manuscript

Comparative evolutionary and phylogenomic analysis of *Avian avulaviruses* 1 to 20

Aziz-ul-Rahman, Muhammad Munir, Muhammad Zubair Shabbir

Please cite this article as: Aziz-ul-Rahman, Munir, M., Zubair Shabbir, M., Comparative evolutionary and phylogenomic analysis of *Avian avulaviruses* 1 to 20, *Molecular Phylogenetics and Evolution* (2018), doi: https://doi.org/10.1016/j.ympev.2018.06.040

Comparative evolutionary and phylogenomic analysis of *Avian avulaviruses* 1 to 20

Aziz-ul-Rahman[1,3], Muhammad Munir[2], Muhammad Zubair Shabbir[3#]

[1]Department of Microbiology University of Veterinary and Animal Sciences, Lahore 54600,

Pakistan https://orcid.org/0000-0002-3342-4462

[2]Division of Biomedical and Life Sciences, Furness College, Lancaster University, Lancaster LA1

4YG United Kingdomhttps://orcid.org/0000-0003-4038-0370

[3]Quality Operations Laboratory University of Veterinary and Animal Sciences 54600 Lahore,

Pakistan https://orcid.org/0000-0002-3562-007X

Corresponding author:

Muhammad Zubair Shabbir

E. mail: shabbirmz@uvas.edu.pk

Abstract

Avian avulaviruses (avulaviruses or AAvVs) infect a wide range of avian species worldwide with variable clinical outcomes and economic impacts. Owing to broad host spectrum, several novel avulaviruses are being reported from both wild and domesticate birds that highlight the potential of the virus to evolve, adapt and emerge in susceptible population. Pathobiological and phylogenetic characterizations of individual avulaviruses are often demonstrated, however, a cumulative and comparative assessment of avulaviruses remains elusive. To assess evolutionary dynamics and potential emergence of novel avulaviruses, we enriched existing databases of all known avulaviruses (specie-type 1-20), and determined their genomics features based on both complete genomes and individual complete genes. While a high nucleotide divergence (up to 65.4%) was observed among avulaviruses, phylogenomic analysis revealed clustering of all avulaviruses into three distinct clades. The major clade (Clade-I) included both oldest and newest avulaviruses (2, 5, 6, 7, 8, 10, 11, 14, 15 and 20) and the second clade (Clade-II) consisted of avulaviruses 1, 9, 12, 13, 16, 17, 18 and 19, whereas the third clade (Clade-III) carried only avulaviruses 3 and 4. Intriguingly, clustering pattern was descriptive for individual gene-based analysis, however, the hemagglutinin-neuraminidase (*HN*) and polymerase (*L*) genes showed clear and discrete branching patterns similar to complete genome-based clustering. Therefore, we propose the use of *HN, or L* genes or complete genome to study epidemiological aspects of the avulaviruses. Genomic and residue characteristics of all genes indicated a continuous evolution of the virus, and substitutions in biologically important motifs warrant future investigations to assess their roles in the pathobiology of the virus. Taken together, this comprehensive analysis of all known avulaviruses ascertains continuous monitoring and surveillance of wild/water-fowls and commercial poultry. These findings further our understanding on the evolutionary dynamics and potential emergence of novel avulaviruses and will establish bases to identify potential of wild-bird origin apathogenic viruses to cause infections in commercial poultry.

1. Introduction

Avian avulaviruses (AAvVs or avulaviruses), formerly known as Avian paramyxoviruses (APMVs), are classified in the genus *Avulavirus* of subfamily *Paramyxovirinae* within the family *Paramyxoviridae* under order *Mononegavirales* (Amarashinge et al., 2017). All avulaviruses are enveloped, non-segmented, and single-stranded RNA viruses with a genome length ranging from 13-17 Kbs. The virus genome encodes six non-overlapping structural proteins in the order of 3′-NP-P/V/W-M-F-HN-L-′5, and additionally two non-structural proteins (V/W) may be expressed by RNA-editing mechanism of the phosphoprotein (*P*) gene (Lambs and Park, 2007). Two surface glycoproteins, fusion (F) and hemagglutinin-neuraminidase (HN), carry fusogenic and hemagglutinin capabilities, respectively. The existence of mono-/multi-basic amino acids at the F protein cleavage site (Fcs) is a key indicator of virulence and based on the nature of Fcs, avulaviruses are classified into velogenic (highly pathogenic), mesogenic (intermediate pathogenic) or lentogenic (low pathogenic) strains (Liu et al., 2018). The HN protein plays a prime role in tissue tropism and contributes significantly in the virulence (Huang et al., 2004). The nucleoprotein (NP), phosphoprotein and large polymerase (L) proteins act as polymerase co-factors and viral RNA-dependent RNA-polymerase, which governs the virus replication (Lamb and Parks, 2007). The envelope of the virus is underlined by a matrix protein (M), which helps in budding and assembly of viral particles (Lamb and Parks, 2007). All avulaviruses genes are annotated by conserved gene-start (GS) and gene-end (GE) sequences which orchester transcription in a sequential manner by start-stop mechanism (Lamb and Parks, 2007). Between coding regions of genes, there exist non-coding intergenic sequences (IGS) at gene boundaries, which are not copied to mRNAs (Lamb and Parks, 2007). Contribution of GE, GS and IGS have been defined in the replication kinetics of avulavirus type I, however, importance of structural and functional domains of each of individual gene features along with potential influence of substitutions at substantial and conserved motifs in other avulaviruses warrant future investigations.

A number of avulaviruses have been reported from different parts of the globe, implying the presence of additional, yet-unreported, novel strains in potential natural reservoirs. In this regard, a total of 20 avulaviruses have been reported in public database. Among them, the avulaviruses 1-9 were identified before the 80s; 10-13 were reported up to 2015, whereas 14, 15, 16, 17, 18, 19 and 20 are novel avulaviruses. These novel and emerging avulaviruses are reported recently from wild/water fowls (Thomazelli et al., 2017; Lee et al., 2017; Neira et al., 2017; Karamendin et al., 2017). Interestingly, these avulaviruses (10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 in particular) were isolated during avian influenza (AI) surveillance in wild birds (Miller et al., 2010; Briand et al., 2010; Terregino et al., 2013; Yamamoto et al., 2015; Thampaisan et al., 2017; Thomazelli et al., 2017; Lee et al., 2017; Neira et al., 2017; Karamendin et al., 2017). Therefore, opportunities gained through AI surveillance have significantly elucidated the ecology of avulaviruses in their natural reservoirs; the wild/water-fowl species. Understanding the spectrum of avian avulaviruses proposed an unexplored potential of waterfowls in transmitting and causing infection in domesticated poultry (Oslen et al., 2006).

The emergence of novel avulaviruses, year-to-year on-going variations in genome, and the virulent, either low or high, implies that distinct avulaviruses are simultaneously evolving among a wide range of avian hosts, across the globe (Miller et al., 2009; Aldous et al., 2014). A vast majority of previously published literature is limited to biologic and genomic characteristics of individual avulaviruses, particularly avulaviruses 1 (Kumar et al., 2008; Palduri et al., 2009; Samuel et al., 2009; Subbiah et al., 2010; Abolnik et al., 2012; Samel et al., 2010; Miller et al., 2010; Briand et al., 2012; Shabbir et al., 2012a; Shabbir et al., 2012b; Munir et al., 2012a; Munir et al., 2012b; Tian et al., 2012; Shabbir et al., 2013; terregino et al., 2013; Yamamoto et al., 2015; Akhtar et al., 2016; Shabbir et al., 2016; Shabbir et al., 2018; Thampaisarn et al., 2017; Thomazelli et al., 2017; Lee et al., 2017; Neira et al., 2017; Karamendin et al., 2017). Therefore, it is imperative to comprehensively analyse all known avian avulaviruses (specie-type 1-20) and to comparatively assess inter-avulaviruses differences based on not only complete genome but also individual genes.

These evolutionary assessments on the genetic diversity of circulating avulaviruses particularly newly identified isolates will provide baseline information towards genome characteristics and evolutionary dynamics of avulaviruses across the globe.

2. Methodology

2.1. Sequence database

The full-length nucleotide sequences of representative avulaviruses (specie-type 1-20) were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). In cases where essential information in public domains were missing regarding newly reported strains of avulaviruses, data was acquired from a member of The International Committee on Taxonomy of Viruses (ICTV) (Dr B. Rima, personal communication, April 2018). Except for avulaviruses 1, where we retrieved a total of five sequences each representing diverse regions of the world, we obtained all available complete genome sequences of avulaviruses 2 to 20 for phylogenetic, comparative residues substitutions at important motif/s and recombination analysis (Table 1). However, for all other analysis including genomic features, percent nucleotide identity or divergence, CDS polymorphism and natural pressure selection sites for synonymous and non-synonymous substitution sites, one representative strain from each avulaviruses was considered. The recovered sequences are available in the GenBank database under accession numbers: Avian avulavirus 1; KU885948, Avian avulavirus 2; HM159993, Avian avulavirus 3; EU403085, Avian avulavirus 4; JX133079, Avian avulavirus 5; GU206351, Avian avulavirus 6; JX522537, Avian avulavirus 7; FJ231524, Avian avulavirus 8; FJ215863, Avian avulavirus 9; EU910942, Avian avulavirus 10; HM755886, Avian avulavirus 11; JQ886184, Avian avulavirus 12; KC333050, Avian avulavirus 13; KU646513, Avian avulavirus 14; KX258200, Avian avulavirus 15; KX932454 Avian avulavirus 16; KY511044, Avian avulavirus 17; KY452442, Avian avulavirus 18; KY452443, Avian avulavirus 19; KY452444, Avian avulavirus 20; MF033136. Here, it is important to indicate that only one isolate and subsequent complete genome sequence data is available so-far in public database for avulaviruses 7, 9, 11, 12, and 14-20 (Table 1).

2.2. Phylogenetic, evolutionary and comparative residue analysis

Clustal W algorithm was used to align the recovered sequences in BioEdit® version 5.0.6 (Hall, 1999). After alignment, phylogenetic relationships between whole genomes of virus strains as well as complete *NP, P, M, F, HN* and *L* genes was determined using distance-based neighbour-joining method in MEGA® version 6.0 (Tamura et al., 2013). In addition, in order to reveal the genus based phylogenetic relationship, previously reported full-length genomes of viruses representing closely related genera within the family *Paramyxovirinae* (*Metapneumovirus*, *Respiroviruse* and *Aquaparamyxovirus*) were also retrieved from NCBI database and used as out-groups. For phylogenomic analysis of the complete genome, Splits Tree Program v4.95 was employed using Neighbor-Net graph method based on the pairwise distance estimated by uncorrected *p*-distance and angle split transformation settings (Huson and Bryant, 2006). The reliability of representative avulaviruses isolates for nucleotide sequence comparison (identity and divergence) was confirmed using Pairwise Sequence Comparisons (PASC). The PASC analysis was performed on the entire genome sequences of all avulaviruses strains using MEGA® version 6.0 (Tamura et al., 2013). The conserved functional domains and substitutions in previously reported significant motifs were predicted through UniProt analysis available at https://www.expasy.org/. The RNAfold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) was used online to predict potential RNA-editing sites in the *P* gene (Hofacker, 2003).

2.3. Nucleotide diversity and evidence of nature of the selection

The nucleotide diversity among coding DNA sequences (CDS) of representative 20 avulaviruses was assessed for DNA polymorphism on the basis of variable sites and mutations, and average numbers of pairwise nucleotide differences using DnaSP version 5.10.01 (available at http://www.ub.es/dnasp) (Librado and Rozas, 2009). To assess the departure from neutrality in all isolates, Tajima's D statistical method was applied (Tajima, 1989). Based on synonymous and non-synonymous substitution sites, Data-monkey adaptive evolution server (http://www.datamonkey.org/) was applied to evaluate the nature of pressure selection among CDS

of all isolates (Delport et al., 2010). The positive and negative selection sites under natural selection were determined through three different genetic algorithms including Single Likelihood Ancestor Counting (SLAC), Fixed Effect Likelihood (FEL) and Internal Branch Fixed Effect Likelihood (IFEL) with $p= 0.05$.

2.4. Recombination analysis

The inter- and intra-strain recombination analysis of CDS of all representative 20 avulaviruses were conducted using three different tools known as SimPlot version 3.5.1 (Ray, 2013), GARD (http://www.datamonkey.org/GARD) and RDP version 4.70 (Martin et al., 2015). The CDS of vaccine strain, LaSota isolate (accession number: AY845400), was used as a query isolate in SimPlot. Percentage identity of query sequence to a panel of reference CDS sequences of known avulaviruses was determined. BootScan analysis and maximum χ^2 method (implementing GARD online) were used to assess the likelihood of a locus for recombination events and detection of putative breakpoint within avulaviruses (Salminen et al., 1995). Utilizing several recombination detection methods into single suite of tool, The RDP package is considered a fast, simple and sensitive method for identification of putative recombination breakpoints. Six various recombination algorithm methods (RDP, GENECONV, BootScan, MaxChi, Chimaera and SiScan) were used to identify putative recombinant and parent isolates at p <0.001 value. Putative recombination events were assumed to have occurred only/true recombinant, when they were consistently identified by at least four of the above-mentioned six algorithms at a probability threshold of 0.05.

3. Results

3.1. Genomic features of avulaviruses

The complete genome length of all avulaviruses ranged from 14,904 to 17,412 nt, consisting of six structural proteins genes in the order of 3′-*NP-P-M-F-HN-L-* ′5. In addition to conventional pattern of these genes, AAvV 6 carries a small hydrophobic (SH) protein, which is otherwise a characteristic of metapneumoviruses (Chang et al., 2001). Two non-structural proteins, namely

V/W, were found only in 2, 3, 5, 11, 14, 15 and 20 avulaviruses. Among all species-types, avulavirus 11 showed a maximum genome length of 17,412 nt followed by 5 (17,262 nt), 3 (16,272 nt), 6 (16,236 nt), 13 (15,996 nt), 20 (15,786), 7 (15,480 nt), 10 (15,456 nt), 14 (15,444 nt), 9 (15,438 nt), 8 (15,342 nt), 12 (15,312 nt), 1 (15, 172 nt), 16 (15,180 nt), 4 (15,054 nt), 19 (15,017 nt), 15 (14,952 nt), 18 (14,931), 17 (14,926 nt) and 2 (14,904 nt). An additional small hydrophobic (SH) region between *F* and *HN* genes, with a nucleotide length of 629 nt, was exclusive to specie-type 6. All avulaviruses shared similar length of leader (55 nt), however, the size of the trailer varied from 17 nt to 776 nt. It was found to be 17 nt long for avulavirus 4, 47 nt for 9 and 16, 54 nt for 6, 81 nt for 20, 114 nt for 1, 127 nt for 7, 145 nt for 17, 154 nt for 2, 171 nt for 8, 204 nt for 12, 206 nt for 10, 226 nt for 18, 254 nt for 19, 277 nt for 14, 552 nt for 5, 707 nt for 3 and 776 nt for 13. Few variations in gene-start (GS) and gene-end (GE) nucleotide sequences (negative sense) were also observed among all avulaviruses. A comparative analysis of whole genome of all known avulaviruses has been summarized in Table 2.

3.2. Phylogenetic and evolutionary analysis

The phylogenetic analysis revealed a distinct and characteristic classification pattern. All avulaviruses clustered together in a genus *Avulavirus,* when compared with other closely related viruses as an out-group (Fig. 1). Based upon relatedness among genome characteristics, three main clades were observed within the genus *Avulavirus.* The Clade-I constituted a biggest group of isolates and consisted of ten avulaviruses including 2, 5, 6, 7, 8, 10, 11, 14, 15 and 20 strains, whereas, eight avulaviruses including 1, 9, 12, 13, 16, 17, 18 and 19 strains were clustered together in clade-II. On the other hand, only two avulaviruses (3 and 4) clustered in clade-III (Fig. 1 and 2). As expected, different isolates originating from same avulavirus were also clustered within the same clade. Additionally, the Split Tree analysis was conducted to verify the clustering pattern, and a pattern similar to phylogenetic analysis was noticed. Moreover, a comparable genetic diversity was also noticed by PASC analysis (Table 3, Fig. 2). Based on these clustering patterns, it is plausible to propose that avulaviruses belonging to clade-III are ancestral, and Clade-I and II represent high

inter-clade similarities compared to clade-III. Additionally, identification of novel avulaviruses is not periodic, merely represents evolutionary links and these viruses are named according to their identification or reporting patterns. Therefore, similar to classification of avulaviruses 1, the nomenclature of all reported avulaviruses needs revision.

We next assessed a comprehensive analysis of individual genes to delineate the epidemiological association among avulaviruses. While minor variations were observed in *NP, P, M, F* genes-based clustering patterns (Fig. 3A, 3B, 3C, 3D, 3E, 3F), phylogenetic analysis of individual *HN* and *L* genes showed similar clustering patterns to complete genome analysis (Fig. 3). This comparable clustering of avulaviruses highlights the need to consider both *HN* and complete genome (in addition to *F* gene which is currently being used) to effectively establish epidemiological linking between known and future isolates.

Direct comparison of nucleotide identity between avulaviruses 1-20 indicated a maximum (88.4%) nucleotide identity between avulaviruses 17 and 18, followed by 83.6% between 18 and 19, 82.9% between 17 and 19, 79.7% between 10 and 20, 79.4% each between 8 and 20, and between 2 and 20, 77.3% between 15 and 20, 71.9% between 1 and 19, 71.8% between 1 and 18, 71.7% between 12 and 18, 71.5% each between 12 and 17, and between 18 and 16, 71.3% 16 and 19, 70.6% each between 12 and 19, and between 16 and 17, 70.4% each between 13 and 17, and between 13 and 18, 69.6% between 9 and 19, 69.3 between 9 and 17, 68.3% between 9 and 18, 66.1% between 11 and 20, 65.4% between 5 and 20, 65.1% 7 and 20, 64.7% 1 and 16, 64.5% between 14 and 20, 60.8% between 6 and 20 (Table 3). On the other hand, a high nucleotide identity (96-99.8%) within an avulavirus specie-type was observed (data not shown).

3.3. Comparative residue analysis

3.3.1. F and HN protein analysis

A typical cleavage site of monobasic or polybasic residues was identified in all avulaviruses (Fig. 4), however it was enriched with arginine (R) in avulaviruses 1 and 3 whereas a high number of lysines (K) was found in avulavirus 5. Next to the cleavage motif, phenylalanine (F) was observed

in avulaviruses 1, 2, 4, 5, 7 and 11, and isoleucine (I) in avulaviruses 9, 10, 17 and 19. In rest of all reported avulaviruses, a leucine (L) residue was observed in the corresponding amino acid position. Additional substitutions were noted in the conserved fusion peptide motif of all avulaviruses except for avulaviruses 12, 13 and 16. Non-synonymous substitutions were also observed in hypervariable regions of fusion protein of all avulaviruses; avulaviruses 1 and 16, 17 and 18 were most identical whereas avulavirus 15 was found to be most divergent (Table 4).

Regarding the HN protein, a highly conserved hexa-peptide motif was identified among all avulaviruses. However, variations were observed in the hydrophobic signal anchor domain in all avulavirues, a maximum identity was noticed among avulaviruses 1, 8 and 16. The hemagglutinin active motif-I was highly conserved in all avulaviruses except a replacement of phenylalanine (F) with tyrosine (Y) was noticed in avulavirus 16 at first position of the motif. Similar patterns of substitutions at two different positions was identified for hemagglutinin active motif-II; glycine (G) was replaced by alanine (A) at first position in avulaviruses 6 and 8, while alanine (A) was observed instead of glycine/serine (G/S) at fourth position in avulaviruses 2, 3, 9, 12 and 17. Residues in hypervariable regions of HN protein were similar for avulaviruses 1, 9, 16, whereas a number of non-synonymous substitutions were observed for the rest of avulaviruses with significant divergence in avulaviruses 11, 15 and 20 (Table 5).

3.3.2. NP, P, and L proteins analysis

A highly conserved N-self-assembly motif ($F-X_4YX_3YXYAMG$; where X denoted any amino acid and Y denoted an aromatic amino acid) was observed in NP protein of all avulaviruses. The motif was highly conserved in majority of avulaviruses such as avulaviruses 3, 5, 6, 7, 8, 9, 10, 11, 14, 15, 17, 18, 19 and 20. However, in avulaviruses 2, 4 and 12, the tyrosine (Y) was replaced by phenylalanine (F) at the first position of motif. In avulaviruses 1, 13 and 16, same substitution was repeated at 4th position. Two additional unique substitutions (tyrosine to phenylalanine at 12th position and alanine to serine at position 13 of the motif) were observed in avulavirus 13. The pattern of RNA editing (negative sense) in phosphoprotein was identical for avulaviruses 1, 2, 4, 5,

6, 7, 8, 9, 10, 12, 13, 15, 17, 18, 19 and 20, whereas variations were observed in avulaviruses 3, 11, 14 and 16 (Table 6). These variations may highlight the magnitude of non-structural proteins expression, which can reflect upon the variation in pathogenicity among different avulaviruse. The four catalytic domains participate in the synthesis of co-factors for RNA-dependent RNA-polymerase and one ATP binding motif was conserved in large/polymerase (L) proteins. Domain -I was found to be identical in avulaviruses 1, 3, 9, 16 and 12, 13. For each of avulaviruses 1 and 16, 2 and 15, 5, 7, 8 and 12, and 17and 18, the domain-II was identical, whereas variations were identified in rest of the avulaviruses. Similarly, domain-III was identical in each of avulaviruses 1, 9, and 16, 12 and 13, and 17 and 18, whereas, domain-IV was identical in avulaviruses 1 and 16. Involved in polymerase activity, a motif (QGDNQ) between 4th to 8th positions was found to be highly conserved in domain-III of all sequences. The avulaviruses 2, 5, 8, 9,10,11,12,13,14, 15 and 20 carry identical ATP binding motif, avulaviruses 1, 17, 18 and 19 were similar to each other, while same motif had lysine (K) replaced by arginine (R) in avulaviruses 3 and 4 (Table 6).

3.3.3. M protein analysis

The M protein of all known avulaviruses comprised of a highly conserved potential late domain motif (FPIV) except avulaviruses 15, 17, 18 and 20 where two variations (I→L at 3rd position and V→I at 4th position) were observed. The functional contribution of these mutations in selective avulaviruses remained unaddressed. Additionally, a putative bipartite motif and a nuclear localization signal sequence in the M protein were identified in all representative avulaviruses, indicating the functional importance of these features in the functional regulation of viruses. The motif was much enriched with arginine (R) and lysine (K) residues in avulaviruses 1, 3, 16 and 20 (Table 7).

3.4. Nucleotide diversity and evidence of selection sites

The average nucleotide diversity for complete genome sequences of avulaviruses 1 to 20 was found to be 0.51501. On the other hands, the average nucleotide differences among all haplotypes was found to be k=6682.247. Based on the individual coding genes, DnaSP analysis

revealed a total of 24,347 mutations within 11,381 sites. Among these mutations, a total of 1,211 were monomorphic sites whereas 10,170 were polymorphic. The polymorphic sites consisted of 676 singleton variable sites with 9,494 parsimony informative sites. Having a variance (0.00025) and standard deviation (0.016) for haplotype diversity, the Tajima's D value was found to be negative for all genes with p > 0.10 (Table 8). While genetic diversity of avian avulaviruses primarily represents these changes, the functional implications of these mutations would shed light on specific roles on virus evolution and pathogenicity.

Analysis of complete genome sequences showed higher genetic diversity for the *F, HN* and *L* genes compared to rest of genes in avulaviruses. A hotspot event was also found at 5′ UTR of *F* gene and 3′ UTR of *HN* gene (Fig. 5). Datamonkey analysis of CDS for positive and negative selection sites are summarized in Table 9. None of study gene had mean dN/dS greater than 1 at *p* < 0.05; highest mean dN/dS was observed in *F* gene followed by *P, HN, L, M* and *NP* genes. Positive selection sites with codon position were inferred by three different statistical approaches (SLAC, FEL, IFEL). Taken together, at *p* < 0.05, the *F* gene showed highest positive selection sites followed by *HN* and *L* genes, while highest negative selection sites were found in *L* gene. Based on the dN/dS value, plots against codon positions for individual genes were plotted using SLAC statistical approach (Fig. 6).

3.5. Recombination analysis

Different algorithms (including SimPlot, GARD and RDP) were employed for the detection of putative recombination event. However, no exchange of sequence or putative recombination event between studied avulaviruses was detected. The data outcomes are not presented herein due to lack of any significant result.

4. Discussion

We presented a comparative genomic and evolutionary analysis of so-far known avulaviruses originating from diverse avian hosts, and reported from all geographical locations across the globe. In order to comprehensively map conserved motifs and substitutions in the coding

regions among different avulaviruses, we used whole genome sequences, which represent each avulavirus and provide high-resolution information on genetic diversity of the group. This is important because a specific gene might not evolve at the same rate as that of the whole genome (Miller et al., 2009). Also, we predicted and analysed open reading frames (ORFs) of individual genes of representative avulaviruses. This combinatorial and comprehensive analysis can provide a precise snapshot of virus evolution and design of epidemiological investigations in the future (Valdazo-Gonzalez et al., 2012).

Except for avulavirus 6 that had an additional short hydrophobic (SH) region between *F* and *HN* genes (6542 nt-6970 nt in the whole genome sequence), the genome of all avulaviruses carried six coding genes; a characteristic of all members of genus under *Paramyxoviridae* (Lamb and Park, 2007). The length of leader sequence of avulaviruses was conserved among all isolates however, trailer region was found variable in length (Lamb and Park, 2007). Different genomes of avulaviruses indicate potential virulence of subject strain because insertion of specific nucleotides at the C-terminal of *HN* gene negatively impact on its virulence (Romer-Oberdorfer et al., 2003). Using individual gene (*NP, P, M, F, HN* and *L*)-based phylogeny, we additionally presented a close relationship among avulaviruses 1, 9, 12, 13, 16, 17, 18, and 19 (Neira et al., 2017; Karamendin et al., 2017).

The analysis of complete genome nucleotide identity is a simple method for identification of novel avulavirus, and up to 25% divergence may be suggestive of novel virus or strain (Terregino et al., 2013). This is evident in pairwise comparison of whole genome sequences of all avulaviruses where a substantial divergence was observed. The divergence in percentage nucleotide identity among inter-avulaviruses can also be a marker of novel strain and may highlight the evolutionary distance of avulaviruses. In this context, the complete genomes of all avulaviruses showed a high nucleotide divergence except in avulaviruses 1, 9, 12, 16, 17, 18 and 19, which revealed a high nucleotide identity (70.6%-88.4%) compared to rest of avulaviruses. Similar findings have been reported in previous studies where maximum nucleotide homology among these avulaviruses (1, 9, 12, 16, 17,

18 and 19) has been suggested (Neira et al., 2017; Karamendin et al., 2017). In this context, previous studies also observed the existence of evolution in avulaviruses (Miller et al., 2009; Xiao et al., 2009; Bui et al., 2014). Susceptibility of diverse host species to all avulaviruses also supports transmission and evolution of avulaviruses (Dimitrov et al., 2016), reasoning the emergence of novel avulaviruses (14, 15, 17, 18, 19 and 20) from various wild/water-fowls in the recent years (Thampaisarn et al., 2017; Thomazelli et al., 2017; Lee et al., 2017; Neira et al., 2017; Karamendin et al., 2017). It is important to indicate that two avulaviruses (MF594548 and KY452442) are reported simultaneously from two different host species and geographical regions. One of these (KY452442) has been now officially notified as avulavirus 17. Since both carried a high nucleotide divergence (31.7%) and difference in subsequent phylogenetic analysis, therefore re-annotation of MF594548 as avulavirus 21 or other appropriate is needed (data not shown).

The conserved sequences in the F and HN proteins play a defined role in the fusion, attachment and release from the cells (Lamb and Parks, 2007). The cleavage site of the F protein acts as key determinant of virulence and existence of phenylalanine/leucine residue assists in classification of virulent and avirulent strains (Lamb and Parks, 2007). Reference to known sequence-based criteria for determining the potential virulence of an isolate, sequences those have phenylalanine residue at position next to cleavage motif doesn't require exogenous protease for *in vitro* replication and, therefore, are considered to be highly virulent. On the other hand, those isolates carrying leucine or isoleucine immediately preceding the cleavage site require protease for efficient replication and are considered as avirulent (Lamb and Parks, 2007). This could be the reason that virulent viruses had a wide range of tissue tropism than avirulent strains. Nevertheless, this may not be true for all avulaviruses since we observed F residue next to cleavage motif in representative strains of avulaviruses 2, 4, 5, 7 and 11 that were found lentogenic or avirulent in biological assessment (Subbaih et al., 2010; Samuel et al., 2010; Abolnik et al., 2012; Briand et al., 2012). The presence of a high number of arginine (R) and lysine (K) with phenylalanine (F) residue proceeding to cleavage motif in avulavirus 5 also creates a controversy. This is because the representative strain

has been reported as avirulent in chickens but highly virulent (100% mortality) in budgerigars (Samuel et al., 2010). Together, this provides evidences for a difference in host range susceptibility not ameliorated by presence of polybasic cleavage site that otherwise confer virulence in chicken. That beings said, one can speculate that cleavage site sequence cannot predict strain phenotype either *in vitro* or *in vivo* unless biological assessment is made. Given the fact that a limited data is available on biological assessment of all avulaviruses, the current anticipation to assess the pathogenicity of an isolate based on the F protein cleavage site may be misleading and may subvert the current health and safety measures in laboratory conditions. In fact, there are evidences that presence of velogenic-like F protein cleavage site fail to induce clinical disease in chicken (Panda et al., 2004). This ascertains further investigation on relationship between genome structure and *in vivo* pathogenicity assessment while considering potential variation in susceptibility of different avian species.

The glycine (G) residue in fusion peptide motif was found conserved among all avulaviruses. The motif is critical for membrane fusion activity (Morrison, 2003), where replacement of G residue can affect this function of the protein (Lamb and park, 2007). Likewise, hydrophobic heptad regions were also found conserved among all avulaviruses. The regions are essential for virus fusion and, potential substitution can produce functional changes in the integrity of the F protein (Ayllon et al., 2010). Involved in attachment at cellular surface, a highly conserved sialic acid binding sequence (NRKSC) was conserved in HN protein of all avulaviruses (Mirza et al., 1994). A similar hydrophobic heptad region in the HN that act as structural motif in stalk domain and is responsible for mediating protein-protein interactions (Lupas, 1996). Some conserved regions on polymerase-associated proteins (NP, P and L proteins) are known for their activities such as viral replication and survival in host cells (Lamb and Parks, 2007). A motif in NP protein, known as N self-assembly region, was found conserved in all avulaviruses (Steward et al., 1993; Lamb and Kolakofsky, 2001). The motif is considered important for interaction of NP with N-monomers of RNA during genomic RNA binding (Yu et al., 1998) and was thought to be required for NP-NP cross self-interaction

(Mebatsion et al., 2002). The P protein has a significance to form a complex with both NP and L proteins and play a critical role in RNA synthesis (Lamb and Parks, 2007). All avulaviruses contain a RNA editing site where an insertion of one/two G residue at C-terminus would produce mRNA encoding V/W protein due to internal shift of ORF (Steward et al., 1993; Lamb and Parks, 2007). However, the expression of the V or W proteins has been only confirmed in limited number of avulaviruses. Owing to the fact that V protein of NDV contributes to the pathogenicity of NDV (Huang et al., 2004), understanding the level and dynamics of these non-structural proteins might shed light on the differential pathogenicity of different avulaviruses. Apart from F and HN proteins, the L protein of avulaviruses genomes is also a determinant of virulence (Route and Samal, 2008). The L protein acts as RNA dependent RNA polymerase and contains several conserved domains responsible for transcriptional activity (Wise et al., 2004), especially the sequence QGNDQ from domain-III (Malur et al., 2002). Any substitution in this motif can abolish the polymerase activity of L protein, as seen in an experimental study related to *in vitro* replication of rabies virus (Schnell and Conzelmann, 1995). A putative ATP binding motif ($KX_{21}GXGXG$) in L protein was also found conserved in all avulaviruses and other paramyxoviruses (Poch et al., 1990; Harcourt et al., 2001).

The M protein is considered to be the central organizer of viral morphogenesis and has affinity for ionic interaction with acidic NP protein and hydrophobic tails of F and HN proteins for budding of viruses (Peeples, 1991; Lamb and Parks, 2007). A late domain known as protein-protein interaction motif carrying "FPIV" residues, essential for virus particles budding, has been reported in all avulaviruses (Schmitt et al., 2005), highlighting its importance in the life cycle of avulaviruses. A conserved motif bipartite clustering motif is thought to serve as a nuclear localization signal motif (Peeples, 1991), which was found in the C-terminus of the M protein in avulaviruses (Schmitt e al., 2005). To date, the potential influence of all identified substitutions in the functionality of all proteins is scarce. Future investigations are required to gather detailed information on the contribution of these variations in conserved domains in the intriguing complexities and diversities among avulaviruses.

Nucleotide diversity can help to measure the degree of genetic variation (polymorphism in nucleotide sequence) within a dataset (Nei and Tajima, 1981), where substitution rate is considered as a prime parameter to elucidate virus evolution. The average number of nucleotide difference among whole genome of all avulaviruses was found to be 6682.247. Further, the evolving nature of avulaviruses can be assessed by the nucleotide diversity (0.515101), variance (0.00025) and standard deviation (0.016) of haplotype diversity. This corresponds to distinct feature of RNA viruses due to lack of proofreading activity of reverse transcriptase (Duffy et al., 2008). The DnaSP based nucleotide diversity analysis revealed *F* and *HN* genes as the most diverse genes compared to rest of genes in avaulavirsus. Though it may require further research, the substitutions in both genes may have influenced adaptability and pathogenicity of avulaviruses to different susceptible hosts, such as observed for *HA* gene in H1N1 influenza virus (Tavakoli et al., 2015) and *H* gene in PPR virus (Sahu et al., 2017). Additionally, since these proteins are highly immunogenic, and surface glycoproteins, these remain under high immunological pressure. These features reason the diverse genetic nature of F and HN proteins.

Datamonkey is a well-known web-server for rapid detection of positive selection sites in aligned gene sequences (Pond and Frost, 2005). The statistical calculation of non-synonymous and synonymous (dN-dS) mutations a considerable tool to understand molecular evolution in CDS across closely related and yet diverged isolates. Therefore, Tajima's D method was used as neutrality test to determine the significance of positive selection sites (Fay and Wu, 2003). Positive selection sites were found in *F* and *HN* genes by SLAC analysis, in *NP, P, F, HN, L* genes by FEL analysis and in *P, M, F, HN,* and *L* genes by IFEL analysis. These sites were found as non-significant with less than 1 ratio by Tajima's D statistics, seldom happens in structural domains of genome however, the impact of such positive selection sites with lower level of sequence diversity may cause the emergence of variant (Yang et al., 2000). According to the neutral theory of molecular evolution, such type of molecular variations, which arise via spontaneous mutations, has

no influence on individual's fitness (Fay and Wu, 2003). However, the biological significance of these sites is still unknown and needs to be explored in future.

Multiple vaccination regimens coupled with mass employment of live avulavirus 1 vaccine (e.g., LaSota) increases the probability of its spill-over (Devlin et al., 2016; Rohaim et al., 2017). Such spill-over may enhance additional events such as reversion of virulence and recombination with wild-type strain. A study reported the spill-over of LaSota strain into wild birds (Snoeck et al., 2013) because LaSota is most common vaccine strain being used worldwide (Martinez et al., 2017, Akhtar, et al., 2017) and, therefore, has every chance to be shed in environment from vaccinates (Rohaim et al., 2017). Previous studies have suggested that genetic evolution may be a consequence of evasion of immune response induced by vaccine strains or subsequent infection (Qin et al., 2008, Younus, et al., 2017). Therefore, selective pressure may favour host adaptation and emergence of different genetic variants that can be transmitted to various avian species (Dimitrov et al., 2016). This has given the evidences of interaction of vaccine strains and field circulating avulaviruses, which can facilitate the emergence of novel mutant strain. We used LaSota strain as query sequence to identify the occurrence of putative recombination events while performing recombination analysis. However, no putative breakpoints were observed in our analysis using SimPlot, GARD or RDP. Contrary to our predictions, a limited number of previous studies have reported an occurrence of natural recombination within avulaviruses (Satharasinghe et al., 2016) and, proposed that recombination may facilitate emergence of a novel/new mutant strain (Yin et al., 2011). Based on these observations, it is evident that recombination in avulavirues is a rare event probably due to non-segmented nature of the virus, apathogenic potential in wild birds, and stability of the genome. Future research is warranted not only to assess the impact of reported natural recombination on the evolution of the virus but also to define models that specifically predict such recombination events in avulalviruses.

5. Conclusion

We presented the most comprehensive phylogenomic and evolutionary analysis of whole genome and individual genes of so-far known avulaviruses (1-20) reported from diverse avian species around the globe. The findings of current study revealed the genetic diversity and evolutionary aspects of different specific strains of avulaviruses.The presented information is expected to establish foundations for future investigations on virus epidemiology, emergence of novel avulaviruses and potential implications of wild/water-fowl origin pathotypes in susceptible commercial poultry.

Acknowledgment

Author's Contribution

Conceived and designed the work: AR, MM, MZS; Data analysis: AR, MZS; Draft writing and editing: AR, MM, MZS

Conflict of Interest

Figure Legends

Fig. 1. Complete genome sequence based phylogenetic analysis of study-included isolates representing avulaviruses 1-20

Fig. 2. Complete genome sequence based evolutionary network of study-included isolates representing Clade-wise distribution of avulaviruses 1-20. The number on each branch corresponds to sequenced isolate of each avulavirus and relevant accession numbers are given in methods.

Fig. 3. Comparative phylogenetic analysis of individual complete gene sequences of avulaviruses 1-20. Phylogeny for each gene is represented accordingly i.e. 3A (NP gene), 3B (P gene), 3C (M gene), 3D (F gene), 3E (HN gene), and 3F (L gene). Different colours indicate distinct specie-types of avulaviruses mentioned in Figure 1.

Fig. 4. Residue analysis of avulaviruses 1-20 for cleavage site pattern in the fusion protein at position 112 - 117. Constructing multiple sequence alignment, the diversity and/or conserveness of residues at each position of the cleavage site were analysed through WebLogo 3.1 (accessible at http://weblogo.threeplusone.com/create.cgi).

Fig. 5. Polymorphism for nucleotide diversity of the whole genome sequences of avulaviruses 1-20. The analysis was conducted through DnaSp Ver. 6.10.01 (accessible at http://www.ub.edu/dnasp/) and plotted using Pi value.

Fig. 6. Differences between synonymous and non-synonymous substitutions (dN-dS values) and the codon position of genes of avulaviruses 1-20. The negative values indicate synonymous substitutions while the positive values indicate non-synonymous substitutions.

References

Abolnik, C., de Castro, M., Rees, J., 2012. Full genomic sequence of an African avian paramyxovirus type 4 strain isolated from a wild duck. Virus Genes 45(3),537-541.

Akhtar, S., Muneer, M.A., Muhammad, K., Tipu, M.Y., Anees, M., Rashid, I., Raza-ur-Rehman., Hussain, I., 2017. Molecular Characterization and Epitope Mapping of Fusion (F) and Hemagglutinin (HN) Genes of Avian Paramyxovirus Serotype I from Peacocks in Pakistan. Pakistan J. Zool, 49,755-759.

Aldous, E.W., Fuller, C.M., Ridgeon, J.H., Irvine, R.M., Alexander, D.J., Brown, I.H., 2014. The Evolution of Pigeon Paramyxovirus Type 1 (PPMV- 1) in Great Britain: A Molecular Epidemiological Study. Transbound Emerg Dis. 61(2),134-9.

Amarasinghe, G.K., Bào, Y., Basler, C.F., Bavari, S., Beer, M., Bejerman, N., Blasdell, K.R., Bochnowski, A., Briese, T., Bukreyev, A., Calisher, C.H., 2017. Taxonomy of the order Mononegavirales: update 2017. Arch.Viro. 162(8),2493-2504.

Ayllón, J., Villar, E., Muñoz-Barroso, I., 2010. Mutations in the ectodomain of Newcastle disease virus fusion protein confer a hemagglutinin-neuraminidase-independent phenotype. J. Virol. 84(2),1066-75.

Briand, F.X., Henry, A., Massin, P., Jestin, V., 2012. Complete genome sequence of a novel avian paramyxovirus. J. Viro. 86(14),7710-7710.

Bui, V.N., Mizutani, T., Nguyen, T.H., Trinh, D.Q., Awad, S.S.A., Minoungou, G.L., Yamamoto, Y., Nakamura, K., Saito, K., Watanabe, Y., Runstadler, J., Huettmann, F., Ogawa, H., Imai, K., 2014. Characterization of a genetic and antigenic variant of avian paramyxovirus 6 isolated from a migratory wild bird, the red-necked stint (*Calidris ruficollis*). Arch. Virol. 159,3101–3105.

Chong, Y.L., Padhi, A., Hudson, P.J., Poss, M., 2010. The effect of vaccination on the evolution and population dynamics of avian paramyxovirus-1. PLoS pathogens 22,6(4),e1000872.

Delport, W., Poon, A.F., Frost, S.D., Kosakovsky, P.S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26(19),2455–2457. doi:10.1093/bioinformatics/btq429

Devlin, J.M., Vaz, P.K., Coppo, M.J., Browning, G.F., 2016. Impacts of poultry vaccination on viruses of wild bird. Curr. Opin. Virol. 19,23-9.

Dimitrov, K.M., Ramey, A.M., Qiu, X., Bahl, J., Afonso, C.L., 2016. Temporal, geographic, and host distribution of avian paramyxovirus 1 (Newcastle disease virus). Infect. Genet. Evol. 39,22-34.

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9(4),267–276. doi:10.1038/nrg2323

Fay, J.C., Wu, C.I., 2003. Sequence divergence, functional constraint, and selection in protein evolution. Annu. Rev. Genom. Hum. Genet. 4,213–235. doi:10.1146/annurev.genom.4.020303.162528

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. InNucleic acids symposium series Vol. 41, No. 41, pp. 95-98).

Harcourt, B.H., Tamin, A., Halpin, K., Ksiazek, T.G., Rollin, P.E., Bellini, W.J., Rota, P.A., 2001. Molecular characterization of the polymerase gene and genomic termini of Nipah virus. J. Virol. 287,192-201.

Hofacker, I.L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31,3429–3431.

Huang, Z., Panda, A., Elankumaran, S., Govindarajan, D., Rockemann, D.D., Samal, S.K., 2004. The hemagglutinin-neuraminidase protein of Newcastle disease virus determines tropism and virulence. J. Virol. 78,4176-4184.

Huson, D. H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Bio. Evol. 23(2),254–67.

Karamendin, K., Kydyrmanov, A., Kasymbekov, Y., Asanova, S., Daulbayeva, K., Seidalina, A., Khan, E., Harrison, S.M., Carr, I.M., Goodman, S.J., Moldakozhayev, A., 2017. Novel avian paramyxovirus isolated from gulls in Caspian seashore in Kazakhstan. PloS one, 12(12), p.e0190339.

Kumar, S., Nayak, B., Collins, P.L., Samal, S.K., 2008. Complete genome sequence of avian paramyxovirus type 3 reveals an unusually long trailer region. Virus Res. 137(2),189-197.

Lamb, R., Parks, G., 2007. Paramyxoviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E., eds. Philadelphia: Lippincott Williams & Wilkins, pp. 1449–1496.

Lamb, R.A., Kolakofsky, D., 2001. Paramyxoviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M., Griffin, D.E., Martin, M.A., Lamb, R.A., Roizman, B., Straus, S.E., (Eds.), Fields Virology, 4th ed. Lippincott Williams and Wilkins, Philadelphia, PA, pp. 1305–1340.

Lee, H.J., Kim, J.Y., Lee, Y.J., Lee, E.K., Song, B.M., Lee, H.S., Choi, K.S., 2017. A novel avian paramyxovirus (putative serotype 15) isolated from wild birds. Front. Microbiol. 8.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11),1451–1452. doi:10.1093/bioinformatics/btp187

Liu, H., de Almeida, R.S., Gil, P., Albina, E., 2018. Cleavage site of Newcastle disease virus determines viral fitness in persistent infection cells. Vet. Micro. 216,123-31.

Lupas, A., 1996. Coiled coils: new structures and new functions. Trends Biochem. Sci. 21(10),375-82.

Malur, A.G., Choudhary, S.K., De, B.P., Banerjee, A.K., 2002. Role of a highly conserved NH (2)-terminal domain of the human parainfluenza virus type 3 RNA polymerase. J. Virol. 76,8101–8109.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1(1).

Martinez, J.C., Chou, W.K., Berghman, L.R., Carey, J.B., 2017. Evaluation of the effect of live LaSota Newcastle disease virus vaccine as primary immunization on immune development in broilers. Poul. Sci. J.

Mebatsion, T., Koolen, J.M.M., De Vaan, T.C.L., De Haas, N., Braber, M., Romer-Oberdorfer, A., Van den Elzen, Marel, P., 2002. Newcastle Disease Virus (NDV) marker vaccine: An immunodominant epitope on the nucleoprotein gene of NDV can be deleted or replaced by a foreign epitope. J. Virol. 76,10138-10146.

Miller, P.J., Afonso, C.L., Spackman, E., Scott, M.A., Pedersen, J.C., Senne, D.A., Brown, J.D., Fuller, C.M., Uhart, M.M., Karesh, W.B., Brown, I.H., 2010. Evidence for a new avian paramyxovirus serotype 10 detected in rockhopper penguins from the Falkland Islands. J. Viro. 84(21),11496-11504.

Miller, P.J., Kim, L.M., Ip, H.S., Afonso, C.L., 2009. Evolutionary dynamics of Newcastle disease virus. Virology 391(1),64-72.

Mirza. A.M., Deng, R., Iorio, R.M., 1994. Site-directed mutagenesis of a conserved hexapeptide in the paramyxovirus hemagglutinin-neuraminidase glycoprotein: effects on antigenic structure and function. J. Virol. 68,5093–5099.

Morrison, T.G., 2003. Structure and function of a paramyxovirus fusion protein. Biochim. Biophys Acta 1614,73–84.

Munir, M., Cortey, M., Abbas, M., Qureshi, Z.A., Afzal, F., Shabbir, M.Z., Khan, M.T., Ahmed, S., Ahmed, S., Baule, C., Ståhl, K., Zohari, S., Berg, M., 2012b. Phylogenetic analysis and biological characterization of a novel genetic group of Newcastle Disease Virus (NDV) isolated from outbreaks in commercial poultry farms and from healthy backyard poultry flocks. Infec. Genet. Evol. 12,1010-1019.

Munir, M., Shabbir, M.Z., Yaqub, T., Shabbir, M.A.B., Mukhtar, N., Khan, M.R., Berg, M., 2012a. Complete genome sequence of a velogenic neurotropic avian paramyxovirus 1 isolated from Peacocks (*Pavo cristatus*) in a wildlife park in Pakistan. J. Viro. 86(23),13113-13114.

Nei, M., Tajima, F., 1981. DNA polymorphism detectable by restriction endonucleases. Genetics 97(1),145–163.

Neira, V., Tapia, R., Verdugo, C., Barriga, G., Mor, S., Ng, T.F.F., García, V., Del Río, J., Rodrigues, P., Briceño, C., Medina, R.A., 2017. Novel avulaviruses in penguins, Antarctica. Emerg. Infec. Dis. 23(7),1212.

Olsen, B., Munster, V.J., Wallensten, A., Waldenström, J., Osterhaus, A.D., Fouchier, R.A., 2006. Global patterns of influenza A virus in wild birds. Science 312(5772),384-8.

Paldurai, A., Subbiah, M., Kumar, S., Collins, P.L., Samal, S.K., 2009. Complete genome sequences of avian paramyxovirus type 8 strains goose/Delaware/1053/76 and pintail/Wakuya/20/78. Virus Res. 142(1-2),144-153.

Panda, A., Elankumaran, S., Krishnamurthy, S., Huang, Z., Samal, S.K., 2004. Loss of N-linked glycosylation from the hemagglutinin-neuraminidase protein alters virulence of Newcastle disease virus. J. Viro. *78*(10),4965-4975.

Peeples, M.E., 1991. Paramyxovirus M proteins: pulling it all together and taking it on the road. Kingsbury, D.W., ed. Plenum press, New York, pp. 427–456.

Poch, O., Blumberg, B.M., Bougueleret, L., Tordo, N., 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. 71,1153-1162.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21(10),2531–2533. doi:10.1093/bioinformatics/bti320

Qin, Z., Sun, L., Ma, B., Cui, Z., Zhu, Y., Kitamura, Y., Liu, W., 2008. F gene recombination between genotype II and VII Newcastle disease virus. Virus research. 131(2):299-303.

Ray, S.C., 2003. SimPlot for Windows (version 3.5.1). Baltimore, MD: Available online at: http://sray.med.som.jhmi.edu/SCRoftware/ (Accessed June 2015)

Rohaim, M.A., El Naggar, R.F., Helal, A.M., Hussein, H.A., Munir, M., 2017. Reverse spillover of avian viral vaccine strains from domesticated poultry to wild birds. Vaccine 35(28),3523-7.

Römer-Oberdörfer, A., Werner, O., Veits, J., Mebatsion, T., Mettenleiter, T.C., 2003. Contribution of the length of the HN protein and the sequence of the F protein cleavage site to Newcastle disease virus pathogenicity. J. Gen. Virol. 84(11),3121-9.

Rout, S.N., Samal, S.K., 2008. The large polymerase protein is associated with the virulence of Newcastle disease virus. J. Virol. 82,7828-7836.

Sahu, A.R., Wani, S.A., Saminathan, M., Rajak, K.K., Sahoo, A.P., Pandey, A., Saxena, S., Kanchan, S., Tiwari, A.K., Mishra, B., Muthuchelvan, D., 2017. Genome sequencing of an Indian peste des petits ruminants virus isolate, Izatnagar/94, and its implications for virus diversity, divergence and phylogeography. Arch. Virol. 162(6),1677-93.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retroviruses 11(11),1423-5.

Samuel, A.S., Kumar, S., Madhuri, S., Collins, P.L., Samal, S.K., 2009. Complete sequence of the genome of avian paramyxovirus type 9 and comparison with other paramyxoviruses. Virus Res. 142(1-2),10-18.

Samuel, A.S., Paldurai, A., Kumar, S., Collins, P.L., Samal, S.K., 2010. 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 (2),e9269.

Satharasinghe, D.A., Murulitharan, K., Tan, S.W., Yeap, S.K., Munir, M., Ideris, A., Omar, A.R., 2016. Detection of inter-lineage natural recombination in avian paramyxovirus serotype 1 using simplified deep sequencing platform. Fron. Microbiol. 7.

Schmitt, A.P., Leser, G.P., Morita, E., Sundquist, W.I., Lamb, R.A., 2005. Evidence for a new viral late domain core sequence, FPIV, necessary for budding of a paramyxovirus. J. Virol. 79,2988–2997.

Schnell, M.J., Conzelmann, K.K., 1995. Polymerase activity of in vitro mutated rabies virus L protein. J. Virol. 214,522–530.

Shabbir, M.Z., Abbas, M., Yaqub, T., Mukhtar, N., Subhani, A., Habib, H., Sohail, M.U., Munir, M., 2012a. Genetic analysis of Newcastle disease virus from Punjab, Pakistan. Virus Genes 46,309–315.

Shabbir, M.Z., Akhtar, S., Tang, Y., Yaqub, T., Ahmad, A., Mustafa, G., Alam, M.A., Santhakumar, D., Nair, V., 2016. Infectivity of wild bird origin Avian Paramyxovirus serotype 1 and vaccine effectiveness in chickens. J. Gen. Viro. 97 (12),3161-3173. DOI: 10.1099/jgv.0.000618

Shabbir, M.Z., Goraya, M.U., Abbas, M., Yaqub, T., Shabbir, M.A.B., Ahmad, A., Anees, M., Munir, M., 2012b. Complete genome sequencing of a velogenic viscerotropic avian paramyxovirus 1 isolated from pheasants (*Pucrasia macrolopha*) in Lahore, Pakistan. J. Viro. 86(24),13828-13829.

Shabbir, M.Z., Nissly, R.H., Ahad, A., Rabbani, M., Chothe, S.K., Sebastian, A., Albert, I., Jayarao, B.M., Kuchipudi, S.V., 2018. Complete Genome Sequences of Three Related Avian Avulavirus 1 Isolates from Poultry Farmers in Pakistan. Genome announc. 6(18), e00361-18.

Shabbir, M.Z., Zohari, S., Yaqub, T., Nazir, J., Shabbir, M.A.B., Mukhtar, N., Shafee, M., Sajid, M., Anees, M., Abbas, M., Khan, M.T., Ali, A.A., Ghafoor, A., Ahad, A., Channa, A.A., Anjum, A.A., Kalhoro, N.H., Ahmad, A., Goraya, M.U., Iqbal, Z., Khan, S.A., Aslam, H.B., Zehra, K., Sohail, M.U., Yaqub, W., Berg, M., Munir, M., 2013. Genetic diversity of Newcastle disease virus in Pakistan: A countrywide perspective. Viro. J. 10,170 – 179.

Snoeck, C.J., Owoade, A.A., Couacy-Hymann, E., Alkali, B.R., Okwen, M.P., Adeyanju, A. T., Komoyo, G.F., Nakouné, E., Le Faou, A. and Muller, C.P., 2013. High genetic diversity of Newcastle disease virus in poultry in west and central Africa: cocirculation of genotype XIV and newly defined genotypes XVII and XVIII. J. Clin. Microbiol. 51,2250–2260. doi:10.1128/JCM.00684-13

Steward, M., Vipond, I.B., Millar, N.S., Emmerson, P.T., 1993. RNA editing in Newcastle disease virus. J. Gen. Virol. 74,2539–2547.

Subbiah, M., Nayak, S., Collins, P.L., Samal, S.K., 2010. 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 Res. 152(1-2),85-95.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3),585–595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30,2725–2729.

Tavakoli, F., Moattari, A., Shamsi, S.A.M,, Kadivar, M.R., Khodadad, N., Pirbonyeh, N., Emami, A., 2015. Antigenic Variation of the Haemagglutinin Gene of the Influenza A (H1N1) pdm09 Virus Circulating in Shiraz, February-April 2013. Iran J. Immunol. 12(3),198–208.

Terregino, C., Aldous, E.W., Heidari, A., Fuller, C.M., De Nardi, R., Manvell, R.J., Beato, M.S., Shell, W.M., Monne, I., Brown, I.H., Alexander, D.J., Capua, I., 2013. Antigenic and genetic analyses of isolate APMV/wigeon/Italy/3920-1/2005 indicate that it represents a new avian paramyxovirus (APMV-12). Arch. Virol. 158,2233–2243.

Thampaisarn, R., Bui, V.N., Trinh, D.Q., Nagai, M., Mizutani, T., Omatsu, T., Katayama, Y., Gronsang, D., Le, D.H., Ogawa, H., Imai, K., 2017. Characterization of avian paramyxovirus serotype 14, a novel serotype, isolated from a duck fecal sample in Japan. Virus Res. 228,46-57.

Thomazelli, L.M., de Araújo, J., Fabrizio, T., Walker, D., Reischak, D., Ometto, T., Barbosa, C.M., Petry, M.V., Webby, R.J., Durigon, E.L., 2017. Novel avian paramyxovirus (APMV-15) isolated from a migratory bird in South America. PloS one 12(5),e0177214.

Tian, Z., Chai, H., Li, F., Sun, J., Chen, G., Hu, X., Hua, Y., Xiang, W., 2012. Complete nucleotide sequence of avian paramyxovirus type 6 strain JL isolated from mallard ducks in China. J. Viro. 86(23),13112-13112.

Valdazo-Gonzalez, B., Polihronova, L., Alexandrov, T., Normann, P., Knowles, N.J., Hammond, J.M., Georgiev, G.K., Ozyoruk, F., Sumption, K.J., Belsham, G.J., King, D.P., 2012. Reconstruction of the transmission history of RNA virus outbreaks using full genome sequences:

foot-and-mouth disease virus in Bulgaria in 2011. PLoS One 7(11),e49650. doi:10.1371/journal.pone.0049650

Wise, M.G., Sellers, H.S., Alvarez, R., Seal, B.S., 2004. RNA-dependent RNA polymerase gene analysis of worldwide Newcastle disease virus isolates representing different virulence types and their phylogenetic relationship with other members of the paramyxoviridae. Virus Res. 104,71–80.

Xiao, S., Paldurai, A., Nayak, B., Subbiah, M., Collins, P.L., Samal, S.B., 2009. Complete genome sequence of avian paramyxovirus type 7 (strain Tennessee) and comparison with other paramyxoviruses. Virus Res. 145,80–91.

Yamamoto, E., Ito, H., Tomioka, Y., Ito, T., 2015. Characterization of novel avian paramyxovirus strain APMV/Shimane67 isolated from migratory wild geese in Japan. J.Vet. Med. Sci. 77(9),1079-1085.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1),431–449.

Yin, Y., Cortey, M., Zhang, Y., Cui, S., Dolz, R., Wang, J., Gong, Z., 2011. Molecular characterization of Newcastle disease viruses in Ostriches (*Struthio camelus L.):* further evidences of recombination within avian paramyxovirus type 1. Vet. Microbiol. 149(3),324-9.

Younus, M., Maqbool, A., Khan, I., Umar, S., 2017. Pathological Alterations during Co-Infection of Newcastle Disease Virus with Escherichia coli in Broiler Chicken. Pak. J. Zoo. 1,49(6).

Yu, M., Hansson, E., Shiell, B., Michalski, W., Eaton, B.T., Wang, L.F., 1998. Sequence analysis of the Hendra virus nucleoprotein gene: comparison with other members of the subfamily Paramyxovirinae. J. Gen. Virol. 79,1775–1780.

Table 1. A brief of GeneBank derived complete genome sequence detail of avulaviruses specie-types 1 – 20 used in this study

Avian avulaviruses	Accession number	Year of isolation	Bird specie	Country
Avian avulavirus 1	KU885948	2014	Peacock	Pakistan
	HQ697254	2010	Chicken	Indonesia
	JQ247691	1971	Chicken	USA
	KR074406	2005	Chicken	Malaysia
	KJ782375	1997	Goose	China
Avian avulavirus 2	HM159993	2006	Chicken	England
	HQ896023	1999	Chicken	China
	EU338414	1956	Chicken	USA
	HM159994	1980	Gadwall	Kenya
	HQ896024	2001	Chicken	China
Avian avulavirus 3	EU403085	1975	Parakeet	Netherland
	EU782025	1968	Turkey	USA
Avian avulavirus 4	JX133079	2010	Egyptian Goose	South Africa
	EU877976	2006	Mallard Duck	South Korea
	KU601399	2015	*Uris aalge*	Russia
	KC439346	2012	Duck	China
	JN571485	2007	Mallard Duck	Belgium
	KY681684	2012	Mallard Duck	South Korea
Avian avulavirus 5	GU206351	2010	Budgerigar	Japan
	LC168750	1975	Budgerigar	Japan
Avian avulavirus 6	JX522537	2012	Mallard Duck	China
	KP762799	2013	Red Crested Pochard	Kazakhstan
	EU622637	1977	Duck	Hong Kong
	AY029299	1998	Duck	Taiwan
	KF267717	2011	Mallard Duck	China
	KT962980	2009	Tea Duck	Russia
	EF569970	2003	Goose	Russia
Avian avulavirus 7	FJ231524	1975	Dove	USA
Avian avulavirus 8	FJ215863	1976	Goose	USA
	FJ5215864	1978	Pintail Duck	Japan
	MF448515	2013	Little Stint	Kazakhstan
	JX901129	1978	Pintail Duck	Japan
	FJ619036	1976	Canada Goose	USA
	MF448514	2013	Whooper Swan	Kazakhstan
Avian avulavirus 9	EU910942	1978	Domestic Duck	USA
Avian avulavirus 10	HM755886	2007	Penguin	Falkland Island
	HM147142	2007	Penguin	Falkland Island
	HM755887	2007	Penguin	Falkland Island
	HM755888	2007	Penguin	Falkland Island
Avian avulavirus 11	JQ886184	2010	Common Snipe	France
Avian avulavirus 12	KC333050	2005	Wigeon	Italy
Avian avulavirus 13	KU646513	2013	Wild Goose	Kazakhstan
	KX119151	2011	White Fronted Goose	Ukraine
	LC041132	2000	Anseriform spp	Japan
Avian avulavirus 14	KX258200	2011	Duck	Japan
Avian avulavirus 15	KX932454	2012	Sandpiper	Brazil
Avian avulavirus 16	KY511044	2014	Wild Duck	South Korea
Avian avulavirus 17	KY452442	2014	Penguin	Antarctica
Avian avulavirus 18	KY452443	2014	Penguin	Antarctica
Avian avulavirus 19	KY452444	2014	Penguin	Antarctica
Avian avulavirus 20	MF033136	2014	Gull	Kazakhstan

Table 2. A comparison of distinct features of complete genome sequences of avian avulaviruses 1 - 20

Regions	AAvV 1	AAvV 2	AAvV 3	AAvV 4	AAvV 5	AAvV 6*	AAvV 7	AAvV 8	AAvV 9	AAvV 10	AAvV 11	AAvV 12	AAvV 13	AAvV 14	AAvV 15	AAvV 16	AAvV 17	AAvV 18	AAvV 19	AAvV 20
Leader	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
GS[a]	UGC_3AUCU_2	C_5GCUGU	$UC_2UCG_2U_2$	$CAC_4U_2C_2$	$C_5U_2N_3$	CUC_5U_2C	CUC_3NCUN_2	C_5GCUG_2	UGC_3AUCU_2	UC_4GCUGN	CGC_3GCU_2N	UGC_3GUCU_2	UGC_3GUCU_2	CUC_5N_2	NA	UGC_3NUCU_2	NA	NA	NA	NA
NP	1,470	1,347	1,374	1,551	1,398	1,389	1,392	1,386	1,470	1,374	1,368	1,482	1,482	1,392	1,367	1,476	1,482	1,572	1,548	1,380
P	1,188	1,167	1,161	1,364	1,341	1,293	1,185	1,218	1,260	1,218	1,287	1,218	1,194	1,227	1,247	1,200	1,209	1,179	1,134	1,296
M	1,095	1,110	1,191	1,293	1,099	1,101	1,092	1,110	1,095	1,110	1,116	1,095	1,101	1,092	1,124	1,095	1,152	1,098	1.152	1,131
F	1,662	1,611	1,632	1,891	1,635	1,620	1,620	1,632	1,656	1,647	1,689	1,641	1,638	1,626	1,707	1,656	1,626	1,632	1,725	1,611
HN	1,714	1,743	1,734	1,914	1,725	1,842	1,710	1,734	1,740	1,728	1,752	1,845	1,740	1,743	1,739	1,857	1,800	1,776	1,764	1,725
L	6,615	6,729	6,597	6,834	6,792	6,726	6,684	6,717	6,609	6,711	6,756	6,609	6,600	6,743	6,650	6,609	6.609	6,624	6.654	6,729
IGS[b]	2,1,1,31,47	7,7,23,9,3	55,62,31,34,59	9,34,14,37,42	57,42,21,31,4	57,42,21,48,28,4	24,13,6,11,15	2.30,25,8,1	19,6,30,22,0	2,9,19,51,51	15,6,37,11,64	2,8,11,27,42	14,1,2,14,25	2,16,36,15,33	NA	2,1,6,0,14	NA	NA	NA	8,9,11,9,19
GE[c]	$A_2U(C/U)_2U_6$	$A_2U_2CU_6$	$A_2U_2AU_6$	$A_2U_2A_2U_5$	$A_2U(A/U)NU_5$	A_2U2AU_{5-7}	$A_2UN_2U_3NU_{1-3}$	$A_2U_2CU_6$	A_2UNU_6	$A_2UN_2U_{5-6}$	$A_2UN_2U_6$	$A_2UN_2U_2NU_{3-4}$	$A_2U_2CU_{5-6}$	$A_2UN_3U_3NU_{1-2}$	NA	$A_2UN_2U_{5-6}$	NA	NA	NA	NA
Trailer	114	154	707	17	552	54	127	171	47	206	402	204	776	277	NA	47	145	226	254	81
Total Length (bp)	15,192	14,904	16,272	15,054	17,262	16,236	15,480	15,342	15,438	15,456	17,412	15,312	15,996	15,444	14,952	15,180	14,926	14,931	15,017	15,786

*AAvV 6 has an additional SH region of 629 bp; exclusive in whole comparative analysis and contains six IGS as compared to other avulaviruses

a: Gene Start, **b**: Intergenic Sequence, **c**: Gene End, NA: Information not available

Table 3. Percentage nucleotide identity and divergence of complete genome sequences of avian avulaviruses 1 - 20

AAvVs	AAvV 1	AAvV 2	AAvV 3	AAvV 4	AAvV 5	AAvV 6	AAvV 7	AAvV 8	AAvV 9	AAvV 10	AAvV 11	AAvV 12	AAvV 13	AAvV 14	AAvV 15	AAvV 16	AAvV 17	AAvV 18	AAvV 19	AAvV 20
AAvV 1		57.7	63.4	61.8	63.2	60.0	57.9	60.1	44.4	58.8	63.0	46.9	49.4	60.2	53.9	35.3	28.5	28.2	28.1	46.2
AAvV 2	42.3		63.8	62.1	59.2	57.0	52.8	44.1	59.7	46.2	58.8	59.8	60.6	55.5	46.9	59.2	42.8	43.4	43.7	20.6
AAvV 3	36.4	36.2		57.1	64.4	65.0	63.4	63.4	63.4	64.5	64.0	63.7	62.1	63.5	62.6	63.8	47.3	46.1	46.9	45.4
AAvV 4	38.2	37.9	42.9		64.7	63.9	62.3	62.6	62.5	62.6	65.4	63.0	63.3	62.2	57.0	62.7	46.8	46.4	46.8	47.3
AAvV 5	36.8	40.8	35.6	35.3		56.9	58.1	59.2	62.8	58.5	48.8	62.6	62.8	55.7	59.3	63.0	44.0	42.9	43.3	34.6
AAvV 6	40.0	43.0	35.0	36.1	43.1		57.0	57.0	61.3	55.6	59.4	61.6	62.7	50.3	50.3	61.7	47.8	47.3	47.8	39.2
AAvV 7	42.1	47.2	36.6	37.7	41.9	43.0		52.5	59.0	52.7	57.7	59.1	59.0	55.4	54.1	58.9	43.7	44.3	45.2	34.9
AAvV 8	39.9	55.9	36.6	37.4	40.8	43.0	47.5		58.8	45.4	58.5	59.5	60.1	54.1	46.7	58.9	43.5	43.2	43.4	20.6
AAvV 9	55.6	40.3	36.6	37.5	37.2	38.7	41.0	41.2		58.7	62.6	47.5	50.9	59.8	59.6	43.6	30.7	29.7	30.4	43.6
AAvV 10	41.2	53.8	35.5	37.4	41.5	44.4	47.3	54.6	41.3		57.7	59.1	58.9	54.5	45.7	59.1	44.2	43.2	44.5	20.3
AAvV 11	37.0	41.2	36.0	34.6	51.2	40.6	42.3	41.5	37.4	42.3		62.5	62.0	59.0	58.3	62.5	43.9	43.5	44.5	33.9
AAvV 12	53.1	40.2	36.3	37.0	37.4	38.4	40.9	40.5	52.5	40.9	37.5		42.2	59.7	59.5	47.1	28.5	28.7	29.4	43.8
AAvV 13	50.6	39.4	37.9	36.7	37.2	37.3	41.0	39.9	49.1	41.1	38.0	57.8		60.3	52.2	48.9	29.6	29.6	28.9	44.8
AAvV 14	39.8	44.5	36.5	37.8	44.3	49.7	44.6	45.9	40.2	45.5	41.0	40.3	39.7		55.7	59.6	44.5	43.4	43.7	35.5
AAvV 15	46.1	53.1	37.4	43.0	40.7	49.7	45.9	53.3	40.4	54.3	41.7	40.5	47.8	44.3		59.4	42.5	43.1	43.6	22.7
AAvV 16	64.7	40.8	36.2	37.3	37.0	38.3	41.1	41.1	56.4	40.9	37.5	52.9	51.1	40.4	40.6		29.4	28.5	28.7	44.8
AAvV 17	71.5	57.2	52.7	53.2	56.0	52.2	56.3	56.5	69.3	55.8	56.1	71.5	70.4	55.5	57.5	70.6		11.6	17.1	44.5
AAvV 18	71.8	56.6	53.9	53.6	57.1	52.7	55.7	56.8	68.3	56.8	56.5	71.7	70.4	56.6	56.9	71.5	88.4		16.43	42.9
AAvV 19	71.9	56.3	53.1	53.2	56.7	52.2	54.8	56.6	69.6	55.5	55.5	70.6	70.1	56.3	56.4	71.3	82.9	83.6		44.4
AAvV 20	53.8	79.4	54.6	52.7	65.4	60.8	65.1	79.4	56.4	79.7	66.1	56.2	55.2	64.5	77.3	55.2	55.3	57.1	55.6	

Nucleotide identity percentage are given in below diagonal, whereas percent divergence values are presented in above diagonal

Table 4. A multiple sequence alignment based comparison of residue patterns at important motif/s in the open reading frame of F protein using different specie-types of avulaviruses 1-20

AAvVs		Cleavage site	Exogenous protease requirement	Fusion peptide (FI**GA**VI**G**SI**ALG**V**AT**AAQI**TAA**A**AL**I)	Hypervariable Regions		
					HRa (QANQNAANILRLKESIAATNEAVHEVTNGLSQLSVAVGKMQQF)	HRb (LITGYPILYDSQTQLLGIQVNLPSVGNLNNMR)	HRc (NNSISNALDKLTESNSKLDKVNVRLTSTSA)
AAvV 1	KU885948	^{112}**RRQKR** ↓F^{117}	-	^{117}................................^{142}	^{143}...^{185}	^{268}.............................^{299}	^{471}...........................^{500}
	HQ697254	^{112}**RRQKR** ↓F^{117}	-	^{117}................................^{142}	^{143}.......A....................I......^{185}	^{268}.............................^{299}	^{471}...........................^{500}
	JQ247691	^{112}**RRQKR** ↓F^{117}	-	^{117}...............S......^{142}	^{143}...^{185}	^{268}.............................^{299}	^{471}...........................^{500}
	KR074406	^{112}**RRQKR** ↓F^{117}	-	^{117}................................^{142}	^{143}.............V..................^{185}	^{268}.............................^{299}	^{471}..........V...............^{500}
	KJ782375	^{112}**RRQKR** ↓F^{117}	-	^{117}...............S......^{142}	^{143}...^{185}	^{268}.............................^{299}	^{471}..........V...............^{500}
AAvV 2	HM159993	^{93}KPAS**R** ↓F^{98}	-	^{98}.V..I...V.............V..N^{123}	^{125}..QE..R..W.......KK..A.L.LKD..ATTAI. LD.V.K.^{167}	^{250}.....QIVG..PEL....SY...SEVT GV.^{281}	^{453}..TVNR...ELIK...EW.NA..P.. V.NTS^{482}
	HQ896023	^{93}KPAS**R** ↓F^{98}	-	^{98}.V..I..............V..N^{123}	^{125}..QE......W.......KK..A..L.LKD..ATTAI.L D.V.K.^{167}	^{250}.....QIVG..PEL....SY...SEVT GV.^{281}	^{453}..TV....DELIK....EW.NA..P..V NNTS^{482}
	EU338414	^{93}KPAS**R** ↓F^{98}	-	^{98}.V..I................V..N^{123}	^{125}..QE..R..W.......KK..A..L.LKD..ATTAI. LD.V.K.^{167}	^{250}....QIVG....L...SY...SEVT GV.^{281}	^{453}..TVNRVDELIK...EW.NA..P.. VNNTS^{482}
	HM159994	^{93}KPAS**R** ↓F^{98}	-	^{98}.V..I...V.............V..N^{123}	^{125}..QE..R..W..........A..L.LKD..ATTAI.LD .V.K.^{167}	^{250}....QIVG..PEL....SY...SEVT GV.^{281}	^{453}..TVNRVDELIK...EW.NA..P.. VNNTS^{482}
AAvV 3	EU403085	^{96}**R**P**R**G**R** ↓L^{101}	+	^{101}LF.PI..........S......I..V^{126}	^{133}R.Q...ND..A..NALQSS...IRQL.Y.QD ...L.IS.I.KA^{175}	^{258}.LS.QV.G..PSV.G.I.R....MR TQKIDRAL^{289}	^{461}QDHANA..A.VE..DKY.RA.T GGNY.NWY^{490}
	EU782925	^{96}**R**P**R**G**R** ↓L^{101}	+	^{101}LF.PI..........S......I..V^{126}	^{133}R.Q...ND..A..NALQSS...IRQL.Y.QD K.LL.IS.I.KA^{175}	^{258}.LS.QV.G..PSV.G.I.R....MR T....RAL^{289}	^{461}Q....NA..A.VE..DKY.RA.TGG NY.NWY^{490}
AAvV 4	JX133079	^{115}DIQP**R** ↓F^{120}	-	^{120}....I.ATG..A....V.E...Q..S^{145}	^{147}.SKT..Q.....RD...Q...Q..F.ISQ..EATAT VLS.L.AE^{189}	^{276}..QSQVVG.....Y....V.R....VR IQEVQ.T.^{307}	^{479}.S..EA.REQ.EL..QI.SRI.P.IVN DES^{508}
	EU877976	^{115}DIQP**R** ↓F^{120}	-	^{120}....I.ATG..A....V.E...Q..S^{145}	^{147}.SKT..Q....K.RD..Q...Q..F.ISQ..EATA TVLS.L.AE^{189}	^{276}..QSQVVG....Y...V.R....VR IQEVQ.T.^{307}	^{479}.S..EA.REQ.EL..QI.SRI.P.IVN DES^{508}
	KU601399	^{115}DIQP**R** ↓F^{120}	-	^{120}....I.ATG..A....V.E...Q..S^{145}	^{147}.SKT..Q....K.RD..Q...Q..F.ISQ..EATA TVLS.L.AE^{189}	^{276}..QSQVVG....Y...V.R....VR IQEVQ.T.^{307}	^{479}.S..EA.REQ.EL..QI.SRI.P.IVN DES^{508}
	KC439346	^{115}DIQP**R** ↓F^{120}	-	^{120}....I.ATG..A....V.E...Q..S^{145}	^{147}.SKT..Q....K.RD..Q...Q..F.ISQ..EATA TVLS.L.AE^{189}	^{276}..QSQVVG....Y...V.R....VR IQEVQ.T.^{307}	^{479}.S..EA.REQ.EL..QI.SRI.P.IVN DES^{508}
AAvV 5	GU206351	^{105}**KRKKR** ↓F^{110}	-	^{110}.V.....AA...L......V.A....N^{135}	^{136}..QE..R.......N..QK.....M.LKDAVG.TA ..ID.T.A.^{178}	^{261}..Q.RITQ...MN....I....QI..IST VSG..^{292}	^{464}.T.L.KVESLIK.....I..S..PQILN VKT^{493}
	LC168750	^{105}**KRKKR** ↓F^{110}	-	^{110}.V.....AA...L......V.A....N^{135}	^{136}..QE..R.......N..QK.....M.LKDAVG.TA .....T.A.^{178}	^{261}..Q.RITQ...MN....I....QI..IST V.G..^{292}	^{464}.T.L.KVESLIK....I..S..PQIL..K T^{493}
AAvV 6	JX522537	^{113}APEP**R** ↓L^{118}	+	^{118}L....I..T....L......V...L..N^{143}	^{145}..QE..RA..N.....TK.....L.LKDATG.IAI. LD.T.R.^{187}	^{270}..N.RI.Q..QAS.IMV...SV..IS SI.GL.^{301}	^{473}.Q.L...VE.LIDQ...QL.NR..PNIV NNT.^{502}
	KP762799	^{113}APEP**R** ↓L^{118}	+	^{118}L....I.......L.....V...L..N^{143}	^{145}..QE..RA..N.....K.....L.LKDATG.IAI. LD.T.R.^{187}	^{270}..N.RI.Q..QAS.IMV...SV..IS SISGL.^{301}	^{473}.Q.L.....LIDQ..QL.NR..PNIVN NT.^{502}
	EU622637	^{113}APEP**R** ↓L^{118}	+	^{118}L....I..T.........V...L..N^{143}	^{145}..QE..RA..N.....TK.....L.LKDATG.IAI. L..T.R.^{187}	^{270}..N.RI.Q..QAS.IMV...SV..IS ..SGL.^{301}	^{473}.Q.L..VE.LIDQ..QL.NR..PNIV ..T.^{502}
	AY029299	^{113}APEP**R** ↓L^{118}	+	^{118}L....I..T.........V...L..N^{143}	^{145}..QE..RA..N.....TK.....L.LKDATG.IAI. LD.T.R.^{187}	^{270}..N.RI.Q..QAS.IMV...SV..IS SISGL.^{301}	^{473}.Q.L..VE.LIDQ.....NR..PNIVN NT.^{502}

30

AAvV 7	FJ231524	^{101}LPSS**R** ↓F^{106}	-	^{106}.A.LV..A.......S.......V..T^{131}	^{133}K.Q...QE.I..RD..QN.IN..NDI.V...SIG..LS.V.NY^{175}	^{258}..SAQIVSF.K..GIAILY.T...I AT.SGS.^{289}	^{461}.....S.T.N.QKAQES.N.IIPAA..SWL^{490}
AAvV 8	FJ215863	^{98}YPQT**R** ↓L^{103}	+	^{103}L..............S.....V..K^{128}	^{130}..QD..R...A...ALSK.....K.LSS..Q.TAI.L..I.S.^{172}	^{255}....QV.G....Y...V....Y..ISE VTGV.^{286}	^{458}..TVNRVE.LIHQ.EEW.A..P.I VNNTT^{487}
	FJ215864	^{98}YPQT**R** ↓L^{103}	+	^{103}L...................V..K^{128}	^{130}..QD..R...A...ALSK.....K.LSS..Q.TAI.L..I.S.^{172}	^{255}....QV.G....Y...V....Y..ISE VTGV.^{286}	^{458}..TVNRVE.LIHQ.EEW.A..PHI VNNTT^{487}
	MF448515	^{98}YPQT**R** ↓L^{103}	+	^{103}L..................S.....V..K^{128}	^{130}..QD..R...A...A..K.....K.LSS..Q.TAI.L..I.S.^{172}	^{255}....QV......Y...V....Y..ISEV TGV.^{286}	^{458}..TVNRVE.L.....EEW.A..PHIV NNTT^{487}
	FJ619036	^{98}YPQT**R** ↓L^{103}	+	^{103}L...................V..K^{128}	^{130}..QD..R...A....LSK.....K.LSS..Q.TAI.L..I.S.^{172}	^{255}....Q.G....Y....V....Y..ISEV TGV.^{286}	^{458}....NRVE.L..Q.EEW.A..PHIV NNTT^{487}
AAvV 9	EU910942	^{104}I**R**EG**R** ↓I^{109}	+	^{109}IF..IL.GV.............I..^{134}	^{136}...E..K....I.D..TK.....RD...V..TI....L.D.^{178}	^{261}...Q.....ES.I.A....S....ISD.R GV.^{292}	^{464}.S.L.S..E..D..DAA.S....H......^{493}
AAvV 10	HM755886	^{158}**K**PSQ**R** ↓I^{163}	+	^{163}I...I...V.........V...L.MT^{188}	^{127}..Q...R..WK.....KN..Q..L.LKD..Q.SAI.LD.V.S.^{169}	^{252}....QV.G...AD.I.L....SY...S RVQGV.^{283}	^{455}..T.TKVE.LISK..DW.S...PT.I. NDT^{484}
	HM147142	^{158}**K**PSQ**R** ↓I^{163}	+	^{163}I...I...V.............L.MT^{188}	^{127}..Q...R..WK.....KN..Q..L.LKD..Q.SAI.L..V.S.^{169}	^{252}....Q..G...AD.I.L....SY...SR VQGV.^{283}	^{455}..T.TK...LISK..DW.S...PT.I.N.T^{484}
AAvV 11	JQ886184	^{113}SGTK**R** ↓F^{118}	-	^{118}.V......V.......S......LAM^{143}	^{145}..QD..KA.WK...A.SS..Q..L.LKE.VNT.G...D.I.GY^{187}	^{270}..N.QIVS..ASI.TII.K.SI..IS S.SRFS381	^{473}.QTLDKVGQLINT..QI.ASL.P K.VNNTS^{502}
AAvV 12	KC333050	^{105}G**R**EP**R** ↓L^{110}	+	^{110}LV.I...V.............T...^{135}	^{137}......E..A..AKGL.......TDL.K.VGS.AI G...L.DY^{179}	^{262}..K.R..M..DGNKI.....T.....RI .GA.^{293}	^{465}.Q.LDS..ENIK....Y.S....K.I.S..^{494}
AAvV 13	KU646513	^{102}V**R**EN**R** ↓L^{107}	+	^{107}LV.I...V.............T....^{132}	^{134}.........A..AN....K......TDL.E..GT.AIG ...L.DY^{176}	^{259}..K.M..M..DANK..AV..E....I. K..GA^{290}	^{462}.Q..DS.KENIA....KF.SQ...K.L. S..^{491}
	KX119151	^{102}V**R**EN**R** ↓L^{107}	+	^{107}LV..I.................T...^{132}	^{134}.....A..AN.....TDL.E..GT.AIG ...L.DY^{176}	^{259}..K.M.....DANK..AV..E....I. K..GA^{290}	^{462}.Q..DS.KE..A....KF.SQ...K.L.S ..^{491}
	LC041132	^{102}V**R**EN**R** ↓L^{107}	+	^{107}LV..I.................T...^{132}	^{134}.....A..AN......TDL.E..GT.AIG... L..Y^{176}	^{259}..K.M..M..DANK..AV..E....I. K..GA^{290}	^{462}.Q..DS.KENIA...KF.SQ...K.L. S..^{491}
AAvV 14	KX258200	^{98}T**R**EG**K** ↓L^{103}	+	^{103}L...I..TA...L.....V...I.E^{128}	^{129}..QD..RA..T.....RN..N..S.LKT...EV.I.L S.T.DY^{171}	^{254}.V..RIVQ..TAD..VI...SI...S T.SGY.^{285}	^{457}...NK.E.YIDK..EL.NR...DIV NDT.^{486}
AAvV 15	KX932454	^{95}VP**K**E**R**↓L^{100}	+	^{100}LV..I..TV......S.......V.VA^{125}	^{130}N.KDIWKLKNAILSTNE.VL.LKTGL QQTAIA.DKIQDYINNE^{172}	^{255}QVVD.DPASQILIIQVSYPS ISRLSDIRATEL^{286}	^{458}.RVEDLIHESESWLSRVNPK LISNTAIIVL^{487}
AAvV 16	KY511044	^{110}LVQA**R** ↓L^{115}	+	^{115}LV.I...V......S......^{140}	^{141}...K.E...K.Q..........D.....A.......D.^{183}	^{266}..S.S.............T....SS......^{297}	^{469}....N.....IE...QI.ES...S....NA^{498}
AAvV 17	KY452442	^{106}GIQS**R**↓I^{111}	+	^{111}IV..IL.GV......S.....G...A^{136}	^{137}SR...E...K..Q......D..Q..IAG.RE.VI.I ....DY^{179}	^{262}..Q.Q..G...EK....ILS.SV..IN AVD.L.^{293}	^{465}.Q..T..QASINK..QI.EG.....IQV P.^{494}
AAvV 18	KY452443	^{108}AAQS**R**↓L^{113}	+	^{113}L...I..GV......S.....GI..A^{138}	^{139}.SK...E...K..QA..N..N..Q.LITSQQEV. T.L..I.DY^{181}	^{264}..Q.Q..A..AEY.I.V..A.SI..INT V..L.^{295}	^{467}...LT..QAAVDK..EI.K......I.E AP^{496}
AAvV 19	KY452444	^{139}**R**GQA**R**↓I^{144}	+	^{144}IV..IL.GV......S......M..V^{169}	^{170}.TE...K...K..K....E..N..Q.II..QQ..GI.I..I. DY^{212}	^{295}..K.Q..G....E.....L.TVRI..ISK IK...^{326}	^{498}.A.....QTA.DR..AI.NG....S.IR KDS^{527}
AAvV 20	MF033136	^{96}EQQA**R**↓L^{101}	+	^{101}L...IV.TV......S.......L.VN^{126}	^{127}..QD..RQ.WK....A.LK..D..L.LKE.....TA ..LD.V.S.^{169}	^{252}.V.QI.G..ASS.I.L....SY...SE ITGV.^{383}	^{455}..TVNKVE.LINQ.NGW..S..PK .MSNTT^{484}

Table 5. A multiple sequence alignment based comparison of residue patterns at important motif/s in the open reading frame of HN protein using different specie-types of avulaviruses 1-20

AAvVs		Hydrophobic signal anchor domain (FRIAVLLLMIMILAISAAAL)	Hexapeptide motif (NRKSCS)	Hemagglutinin active motif-I (FX2YG2V/L/I)	Hemagglutinin active motif-II (GAEG/SRI/V/L)	Hypervariable Regions		
						HRa (LSSSQDVIDRIYKQV)	HRb (LLNTESIIMNAITSL)	HRc (TDPYPLIFHRNHTLRGVFGTMLDDGQA)
AAvV 1	KU885948	25.........45	234 239	314 320	399 404	74.........88	96.........110	471......I......500
	HQ697254	25.........45	234 239	314 320	399 404	74...YE...88	96.........110	471.............500
	JQ247691	25......VI........45	234 239	314 320	399 404	74......E...88	96.........110	471...R.............500
	KR074406	25.........I........45	234 239	314 320	399 404	74.........88	96...V......110	471.............V......500
	KJ782375	25......VI........45	234 239	314 320	399 404	74......E......88	96.........110	471.............V......500
AAvV 2	HM159993	24...LT.SIGVVC...NI.K.44	235 240	315 321	398...A..403	73.KVPVNQ.NDMFRI.87	95MTS..KE.TSQVGF.109	478Q.IW.IS........TAH.SQVVWV.QY.EAFY.507
	HQ896023	24...LT.S..VVC...NI.K.44	235 240	315 321	398...A..403	73.KVPVNQ.N.MFRI.87	95MTS.QKE.TS.GF.109	478Q.IW.IS........TAH.SQV...QY.EAFY.507
	EU338414	24...LT.SI..VC...NI.K.44	235 240	315 321	398...A..403	73.KVP....NDMFRI.87	95MTS.QKE.TSQVGF.109	478Q.IW.IS........TAH.SQVVWV.QY.E..Y.507
	HM159994	24...LT.SI.VVC...NI.K.44	235 240	315 321	398...A..403	73.KVPV..Q.NDM.RI.87	95MTSLQKE.TS.VGF.109	478Q.IW.IS........TAH.SQV..V.QY.EAFY.507
AAvV 3	EU403085	24Y.ATT.S.NT.VV..ISI.L44	237 242	317 323	402...A..407	77.ADLESQLRE.RRDT.91	99ID...NL.LTTLA.I113	480..VW.TKN..NNDAQDPNLFYTVY.NNSTR507
	EU782925	24Y.....S.NT.VV..ISI.L44	237 242	317 323	402...A..407	77.ADLE...RE.RRDT.91	99ID...NL...LA.I113	480..VW..TKN...AQDPNLFYTVY...STR507
AAvV 4	JX133079	25Y.VVSL...VSA.I..IVI.45	256 261	335 341	422 427	101IT.IMTDTLDTRNAA115	123.NSL.ANLLS.LGGN137	502S.IW..SL...NNNTSDSIFA...QG..T528
	EU877976	25Y...SL...VSA.I..IVI.45	256 261	335 341	422 427	101IT.IM...LDTR.AA115	123.NSL.ANLLS.L.GN137	502S.IW..SL...NNNTSDSIFA.T.Y.QGKTT528
	KU601399	25Y.VVSL...VSA.I..IVI.45	256 261	335 341	422 427	101IT.IMTDTLDTRNAA115	123...L.ANLLS.LGGN137	502S.I...SL...NNNT..IFA.TMY.Q...T528
	KC439346	25Y...SL...VSA.I..IVI.45	256 261	335 341	422 427	101IT.IMTDTLDTRNAA115	123....ANLLS.L..N137	502S.I...SL...N...SDSIFA.TMY.QGKTT528
AAvV 5	GU206351	23..LI..T.LCL.ACT.VSQ.43	234 239	314 320	397 402	72ILNQNIKE.L..RE.86	94.DRVTVEVGT.VNQI108	477Q.LWAIS....AGETLSEMTF.GY.EASTQ506
	LC168750	23..LI....LCL.ACT.VSQ.43	234 239	314 320	397 402	72ILNQNIKE.L..RE.86	94.DRVTVEVGT.VNQI108	477Q.L..IS....AGETL...TF.GY.EASTQ506
AAvV 6	JX522537	29Y.V....DTTV.CVGIV.V49	240 245	320 326	403A..408	78W.DTHQKVNS.F.E.92	100.DKMQV..GT.VNII114	483Q.VWVVS....IGSQSNETV.V.GY..AAA.512
	KP762799	29Y.V.T..D.TV.CV.IV.V49	240 245	320 326	403A..408	78W.DTH....S.F.E.92	100.DKMQVEMGT.VNII114	483Q.VWVVS....IGSQS..V.V.GY..AAA.512
	EU622637	29Y.VI...DT...CV.IV.V49	240 245	320 326	403A..408	78W.DTHQK.NS.F.E.92	100.DKMQ...GT.VNII114	483Q.VWVVS....I...SNETV.V.GY..AAA.512
	AY029299	29Y.VIT..D.TV..VGIV.V49	240 245	320 326	403A..408	78W.DT...VNS.F.E.92	100.DKMQVE.GT.VNII114	483Q.VWVVS....IGSQSNETV.V.GY......512
AAvV 7	FJ231524	25Y.VTI..DVV..SV.ALIS.45	225 230	304 310	389 394	62T.DI.AKVSS..RSN96	84.DQINQA.SSSARQ198	469Q.LWV.Y....DLGKLENTTAV.LY.NSAVG498
AAvV 8	FJ215863	25.WS...ITTGC..L.ISI.45	236 241	316 322	399A..404	74.RNPINQ.ND.FRI.88	96VT.IQKDLASQFNM.110	479S.IW..T....SSSSLPSIIWI.QY..APVR508
	FJ215864	25.WS...ITTGC..L.ISI.45	236 241	316 322	399A..404	74.RNPINQ.ND.FRI.88	96VT.I..DLASQFNM.110	479S.IW..T....SSSSLPSIIWI.QY..APVR508
	MF448515	25..WS...IT..C..L.ISI.45	236 241	316 322	399A..404	74.RNP....ND.FRI.88	96VT.IQKDLASQFNM.110	479S.I...T....SS..LPSI..I.QY..A..R508
	FJ619036	25..WS...ITT.C..L.ISI.45	236 241	316 322	399A..404	74.RNP..Q.ND.FRI.88	96VT.IQKDL..QFNM.110	479S.IW..T....SS.SLPSIIWI.Q...A.VR508
AAvV 9	EU910942	25..VVS...GFTS.VLTAC..45	234 239	314 320	399A..404	74.LMS.IGNK....88	96......TL.S...110	471.A...F....WSED.KVN..Y.M....ITS500
AAvV 10	HM755886	25..VTTITMVVTCVVL..II.45	236 241	316 322	399 404	74.TVPNNQ.GE.F.I.88	96VSSSQQA.AGQ.GM.110	479.MW.IS....ITG.ISDYAWISHY..APTS508
	HM147142	25..V....MVVT..VL..II.45	236 241	316 322	399 404	74.TVPNNQ.GE.F.I.88	96VSS..QA.A.Q.GM.110	479..MW.IS....ITG.ISDYA....Y..APTS508
AAvV 11	JQ886184	24..VVLII.Q.TM.C..IS.44	235 240	315 321	400 405	73.NPLLSYLPG.NRE.87	95.DKIQQSATSE.NR.109	480Q.LW..SPLAIEN.TA.NPTFA.AF.NAFTT509
AAvV 12	KC333050	25..VCMII.ALSS.CVTV...45	234 239	314 320	399A..404	74V.RLE.TSQK..R..88	96Q..M.TN.L......110	471.A...V.....S.S.DIVA.Y.MQ.AA.T.500
AAvV 13	KU646513	25...LTIV.CLTSIG.GIP..45	234 239	314 320	399 404	74IDE.INTEQK..R..88	96Q..M..N.LS..A..110	471.G..IV......S.SIAA.Y.MQ.N.VTN500
	KX119151	25...L.IV.CLTS...GIP..45	234 239	314 320	399 404	74I...INTEQK..R..88	96Q..M.....S..A..110	471.G..IV......S..SI...Y.MQ.N.VTN500
	LC041132	25...L..V.CLTSIG.GIP..45	234 239	314 320	399 404	74I.E.INT..K..R..88	96Q..M..N..S..A..110	471.G..IV......S.SIAA.Y.MQ.N.V.N500
AAvV 14	KX258200	23..TLT.I.NLA..SVTI.SI43	234 239	314 320	397 402	72..D.VQQEKM...E.86	94.DKIQTDVGTSVAQI108	477..LFSVT....IGSTTDKDTY..VY..SATERKD506

32

AAvV 15	KX932454	^{27}Y.AVS.IFQ.L.FSLVLTNVI^{47}	^{238}.......^{243}	^{318}.......^{324}	^{401}.S.N..^{406}	^{76}.ETPLNQ.ND.FRLT^{90}	^{98}MNTMTRE.TSQLNI.^{112}	^{481}A.IW..NNPAEPSKIIWVSHY.NSEVG.EF^{510}
AAvV 16	KY511044	^{25}...T..VVSVIT.GLT..S.^{45}	^{234}.......^{239}	^{314}Y.......^{320}	^{399}.......^{404}	^{74}IGGM...L..T....^{88}	^{96}...M..T........^{110}	^{471}............AKD.SI...Y.M...N.VT.^{500}
AAvV 17	KY452442	^{56}..VSAIA.LLT..GL.I.TFC^{76}	^{265}.......^{270}	^{345}.......^{351}	^{430}...A..^{435}	^{105}ITAIL.YCQK....S^{119}	^{127}.SAMQTS.LQSLSA.^{141}	^{502}..A..I..SADKKPI.....Y.NHRSD^{528}
AAvV 18	KY452443	^{45}Y.ALSIMMLTTV.GLAV.IFC^{65}	^{254}.......^{259}	^{334}.......^{340}	^{419}.......^{424}	^{94}VVEVLE.ESK..R.S^{108}	^{116}.SAMQTS..QSLSA.^{130}	^{491}..A..IV.SASKAPL.....Y..HRFD^{517}
AAvV 19	KY452444	^{26}..LTC.AMLVTI.SLN.VLVV^{46}	^{235}.....^{240}	^{315}.....^{321}	^{400}.......^{405}	^{75}.KGIDEGVGK....T^{89}	^{97}.NSF.TNV.QGLS.I^{111}	^{472}..GF..V.DNSGNV.A.Y.MY.S.KTQ^{498}
AAvV 20	MF033136	^{23}...TS.I.LLL.TGMTI.G.V^{43}	^{234}.......^{239}	^{314}.......^{320}	^{397}.......^{402}	^{72}.TV.NNQ.ND.F.I.^{86}	^{94}VSSIQQD.SSQFNK.^{108}	^{477}..LW.TESSNASNIVWVGQY..AITE^{503}

The highly significant conserved residues are bold and underlined. "." Shows the identical residue

33

Table 6. A multiple sequence alignment based comparison of residue patterns at important motif/s in the open reading frames of polymerase-associated proteins (NP, P, and L) using different specie-types of avulaviruses 1-20

AAvVs		NP Protein N-N assembly motif (FX4YX3YXYAMG)	P Protein RNA editing motif* (UUUUUCCC)	L Protein Domain-I (FXTTDLQKYCLN WRYQS/T)	Domain-II (FIVSXRGGIEGLCQK LWTMISIAAI)	Domain-III (CMVQGDNQVIA VTR)	Domain-IV (FDGAILSQVLKNSS KL)	ATP binding motif (KX21GXGXG)
AAvV 1	KU885948	^{322}............F...^{336}	....	^{637}.................^{653}	^{709}Y.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...^{1782}
	HQ697254	^{322}............F...^{336}	....	^{637}.................^{653}	^{709}Y.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...^{1782}
	JQ247691	^{322}............F...^{336}	....	^{637}.................^{653}	^{709}Y.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...^{1782}
	KR074406	^{322}............F...^{336}	....	^{637}.................^{653}	^{709}Y.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...^{1782}
	KJ782375	^{322}............F...^{336}	....	^{637}.................^{653}	^{709}Y.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...^{1782}
AAvV 2	HM159993	^{324}....F.........^{338}	....	^{661}.............G.^{677}	^{733}.................SI.^{757}	^{770}A.............^{783}	^{840}WE.RT...S...AT..^{845}	^{1796}.............G^{1822}
	HQ896023	^{324}....F.........^{338}	....	^{661}.............G.^{677}	^{733}.................SI.^{757}	^{770}A.....I.^{783}	^{840}WE..T...S...AT..^{845}	^{1796}.............G^{1822}
	EU338414	^{324}....F.........^{338}	....	^{661}.............G.^{677}	^{733}.................SI.^{757}	^{770}A.....I.^{783}	^{840}WE..T...S...AT..^{845}	^{1796}.............G^{1822}
	HM159994	^{324}....F.........^{338}	....	^{661}.............G.^{677}	^{733}.................SI.^{757}	^{770}A.....I.^{783}	^{840}WE..T...S...AT..^{845}	^{1796}.............G^{1822}
AAvV 3	EU403085	^{322}.............^{336}	AA......	^{632}.................^{648}	^{704}....A.......M..II.IS.^{728}	^{741}S........G...^{754}	^{811}YP.RV.P....IR..^{826}	^{1752}R......^{1778}
	EU782925	^{322}.............^{336}	AA......	^{632}.................^{648}	^{704}....A.......M..II.IS.^{728}	^{741}S........G...^{754}	^{811}Y..R..P.L......^{826}	^{1752}R......^{1778}
AAvV 4	JX133079	^{322}....F.........^{336}	....	^{644}Y...........H..^{661}	^{716}..I...........TG..^{740}	^{753}ATL......L.I.K^{766}	^{823}.N.S....C....F..^{838}	^{1767}R......^{1793}
	EU877976	^{322}....F.........^{336}	....	^{644}Y.....KH...^{661}	^{716}..I.......Q...TG..^{740}	^{753}ATL......L.I.K^{766}	^{823}.N.S....C....F..^{838}	^{1767}R......^{1793}
	KU601399	^{322}....F.........^{336}	R......	^{644}Y.....KH...^{661}	^{716}..I.......Q...TG..^{740}	^{753}ATL......L.I.K^{766}	^{823}.N.S....C....F..^{838}	^{1767}R......^{1793}
	KC439346	^{322}....F.........^{336}	....	^{644}Y........KH..^{661}	^{716}..I.......Q...TG..^{740}	^{753}ATL......L.I.K^{766}	^{823}.N.S....C....F..^{838}	^{1767}R......^{1793}
AAvV 5	GU206351	^{324}.............^{338}	....	^{688}...........P.^{704}	^{760}.........M....S..^{780}	^{797}S......A.I..^{810}	^{867}.E.R....G...AA..^{882}	^{1819}............G^{1845}
	LC168750	^{324}.............^{338}	....	^{688}...........P.^{704}	^{760}.........M....S..^{780}	^{797}S......A.I..^{810}	^{867}.E.R....G...AA..^{882}	^{1819}............G^{1845}
AAvV 6	JX522537	^{324}.............^{338}	....	^{664}...............TV^{680}	^{736}.........M......^{760}	^{773}S......A.I.T^{786}	^{843}.E.K....L...A.R.^{848}	^{1797}....S..^{1823}
	KP762799	^{324}.............^{338}	....	^{664}...............TV^{680}	^{736}.........M......^{760}	^{773}S......A.I.T^{786}	^{843}.E.K....L...A.R.^{848}	^{1797}....S..^{1823}
	EU622637	^{324}.............^{338}	....	^{664}...............TV^{680}	^{736}.........M......^{760}	^{773}S......A.I.T^{786}	^{843}.E.K....L...A.R.^{848}	^{1797}....S..^{1823}
	AY029299	^{324}.............^{338}	....	^{664}...............TV^{680}	^{736}.........M......^{760}	^{773}S......A.I.T^{786}	^{843}.E.K....L...A.R.^{848}	^{1797}....S..^{1823}
AAvV 7	FJ231524	^{324}.............^{338}	....	^{655}Y.........E.^{671}	^{727}.........M....S..^{751}	^{764}S........I..^{777}	^{834}.E.RV...I...A...^{839}	^{1784}..C...^{1810}
AAvV 8	FJ215863	^{324}.............^{338}	....	^{661}.....E...Q...S.^{677}	^{734}.........M....S..^{758}	^{770}A.........I.K^{783}	^{840}.E.R.......AA...^{845}	^{1796}............^{1822}
	FJ215864	^{324}.............^{338}	....	^{661}.....E...Q...S.^{677}	^{734}.........M....S..^{758}	^{770}A.........I.K^{783}	^{840}.E.R.......AA...^{845}	^{1796}............^{1820}
	MF448515	^{324}.............^{338}	....	^{661}.....E...Q...S.^{677}	^{734}.........M....S..^{758}	^{770}A.........I.K^{783}	^{840}.E.R.......AA...^{845}	^{1796}............^{1820}
	FJ619036	^{324}.............^{338}	....	^{661}.....E...Q...S.^{677}	^{734}.........M....S..^{758}	^{770}A.........I.K^{783}	^{840}.E.R.......AA...^{845}	^{1796}............^{1820}
AAvV 9	EU910942	^{322}.............^{336}	....	^{637}Y.............^{654}	^{709}.........T...S..^{733}	^{746}.................^{759}	^{816}K..K.....A...AA.^{831}	^{1759}............^{1785}
AAvV 10	HM755886	^{324}.............^{338}	....	^{661}.............A.^{677}	^{733}.........M....SI.^{757}	^{770}A.........GI..^{783}	^{840}WE.R....L...AT..^{845}	^{1796}............^{1820}
	HM147142	^{324}.............^{338}	....	^{661}.............A.^{677}	^{733}.........M....SI.^{757}	^{770}A.........GI..^{783}	^{840}WE.R....L...AT..^{845}	^{1796}............^{1820}
AAvV 11	JQ886184	^{324}.............^{338}	.C..AGU.	^{669}.............E.^{685}	^{741}.........M....S..^{765}	^{778}S.........I.K^{791}	^{849}.E.K....S...AL..^{854}	^{1802}............^{1828}
AAvV 12	KC333050	^{322}....F.........^{336}	....	^{635}.....S.........^{651}	^{707}.........S....S..^{731}	^{744}.........S.K^{757}	^{814}K.R.......AA.I^{829}	^{1756}............^{1782}
AAvV 13	KU646513	^{322}...........FS.^{336}	....	^{635}.....S.........^{651}	^{707}.........S......^{731}	^{744}.........S.K^{757}	^{814}K.R......M...AA..^{829}	^{1756}............^{1782}
	KX119151	^{322}...........FS..^{336}	....	^{635}.....S.........^{651}	^{707}.........S......^{731}	^{744}.........S.K^{757}	^{814}K.......M...AA..^{829}	^{1756}............^{1782}
	LC041132	^{322}...........FS.^{336}	....	^{635}.....S.........^{651}	^{707}.........S......^{731}	^{744}.........S.K^{757}	^{814}K.......M...AA..^{829}	^{1756}............^{1782}
AAvV 14	KX258200	^{324}.............^{338}	A......	^{667}....T........P.^{683}	^{739}.........M.S.I...^{763}	^{776}A.......TM.I.K^{789}	^{846}.E.RV...G...A.R.^{851}	^{1800}............^{1826}
AAvV 15	KX932454	^{334}.............^{336}	....	^{661}.........S.^{677}	^{733}.................SI.^{757}	^{770}A.........I..^{783}	^{840}WE.R....I...AT..^{845}	^{1796}............^{1820}
AAvV 16	KY511044	^{334}............F.^{336}	.....U..	^{637}.................^{653}	^{709}.................^{733}	^{746}.................^{759}	^{816}K.................^{831}	^{1756}..A...S^{1782}

34

AAvV 17	KY452442	^{322}.............^{336}	……..	^{639}.M….E…T…..V^{655}	^{711}.....................S..^{735}	^{748}.........L…K^{761}	^{818}K..K….I…A…^{833}	^{1758}..A…^{1784}
AAvV 18	KY452443	^{352}.............^{366}	……..	^{645}.L….E…T…..V^{661}	^{717}.....................S..^{741}	^{754}.........L…K^{767}	^{824}K..R….I…A…^{839}	^{1764}..A…^{1790}
AAvV 19	KY452444	^{342}.............^{356}	……..	^{656}.L……..T……^{672}	^{728}.I…………..S…….^{752}	^{765}…………..K^{778}	^{835}K..R….M…A…^{850}	^{1775}..A…^{1801}
AAvV 20	MF033136	^{324}…..…....…^{338}	……..	^{663}.L…………S.^{679}	^{735}….P……………..SI.^{529}	^{772}A……..…..I..^{785}	^{842}WE.R….I…AA.V^{857}	^{1798}……^{1824}

The highly significant conserved residues are bold and underlined. "." Shows the identical residue. *= Negative sense Sequence

35

Table 7. A multiple sequence alignment based comparison of residue patterns at important motif/s in the open reading frame of M protein using different specie-types of avulaviruses 1-20

AAvVs		M Protein	
		M late domain (FPI/LV)	Bipartite nuclear localization motif
AAvV 1	KU885948	$^{23}....^{26}$	^{247}**KKRKK**VTFD**K**IEG**K**I**RR**^{263}
	HQ697254	$^{23}....^{26}$	^{247}**KKGKK**VTFD**K**IEG**K**I**RR**^{263}
	JQ247691	$^{23}....^{26}$	^{247}**KKGKK**VTFD**K**IEG**K**I**RR**^{263}
	KR074406	$^{23}....^{26}$	^{247}**KKGKK**VTFD**K**IEG**K**I**RR**^{263}
	KJ782375	$^{23}....^{26}$	^{247}**KKGKK**VTFD**K**IEG**K**I**RR**^{263}
AAvV 2	HM159993	$^{22}....^{25}$	^{248}**KK**TNA**K**GES**R**TISNLEG ^{264}
	HQ896023	$^{22}....^{25}$	^{248}**KK**TNA**K**GES**R**TISNLEG ^{264}
	EU338414	$^{22}....^{25}$	^{248}**KK**TNA**K**GES**R**TISNLEG ^{264}
	HM159994	$^{22}....^{25}$	^{248}**KK**TNA**K**GES**R**TISNLEG ^{264}
AAvV 3	EU403085	$^{23}....^{26}$	^{253}**KR**TA**K**Q**RR**TPSEI**K**V**R**^{271}
	EU782925	$^{23}....^{26}$	^{253}**KR**TA**K**Q**RR**TPSEI**K**V**R**^{271}
AAvV 4	JX133079	$^{22}....^{25}$	^{248}**RK**GNM**R**TLSQAAD**K**V**RR**^{264}
	EU877976	$^{22}....^{25}$	^{248}**RK**GNM**R**TLSQAAD**K**V**RR**^{264}
	KU601399	$^{22}....^{25}$	^{251}**RK**GNM**R**TLSQAAD**K**V**RR**^{267}
	KC439346	$^{22}....^{25}$	^{251}**RK**GNM**R**TLSQAAD**K**V**RR**^{267}
AAvV 5	GU206351	$^{22}....^{25}$	^{246}**RK**GAD**R**SVLQI**K**E**K**V**RK**^{262}
	LC168750	$^{22}....^{25}$	^{246}**RK**GAD**R**SVLQI**K**E**K**V**RK**^{262}
AAvV 6	JX522537	$^{22}....^{25}$	^{246}**RR**GVD**R**SVENI**R**N**K**V**R**A^{262}
	KP762799	$^{22}....^{25}$	^{246}**RR**GVD**R**SVENI**R**N**K**V**R**A^{262}
	EU622637	$^{22}....^{25}$	^{246}**RR**GVD**R**SVENI**R**N**K**V**R**A^{262}
	AY029299	$^{22}....^{25}$	^{246}**RR**GVD**R**SVENI**R**N**K**V**R**A^{262}
AAvV 7	FJ231524	$^{22}....^{25}$	^{242}ASG**K**P**R**SLEDM**RKK**V**R**D^{268}
AAvV 8	FJ215863	$^{22}....^{25}$	^{248}**KK**TSS**K**G**K**P**R**TLDEL**K**T^{264}
	FJ215864	$^{22}....^{25}$	^{248}**KK**TSS**K**G**K**P**R**TLDEL**K**T^{264}
	MF448515	$^{22}....^{25}$	^{248}**KK**TSS**K**G**K**P**R**TLDEL**K**T^{264}
	FJ619036	$^{22}....^{25}$	^{248}**KK**TSS**K**G**K**P**R**TLDEL**R**T^{264}
AAvV 9	EU910942	$^{23}....^{26}$	^{247}**KR**G**KK**VTFE**K**LEE**K**I**RR**^{263}
AAvV 10	HM755886	$^{22}....^{25}$	^{248}**KK**TNA**K**GEA**R**TLVNLQE^{264}
	HM147142	$^{22}....^{25}$	^{248}**KK**TNA**K**GEA**R**TLVNLQE^{264}
AAvV 11	JQ886184	$^{22}....^{25}$	^{251}M**K**GTL**R**SIQDVIE**K**V**R**^{267}
AAvV 12	KC333050	$^{23}....^{26}$	^{246}**K**HGN**K**LAME**R**LEN**K**I**RR**^{262}
AAvV 13	KU646513	$^{23}....^{26}$	^{246}**K**GN**K**ISVD**K**LEL**K**I**RR**M^{262}
	KX119151	$^{23}....^{26}$	^{246}**K**GN**K**ISVD**K**LEL**K**I**RR**M^{262}
	LC041132	$^{23}....^{26}$	^{246}**K**GN**K**ISVD**K**LEL**K**I**RR**M^{262}
AAvV 14	KX258200	$^{22}....^{25}$	^{242}**RR**GET**R**TVENL**K**E**K**V**RR**^{268}
AAvV 15	KX932454	$^{22}...I^{25}$	^{256}A**R**GEI**R**TLE**K**ISD**K**I**R**A^{272}
AAvV 16	KY511044	$^{23}....^{26}$	^{247}**KK**G**KK**ISFD**K**LER**K**I**RR**^{263}
AAvV 17	KY452442	$^{42}..LI^{45}$	^{265}**KK**GG**KR**HTIEVAE**K**I**RR**^{281}
AAvV 18	KY452443	$^{24}..LI^{27}$	^{247}**KK**GG**KK**HSIEIAN**K**V**RR**^{263}
AAvV 19	KY452444	$^{42}..L.^{45}$	^{266}**K**SG**KR**MPIE**K**VAE**K**V**RR**^{282}
AAvV 20	MF033136	$^{22}..LI^{25}$	^{248}**KRR**GS**R**SLEAI**K**D**K**V**RK**^{266}

The highly significant conserved residues are bold and underlined. "." Shows the identical residue

Table 8. A brief summary of the DnaSP-derived polymorphism analysis for individual complete genes (CDS) of avian avulaviruses 1-20

Parameters	NP	P	M	F	HN	L
Total no. of Mutation	2,303	1,742	2,378	3,024	2,345	12,555
Invariable Monomorphic Sites	206	43	21	67	44	830
Variable Polymorphic Sites	1,041	687	983	1,261	924	5,274
Singleton Variable Sites	79	27	89	112	44	325
Parsimony Informative Sites	962	660	894	1,149	880	4,949
No. of Haplotypes (h)	20	20	20	20	20	20
Nucleotide Diversity (Pi)	0.45902	0.56637	0.55	0.53875	0.57166	0.50181
Average no. of pairwise nucleotide difference (k)	572.395	413.447	560.663	715.458	553.368	3063.047
Tajima' D	-0.49587	-0.66234	-0.68593	-0.67382	-0.68283	-0.56435

Whole genome analysis showed 0.00025 value of variance of haplotype diversity and 0.016 value of Standard Deviation of haplotype diversity. The divergence time was estimated as T= 6.734 with value of X-Square (0.0136) and significant *p*-value (0.0136*) * 0.01< *p* <0.05, **0.001< *p* <0.01, *** *p* <0.001.

Table 9. A brief summary of the Datamonkey-derived natural pressure selection site/s analysis for individual complete genes (CDS) of avian avulaviruses 1-20

Parameters	NP	P	M	F	HN	L
Mean dN/dS	0.226926	0.611799	0.311001	0.680179	0.374762	0.33581
Single Likelihood Ancestor Counting (SLAC)						
No. of positive and negative selection sites along with codon position	0 Positive sites 279 Negative sites	0 Positive site 19 Negative sites	0 Positive site 37 Negative sites	2 Positive site (91,201) 0 Negative site	2 Positive sites (308,506) 68 Negative sites	0 Positive site 43 Negative sites
Fixed Effect Likelihood (FEL)						
No. of positive and negative selection sites along with codon position	1 Positive site (458) 323 Negative sites	4 Positive sites (130,195,313,417) 39 Negative sites	0 Positive site 175 Negative sites	9 Positive sites (22,91,111,281,295,30 2,347,352,401) 17 Negative sites	2 Positive sites (308,506) 171 Negative sites	8 Positive sites (165,1625,1743,1844,187 5,1995,2048,2165) 563 Negative sites
Internal Branch Fixed Effect Likelihood (IFEL)						
No. of positive and negative selection sites along with codon position	1 Positive site (458) 217 Negative sites	5 Positive sites (122,130,182,359,401) 34 Negative sites	1 Positive site (90) 81 Negative sites	15 Positive sites (49,60,91,111,156,220 ,322,333,340,388,393, 417,428,465,529) 10 Negative sites	8 Positive sites (29,78,152,220,308, 322,485,506) 73 Negative sites	13 Positive sites (52,279,994,165,1625,16 80,1743,1875,1994,1995, 2002,2048,2103) 311 Negative sites

38

A

B

C

D

E

F
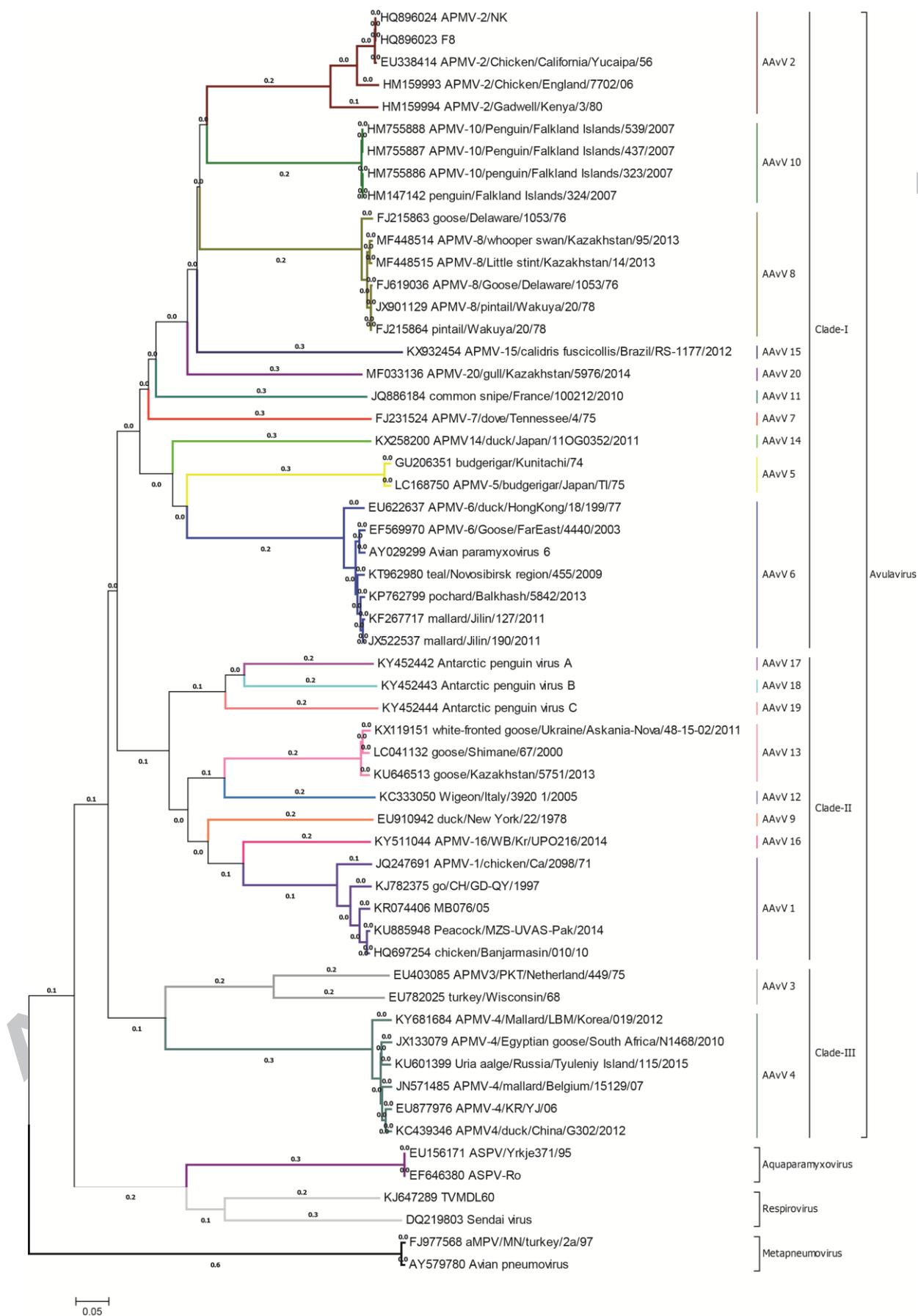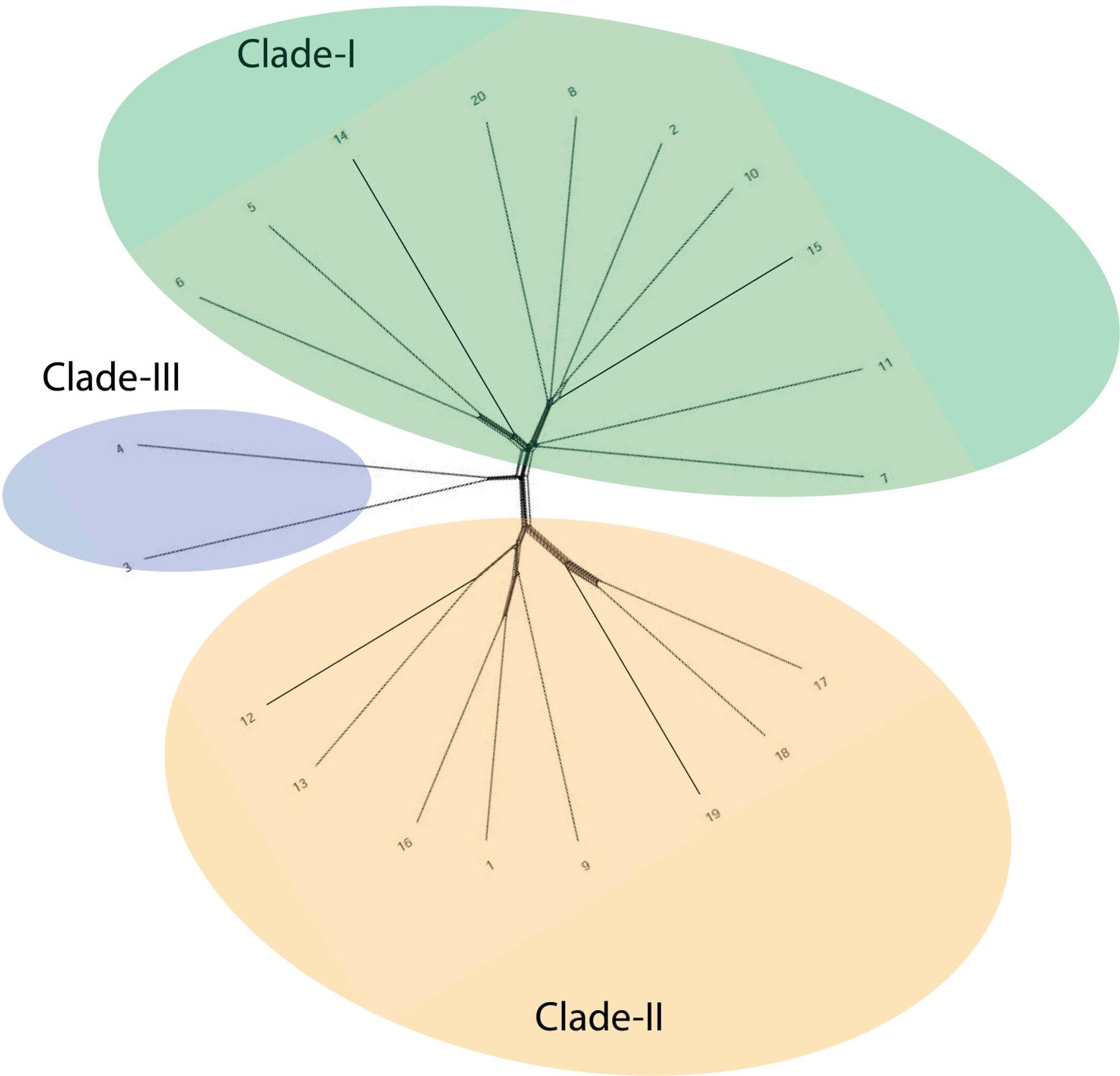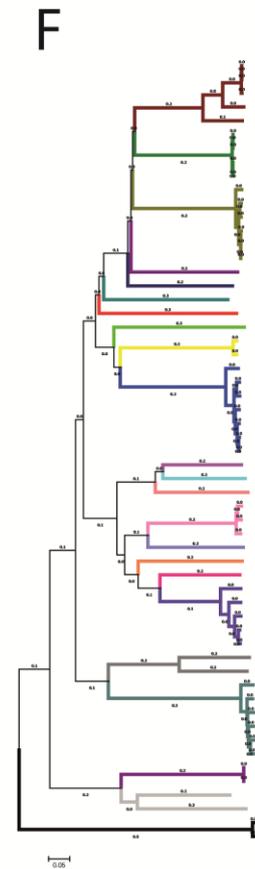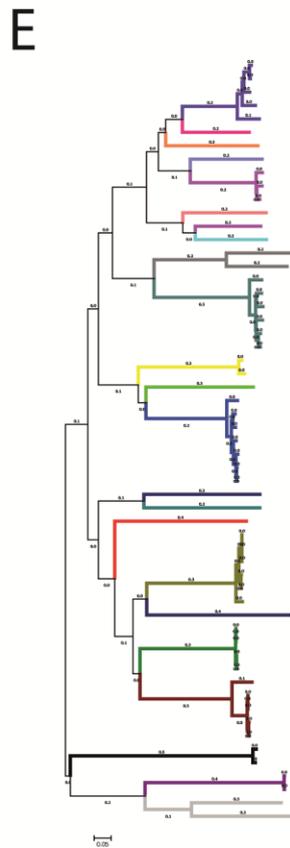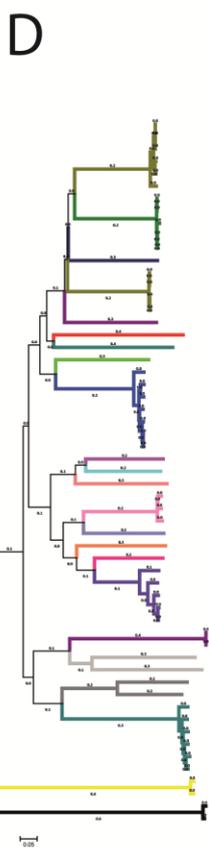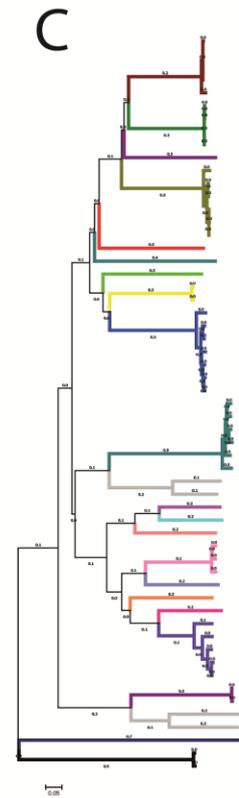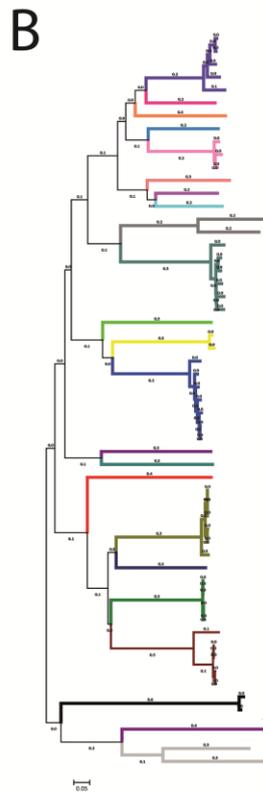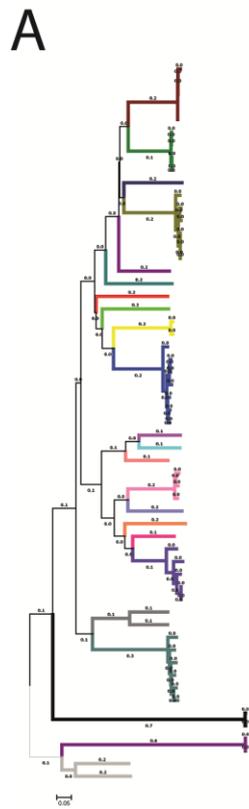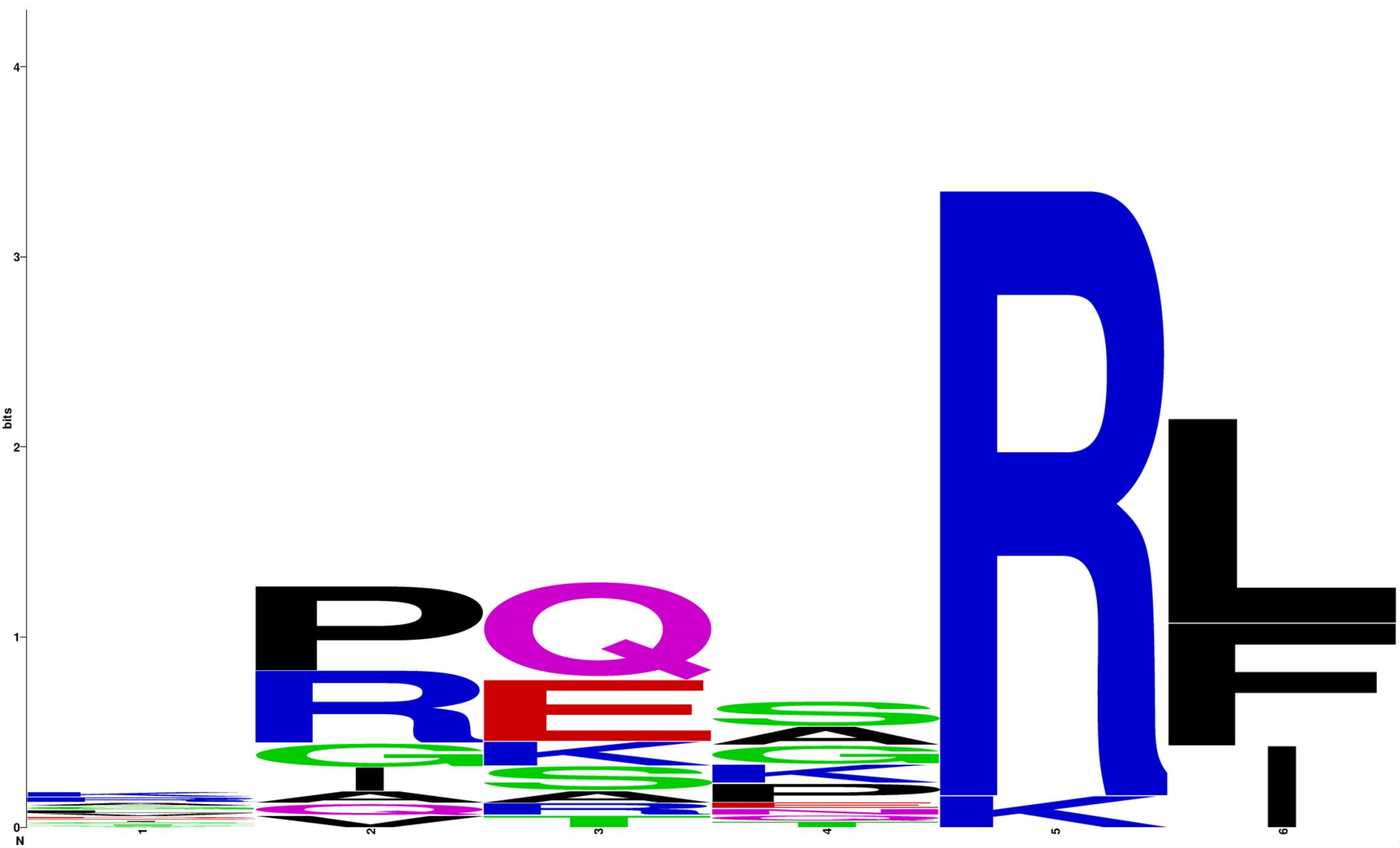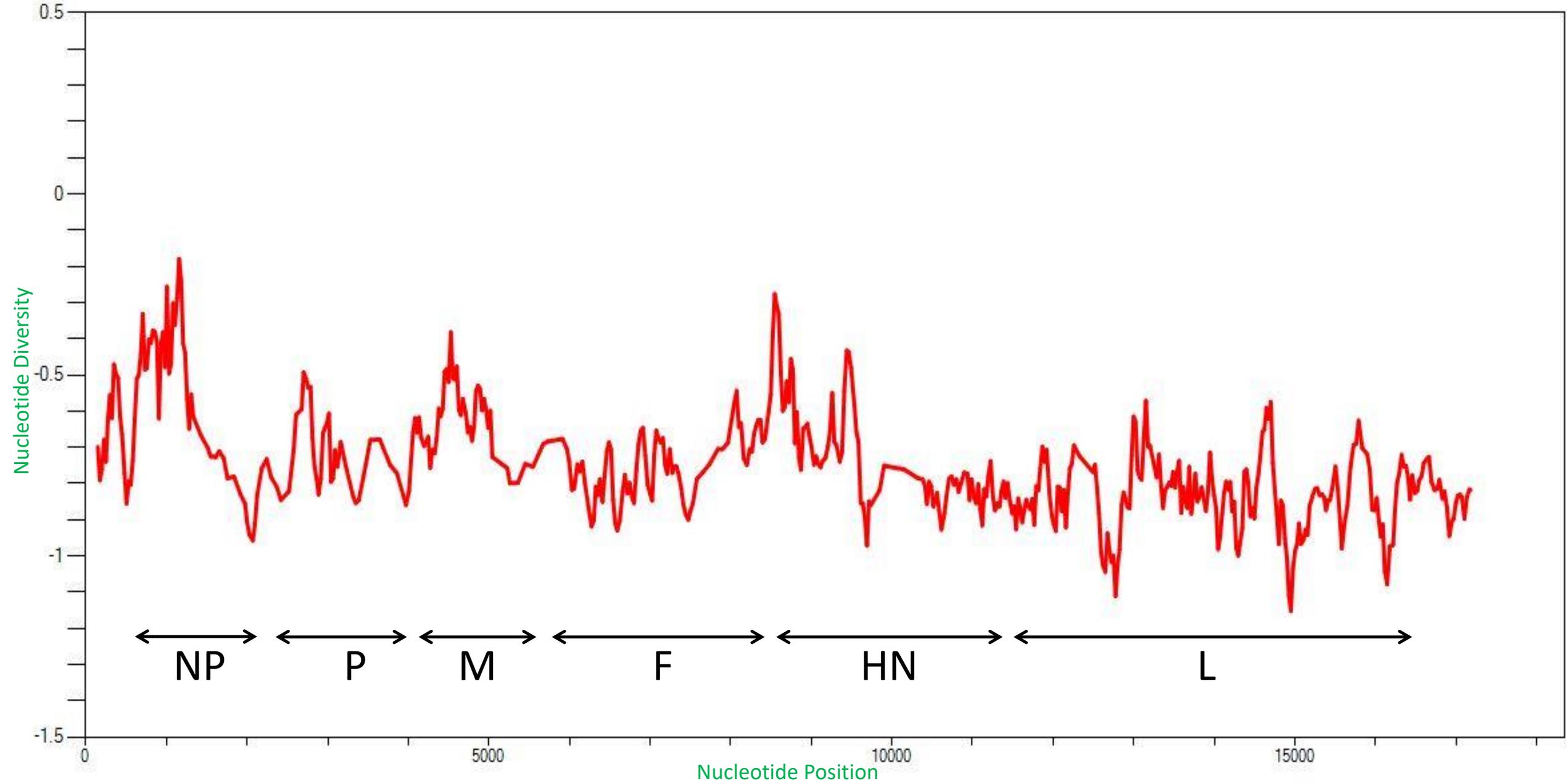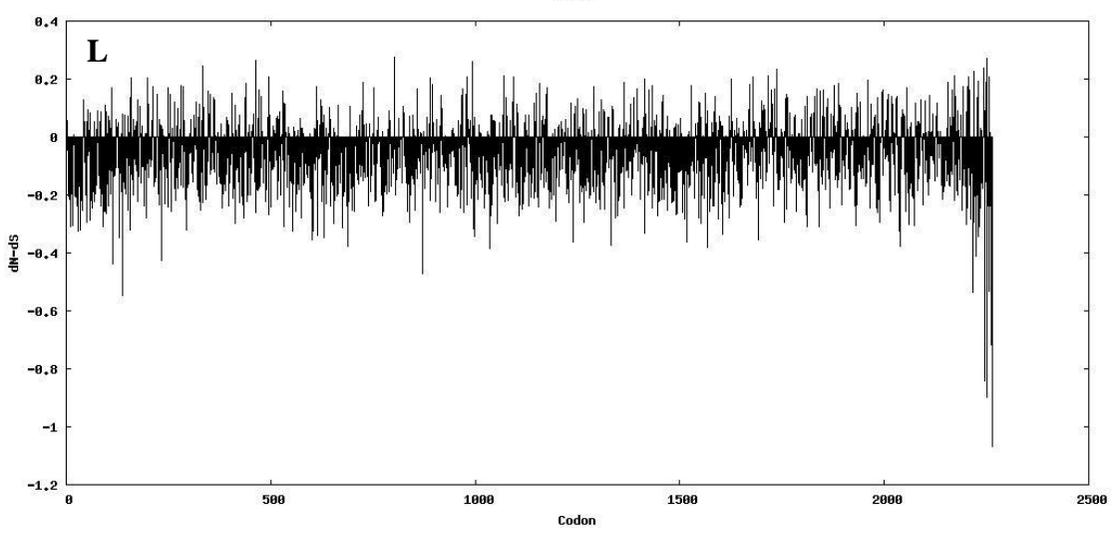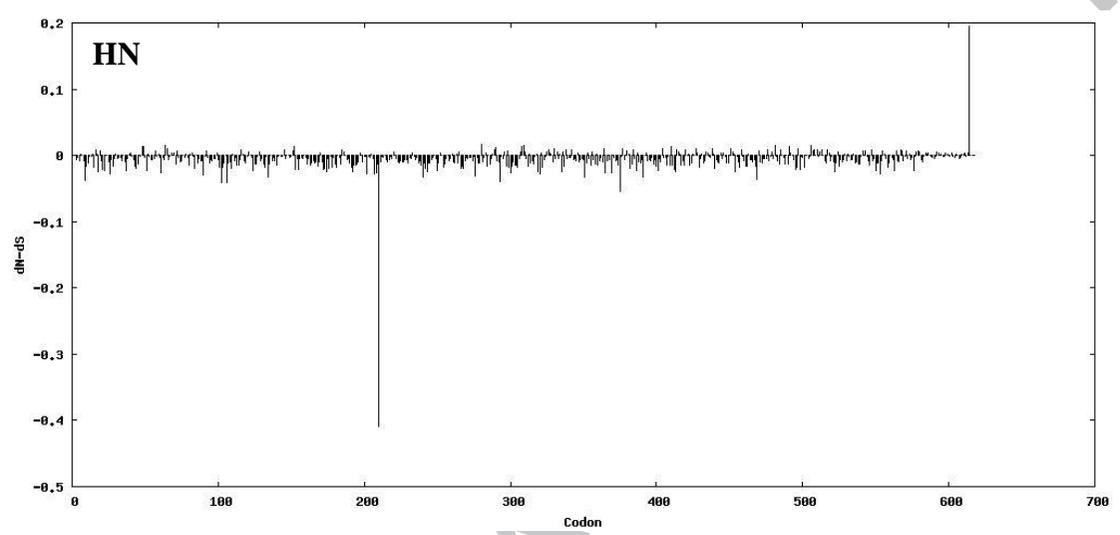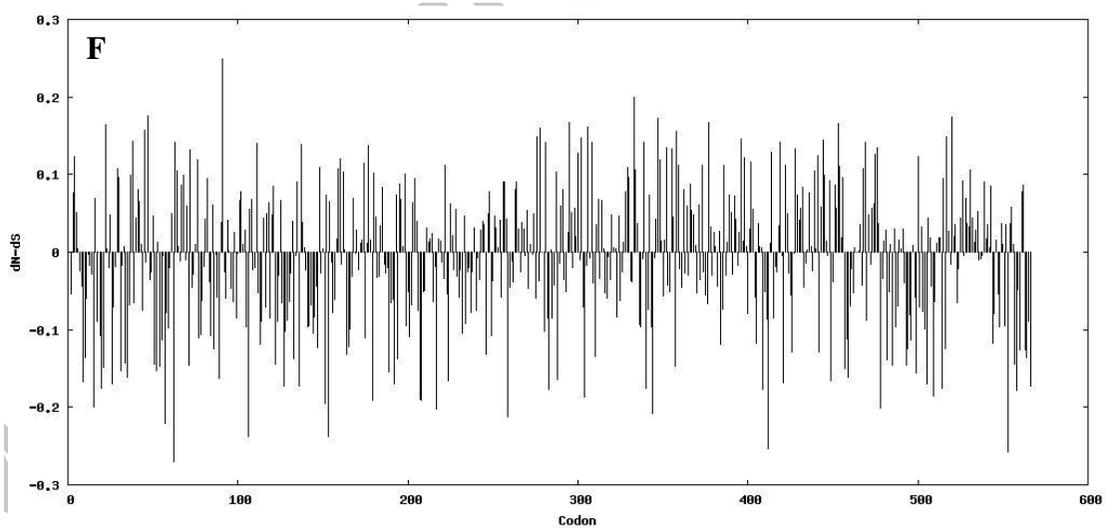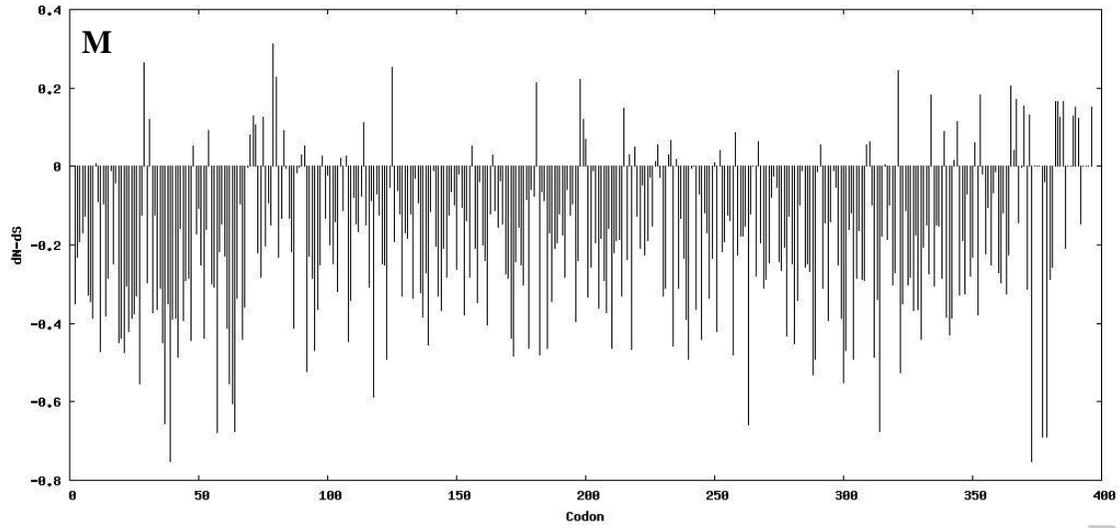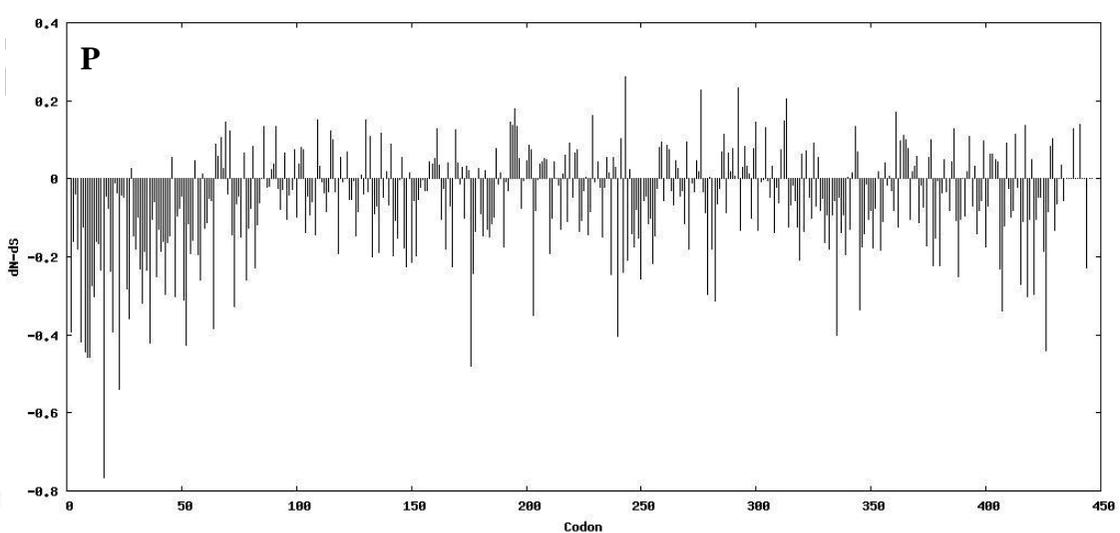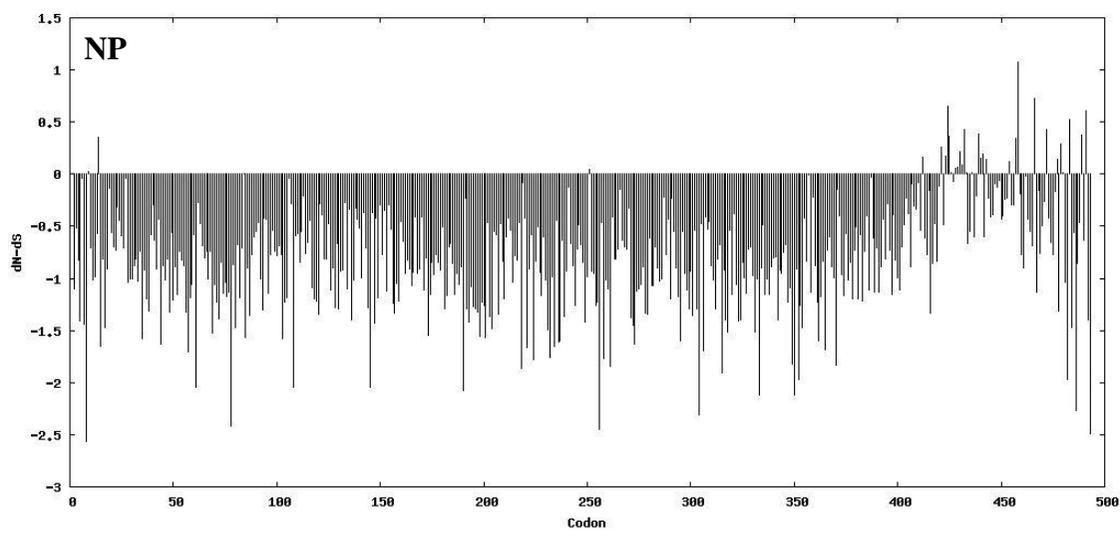
Highlights

- Novel *Avian avulaviruses* (14–20 in particular) have been reported recently that highlight the potential of virus to evolve in the environment

- While biologic and genetic characterization of some of these individual avulaviruses are often demonstrated, a comparative genomic assessment of all avulaviruses would reveal a basis for evolutionary dynamics and future emergence of novel strain around the globe

- For the first time, we demonstrated genomic, residue and evolutionary characteristics of representative strain of each of avulaviruses (specie-type 1 -20)