A comparative genomic and evolutionary analysis of circulating strains of *Avian avulavirus* 1 in Pakistan

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;[2]Division of Biomedical and Life Sciences, Lancaster University, Lancaster LA1 4YG United Kingdom; [3]Quality Operations Laboratory University of Veterinary and Animal Sciences Lahore 54600, Pakistan

***Corresponding authors:** Both authors Aziz-ul-Rahman (drazizangel@gmail.com) and Muhammad Zubair Shabbir (shabbirmz@uvas.edu.pk) equally contributed to this work**.**

[1]ORCID: 0000-0002-3342-4462

[2]ORCID**:** 0000-0002-7889-1398

[3]ORCID: 0000-0002-3562-007X

Abstract

Newcastle disease, caused by *Avian avulavirus* 1 (AAvV 1), is endemic to many developing countries around the globe including Pakistan. Frequent epidemics are not uncommon even in vaccinated populations and are largely attributed to the genetic divergence of prevailing isolates and their transmission in the environment. With the strengthening of laboratory capabilities in Pakistan, a number of genetically diverse AAvV 1 strains have recently been isolated and individually characterized in comparison with isolates reported elsewhere in the world. However, there lacks sufficient comparative genomic and phylogenomic analyses of field circulating strains that can elucidate the evolutionary dynamics over a period of time. Herein, we enriched the whole genome sequences of AAvV reported so far (n= 35) from Pakistan and performed comparative genomic, phylogenetic and evolutionary analyses. Based on these analyses, we found only isolates belonging to genotypes VI, VII and XIII of AAvV 1 in a wide range of avian and human hosts. Comparative phylogeny revealed the concurrent

circulation of avualviruses representing different sub-genotypes such as VIg, VIm, VIIa, VIIb, VIIe, VIIf, VIIi, XIIIb and XIIId. We found that that isolates of genotype VII were more closely associated with viruses of genotype XIII than genotype VI. An inter-genotype comparative residue analysis revealed a few substitutions in structurally and functionally important motifs. Putative recombination events were reported for only one of the captive-wild bird (pheasant)-origin isolates. The viruses of genotype VII had high genetic diversity as compared to isolates from genotypes VI and XIII and, therefore, have more potential to evolve over time. Taken together, the current study provides an insight into the genetic diversity and evolutionary dynamics of AAvV 1 circulating in Pakistan. Such findings are expected to facilitate better intervention strategies for the prevention and control of ND in disease-endemic countries across the globe particularly Pakistan.

Keywords *Avian avulavirus* 1; genomic comparison; phylogenomic analysis; evolutionary dynamic; genetic diversity; genotypes and sub-genotypes

Introduction

Avian avulavirus 1 (AAvV 1) causes a highly contagious disease (Newcastle disease; ND) in commercial and backyard poultry with enormous economic impacts worldwide (Alexander 2003). The virus was recently classified as genus *Orthoavulavirus* within the family *Paramyxoviridae* (Kuhn et al. 2019). It has an enveloped, mono-partite, negative-sense single-stranded RNA genome that is potentially 15186, 15192 or 15198 nucleotides in length (Kolakofsky et al. 2005). The whole genome contains six coding genes: nucleocapsid (*N*), phosphoprotein (*P*), matrix (*M*), fusion (*F*), haemagglutinin-neuraminidase (*HN*) and large polymerase (*L*) protein. These are encoded in the order 3′-*NP-P/V/W-M-F-HN-L*-5′ (Aldous et al. 2003). Based on pathogenicity, the virus is categorized into three major pathotypes including velogenic, mesogenic and lentogenic (Alexander 1998).

Phylogenetic analyses classified the viruses of AAvV 1 into two distinct classes; class-I and class-II. Class-I contains nine distinct genotypes (I-IX) of avirulent strains that share a 15198 nt long genome, while class-II contains at least eighteen distinct genotypes (I-XVIII) of high virulent, low virulent and avirulent strains that have a 15186 and/or 15192 nt long genome (Aldous et al. 2003; Kim et al. 2007). Based on partial sequencing of *F* genes (375bp between 4597-4972 nts), avualviruses of genotypes I, II, VI, VII, XIII are further divided into different sub-genotypes of both virulent and avirulent strains isolated from a wide range of avian species (Shabbir et al. 2012a, b, 2013, 2016; Munir et al. 2012a, b; Miller et al. 2015; Akhtar et al. 2016; Nath and Kumar 2015; Das and Kumar 2017, Xue et al. 2017; Barman et al. 2017; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019).

ND is endemic to many parts of the globe including Pakistan. Similar to any other developing country, there exists a mixed poultry production system in Pakistan comprising of a large population of commercial and domestic poultry. Such a production system, which includes many birds being raised in backyard poultry farms and live bird markets, may favour the emergence of novel avulaviruses that may be divergent enough to be regarded as new sub-genotypes or escape mutants. This is evident from the fact that despite exhaustive application of classical vaccines and biosecurity measures, a number of disease outbreaks are reported each year in multiple avian hosts. In this regard, a vast majority of studies in Pakistan have reported the epidemiology and surveillance of ND in commercial and domestic poultry and wild susceptible avian hosts. Historically, beginning in the 1990s, these studies were limited to sero-surveillance (Numan et al. 2005; Aziz-ul-Rahman et al. 2017); whereas in recent years, partial genome-based studies have been conducted to investigate strain genotyping (Shabbir et al. 2012a, 2013; Munir et al. 2012b; Farooq et al. 2014; Akhtar et al. 2016, 2017; Wajid et al. 2018) while few studies involved in complete genome sequencing and biologic characterization

of field circulating AAvV 1 strains (Munir et al. 2012a; Shabbir et al. 2012b, 2016, 2018; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019).

Notably, each of these isolates are individually characterized based on a comparison of their genomic characteristics with avulaviruses reported elsewhere in the world, instead of countrywide reported complete genome sequences of avulaviruses originating from multiple hosts. Such analysis could provide baseline knowledge in understanding genetic diversity and evolutionary potential of circulating avulaviruses in a particular geographical setting over time. The resulting analysis may not only be used to inform the appropriate intervention for disease control and management, but may also be replicated for other settings across the globe with similar poultry-rearing systems.

Materials and Methods

Database information

To date (April, 2019), a total of 287 genome sequences corresponding to AAvV 1 from Pakistan are publicly available (http://www.ncbi.nlm.nih.gov/). These include complete genome (n = 35) and partial *F* gene (n=252) sequences including the hypervariable region reported from multiple avian hosts including commercial and backyard poultry, pigeons, captive- and migratory wild birds. All essential and relevant information about the complete genome were retrieved from these individual studies for the evolutionary and phylogenomic, comparative residue, and putative recombination analyses (Table 1).

Phylogenetic analysis and evolutionary distance estimation

The whole genome and partial *F* gene sequences were aligned to corresponding sequences of avulaviruses representing different genotypes and sub-genotypes across the globe (GenBank) using ClustalW methods in BioEdit® version 5.0.6 (Hall 1999) for subsequent phylogenetic analysis, estimation of evolutionary distances and prediction of deduced amino acid

substitution sites. To describe the topology of the phylogenetic tree, an analysis of the best fit substitution model was performed in MEGA® version 6.0 software, and the goodness of fit of the individual model was measured by corrected Akaine Information Criterion (AICc) and Bayesian Information Criterion (BIC) (Tamura et al. 2013). To reveal evolutionary dynamics of sub-genotypes, a time calibrated phylogenetic analysis was performed using two different models: the General Time Reversible model (GTR) with discrete gamma distribution (+G) and the Real-Time-Maximum Likelihood model for invariant sites (+I) in sequences. The evolutionary history was inferred by relative divergence time, and the statistical analysis was based on 1000 replication bootstrap values in MEGA® version 6.0 (Tamura et al. 2013). In this analysis, all positions containing gaps and/or missing codon were eliminated with codon positions as 1^{st}, 2^{nd}, 3^{rd} and non-coding.

To determine the nucleotide identity and divergence, Pairwise Sequence Comparisons (PASC) analysis was performed using whole genome sequences in MEGA® version 6.0 (Tamura et al. 2013). The mean inter-population evolutionary diversity (mean evolutionary distance) between sub-genotypes was estimated through PASC analysis using the maximum composite likelihood method (d: Transitions + Transversions model). The rate and pattern of substitutions among sites was modelled with gamma distribution (parameter=1 with homogenous lineage pattern) (Tamura and Kumar 2002). A few partial *F* gene sequences representing different sub-genotypes were used in this analysis. To further assess the evolutionary network of partial *F* genes and, for identification of ancestor isolate and genotype, Splits Tree4 program (version 4.95) was employed using the Neighbour-Net graph method based on pairwise distances estimated by uncorrected *p*-distance and angle split transformation setting (Huson and Bryant 2006).

Comparative residue analysis and selective pressure analysis

ORF Finder and BioEdit were used to predict deduced amino acid sequences of all coding genes. A comparative residue alignment for identification of residue substitutions in conserved and functional motifs of all proteins of isolates from different genotypes (VI, VII, XIII) was created using graphic view option in BioEdit® version 5.0.6 (Hall 1999). Similarly, compared to vaccine strains (LaSota; AF077761, Mukteswar; EF201805), sub-genotype-based residue substitutions were also identified in partial *F* gene sequences containing the hypervariable region using BioEdit. The Datamonkey adaptive evolution server (http://www.datamonkey.org/) was used to evaluate the nature of selection among coding DNA sequences (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 Fast Unbiased Bayesian Approximation (FUBAR) at Bayer factor $p = 0.05$.

Genetic diversity and genome based analysis for vaccine efficacy

The genetic diversity among CDS of avulaviruses included in the study was assessed for genome polymorphism on the basis of variable sites and mutations and the average number of pairwise nucleotide differences identified using DnaSP version 5.10.01 (http://www.ub.es/dnasp) (Librado and Rozas 2009). To demonstrate the departure from neutrality in all isolates, Tajima's D statistical method was used (Tajima 1989). For estimation of synonymous and non-synonymous substitution rates among CDS of all known NDV strains in genotype VI, VII and XIII against vaccine strains (LaSota, Mukteswar), synonymous and non-synonymous analysis program (SNAP v2.1.1) was used *via* an online webserver available at https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html. The ratio and proportion of observed synonymous and non-synonymous substitutions was calculated by a statistical tool integrated into the webserver.

Detection of putative recombination event

For identification of putative breakpoints, recombination analysis was performed on all complete genome sequences of 35 AAvV 1 isolates using four different software, including SimPlot (Ray 2003), GARD (http://www.datamonkey.org/GARD), DAMBE (Xia 2001) and RDP4 version 4.95 (Martin et al. 2015). Initially, all isolates were used at once in order to check the possibility of the occurrence of recombination events. Later on, in order to exclude influence or ambiguousness of sequences shared among strains isolated from the same bird species at a time and reliability of analysis, individual isolates representing each species were used for each possible parental type at a time. Owing to the enhanced accuracy, clarity and reliability of the analysis, outcomes identified by RDP4 were considered conclusive for further interpretation. The RDP4 was preferred because it employs a combination of seven different algorithms (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq) to better unleash putative recombinant and parent isolates at $p<0.001$. Putative recombination events were assumed to have occurred only when they were consistently identified by at least five of the above-mentioned seven algorithms at a probability threshold of 0.05.

Results

Genome characterization of all known *Avian avulavirus* 1

A total of 35 whole genome sequences of AAvV 1 from multiple avian species [Chicken (n=9), Pigeon (n=8), Parrot (n=5), Duck (n=5), Peacock (n=3), and Pheasant (n=2)] and human (n = 3) were acquired from a public database (Table 1) and analysed. As is typical for virulent strains, all isolates showed an identical cleavage motif ($_{112}$RRQKR↓F$_{117}$) in the *F* gene except for one isolate (KX236100), which contained a K to R residue substitution at position 115 ($_{112}$RRQRR↓F$_{117}$). All avian-originated isolate genomes were 15,192 nt in length except for one pigeon-originated isolate (KU885949), which was 15189 nt long and contained deletions at positions 2125, 7103 and 8330 positions. Although these isolates had the same genome length, however, several insertions and deletions at different positions were observed in five

different isolates, including KX236100 (pigeon, genotype VI), JX532092 (peacock, genotype VII), JX854452 (pheasant, genotype VII), JN682210 (chicken, genotype XIII) and JN682211 (chicken, genotype XIII). Important to note is that three of them originated from humans and had genome lengths ranging from 15146 to 15179 nt with several insertions and deletions (Table 1). The PASC analysis for the entire genome revealed 0.01%-15.33% divergence among isolates from genotype VI, VII and XIII. This estimation revealed that avulaviruses of genotype VI and VII had maximum percent divergence (13.99%-15.33%) followed by 14.58%-14.73% between genotype VI and XIII, and 11.5%-12.02% between genotypes VII and XIII (Table 2).

Phylogenetic analysis and evolutionary distance estimation

Based on the whole genome sequences, the phylogenetic analysis clustered all isolates into three distinct genotypes; genotype VI, VII and XIII. Of 35 AAvV 1 isolates, 26 isolates were clustered in genotype VII with those reported from Indonesia (HQ697254), 4 isolates were clustered in genotype VI with those reported from China, and 5 isolates clustered in genotype XIII with isolates from India (Fig. 1). Furthermore, based on partial *F* gene (375 bp) sequences, the phylogenetic analysis revealed that AAvV strains from genotype VII showed continuous evolution and clustered into five distinct sub-genotypes (VIIa, VIIb, VIIe, VIIf, VIIi) (Fig. 2A). The AAvVs from genotype VI were clustered into two distinct sub-genotypes (VIg and VIm) (Fig. 2B), whereas the avulaviruses from genotype XIII (n =5) clustered into two sub-genotypes (XIIIb and XIIId) (Fig. 2C). The evolutionary network pattern of under study-isolates revealed distinct genetic diversity and diverse evolutionary relationships among sub-genotypes VI, VII and XIII (Fig. 3).

In Pakistan, all known avulaviruses of genotype VI belonged to only two sub-genotypes (VIg, VIm) and demonstrated 7.7% genetic distance with an overall range of 8.4%-11.3% and 7.7%-11.9% with other sub-genotypes, respectively (Table 3). Avulaviruses from VIIa revealed a high genetic distance (40.1%) with viruses from VIIe followed by 35.4% between viruses from

VIIe and VIIi, 33.9% between viruses from VIIe and VIIf, 30.6% between viruses from VIIb and VIIe, 25.1% between viruses from VIIa and VIIb, 21.5% between viruses from VIIa and VIIf, 19% between viruses from VIIb and VIIf, 16.2% between viruses from VIIf and VIIi, 15.2% between viruses from VIIb and VIIi and 6.8% between viruses from VIIa and VIIi (Table 4). Similar to genotype VI, avulaviruses of genotype XIII also belong to two sub-genotypes (XIIIb, XIIId) and shared 3.9% genetic distance to each other (Table 5).

Comparative residue analysis

Comparative residue analysis of all proteins of avulaviruses from genotype VI, VII and XIII revealed conservation in all functional motifs at the protein level. Few substitutions among strains from three genotypes were also found with the maximum number of substitutions in the *L*-gene (n=48), followed by the *F* gene (n=32), *HN* gene (n=28), *P* gene (n=16), *M* gene (n=13) and *NP* gene (n=11) (Table 6). The inter-genotype comparative residue analysis revealed close relatedness between isolates from genotypes VII and XIII as compared to isolates from genotype VI. Using LaSota and Mukteswar as reference strains, partial *F* gene sequences of avulaviruses included in this study representing different sub-genotypes were compared to each other and numerous substitutions were found at different positions. The cleavage site for all the strains was found to be identical except for a pigeon-originated isolate (KY042139, sub-genotype VIg) which was identical to a mesogenic vaccine strain (EF201805) (Table 7).

Genetic diversity and genome-based analysis for vaccine efficacy

Based on individual coding genes, the DnaSP analysis revealed a high mutation rate in AAvVs isolates of genotype VII but high haplotype diversity was observed in CDS of isolates belonging to the VI genotype. Nucleotide diversity was found to be relatively higher in CDS of isolates belonging to genotype VI as compared to VII and XIII. For estimation based on the theory of neutrality in all CDS of isolates from individual genotypes, the Tajima's D value was found to be negative for all genes with $p > 0.10$ except *P* and *F* genes of genotype XIII (Table

8). An investigation of the synonymous and non-synonymous substitution rate against the vaccine strains revealed that CDS of isolates from all genotypes showed high synonymous and non-synonymous substitution rates when the LaSota strain was used a query compared to the Mukteswar strain (Table 9 and 10).

Selective pressure analysis

In the selective pressure analysis, none of coding genes from genotype VI had a mean dN/dS greater than 1 at $p < 0.05$ where the highest mean was observed for the *P* gene (0.37830) followed by the *HN* (0.24668), *L* (0.13024), *F* (0.10890), *M* (0.10332) and *NP* (0.05012) genes. Using three different statistical approaches (SLAC, FEL and IFEL), two positive selection sites were inferred by SLAC in the *P* gene, whereas one was found in the *F* gene. The FEL algorithm revealed three positive selection sites in the *F* gene, two in the *P* gene and one in the *NP* gene. Similarly, using the IFEL algorithm, three positive sites were found in the *P* gene, two in the *F* gene and one in the *L* gene. All coding genes of avulaviruses from genotype VII had a mean dN/dS less than 1 at $p < 0.05$ in the order of *F* (0.38894) > *HN* (0.29408) > *P* (0.28340) > *L* (0.13761) > *P* (0.098717) and > *M* (0.01944). Using three different statistical algorithms, SLAC revealed two positive selection sites in the *F* gene and only one in the *NP* gene. The FEL algorithm revealed seven positive selection sites in the *F* gene, five in the *L* gene and one each in the *HN* and *P* genes. In comparison, using the IFEL algorithm, numbers of positive selection sites as high as eleven in the *L* gene, nine in the *F* gene, three in the *HN* gene, two in the *M* gene, and one in the *P* gene were observed. Similar to genotype VI and VII, none of the coding genes from isolates belonging to XIII had a mean dN/dS greater than 1 at $p < 0.05$ in the order *P* (0.35729) > *M* (0.16906) > *HN* (0.16462) > *F* (0.14771) > *L* (0.10332) > *NP* (0.09096). Thus, only one positive selection site was found in the *F* gene using the SLAC algorithm, while eleven positive sites were observed in the *L* gene, five in the *F* gene, three in the *HN* and one in each the *NP* and *P* genes using the FEL algorithm. On the other hand, thirteen

positive sites in *L* gene, seven in *F* gene, three in *HN* gene, two in *NP* gene and one in each *P* and *M* gene were inferred using the IFEL algorithm (Table 11).

Recombination Analysis

To identify novel recombinant events among isolates, all existing whole genomes of AAvVs were subjected to a detailed recombination analysis. However, only two putative recombination events (between 1008-3474 and 8818-9592 nt) (Fig. 4A, B) were observed in pheasant-originated isolate (JX854452) with chicken-originated isolates. Noteworthy, both putative recombination events were observed in the same isolate sharing genomic sequence with same potential major parent (KP776462) from genotype VII and potential minor parent (JN682210) from genotype XIII. All seven recombination algorithm methods revealed the occurrence of recombination events in this isolate at $p<0.001$ value. The detailed information of inferred breakpoint and *p*-value of algorithm approaches are provided in Table 12.

Discussion

Given the importance of the *F* gene in determining virulence of avulaviruses and conferring the major viral antigen, as well as the fact that current classification systems rely upon phylogenetic analysis of the same gene (Aldous et al. 2003; de Leeuw et al. 2003, 2005; Tsai et al. 2004; Kim et al. 2007), a large number of partial *F* gene sequences from multiple avian hosts were available in GenBank and other public domains. Since *F*-gene-based analysis is insufficient to predict the true evolution of viruses, we performed the first-ever analysis highlighting the genomic and evolutionary aspects of whole genomes of countrywide circulating *Avian avulavirus* 1 strains. We performed this analysis because a specific gene might not evolve at the same rate as does the whole genome (Miller et al. 2009), and, therefore, our analysis can provide more comprehensive information to unleash virus evolution to serve as the basis for designing epidemiological investigations in the future. Additionally, the emergence of new strains in global epizootics and continuous evolution by year-to-year

changes in genomic sequences can be investigated by complete genome analysis of avulaviruses isolated from different susceptible hosts. Because the vast genetic diversity may be favoured by the large variety of susceptible avian species and by the availability of highly mobile, wild/migratory waterfowl as natural reservoirs, continuous monitoring of viral evolution and molecular epidemiology of countrywide circulating avulaviruses is imperative for disease endemic countries such as Pakistan.

Across Pakistan, *Avian avulavirus* 1 strains of genotype VI (Akhtar et al. 2016; Shabbir et al. 2016), VII (Shabbir et al. 2012a, b, 2013, 2016; Munir et al., 2012a, b; Abbas et al. 2014; Farooq et al. 2014; Akhtar et al. 2016; Wajid et al. 2017; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019) and XIII (Miller et al. 2015; Shabbir et al. 2018) have been reported previously from a wide range of hosts. Among these, avulaviruses of genotype XIII have not only been found responsible for outbreaks in poultry (Khan et al. 2010), but also have been implicated in mild respiratory infections in poultry workers (Shabbir et al. 2018). Apart from clinical infection, highly virulent avulaviruses of genotype VII were isolated from clinically healthy backyard poultry birds (Munir et al. 2012b), and asymptomatic wild- and water- fowl (Akhter et al. 2016, 2017; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019).

Previous studies have suggested that the replacement of genotype XIII by genotype VII occurred between 2010 and 2013 (Khan et al. 2010; Miller et al. 2015). However, this claim is not convincing as new evidence suggests that the re-emergence of genotype XIII viruses isolated from humans is linked with poultry production settings (Shabbir et al. 2018). Such findings highlight the existence of genotype XIII viruses in the field and may suggest the need for revision of diagnostic assays that fail to detect these circulating viruses. The phylogenetic and evolutionary comparison of whole genome and partial *F* gene sequences of human-originated genotype XIII to genotype XIIId, which is distinct from the previously reported genotype XIIIb from poultry, suggests continuous evolution of avulaviruses. For genotype VII,

phylogenetic analyses of whole genome and partial *F* gene sequences revealed that the current prevailing strains (since 2015) belong to sub-genotype VIIi and are evolutionarily similar to strains circulating in Israel, China and Indonesia (Shabbir et al. 2013a, b, 2016; Akhtar et al. 2016, 2017; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019). The clustering of Pakistani avulaviruses originating from birds with isolates originating from Israel, China and Indonesia, highlight a continuous evolution and epizootic nature across the boundaries. It is important to note that equally virulent and genetically identical avulaviruses have also been isolated from multiple symptomatic and asymptomatic wild and captive avian species from Pakistan (Shabbir et al. 2012; Munir et al. 2012a, Qamar-un-Nisa et al. 2017, Akhtar et al. 2017), Israel and a public zoo in Mexico (Panshin et al. 2002; Miller et al. 2009). Therefore, the role of wild birds as natural reservoirs in dissemination of the virus cannot be overlooked. Recently, a spillover of AAvV 1 from wild to poultry and vice versa was investigated (Cardenas et al. 2013; Wajid et al. 2018). Most studies have reported outbreaks that occurred over short periods of time and were considered to be caused by accidental spillover of viruses (Vijayarani et al. 2010; Kumar et al. 2013; Dimitrov et al. 2016). Such an evidence of spill-over facilitates further evolution resulting in subsequent emergence of novel variant or escape mutant. These studies point out the significant role of different avian species, kept in captivity in poultry production settings in Pakistan, in virus dissemination. Although the current genetic diversity and evolutionary analyses reveal strong relationships among viruses originating from different avian species including poultry, limited information is available concerning the potential role of these avian species in the dissemination of virus.

The genetic diversity of avulaviruses, evidenced by synonymous and non-synonymous substitutions in residues of coding genes, has a crucial role in their evolution and subsequent adaptation to a wide range of hosts (Aziz-ul-Rahman et al. 2018b). Besides, previous studies showed evidence of evolution in AAvV 1 under vaccine-induced immune pressure (Chong et

al. 2010; Orabi et al. 2017). Considering the influence of vaccination on evolution, the genome sequence of vaccine strains commonly used in the field (LaSota; AF077761 and Mukteswar; EF201805) in Pakistan were selected for investigation of synonymous and non-synonymous substitutions in the current study. Comparative residue analysis of each gene of avulaviruses included in the study of genotype (VI, VII and XIII) with vaccine strains (genotype II and III) showed a number of synonymous and non-synonymous substitutions at varying substitution rates. However, a higher rate of non-synonymous substitution was more evident in avulaviruses of genotype VII and therefore, subsequent protection-challenge experiments are needed. This is important because the emergence of escape mutants over a period of time pose a major threat to the control of avulaviruses using classical vaccine due to decreased protection from challenge (Perozo et al. 2012; Ali et al. 2014; Farooq et al. 2014; Abbas et al. 2014; Umar et al. 2015). In fact, a high rate of non-synonymous substitutions in coding genes may alter the structural and biological function of viral proteins and the virus' subsequent virulence (Reitter et al. 1995). In this regard, a massive or irrational vaccine effort in the field may play a crucial role in the emergence of escape mutants such as those reported previously from vaccinates (Rehmani et al. 2015). The presence of virulent viruses in vaccinates in commercial farms (Rehmani et al. 2015) and their constant evolution over time (Miller et al. 2009) suggests the existence of a high environmental viral load with continuous replication in endemic countries. In addition, vaccine failure is correlated with substantial changes at the nucleotide level and genotype mismatching, which cause differences at essential immunodominant epitopes (Liu et al. 2018). Recently, a few studies on *F* and *HN* genes in avulaviruses of Pakistan revealed substitutions at both nucleotide and amino acid level at biologically and functionally important motifs (Shabbir et al. 2013a, b, 2016; Abbas et al. 2014; Akhtar et al. 2016, 2017; Habib et al. 2018; Aziz-ul-Rahman et al. 2018a, 2019).

Besides residue substitutions, natural pressure selection sites in the genome are influenced by the environment and play a key role in virus evolution. High positive selection sites in CDS may affect the structure and function of the corresponding protein. The statistical calculation of non-synonymous and synonymous (dN/dS) mutations is significant in understanding the molecular evolution of CDS across closely related yet divergent strains. Datamonkey, in this regard, is a well-known web-server for rapid detection of positive (dN>dS) and negative (dN<dS) selection sites in aligned CDS (Pond and Frost, 2005). Evidence of positive selection sites in all coding genes of inter- (genotype VI, VII and XIII) and intra-avulaviruses (AAvV 1-20) identified by SLAC, FEL and IFEL statistical approaches highlighted the underlying mechanism of virus evolution (Aziz-ul-Rahman et al. 2018b). Therefore, the emergence of strains representing novel sub-genotypes is possible in the near future. In our analysis, the mean dN-dS were found to be non-significant with a minimum number of positive selection sites. While this seldom happens in structural domains of the genome, the impact of the combination of such positive selection sites with lower level sequence diversity may cause the emergence of variants (Yang et al. 2000). According to the neutral theory of molecular evolution, types of molecular variation that arise via spontaneous mutations have no influence on the virus' fitness (Fay and Wu 2003). Indeed, such substitutions might represent potential positive selection sites that could shape the virulence and evolution of new strains. However, the biological significance of these sites is still unknown, and it would be an area of interest to investigate the role of substitutions in pathogenicity using a reverse genetics approach.

The occurrence of recombination between strains may also potentially influence the evolution of avulaviruses (Qin et al. 2008; Satharasinghe et al. 2016) and, therefore may facilitate the emergence of novel/new mutant strains (Yin et al. 2011). In the present study, recombination analysis identified two potential recombination events in an avulavirus originating from pheasants with potential major (genotype VII) and minor parent strain (genotype XIII) of

chicken origin. This is not surprising because recombination within avulaviruses of the same genotype (VII) (Han et al. 2008) and between two different genotypes (II and VII) (Qin et al., 2008) has previously been identified. The possibility of recombination between strains of the same and different genotypes highlights the potential impact of recombination on the emergence and evolution of novel avulaviruses. While the occurrence of natural recombination may facilitate the emergence of a novel/new mutant virus (Han et al. 2008; Qin et al., 2008), there is controversy in the interpretation of natural recombination. Previous studies suggested that recombination may play a vital role in AAvV 1 evolution (Zhang et al. 2010; Yin et al. 2011), while others concluded otherwise, referring to it as a sequencing artefact due to proof-reading errors of the polymerase enzyme (Afonso 2008). That being said, recombination-based evolution in AAvVs is an on-going discussion and further studies are needed to understand its significance.

Based on all essential analyses, the current study concluded the concurrent evolution and circulation of avulaviruses of different sub-genotypes (VIg, VIm, VIIa, VIIb, VIIe, VIIf, VIIi, XIIIb and XIIId) in Pakistan, highlighting its evolutionary potential over a period of time. Evidences of positive selection sites and putative recombination events suggest increased genetic diversity among avulaviruses originating from multiple avian hosts. Taken together, this evidence may not only suggest the potential for the emergence of novel variants, but also the possibility of failure in diagnostics and vaccines. The evolutionary knowledge of genetic diversity of field prevailing avulaviruses impacts many aspects, from the broadest investigations of virus taxonomy, to the finest details of molecular epidemiology and vaccine design. Therefore, continuous monitoring and surveillance of viral evolution with necessary periodic updates should be ascertained for disease control interventions in disease-endemic countries.

Acknowledgment

We thank Ms. Juliette Rando (Medicine Department at University of Pennsylvania, USA) for critical reading and necessary edits in the manuscript text.

Compliance with ethical standards

Conflict of interest: The authors declare no conflict of interests.

Ethical approval: This article does not contain studies with animals or humans performed by any of the authors.

Figure legends

Fig. 1 Phylogenetic analysis of so-far reported complete genome sequences of *Avian avulavirus* 1 (highlighted with blue colour) from Pakistan. The neighbour-joining method with 1000 bootstraps was used for analysis of evolutionary relationship between and within isolates from different genotypes and representative isolates from worldwide using MEGA 6 software.

Fig. 2 Phylogenetic analysis based on partial *F* gene sequences of so-far reported *Avian avulavirus* 1 (highlighted with blue colour) from Pakistan with representative isolates from worldwide using MEGA 6 software. The phylogeny analysis was conducted for classification of sub-genotypes within genotype VI (**A**), VII (**B**) and XIII (**C**).

Fig. 3 Partial *F* gene sequence based evolutionary network of so-far reported *Avian avulavirus* 1 from Pakistan representing sub-genotypes-wise distribution of isolates.

Fig. 4 The plot display graphically illustrating the evidences underlying the detection of individual user-selected recombination event. The display depicting the coloured coded line representation of recombinant, major parent and minor parent isolates for the recombinant event 1 (**A**) and event 2 (**B**).

References

Abbas M, Shabbir MZ, Khaliq SA et al (2014) Sequence analysis of the hemagglutinin-neuraminidase gene of Newcastle disease virus from Punjab Pakistan. Adv Ani Vet Sci 2:242-247.

Afonso CL (2008) Not so fast on recombination analysis of Newcastle disease virus. J Virol 82:9303-9303.

Akhtar S, Muneer MA, Muhammad K et al (2016) Genetic characterization and phylogeny of pigeon paramyxovirus isolate (PPMV-1) from Pakistan. SpringerPlus 5(1):1295.

Akhtar S, Muneer MA, Muhammad K et al (2017) MoleculaCharacterization 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, Mynn J, Banks J, Alexander D (2003) A molecular epidemiological study of avian paramyxovirus type 1 (Newcastle disease virus) isolates by phylogenetic analysis of a partial nucleotide sequence of the fusion protein gene. Avian Pathol 32:237-255.

Alexander DJ (1998) Newcastle disease virus and other avian paramyxoviruses, p. 156–163. In D. E. Swayne et al. (ed.), A laboratory manual for the isolation and identification of avian pathogens, 4th ed. American Association of Avian Pathologists, Kennett Square, PA.

Alexander DJ (2003) Newcastle disease virus, other avian paramyxoviruses, and pneumovirus infections. In: Saif YM, Barnes HJ, Glisson JR, Fadly AM, McDougald LR, Swayne DE, (eds) Disease of poultry, 11edn. Iowa State University Press, Ames, pp 63–87.

Ali M, Muneer B, Hussain Z et al (2014) Evaluation of efficacy of killed and commercially available live Newcastle disease vaccine in broiler chickens in Pakistan. J Ani Plant Sci 24:1663-1667.

Aziz-ul-Rahman, Habib M, Riaz T et al (2017) Seroprevalence of newcastle disease virus (NDV) in commercial and domesticated birds: Pakistan during current surge of NDV. J Inf Mol Biol 4(4):54-59.

Aziz-ul-Rahman, Munir M, Shabbir MZ (2018b) Comparative evolutionary and phylogenomic analysis of *Avian avulaviruses* 1 to 20. Mol Phyl Evol 127:931-51

Aziz-ul-Rahman, Yaqub T, Imran M, et al (2018a) Phylogenomics and Infectious Potential of *Avian Avulaviruses* Species-Type 1 Isolated from Healthy Green-Winged Teal (Anas carolinensis) from a Wetland Sanctuary of Indus River. Avian Dis 62(4):404-15.

Aziz-ul-Rahman, Yaqub T, Imran M, et al (2019) Sequence analysis and biological characterization of virulent *Avian avulavirus* 1 isolated from asymptomatic migratory fowls. Acta Virologica 63:181-186. https://doi.org/10.4149/av_2019_208

Barman LR, Nooruzzaman M, Sarker RD et al (2017) Phylogenetic analysis of Newcastle disease viruses from Bangladesh suggests continuing evolution of genotype XIII. Arch Virol 162(10):3177-82.

Cardenas GS, Navarro LR, Morales R et al (2013) Molecular epidemiology of Newcastle disease in Mexico and the potential spillover of viruses from poultry into wild bird species. Appl Environ Microbiol 79:4985–4992.

Chong YL, Padhi A, Hudson PJ et al (2010) The effect of vaccination on the evolution and population dynamics of avian paramyxovirus-1. PLoS Pathogens 22;6(4):e1000872.

Das M, Kumar S (2017) Evidence of independent evolution of genotype XIII Newcastle disease viruses in India. Arch Virol 162(4):997-1007.

de Leeuw OS, Hartog L, Koch G, Peeters BPH (2003) Effect of fusion protein cleavage site mutations on virulence of Newcastle disease virus: non-virulent cleavage site mutants revert to virulence after one passage in chicken brain. J Gen Virol 84 (2):475–484.

de Leeuw OS, Koch G, Hartog L et al (2005) Virulence of Newcastle disease virus is determined by the cleavage site of the fusion protein and by both the stem region and globular head of the haemagglutinin–neuraminidase protein. J Gen Virol 86:1759-1769.

Delport W, Poon AF, Frost SD, Kosakovsky PSL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26(19):2455–2457. doi:10.1093/bioinformatics/btq429

Dimitrov KM, Ramey AM, Qiu X et al (2016) Temporal, geographic, and host distribution of avian paramyxovirus 1 (Newcastle disease virus). Infect Genet Evol 39:22-34.

Farooq M, Saliha U, Munir M, Khan QM (2014) Biological and genotypic characterisation of the Newcastle disease virus isolated from disease outbreaks in commercial poultry farms in northern Punjab, Pakistan. Virol Reports 3:30-39.

Fay JC, Wu CI (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

Habib M, Yaqub T, Nazir J et al (2018) Genomic and biological characterization of Newcastle disease viruses isolated from migratory mallards (*Anas platyrhynchos*). Arch Virol 30:1-10.

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

Han GZ, He CQ, Ding NZ, Ma LY (2008) Identification of a natural multi-recombinant of Newcastle disease virus. Virol 371:54-60.

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

Khan TA, Rue CA, Rehmani SF et al (2010) Phylogenetic and biological characterisation of Newcastle disease virus isolates from Pakistan. J Clin Microbiol 48:1892-1894.

Kim LM, King DJ, Curry PE et al (2007) Phylogenetic diversity among low-virulence Newcastle disease viruses from waterfowl and shorebirds and comparison of genotype distributions to those of poultry-origin isolates. J Virol 81:12641-12653.

Kolakofsky D, Roux L, Garcin D, Ruigrok RW (2005) Paramyxovirus mRNA editing, the ‘rule of six’and error catastrophe: a hypothesis. J Gen Virol 86:1869-1877.

Kuhn JH, Wolf YI, Krupovic M et al (2019) Classify viruses—the gain is worth the pain. Nature 566:318-320. doi:10.1038/d41586-019-00599-8

Kumar A, Maan S, Mahajan NK et al (2013) Detection and molecular characterization of Newcastle disease virus in peafowl (*Pavo cristatus*) in Haryana State, India. Indian J Virol 24(3):380-385.

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 RS, Gil P, Majó N, Nofrarías M, Briand FX, Jestin V, Albina E. Can genotype mismatch really affect the level of protection conferred by Newcastle disease vaccines against heterologous virulent strains?. Vaccine. 2018 Jun 22;36(27):3917-25.

Martin DP, Murrell B, Golden M et al (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1(1).

Miller PJ, Haddas R, Simanov L et al (2015) Identification of new sub-genotypes of virulent Newcastle disease virus with potential panzootic features. Infec Genet Evol 29:216-229.

Miller PJ, Kim LM, Ip HS, Afonso CL (2009) Evolutionary dynamics of Newcastle disease virus. Virology 391(1):64-72.

Munir M, Cortey M, Abbas M et al (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 MZ, Yaqub T et al (2012a) Complete genome sequence of a velogenic neurotropic avian paramyxovirus 1 isolated from Peacocks (*Pavo cristatus*) in a wildlife park in Pakistan. J Virol 86(23):13113-13114.

Nath B, Barman NN, Kumar S (2015) Molecular characterization of Newcastle disease virus strains isolated from different outbreaks in Northeast India during 2014–15. Microb Pathog 91:85–91.

Numan M, Zahoor MA, Khan HA, Siddique M (2005) Serologic status of Newcastle disease in broilers and layers in Faisalabad and surrounding districts. Pak Vet J 25(2):55.

Orabi A, Hussein A, Saleh AA et al (2017) Evolutionary insights into the fusion protein of Newcastle disease virus isolated from vaccinated chickens in 2016 in Egypt. Arch Virol 162(10):3069-79.

Panshin A, Shihmanter E, Weisman Y et al (2002) Antigenic heterogeneity among the field isolates of Newcastle disease virus (NDV) in relation to the vaccine strain: 1. Studies on viruses isolated from wild birds in Israel. Comp Immunol Microbiol Infec Dis 25(2):95-108.

Perozo F, Marcano R, Afonso CL (2012) Biological and phylogenetic characterisation of a genotype VII Newcastle disease virus from Venezuela: efficacy of field vaccination. J Clin Microbiol 50:1204-1208.

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

Qamar-un-Nisa, Younus M, Muti-ur-Rehman et al (2017) Pathological Alterations during Co-Infection of Newcastle Disease Virus with Escherichia coli in Broiler Chicken. Pak J Zool 49:1953-1961.

Qin Z, Sun L, Ma B et al (2008) F gene recombination between genotype II and VII Newcastle disease virus. Virus Res 131(2):299-303.

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

Rehmani SF, Wajid A, Bibi T et al (2015) Presence of virulent Newcastle disease virus in vaccinated chickens in farms in Pakistan. J Clin Microbiol 53:1715–8. https://doi.org/10.1128/JCM.02818-14

Reitter JN, Sergel T, Morrison TG (1995) Mutational analysis of the leucine zipper motif in the Newcastle disease virus fusion protein. J Virol 69:5995–6004.

Satharasinghe DA, Murulitharan K, Tan SW et al (2016) Detection of inter-lineage natural recombination in avian paramyxovirus serotype 1 using simplified deep sequencing platform. Fron Microbiol 7:1907.

Shabbir MZ, Abbas M, Yaqub T et al (2012a) Genetic analysis of Newcastle disease virus from Punjab, Pakistan. Virus Genes 46:309–315.

Shabbir MZ, Akhtar S, Tang Y et al (2016) Infectivity of wild bird origin Avian Paramyxovirus serotype 1 and vaccine effectiveness in chickens. J Gen Virol 97(12):3161-3173. DOI: 10.1099/jgv.0.000618

Shabbir MZ, Goraya MU, Abbas M et al (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 MZ, Nissly RH, Ahad A et al (2018) Complete Genome Sequences of Three Related Avian Avulavirus 1 Isolates from Poultry Farmers in Pakistan. Genome announc 6(18):e00361-18.

Shabbir MZ, Zohari S, Yaqub T et al (2013) Genetic diversity of Newcastle disease virus in Pakistan: A countrywide perspective. Viro J 10:170–179.

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

Tamura K, Kumar S (2002) Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol 19:1727–1736.

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729.

Tsai H-J, Chang K-H, Tseng C-H et al (2004) Antigenic and genotypical characterization of Newcastle disease viruses isolated in Taiwan between 1969 and 1996. Vet Microbiol 104(1–2):19–30.

Umar S, Younus M, Arshad HU et al (2015) Recent advances in vaccination strategies and future perspectives for the control of Newcastle disease in Pakistan: A review. Intern J Vet Sci 4:136-141.

Vijayarani K, Muthusamy S, Tirumurugaan KG et al (2010) Pathotyping of a Newcastle disease virus isolated from peacock (*Pavo cristatus).* Trop Anim. Health Prod 42:415–419.

Wajid A, Dimitrov KM, Wasim M et al (2017) Repeated isolation of virulent Newcastle disease viruses in poultry and captive non-poultry avian species in Pakistan from 2011 to 2016. Pre Vet Med 142:1-6.

Wajid A, Dundon WG, Hussain T, Babar ME (2018) Pathotyping and genetic characterization of avian avulavirus-1 from domestic and wild waterfowl, geese and black swans in Pakistan, 2014 to 2017. Arch Virol 2:1-6.

Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92(4):371-3.

Xue C, Cong Y, Yin R et al (2017) Genetic diversity of the genotype VII Newcastle disease virus: identification of a novel VIIj sub-genotype. Virus Genes 53(1):63-70.

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

Yin Y, Cortey M, Zhang Y et al (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.

Zhang R, Wang X, Su J et al (2010) Isolation and analysis of two naturally-occurring multi-recombination Newcastle disease viruses in China. Virus Res 151:45-53.

Table 1: Detailed information of whole genomes sequences of *Avian avulavirus* 1 isolated from different birds in Pakistan

Sr. No.	Accession numbers	Year of isolation	Host	Location	Genome length (bp)	Insertion and deletion of nucleotides in genome	Cleavage site pattern	Genotype	Sub-genotype
1	KX268690	2016	Parakeet	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
2	KX268691	2016	Parakeet	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
3	KX791183	2016	Parakeet	Rawalpindi	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
4	KX791184	2016	Backyard poultry	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
5	KX791185	2016	Backyard poultry	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
6	KX791186	2016	Backyard poultry	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
7	KX791187	2016	Backyard poultry	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
8	KX791188	2016	Backyard poultry	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
9	KY697611	2016	Mallard	Mianwali	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
10	KY697612	2016	Mallard	Mianwali	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
11	MH019281	2015	Human	Lahore	15167	Deletion at (1-14, 8329, 15182-15192) position	$_{112}$RRQKR↓F$_{117}$	XIII	XIIId
12	MH019282	2015	Human	Lahore	15179	Insertion at (1724-1735) deletion at (1-9, 1724-1735, 8329, 15182-15192) position	$_{112}$RRQKR↓F$_{117}$	XIII	XIIId
13	MH019283	2015	Human	Lahore	15146	Deletion at (1-14, 1697-1716, 8329, 15182-15192) position	$_{112}$RRQKR↓F$_{117}$	XIII	XIIId
14	KX268688	2015	Parakeet	Rawalpindi	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
15	KX268689	2015	Parrot	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
16	KY290560	2015	Peacock	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
17	KY290561	2015	Pheasant	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
18	KU845252	2015	Duck	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
19	MF437286	2015	Green-winged teal	Mianwali	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
20	MF437287	2015	Green-winged teal	Mianwali	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
21	KX236100	2015	Pigeon	Lahore	15192	Insertion at 1631 position and deletion at 2015 position	$_{112}$RRQRR↓F$_{117}$	VI	VIm
22	KX496962	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
23	KX496963	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
24	KX496964	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
25	KX496965	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
26	KY042135	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VI	VIm
27	KX236101	2015	Pigeon	Lahore	15192	NO	$_{112}$RRQKR↓F$_{117}$	VI	VIm
28	KU885949	2014	Pigeon	NA	15189	Deletion at 2125, 7103 and 8330 position	$_{112}$RRQKR↓F$_{117}$	VI	VIm
29	KP776462	2014	Commercial Poultry	NA	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
30	KU885948	2014	Peacock	NA	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
31	KM670337	2013	Vaccinated Chicken	NA	15192	NO	$_{112}$RRQKR↓F$_{117}$	VII	VIIi
32	JX532092	2012	Peacock	Lahore	15192	Insertion at 2130 and 7057 position, deletion at 2125 and 7103 position	$_{112}$RRQKR↓F$_{117}$	VII	VIIa
33	JX854452	2011	Pheasant	NA	15192	Insertion at 7057 and 8329 position, deletion at 2125 and 7103 position	$_{112}$RRQKR↓F$_{117}$	VII	VIIa
34	JN682210	2010	Commercial Poultry	Rawalpindi	15192	Insertion at (1686-1690, 8336) and deletion at (1649-1653, 8330) position	$_{112}$RRQKR↓F$_{117}$	XIII	XIIIb
35	JN682211	2010	Commercial Poultry	Rawalpindi	15192	Insertion at 8336 position and deletion at 8330 position	$_{112}$RRQKR↓F$_{117}$	XIII	XIIIb

Table 2: Estimation of percent divergence for complete genomes sequences of so-far reported *Avian avulavirus* 1 from Pakistan

Accession numbers	KX268690	KX268691	KX791183	KX791184	KX791185	KX791186	KX791187	KX791188	KY967611	KY967612	MH019281	MH019282	MH019283	KX268688	KX268689	KY290560	KY290561	KU845252	MF437286	MF437287	KX236100	KX496962	KX496963	KX496964	KX496965	KY042135	KX236101	KU885949	KP776462	KU885948	KM670337	JX532092	JX854452	JN682210
KX268690																																		
KX268691	0.01																																	
KX791183	0.48	0.48																																
KX791184	1.32	1.32	1.17																															
KX791185	1.35	1.35	1.20	0.07																														
KX791186	1.33	1.33	1.18	0.05	0.06																													
KX791187	1.32	1.32	1.17	0.03	0.05	0.03																												
KX791188	1.33	1.33	1.18	0.05	0.06	0.00	0.03																											
KY967611	1.76	1.76	1.65	1.08	1.08	1.08	1.08	1.08																										
KY967612	1.77	1.77	1.65	1.08	1.08	1.09	1.08	1.09	0.03																									
MH019281	11.91	11.8	11.79	11.92	11.95	11.93	11.92	11.93	11.99	12.00																								
MH019282	11.91	11.9	11.8	11.93	11.96	11.94	11.93	11.94	12.00	12.01	0.01																							
MH019283	11.91	11.8	11.79	11.92	11.95	11.93	11.92	11.93	11.99	12.00	0.00	0.01																						
KX268688	0.51	0.51	0.03	1.15	1.18	1.16	1.15	1.16	1.63	1.63	11.79	11.80	11.79																					
KX268689	0.10	0.10	0.43	1.28	1.30	1.28	1.28	1.28	1.71	1.72	11.87	11.88	11.87	0.44																				
KY290560	1.17	1.17	1.04	1.1	1.12	1.10	1.1	1.10	1.45	1.46	11.76	11.76	11.76	1.02	1.12																			
KY290561	1.14	1.14	1.02	1.06	1.09	1.07	1.06	1.07	1.42	1.42	11.73	11.73	11.73	1.00	1.10	0.14																		
KU845252	0.66	0.66	0.32	1.24	1.26	1.24	1.24	1.24	1.58	1.58	11.82	11.83	11.82	0.30	0.60	0.00.92	0.88																	
MF437286	1.75	1.75	1.64	1.07	1.10	1.08	1.07	1.08	0.13	0.11	12.01	12.02	12.01	1.62	1.71	1.44	1.41	1.57																
MF437287	1.74	1.74	1.63	1.06	1.06	1.06	1.06	1.06	0.11	0.09	12.01	12.02	12.01	1.60	1.69	1.43	1.4	1.56	.07															
KX236100	14.32	14.32	14.25	14.41	14.45	14.41	14.39	14.41	14.42	14.43	14.65	14.66	14.65	14.25	14.30	14.21	14.21	14.25	14.45	14.45														
KX496962	0.76	0.76	0.64	0.9	0.93	0.91	0.90	0.91	1.16	1.17	11.54	11.55	11.54	0.62	0.72	0.64	0.62	0.64	1.16	1.14	14.00													
KX496963	01.69	1.69	1.56	1.62	1.65	1.63	1.62	1.63	1.95	1.95	12.01	12.02	12.01	1.54	1.65	1.02	0.99	1.53	1.93	1.92	14.45	1.12												
KX496964	0.76	0.76	0.38	1.37	1.40	1.38	1.37	1.38	1.71	1.72	11.79	11.79	11.79	0.36	0.70	1.24	1.22	0.37	1.71	1.69	14.27	0.64	1.7											
KX496965	0.76	0.76	0.38	1.37	1.40	1.38	1.37	1.38	1.71	1.72	11.79	11.79	11.79	0.36	0.70	1.24	1.22	0.37	1.71	1.69	14.27	0.64	1.7	0.00										
KY042135	14.29	14.28	14.25	14.44	14.48	14.44	14.42	14.44	14.38	14.4	14.72	14.73	14.72	14.25	14.27	14.24	14.24	14.25	14.42	14.41	1.81	14.00	14.41	14.31	14.31									
KX236101	14.28	14.26	14.19	14.31	14.35	14.31	14.29	14.31	14.40	14.41	14.47	14.48	14.47	14.18	14.24	14.18	14.15	14.22	14.41	14.40	3.83	13.94	14.38	14.19	14.19	3.89								
KU885949	14.29	14.29	14.24	14.45	14.48	14.45	14.43	14.45	14.38	14.41	14.58	14.58	14.58	14.23	14.27	14.20	14.20	14.22	14.43	14.42	1.34	13.99	14.44	14.25	14.25	1.85	3.80							
KP776462	1.08	1.08	0.95	1.02	1.04	1.02	1.02	1.02	1.36	1.36	11.67	11.68	11.67	0.93	1.03	0.14	0.11	0.82	1.35	1.34	14.16	0.55	0.98	1.15	1.15	14.18	14.11	14.16						
KU885948	1.43	1.43	1.28	1.36	1.39	1.37	1.36	1.37	1.72	1.73	11.91	11.91	11.91	1.26	1.38	0.74	0.72	1.27	1.71	1.69	14.25	0.89	0.9	1.46	1.46	14.23	14.19	14.25	0.72					
KM670337	1.12	1.12	0.98	1.08	1.11	1.09	1.08	1.09	1.41	1.42	11.70	11.7	11.70	0.96	1.07	0.34	0.31	0.89	1.40	1.39	14.11	0.58	0.82	1.16	1.16	14.16	14.10	14.10	0.3	0.56				
JX532092	5.52	5.52	5.40	5.49	5.51	5.49	5.49	5.49	5.58	5.59	11.60	11.61	11.60	5.38	5.47	5.18	5.17	5.37	5.60	5.60	15.4	4.89	5.62	5.41	5.41	15.3	15.33	15.32	5.1	5.44	5.19			
JX854452	5.33	5.33	5.21	5.26	5.29	5.27	5.26	5.27	5.37	5.38	11.50	11.51	11.50	5.19	5.28	5.01	5.00	05.19	5.37	5.37	15.27	4.69	5.43	5.22	5.22	15.17	15.19	15.19	4.94	5.26	5.01	0.35		
JN682210	12.76	12.74	12.68	12.78	12.81	12.79	12.78	12.79	12.91	12.92	4.57	4.58	4.57	12.68	12.70	12.61	12.59	12.7	12.92	12.92	15.15	12.41	12.82	12.7	12.7	15.19	15.01	15.18	12.53	12.73	12.59	9.36	9.28	
JN682211	12.93	12.91	12.84	12.86	12.89	12.87	12.86	12.87	13.0	13.01	4.43	4.44	4.43	12.84	12.87	12.76	12.74	12.84	13.01	13.02	15.27	12.52	12.98	12.85	12.85	15.26	15.04	15.23	12.67	12.86	12.69	9.85	9.77	2.81

Table 3: Estimation of mean inter-populational evolutionary diversity per site among sub-genotypes of genotype VI *Avian avulavirus* 1 using partial sequences of *F* gene. The genetic distance of viruses from Pakistan origin is bold

Sub-genotype	VIa	VIb	VIc	VId	VIe	VIf	VIg	VIh	VIi	VIj	VIk	VIl	VIm	VIn
VIa (n = 2)														
VIb (n = 3)	0.099													
VIc (n = 3)	0.101	0.096												
VId (n = 2)	0.114	0.118	0.104											
VIe (n = 5)	0.080	0.088	0.093	0.110										
VIf (n = 4)	0.088	0.094	0.104	0.124	0.067									
VIg (n = 5)	0.103	0.098	0.084	0.091	0.101	0.106								
VIh (n = 3)	0.092	0.105	0.117	0.139	0.092	0.101	0.113							
VIi (n = 3)	0.129	0.127	0.116	0.024	0.122	0.136	0.099	0.148						
VIj (n = 2)	0.121	0.128	0.106	0.119	0.106	0.117	0.096	0.134	0.122					
VIk (n = 3)	0.012	0.096	0.097	0.111	0.075	0.082	0.100	0.084	0.126	0.120				
VIl (n = 2)	0.047	0.094	0.093	0.114	0.073	0.084	0.099	0.080	0.126	0.114	0.040			
VIm (n = 6)	0.114	0.119	0.106	0.106	0.116	0.110	**0.077**	0.119	0.117	0.089	0.113	0.112		
VIn (n = 5)	0.057	0.093	0.097	0.116	0.0772	0.084	0.089	0.078	0.124	0.117	0.053	0.051	0.116	

Table 4: Estimation of mean inter-populational evolutionary diversity per site among sub-genotypes of genotype VII *Avian avulavirus* 1 using partial sequences of *F* gene. The genetic distance of viruses from Pakistan origin is bold

Sub-genotype	VIIa	VIIb	VIIc	VIId	VIIe	VIIf	VIIg	VIIh	VIIi	VIIj	VIIk
VIIa (n = 2)											
VIIb (n = 7)	**0.251**										
VIIc (n = 3)	0.300	0.187									
VIId (n = 3)	0.176	0.147	0.231								
VIIe (n = 6)	**0.401**	**0.306**	0.312	0.296							
VIIf (n = 8)	**0.215**	**0.019**	0.231	0.160	**0.339**						
VIIg (n = 4)	0.407	0.306	0.317	0.301	0.002	0.345					
VIIh (n = 4)	0.187	0.172	0.255	0.127	0.367	0.175	0.373				
VIIi (n = 8)	**0.068**	**0.152**	0.244	0.114	**0.354**	**0.162**	0.360	0.136			
VIIj (n = 4)	0.199	0.213	0.278	0.074	0.343	0.209	0.349	0.169	0.148		
VIIk (n = 3)	0.186	0.212	0.263	0.133	0.342	0.192	0.348	0.149	0.135	0.137	

Table 5: Estimation of mean inter-populational evolutionary diversity per site among sub-genotypes of genotype XIII *Avian avulavirus* 1 using partial sequences of *F* gene. The genetic distance of viruses from Pakistan origin is bold

Sub-genotype	XIIIa	XIIIb	XIIIc	XIIId
XIIIa (n = 4)				
XIIIb (n = 4)	0.057			
XIIIc (n = 3)	0.068	0.079		
XIIId (n= 3)	0.045	**0.039**	0.098	

Table 6: Comparative residue analysis of all coding genes of so-far reported *Avian avulavirus* 1 from Pakistan

Proteins	Conserved and Functional Motifs	Genotype	Comparative residue substitutions
NP	N-N self-assembly motif (^{322}FAPAEYAQLYSFAMG^{336})	VI	^{21}A, ^{110}N, ^{142}S, ^{146}I, ^{203}V, ^{288}T, ^{341}I, ^{406}V, ^{421}N, ^{464}S, ^{480}T
		VII	^{21}T, ^{110}S, ^{142}G, ^{146}V, ^{203}I, ^{288}A, ^{341}L, ^{406}M, ^{421}D, ^{464}P, ^{480}S
		XIII	^{21}T, ^{110}N, ^{142}G, ^{146}V, ^{203}V, ^{288}T, ^{341}L, ^{406}M, ^{421}D, ^{464}P, ^{480}S
P	RNA editing motif (GGGAAAAA)	VI	^{10}E, ^{66}S, ^{67}E, ^{83}I, ^{90}T, ^{103}G, ^{126}S, ^{138}S, ^{146}P, ^{159}N, ^{163}R, ^{209}S, ^{258}I, ^{323}T, ^{333}N, ^{343}I
		VII	^{10}D, ^{66}P, ^{67}D, ^{83}A, ^{90}A, ^{103}E, ^{126}N, ^{138}G, ^{146}Q, ^{159}S, ^{163}E, ^{209}N, ^{258}M, ^{323}A, ^{333}H, ^{343}A
		XIII	^{10}D, ^{66}P, ^{67}D, ^{83}A, ^{90}A, ^{103}E, ^{126}N, ^{138}G, ^{146}Q, ^{159}S, ^{163}E, ^{209}N, ^{258}M, ^{323}A, ^{333}H, ^{343}A
M	M late domain (^{23}FPIV^{26}), Bipartite nuclear localization motif (^{247}KKGKKVTFDKIEEKIRR^{263})	VI	^{36}K, ^{53}K, ^{75}V, ^{77}M, ^{105}K, ^{136}Q, ^{142}T, ^{143}A, ^{196}I, ^{213}E, ^{353}R, ^{355}T, ^{363}R
		VII	^{36}Q, ^{53}R, ^{75}A, ^{77}V, ^{105}E, ^{136}R, ^{142}M, ^{143}V, ^{196}V, ^{213}D, ^{353}K, ^{355}A, ^{363}K
		XIII	^{36}Q, ^{53}K, ^{75}V, ^{77}V, ^{105}E, ^{136}R, ^{142}M, ^{143}V, ^{196}V, ^{213}E, ^{353}R, ^{355}T, ^{363}R
F	Single peptide (^{1}MGSKPSIRIPVPLMLITRI^{19}), Cleavage motif (^{112}RRQK/RR^{117}), Fusion peptide (^{117}FIGAVIGSIVALGVATAAQITAAAALI^{142})	VI	^{14}S, ^{25}C, ^{90}N, ^{107}S, ^{121}I, ^{124}G, ^{132}S, ^{176}A, ^{179}I, ^{208}I, ^{246}I, ^{255}I, ^{270}A, ^{272}H, ^{304}G, ^{321}R, ^{342}N, ^{337}H, ^{385}A, ^{402}V, ^{411}H, ^{425}S, ^{445}Q, ^{448}V, ^{452}E, ^{487}R, ^{492}D, ^{494}K, ^{506}V, ^{509}A, ^{522}S, ^{537}M
		VII	^{14}M, ^{25}Y, ^{90}S, ^{107}A, ^{121}V, ^{124}S, ^{132}A, ^{176}S, ^{179}V, ^{208}V, ^{246}M, ^{255}V, ^{270}T, ^{272}Y, ^{304}E, ^{321}K, ^{342}D, ^{337}Y, ^{385}T, ^{402}A, ^{411}N, ^{425}N, ^{445}L, ^{448}I, ^{452}D, ^{487}K, ^{492}N, ^{494}R, ^{506}A, ^{509}V, ^{522}A, ^{537}L
		XIII	^{14}S, ^{25}C, ^{90}S, ^{107}S, ^{121}V, ^{124}S, ^{132}A, ^{176}A, ^{179}V, ^{208}V, ^{246}M, ^{255}V, ^{270}A, ^{272}Y, ^{304}E, ^{321}K, ^{342}N, ^{337}Y, ^{385}A, ^{402}A, ^{411}N, ^{425}S, ^{445}Q, ^{448}I, ^{452}E, ^{487}K, ^{492}N, ^{494}R, ^{506}V, ^{509}A, ^{522}S, ^{537}L
HN	Hydrophobic signal anchor (^{25}FRIAVLLLMIMILAISAAAL^{44}), Hexapeptide motif (^{234}NRKSCS^{239}), Haemagglutinin active motif-I (^{314}FPVYGGL^{320}), Haemagglutinin active motif-II (^{399}GAEGRI^{204})	VI	^{2}D, ^{6}R, ^{7}K, ^{9}V, ^{27}V, ^{36}T, ^{39}F,^{46}H, ^{56}I, ^{57}A, ^{62}M, ^{65}R, ^{75}N, ^{102}M, ^{112}H, ^{182}A, ^{218}R, ^{263}R, ^{289}A, ^{293}R, ^{304}S, ^{323}S, ^{333}N, ^{353}Q, ^{390}V, ^{395}V, ^{431}V, ^{495}E
		VII	^{2}S, ^{6}N, ^{7}R, ^{9}M, ^{27}I,^{36}I, ^{39}I, ^{46}Y, ^{56}L, ^{57}T, ^{62}A, ^{65}K, ^{75}S, ^{102}I, ^{112}Y, ^{182}T, ^{218}K, ^{263}K, ^{289}T, ^{293}K, ^{304}G, ^{323}N, ^{333}K, ^{353}R, ^{390}I, ^{395}I, ^{431}I, ^{495}G
		XIII	^{2}D, ^{6}N, ^{7}R, ^{9}V, ^{27}V, ^{36}I, ^{39}F, ^{46}Y, ^{56}L, ^{57}T, ^{62}M, ^{65}K, ^{75}S, ^{102}I, ^{112}Y, ^{182}T, ^{218}R, ^{263}K, ^{289}T, ^{293}K, ^{304}G, ^{323}N, ^{333}K, ^{353}Q, ^{390}I, ^{395}V, ^{431}I, ^{495}E
L	Domain interact with P protein (^{13}IILPESHLSSPLV^{25}), Domain-I (^{637}FITTDLQYCLNWRYQT^{653}), Domain-II (^{709}YIVSARGGIEGLCQKCWTMISIAAI^{733}), Domain-III (^{746}CMVQGDNQVIAVTR^{759}), Domain-IV (^{816}KDGAILSQVLKNSSKL^{831}), ATP binding motif (K_{21}AXGXG), Polymerase associated motif in domain-III (QGDNQ)	VI	^{12}Q, ^{46}I, ^{99}S, ^{103}V, ^{122}A, ^{149}L, ^{155}N, ^{156}T, ^{206}T, ^{295}S, ^{320}N, ^{337}V, ^{341}I, ^{438}D, ^{497}V, ^{626}L, ^{740}S, ^{759}K, ^{892}S, ^{974}S, ^{1032}A, ^{1079}T, ^{1102}P, ^{1110}N, ^{1174}N, ^{1226}V, ^{1291}I, ^{1371}T, ^{1410}K, ^{1511}A, ^{1523}S, ^{1524}G, ^{1564}D, ^{1579}S, ^{1636}V, ^{1695}R, ^{1701}K, ^{1726}R, ^{1732}L, ^{1757}V, ^{1771}R, ^{1852}S, ^{1998}R, ^{2008}I, ^{2120}R, ^{2127}D, ^{2130}G, ^{2195}A
		VII	^{12}R, ^{46}T, ^{99}A, ^{103}I, ^{122}T, ^{149}P, ^{155}S, ^{156}S, ^{206}V, ^{295}G, ^{320}D, ^{337}A, ^{341}V, ^{438}E, ^{497}I, ^{626}P, ^{740}A, ^{759}R, ^{892}P, ^{974}N, ^{1032}V, ^{1079}M, ^{1102}S, ^{1110}T, ^{1174}S, ^{1226}I, ^{1291}V, ^{1371}M, ^{1410}L, ^{1511}T, ^{1523}N, ^{1524}A, ^{1564}N, ^{1579}N, ^{1636}L, ^{1695}K, ^{1701}Q, ^{1726}L, ^{1732}P, ^{1757}A, ^{1771}H, ^{1852}A, ^{1998}K, ^{2008}T, ^{2120}K, ^{2127}N, ^{2130}S, ^{2195}T
		XIII	^{12}Q, ^{46}T, ^{99}A, ^{103}V, ^{122}T, ^{149}P, ^{155}S, ^{156}S, ^{206}V, ^{295}G, ^{320}D, ^{337}A, ^{341}V, ^{438}D, ^{497}V, ^{626}P, ^{740}S, ^{759}R, ^{892}P, ^{974}N, ^{1032}V, ^{1079}M, ^{1102}S, ^{1110}N, ^{1174}S, ^{1226}I, ^{1291}I, ^{1371}M, ^{1410}L, ^{1511}T, ^{1523}S, ^{1524}A, ^{1564}N, ^{1579}N, ^{1636}L, ^{1695}K, ^{1701}Q, ^{1726}L, ^{1732}P, ^{1757}A, ^{1771}H, ^{1852}A, ^{1998}K, ^{2008}T, ^{2120}K, ^{2127}D, ^{2130}S, ^{2195}A

Table 7: Comparative residue analysis of all sub-genotypes of *Avian avulavirus* 1 with respect of field used vaccine strains (LaSota; AF077761 and *Mukteshwar;* EF201805)

Genotype	Representative isolates	Amino acid position and substitutions																						
		12	**16**	**27**	**31**	**33**	**35**	**38**	**40**	**43**	**46**	**47**	**48**	**50**	**54**	**65**	**68**	**71**	**76-81**	**85**	**88**	**89**	**96**	**109**
II	Vaccine strain (AF077761)	K	I	V	P	L	K	E	C	A	D	A	Y	R	T	R	E	T	GRQGRL	I	G	V	A	K
III	Vaccine strain (EF201805)	-	-	-	-	M	-	-	-	-	E	-	-	-	-	-	-	-	R--R-F	-	S	-	-	N
VI	VIg	-	-	I	-	M	-	-	-	-	E	-	-	-	A	-	G	S	R-KK-F	-	S	-	S	N
	VIm	-	-	-	-	M	-	-	-	-	E	-	-	-	-	-	G	S	R--K-F	-	-	-	S	N
VII	VIIa	-	-	-	-	M	-	G	-	P	E	G	C	S	P	K	G	A	R--K-F	V	S	-	-	N
	VIIb	-	V	-	L	M	R	-	-	-	E	-	-	-	-	K	G	S	R--K-F	V	S	-	-	N
	VIIe	-	V	-	-	M	-	-	-	-	E	-	-	-	-	K	G	S	R--K-F	V	S	-	-	N
	VIIf	R	-	-	-	M	-	-	-	-	E	-	-	-	-	-	G	S	R--K-F	V	S	-	-	N
	VIIi	-	V	-	-	M	-	-	-	-	E	-	-	-	-	K	G	A	R--K-F	V	S	I	-	N
XIII	XIIIb	-	-	-	-	M	-	K/Q	G	-	E	-	-	-	-	-	G	S	R--K-F	V	S	-	-	N
	XIIId	-	-	-	-	M	-	-	-	-	E	-	-	-	-	-	G	A	R--K-F	V	S	-	-	N

Note: The consensus sequences of each sub-genotype was constructed for comparative residue analysis against vaccine strain

Table 8: Polymorphism results for nucleotide diversity of all coding genes (CDS analysis) of *Avian avulavirus* 1 belongs to different genotypes

Parameters	G-VI						G-VII						G-XIII					
	NP	P	M	F	HN	L	NP	P	M	F	HN	L	NP	P	M	F	HN	L
Total no. of Mutation	75	69	50	74	89	264	126	179	121	101	87	396	48	98	41	45	80	383
Total number of sites	1470	1188	1095	1629	1715	6615	1470	1187	1095	1629	1714	6614	1470	1188	1095	1629	1715	6614
Invariable Monomorphic sites	1396	1119	1045	1555	1626	6352	1346	1011	974	1528	1627	6229	1423	1090	1054	1584	1635	6233
Variable Polymorphic Sites	74	69	50	74	89	263	124	176	121	101	87	385	47	98	41	45	80	381
Singleton Variable Sites	69	69	48	71	81	246	15	9	7	26	18	90	33	0	25	1	66	226
Parsimony Informative Sites	5	5	2	3	8	17	109	167	114	75	69	295	14	98	16	44	14	155
No. of Haplotypes (h)	4	4	4	4	4	4	17	10	13	16	15	22	3	2	3	3	3	4
Nucleotide Diversity (Pi)	0.02596	0.02974	0.02314	0.02302	0.02672	0.02036	0.01511	0.02479	0.02043	0.01440	0.01178	0.01210	0.01490	0.04949	0.01790	0.01645	0.02029	0.02782
Average no. of pairwise nucleotide difference (k)	38.167	35.333	25.333	37.500	45.833	134.667	22.212	29.422	22.372	23.458	20.185	80.025	21.900	58.800	19.600	26.800	34.800	184.000
Tajima' D	-0.7001	-0.6387	-0.7401	-0.7409	-0.5841	-0.6803	-1.2868	-1.4740	-1.1567	-0.4449	-0.4470	-0.9118	-0.3732	1.8997*	-0.0306	1.8139*	-0.7112	0.0066

Note: The intraspecific polymorphism divergence time for whole genome sequences was estimated as T= 1.856 for genotype VI, T= 4.944* for genotype VII and T= 2.245 for genotype XIII (* $0.01< p <0.05$, ** $0.001< p <0.01$, *** $p <0.001$).

Table 9: Estimation of synonymous and non-synonymous substitutions rate in coding genes of so-far reported *Avian avulavirus* 1 using LaSota vaccine strain

Parameters	G-VI						G-VII						G-XIII					
	NP	P	M	F	HN	L	NP	P	M	F	HN	L	NP	P	M	F	HN	L
ds	1.0209	0.6064	0.0861	0.7914	0.6011	0.1686	0.9698	0.6089	0.1109	0.8022	0.6004	0.1647	0.8371	0.5373	0.0916	0.7766	0.5833	0.1405
dn	0.0456	0.1106	0.1901	0.0577	0.1081	0.1559	0.0442	0.0920	0.2242	0.0568	0.1247	0.1591	0.0465	0.0840	0.2061	0.0527	0.1262	0.1500
ds/dn	22.3883	5.4842	0.4530	13.7208	5.5600	1.0814	21.9689	6.6048	0.4938	14.1531	4.8165	1.0352	17.9967	6.3635	0.4447	14.7523	4.6290	0.9340
Variance (ds)	0.016500	0.005286	0.000393	0.006032	0.003413	0.000143	0.014460	0.005442	0.000499	0.007238	0.003817	0.000138	0.009550	0.003647	0.000435	0.006409	0.003383	0.000104
Std Deviation (ds)	0.128453	0.072704	0.019826	0.077667	0.058420	0.011972	0.120248	0.073770	0.022339	0.085076	0.061779	0.011733	0.097723	0.060388	0.020851	0.080055	0.058167	0.010219
Variance (dn)	0.000646	0.000496	0.000166	0.000356	0.000340	0.000031	0.000495	0.000435	0.000193	0.000359	0.000390	0.000031	0.000389	0.000315	0.000163	0.000330	0.000360	0.000028
Std Deviation (dn)	0.025424	0.022262	0.012902	0.018877	0.018437	0.005587	0.022242	0.020863	0.013910	0.018960	0.019752	0.005585	0.019719	0.017744	0.012760	0.018162	0.018985	0.005291
ps/pn	12.6040	4.0435	0.4844	8.8034	4.1070	1.0726	12.6930	4.8061	0.5310	9.0242	3.5974	1.0311	11.1802	4.8101	0.4787	9.5153	3.4950	0.9383
Variance (ps)	0.001080	0.001046	0.000312	0.000731	0.000687	0.000091	0.001089	0.001045	0.000372	0.000852	0.000769	0.000089	0.001024	0.000818	0.000341	0.000808	0.000714	0.000071
Std Deviation (ps)	0.032867	0.032348	0.017677	0.027038	0.026210	0.009561	0.032998	0.032322	0.019274	0.029188	0.027739	0.009410	0.031999	0.028609	0.018453	0.028426	0.026720	0.008442
Variance (pn)	0.000042	0.000098	0.000132	0.000043	0.000068	0.000020	0.000495	0.000083	0.000144	0.000042	0.000079	0.000020	0.000042	0.000070	0.000128	0.000042	0.000076	0.000019
Std Deviation (pn)	0.006506	0.009908	0.011503	0.006561	0.008272	0.004462	0.022242	0.009114	0.012005	0.006503	0.008869	0.004482	0.006451	0.008346	0.011293	0.006447	0.008722	0.004380

Abbreviations: ds = synonymous substitutions per synonymous site, dn= non-synonymous substitutions per non- synonymous site, ds/dn = proportion of ds and dn differences, ps/pn = proportion of synonymous (ps) and non-synonymous (pn) differences

Table 10: Estimation of synonymous and non-synonymous substitutions rate in coding genes of so far reported *Avian avulavirus* 1 using Mukstewar vaccine strain

Parameters	G-VI						G-VII						G-XIII					
	NP	P	M	F	HN	L	NP	P	M	F	HN	L	NP	P	M	F	HN	L
ds	0.7490	0.4066	0.1050	0.6606	0.4617	0.1463	0.7929	0.5250	0.1034	0.6290	0.4396	0.1391	0.6755	0.4534	0.1007	0.6775	0.4558	0.1365
dn	0.0286	0.1056	0.1460	0.0482	0.0872	0.1268	0.0278	0.0863	0.1788	0.0409	0.1011	0.1314	0.0360	0.0911	0.1612	0.0415	0.1034	0.1329
ds/dn	26.2265	3.8508	0.7199	13.7010	5.3007	1.1538	28.6792	6.0923	0.5770	15.4478	4.3496	1.0595	18.7682	4.9767	0.6253	16.3237	4.4137	1.0268
Variance (ds)	0.006991	0.002514	0.000494	0.004296	0.002105	0.000123	0.008341	0.004186	0.000466	0.004134	0.002024	0.000115	0.005695	0.002768	0.000483	0.004772	0.002107	0.000102
Std Deviation (ds)	0.083615	0.050140	0.022221	0.065545	0.045885	0.011097	0.091331	0.064696	0.021592	0.064298	0.044984	0.010702	0.075465	0.052607	0.021988	0.069080	0.045906	0.010102
Variance (dn)	0.000172	0.000293	0.000145	0.000213	0.000193	0.000025	0.000200	0.000339	0.000166	0.000168	0.000210	0.000026	0.000205	0.000288	0.000149	0.000210	0.000216	0.000025
Std Deviation (dn)	0.013117	0.017110	0.012056	0.014591	0.013907	0.005020	0.014140	0.018410	0.012874	0.012981	0.014504	0.005108	0.014333	0.016980	0.012217	0.014490	0.014694	0.004992
ps/pn	16.9039	3.1866	0.7392	9.4039	4.1915	1.1393	18.0230	4.6320	0.6057	10.7451	3.5178	1.0541	12.6676	3.9618	0.6502	11.0464	3.5407	1.0245
Variance (ps)	0.000945	0.000850	0.000373	0.000738	0.000614	0.000083	0.001004	0.001021	0.000354	0.000772	0.000627	0.000079	0.000939	0.000820	0.000370	0.000783	0.000625	0.000071
Std Deviation (ps)	0.030738	0.029159	0.019318	0.027162	0.024779	0.009129	0.031693	0.031950	0.018821	0.027793	0.025031	0.008890	0.030646	0.028633	0.019225	0.027985	0.025000	0.008420
Variance (pn)	0.000023	0.000099	0.000110	0.000037	0.000056	0.000017	0.000024	0.000083	0.000126	0.000031	0.000065	0.000018	0.000034	0.000086	0.000114	0.000034	0.000064	0.000017
Std Deviation (pn)	0.004819	0.009950	0.010481	0.006047	0.007512	0.004130	0.004910	0.009096	0.011224	0.005610	0.008071	0.004243	0.005821	0.009257	0.010681	0.005872	0.008002	0.004161

Abbreviations: ds = synonymous substitutions per synonymous site, dn= non-synonymous substitutions per non- synonymous site, ds/dn = proportion of ds and dn differences, ps/pn = proportion of synonymous (ps) and non-synonymous (pn) differences

Table 11: A brief summary of the Datamonkey-derived natural pressure selection site/s analysis for individual complete genes (CDS) of so-far reported *Avian avulavirus* 1

Parameters	G-VI						G-VII						G-XIII					
	NP	P	M	F	HN	L	NP	P	M	F	HN	L	NP	P	M	F	HN	L
Mean dN/dS	0.050124	0.37830	0.01818	0.10890	0.24668	0.13024	0.09871	0.28340	0.01944	0.38894	0.29408	0.13761	0.09096	0.35729	0.16906	0.14771	0.16462	0.10332
Single Likelihood Ancestor Counting (SLAC)																		
No. of positive and negative selection sites along with codon position	0 positive site and 59 negative sites	2 positive sites (28, 147) and 33 negative sites	0 positive site and 45 negative sites	1 positive site (179) and 48 negative sites	0 positive site and 42 negative sites	0 positive site and 61 negative sites	1 positive site (124) and 49 negative sites	0 positive site and 62 negative sites	0 positive site and 51 negative sites	2 positive sites (48, 342) and 64 negative sites	0 positive site and 48 negative sites	0 positive site and 109 negative sites	0 positive site and 43 negative sites	0 positive site and 60 negative sites	0 positive site and 59 negative sites	1 positive site (202) and 81 negative sites	0 positive site and 89 negative sites	0 positive site and 101 negative sites
Fixed Effect Likelihood (FEL)																		
No. of positive and negative selection sites along with codon position	1 positive site (401) and 65 negative sites	2 positive sites (28, 85) and 36 negative sites	0 positive sites and 47 negative sites	3 positive sites (10, 179, 509) and 55 negative sites	2 positive sites (213, 338) and 76 negative sites	0 positive site and 109 negative sites	1 positive site (21) and 61 negative sites	0 positive site and 55 negative sites	0 positive site and 67 negative sites	7 positive site (34, 48, 299, 287, 295, 302, 342) and 89 negative sites	1 positive site (267) and 79 negative sites	5 positive sites (89 237 569, 1191, 1723) and 159 negative sites	1 positive site (168) and 32 negative sites	1 positive site (57) and 93 negative sites	0 positive site and 31 negative sites	5 positive sites (114, 202, 356, 378, 400) and 104 negative sites	3 positive sites (46, 199, 351) and 86 negative sites	11 positive sites (61, 387, 560, 598, 623, 645, 666, 848, 900, 1191,1565) and 112 negative sites
Internal Branch Fixed Effect Likelihood (IFEL)																		
No. of positive and negative selection sites along with codon position	0 positive site and 14 negative sites	3 positive sites (28, 85, 181) and 9 negative sites	0 positive site and 15 negative sites	2 positive sites (10,457) and 13 negative sites	1 positive site (390) and 88 negative sites	1 positive site (1261) and 159 negative sites	0 positive site and 80 negative sites	1 positive site (203) and 63 negative sites	2 positive sites (12, 299) and 88 negative sites	9 positive sites (34, 48, 154, 299, 287, 295, 299, 302, 342) and 97 negative sites	3 positive sites (201, 267, 399) and 85 negative sites	11 positive sites (64, 89 167, 237, 401, 569,784, 982, 1191, 1534, 1723) and 181 negative sites	2 positive sites (168, 296) and 62 negative sites	1 positive site (57) and 57 negative sites	1 positive site (111) and 69 negative sites	7 positive sites (17, 68, 114, 202, 356, 378, 400) and 73 negative sites	3 positive sites (46, 199, 351) and 90 negative sites	13 positive sites (61, 387, 560, 598, 623, 645, 666, 848, 900, 1191,1243, 1406, 1565) and 167 negative sites

Table 12: The statistical outcomes and detailed information of two recombinant events which observed in recombination analysis of so far reported *Avian avulavirus* 1

Analysis	**Breakpoint 1**	**Breakpoint 2**
Potential recombinant isolate	JX854452 (Pheasant/MM20/Pakistan/ 2011)	JX854452 (Pheasant/MM20/Pakistan/ 2011)
Potential major parent	KP776462 (chicken/NDV/Pak/AW-14)	KP776462 (chicken/NDV/Pak/AW-14)
Similarity with major parent	97.4%	98.3%
Potential minor parent	JN682210 (Chicken/BY/Pakistan/2010)	JN682210 (Chicken/BY/Pakistan/2010)
Similarity with minor parent	99%	99.5%
Region derived from major parent	1-1007 nt, 3475-15233 nt	1-8817 nt, 9593-15219 nt
Region derived from minor parent	1008-3474 nt	8818-9592 nt
Beginning breakpoint	1006 nt	8793 nt
Beginning breakpoint 99% C.I	974-1038 nt	8764-8808 nt
Ending breakpoint	3448 nt	9566 nt
Ending breakpoint 99% C.I	3368-3511 nt	9548-9654 nt
Length of recombination event	2442 nt	774 nt
Probability (MC corrected)	8.733 E-160	3.529 E-84
Indication of true recombination		
Methods	**Average *p*-value**	**Average *p*-value**
RDP	$1.729X\ 10^{-164}$	$5.719X\ 10^{-83}$
GENECONV	$5.924X\ 10^{-158}$	$4.769X\ 10^{-86}$
BootScan	$3.884X\ 10^{-161}$	$9.267X\ 10^{-84}$
MaxChi	$1.023X\ 10^{-43}$	$6.312X\ 10^{-20}$
Chimaera	$6.108X\ 10^{-45}$	$1.743X\ 10^{-20}$
SiScan	$8.836X\ 10^{-52}$	$5.074X\ 10^{-22}$
3Seq	$2.247X\ 10^{-13}$	$8.988X\ 10^{-13}$