NEXT-GENERATION SEQUENCING AND NEWCASTLE DISEASE VIRUS DIAGNOSTICS

by

SALMAN LATIF BUTT

(Under the Direction of James B. Stanton and Claudio L. Afonso)

ABSTRACT

Virulent Newcastle disease virus (vNDV) strains cause Newcastle disease (ND), a highly contagious viral disease of birds with significant economic losses to the global poultry industry. The global spread, constant evolution, varying virulence, and the wide host range of NDV are challenges to the control of ND. Since NDV is an RNA virus that has a high mutation rate, and with the historical precedent of NDV causing panzootics, there is an increased threat of panzootics and the potential for evasion of current diagnostic and control measures. The primary goal of this research was to develop and optimize sequencing protocols for detection and thorough characterization of NDV from different type of clinical samples. Therefore, we aimed to develop a target-independent sequencing protocol to use formalin-fixed paraffin-embedded (FFPE) tissues by using next-generation sequencing (NGS). Multiple complete genomes of NDV were obtained from different types of FFPE tissue samples. These genome sequences were used for enhanced phylogenetic resolution of ND outbreaks. Then we aimed to develop a pan-NDV specific sequencing protocol (AmpSeq) by using MinION Nanopore sequencing technology and a data analysis pipeline for detection, virulence prediction and preliminary pathotyping of NDV. Results demonstrated that this AmpSeq protocol was rapid, sensitive and accurately detected all representative genotypes and sub-genotypes of NDV. Overall, these sequencing protocols will help advance the detection and characterization of NDV isolates and aid in ND control measures worldwide.

INDEX WORDS: Newcastle disease virus, Next-generation sequencing, Formalin fixed paraffin embedded, Evolution, Genome, MinION sequencing, Diagnostics, Clinical samples, Bioinformatics.

NEXT-GENERATION SEQUENCING AND NEWCASTLE DISEASE VIRUS DIAGNOSTICS

by

SALMAN LATIF BUTT

DVM, University of Agriculture Faisalabad, Pakistan, 2010

M.Phil, University of Agriculture Faisalabad, Pakistan, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

NEXT-GENERATION SEQUENCING AND NEWCASTLE DISEASE VIRUS DIAGNOSTICS

by

SALMAN LATIF BUTT

Co-Major Professor:	James B. Stanton
	Claudio L. Afonso
Committee:	Corrie C. Brown
	Monique S. França

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2019

DEDICATION

I dedicate my work to Allah (SWT) and to my mother who inspired me over the years and always expected best of me which kept me running this far. I also dedicate my work to my brother, Farrukh Latif Butt who not only put me in water and asked to learn swimming while always standing ashore to offer help and go beyond his means to support me.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisors Drs. James B Stanton and Claudio L Afonso, Corrie C Brown and Monique C Franca all of who tirelessly guided and encouraged me. I would also like to thank Dr. Keith Harris, former Head of the Department of Veterinary Pathology for his support for these years. I am also very grateful to the people at USDA-ARS-SEPRL NDV group, who provided their suggestion and helped me to thrive through my projects. I cannot forget to acknowledge Foreign Fulbright Program for providing me the opportunity for Doctoral studies. Finally, my deepest gratitude to my family for their support throughout these years.

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Newcastle disease

Newcastle disease (ND) is one of the most severe infectious diseases of poultry and is a major cause of economic losses to the poultry industry [1]. In 1926, ND was first described in outbreaks from two different regions of the world: the island of Java in Indonesia [2] and Newcastle-upon-Tyne in England [3]. In the subsequent years, this newly described disease was identified in the Philippines, India, Ceylon, Korea and Japan and the causative agent was named as Newcastle disease virus (NDV) [4]. In the 1930s, a relatively mild respiratory and neurologic disease was first described in California, USA, and named pneumoencephalitis [5]. The etiological agent for pneumoencephalitis was serologically compatible with NDV.

In the next few years, several NDV isolates obtained from chickens around the world were found to produce mild or no disease at all, determining that there were different subtypes according to the grade of pathogenicity, as seen by clinical disease and mortality rates in chickens [6]. Newcastle disease viruses are a highly diverse group of viruses with two distinct classes and 19 genotypes. In addition to this genetic diversity, NDV has a diverse host range, including domestic and wild bird species and these viruses are also diverse in their virulence. This includes low virulent viruses, whose replication is limited to the respiratory and digestive tracts and which typically cause clinically inapparent infections, to highly virulent viruses that cause acute disease with high rates of mortality [1]. Endemicity of this virus in multiple countries is a major challenge to global poultry production and at least four panzootics have been recognized since it was first identified [7]. Based on this impact, it is a notifiable disease to the World Organisation for Animal Health (OIE).

Clinical presentation of Newcastle disease

More than 240 avian species have been reported to be infected with NDV [8]; however, chickens are the most susceptible avian species to this virus. Clinical disease mainly depends on virulence of the NDV strain, tissue tropism of the virus, host species, route of exposure, viral load, immune status and environmental factors. General clinical signs of NDV infection include depression, loss of appetite, severe dehydration, emaciation and fever [6]. Birds infected with velogenic strains may result in sudden and high rates of mortality up to 100%, often with few clinical signs prior to death. The velogenic viscerotropic NDV causes severe respiratory clinical signs, edema, congestion of the face, conjunctivitis, diarrhea, muscular tremors, torticollis, paralysis of limbs and opisthotonos with a potential to cause 100% mortality in adult birds. With velogenic neurotropic NDV infections neurological signs are often more evident. Mesogenic NDV causes respiratory disease in adult chickens, with rare neurological signs, and sometimes death in young chickens. Usually, lentogenic strains of NDV do not cause disease, but on rare occasions, infections may result into respiratory disease in young birds [6].

Newcastle disease virus

Newcastle disease virus, synonymous with avian paramyxovirus-1 (APMV-1) (recently renamed as Avian Avulaviru-1) virions are generally spherical particles of 100–500 nm. NDV is a negative-sense, single-stranded, non-segmented, enveloped RNA virus of the order Mononegavirales, family *Paramyxoviridae* [9]. Historically, ND genotypes circulating between

1930–1960s, have a genome length of 15186, while NDV genotypes which emerged after 1960 have a genome length of 15192 nucleotides (nt). One distinct group of NDV (Class I, discussed below) has a genome length of 15198. NDV genome encodes six structural genes - nucleocapsid protein (NP), phosphoprotein (P), matrix protein (M), fusion protein (F), hemagglutinin-neuraminidase (HN), and RNA-dependent-RNA polymerase (L) [10]. The NP gene is involved in transcription, replication and encapsidation of the viral genome and protects it against RNAse activity. The P gene encodes phosphoprotein, which functions to avoid uncontrolled encapsidation and is essential for transcription and virus replication. The M gene encodes the matrix protein, which is involved in virus assembly. The F gene encodes the fusion protein, which is responsible for fusion of virion to the host cellular membrane allowing viral entry. The HN gene encodes for hemagglutinin-neuraminidase protein, which facilitates receptor-mediated attachment of viral particles to the host cell. The L gene encodes for the large polymerase protein, which is an RNA-dependent RNA polymerase that is responsible for nucleotide polymerization, 5′mRNA capping, methylation, and 3′polyadenylation of mRNAs as required for synthesis of viral mRNAs and viral replication [6, 11, 12]. Additionally, the 3′end of NDV genome has a 55 nt leader sequence and the 5′end has a trailer sequence of 114 nt [13].

Pathotypes of NDV

NDV pathogenicity is widely variable across different avian species and is dependent on viral genetics. NDV pathogenicity determination is an important step in both diagnostics and research to devise disease control strategy. NDV is classified into five recognized pathotypes based on severity and clinical presentation of disease [6, 14]. The “enteric asymptomatic NDV” replicates in the intestinal epithelium without causing any clinical signs related to infection.

Lentogenic NDV causes mild respiratory disease. Mesogenic NDV is characterized by low mortality rates in young birds and acute respiratory infection with occasional neurological signs. Viscerotropic velogenic NDV (vvNDV) produces acute disease with high rates of mortality and hemorrhages in multiple organs, especially the gastrointestinal tract. It may produce neurological signs such as torticollis and tremors. Neurotropic velogenic NDV (nvNDV) cause more severe neurologic signs and high mortality with fewer hemorrhagic lesions in the gastrointestinal tract [6].

Molecular determinants of pathogenicity

The genetic analysis of F gene sequence has revealed that virulence of NDV is primarily a function of the F gene sequence, with the amino acid sequence of the F protein cleavage site as the most predictive of potential pathogenicity of NDV strains (with the exception of Pigeon Paramyxovirus-1 in which F protein cleavage site is not always correlated with virulence) [15, 16]. For active viral infection, the inactive F protein precursor (F0) undergoes post-translational cleavage at site 117 into F_1 and F_2 by host cell proteases [11, 17]. A polybasic amino acid (3 or more) sequence at this cleavage site ($_{113}$R-Q-R/K-R ↓ F_{117}) results in velogenic and mesogenic NDV, as the ubiquitous furin-like endoproteases can mediate this cleavage. However, lentogenic NDV strains have a monobasic cleavage motif due to a leucine at position 117 ($_{113}$K/R-Q-G/E-R ↓ L_{117}), which limits cleavage to trypsin, primarily present in the respiratory and gastrointestinal tissues [11, 15, 18-20]. The importance of fusion protein cleavage site in pathogenicity of NDV has been demonstrated by several experiments in which the F gene of lentogenic NDV was mutated to have the polybasic amino acid sequence, which resulted in increased virulence [21-24]. However, when the F gene from velogenic NDV was inserted into a mesogenic NDV, there was no increase in the virulence, suggesting that there may be involvement of other virulence determinants of NDV. The HN protein may contribute to the virulence of virus [25-27]; however, the insertion of the HN proteins from the virulent strain Beaudette C into a LaSota Backbone did not increase the virulence of the virus [28] and neither did the insertion of the HN gene from virulent strains into a recombinant mesogenic NDV [26].

Newcastle disease virus genotype classification

All avian paramyxoviruses 1 (APMV-1) belong to single serotype; however, only virulent strains of APMV-1 cause ND. Based upon the genetic analysis of NDV genome sequences and a proposed unified classification system based on complete coding sequence of F gene, there are two distinct genetically divergent groups of NDV. Class I isolates belong to single genotype (Genotype I), whereas class II isolates are further divided into 18 genotypes with their further sub-genotypes [29]. Diel at al., proposed a minimum of 10% mean nucleotide distance of coding sequencing of F gene is required to be assigned to a different genotype and 3 % mean nucleotide difference to be assigned to a different sub-genotype [29].

Temporal, geographic and host range

In general, class I NDVs have been isolated from waterfowl, chickens, and shorebirds and class II NDVs have been repeatedly isolated from commercial and backyard chickens [7]. Although some of the genotypes of NDV are highly mobile and have been isolated across different continents (class I; class II genotype I, II, V, VI and VII), others have more limited geographical distributions (class II genotypes XI, XIII, XIV, XVIII). NDV genotypes have been isolated from chicken, duck and geese, shore birds and have been responsible for disease outbreaks in wild birds [7]. Predominantly, the low virulent class I viruses have been isolated from waterfowl and shorebirds [30] and mesogenic viruses of class II genotypes V or VI have been isolated from cormorants and pigeons, compared to velogenic NDV (class II genotypes V-X) which are usually recovered from vaccinated commercial poultry [31].

Detection of NDV

Virus isolation in specific pathogen free chicken embryonated eggs, virus neutralization assay for antigenic characterization, Intracerebral Pathogenicity Index (ICPI) for pathogenicity and RT-qPCR assays for sample screening purposes have been used over the past many years [32]. Conventional diagnostic tests that do not rely on nucleic acid such as hemagglutination and hemagglutination inhibition test to identify the virus as NDV in general and monoclonal antibodies (mAbs) have been used for detection and characterization of specifically class II NDV isolates. Other serological assays such as ELISA are also in use for determining antibody status as a result of vaccination in commercial poultry, but they have a limited role in surveillance studies. Virus isolation coupled with hemagglutination and hemagglutination inhibition assay is the "gold standard" in NDV diagnostics; however, these tests require samples containing live viruses and up to ten days to complete the procedure from sampling to results [33].

Pan-NDV RT-qPCR assays

Effective control of ND is dependent on rapid, sensitive, and specific diagnostic testing, which are typically oriented towards detection, genotyping, or prediction of virulence but most assays cannot accomplish all three tasks. Until recently, rapid diagnostic assays based on reverse-transcription quantitative PCRs (RT-qPCR) targeting different regions of NDV genome (M, L, F genes) have been used to detect NDV, depending upon the purpose of the assay [32, 34, 35]. The M gene assay is used to identify APMV-1 from diagnostic clinical samples, while the F gene assay is used to predict virulence (see “Molecular determinants of pathogenicity” above). These were developed to mainly detect class II NDV genotypes; however, the genetic variation also predisposes these PCR-based assays to false negative results due to mismatches between primers/probes and targeted nucleic acid sequences [36-38]. These assays were developed in response to the 2002 disease outbreak in California and were used for screening commercial poultry flocks. A new multiplexed M and L gene RT-qPCR (L-TET) was also developed to broadly detect class I and II NDV viruses, as L gene is more conserved region of NDV genomes. Although L gene targeted RT-qPCR assays have been developed for NDV detection, amplicons from this gene do not provide the virulence of detected NDVs [35].

RT-qPCR assays for virulent NDV isolates

Since the main virulence determinant is located in the F gene, it is critical to determine the F gene cleavage site [15]; thus, the F gene RT-PCR was developed to detect virulent NDV in clinical samples. Although the F gene RT-PCR has been widely used as it was field validated during 2002 NDV outbreak in California, this assay failed to detect variants of NDV, i.e., pigeon paramyxovirus-1 (PPMV-1, an antigenic and host variant of avian paramyxovirus 1; genotype Vi). Therefore, redesigning and validation of primers for PPMV-1 has been performed [37]. For these reasons, genotypic characterization of NDV is commonly done by sequencing and phylogenetic analysis of the complete F gene [7, 29]. Complete genome characterization may allow more reliable evolutionary and epidemiological studies.

Sanger sequencing for genotyping of NDV

Genotyping of NDV is commonly achieved through sequencing of the coding sequence of the F gene [29], which also allows for prediction of virulence. Preliminary genotyping can be accomplished through partial F gene sequencing (i.e., variable region) [30]. As discussed above, the rapid RT-qPCRs avoid the highly variable F gene and instead target more conserved regions of the genome (i.e., M and L genes) [32, 34, 35]. However, while this increases the applicability of these assays across genotypes, these assays lack applicability for virulence and genotypic determination. For example, while fusion-based assays can be used for detection [32], the variability of this region, which makes it useful for genotyping, hinders the universal applicability of any single primer set [37, 38]. Most common and current methods for detection, which include genotyping and virulence prediction, rely on Sanger sequencing. However, due to lack of multiplexing capability and limited sequencing depth, and potential for mixed infections, there is room to improve sequence-based diagnostic assays.

Next generation sequencing

DNA sequencing technology has been an important tool for virus characterization by providing information for molecular diagnosis. The high capacity of RNA viruses to mutate may result in a failure of target-specific PCR-based assays [37, 38]. Unlike traditional virus isolation and PCR-coupled Sanger sequencing, sequence-independent amplification will allow discovery of novel and unexpected viruses bypassing the isolation step [39]. Over the past few years with the growing applicability of next-generation sequencing (NGS) technology, different sequencing platforms have been developed and have become available. [40, 41]. This high throughput technology produces enormous amounts of sequencing data and provides deep insights into novel pathogens, analysis of intra-host genetic variation, and molecular epidemiology of viral infections [42]. Recently, application of metagenomic NGS has identified avian gamma coronavirus, orthoreovirus and a picobirnavirus in guinea fowl with fulminating disease [43, 44]. Similarly NGS of enteric virome in turkey and chicken led to the discovery of novel avian picobirnaviruses and many undescribed picornaviruses, which might be helpful in understanding enteric disease performance-related problems [44]. Application of NGS revealed Poecivirus, a novel picornavirus, as candidate pathogen for the causative agent of avian keratin disorder in black capped chickadees [45]. Additionally, mammalian-like astroviruses were also detected from European rollers (carnivorous bird) [46]. These advancements of in NGS have significantly improved the ability to sequence full-length genomes using a non-targeted approach [47].

Next-generation sequencing and NDV

The use of random primers in a non-targeted sequencing approach is ideal for RNA viruses, which rapidly mutate resulting in the problems listed above for PCR and Sanger sequencing. Furthermore, recent studies have shown that the cost of NDV complete genome sequencing by NGS can be reduced approximately 10-fold by multiplexing samples in a single sequencing run using Illumine Miseq instrument [48]. Previously, NDV detection and genome characterization by NGS have been done on egg-passaged live virus [48]. There are limited reports on the application of NGS with formalin-fixed paraffin-embedded (FFPE) samples for infectious disease studies [49-52]. Recently, we have published a retrospective study describing the molecular evolution of pigeon-adapted NDV in the U.S. using archived FFPE tissue samples from wild pigeons [53]. This approach, however, has not been tested to identify minor evolutionary changes or to trace the epidemiology of closely related isolates, as is often the case in endemic countries involved in intensive production of poultry. For example, recently reports showed that virulent NDVs have been circulating in Iran, a neighboring country of Pakistan, for the last two decades. A novel virulent strain of NDV has been recently reported and epidemiological information indicates its relatedness to the NDVs circulating in neighboring countries of Pakistan [54, 55].

Nanopore sequencing and virus diagnostics

Rapid advances in nucleic acid sequencing, have led to different sequencing platforms [56, 57] being widely applied for identification of novel viruses [60], whole genome sequencing [50], transcriptomics, and metagenomics [58, 59]. High capital investments and relatively long turnaround times have limited the widespread use of these next-generation sequencing (NGS) platforms, especially in developing countries [59]. Recent improvements in third-generation sequencing, including those introduced by Oxford Nanopore Technologies (ONT) [60], increase the utility of high-throughput sequencing as a useful tool for surveillance and pathogen characterization [61]. Nanopore sequencing technology is based on the threading of DNA through synthetic protein nanopores [62]. Among the transformative advantages of ONT's sequencing technology are the ability to perform real-time sequence analysis with a short turnaround time [63], the portability of the MinION device, the low startup cost compared to other high-throughput platforms, and the ability to sequence up to several thousand bases from individual RNA or DNA molecules. The MinION device has been successfully used to evaluate antibiotic resistance genes from several bacterial species [64, 65], sequence complete viral genome of an influenza virus [66] and Ebola virus [67], and detect partial viral genome sequences (e.g., Zika virus [68] and poxviruses [63] by sequencing PCR amplicons). Based on the rapid developments in techniques, these DNA sequencing technologies will take infectious disease diagnostics a step further and expand the ability to perform rapid, unbiased vial diagnostics.

Virus detection and bioinformatics

There is an increased dependence of data analysis on computational tools for accurate interpretation of sequencing data produced by the high throughput sequencing platforms [69]. The detailed review of bioinformatics tool for NGS [70] and for Nanopore sequencing [62] data have been reported before. The application of these tools is primarily dependent on the research question about the available sequencing data. Primarily, processing of sequencing data starts with generation of sequences, mapping and aligning of individual sequence reads to a reference and or de novo assembly in pathogen discovery situations. As the goal of sequencing of nucleic acid from clinical samples is, the precise determination of microbial population in samples, both, read-based and de novo assembly-based classifications have been used for taxonomic classification of microbial sequencing data. Read-based metagenomic classification software has been used for identification of microbial species from high-throughput sequencing data [71, 72]. There are multiple open-source and private software which are either standalone tools for metagenomic analysis of sequencing data, for example BLASTn [73], MEGAN [74] and Kraken [75]. Although the utility of these tools warrant the accuracy of results when input sequencing data is from next-generation sequencing platforms where the sequencing error is very low (<0.01%), however, the single-read error rate of nearly 10% of MinION sequencing [76] may limit the accuracy of this approach for Nanopore sequencing data, especially when attempting to obtain subspecies level differentiation [71]. De novo approaches that use quality-based filtering and clustering of reads, or use consensus-based error correction of Nanopore sequencing reads have been reported [77]. There are many web-based, open source data analysis platforms, such as Galaxy [78] and commercial software such as Geneious®, which have integrated multiple computational tools to get meaningful results from sequencing data. Different research groups have developed home based optimized computational pipelines for NGS and MinION sequencing data analysis [48, 61, 79, 80].

The global spread, constant evolution, varying virulence, and the wide host range of NDV are challenges to the control of ND [7]. The wide host range along with the enormous genetic diversity of NDV creates challenges for current rapid diagnostic assays and may pose an increased threat of panzootics. Reliable and affordable epidemiologic tools are needed to improve the management of ND and other infectious diseases.

In the first study, the aim was to sequence and characterize the genomes of NDV directly from clinical FFPE tissue samples by using NGS and to use the obtained data to infer the epidemiology of closely related isolates. This method did not include any procedure to target NDV specifically (e.g., sequence-based capture) or to enrich for viral nucleotides generally (e.g., depletion of host ribosomal RNA). This study demonstrates the ability of NGS, assisted by a customized bioinformatics pipeline, to assemble nearly complete NDV genomes from FFPE tissues and that those genomes are suitable for improved phylogenetic differentiation between closely related NDV isolates.

In the second study, the aim was to develop and optimize a specific, sensitive, and rapid protocol, using the MinION sequencer, to detect representative isolates from all currently circulating (excluding the Madagascar-limited genotype XI) genotypes of NDV. This protocol was also tested on clinical swab samples collected from chickens during disease outbreaks.

Additionally, a Galaxy-based, de novo AmpSeq workflow is presented that results in accurate final consensus sequences allowing for accurate genotype and virulence prediction.

References

1. Miller PJ, Koch G. Newcastle disease. In: Swayne DE, Glisson JR, McDougald LR, Nolan LK, Suarez DL, Nair V, editors. Diseases of poultry. 13th ed. Hoboken, New Jersey: Wiley-Blackwell; 2013. pp. 89-138.
2. Alexander D (2000) Newcastle disease and other avian paramyxoviruses. Revue Scientifique et Technique-Office International des Epizooties 19:443-55.
3. Doyle T (1927) A hitherto unrecognized disease of fowls due to a filter-passing virus. J Com Pathol Ther 40:144-69.
4. Doyle T (1935) Newcastle disease of fowls. J Comp Pathol Ther 48:1-20.
5. Beach J (1944) The neutralization in vitro of avian pneumoencephalitis virus by newcastle disease immune serum. Science 100:361-62.
6. Senne DAD (2008) Newcastle disease. Diseases of poultry, 12th ed Y M Saif A M Fadly J R Glisson L R McDougald L K Nolanand D E Swayne eds Iowa State University Press, Ames, IA:75-100.
7. Dimitrov KM, Ramey AM, Qiu X, Bahl J, Afonso CL (2016) Temporal, geographic, and host distribution of avian paramyxovirus 1 (newcastle disease virus). Infect Genet Evol 39:22-34.
8. Kaleta EF, Baldauf C. Newcastle disease in free-living and pet birds. Newcastle disease: Springer; 1988. pp. 197-246.
9. Amarasinghe GK, Ceballos NGA, Banyard AC, Basler CF, Bavari S, et al (2018) Taxonomy of the order mononegavirales: Update 2018. Arch Virol:1-12.

10. Chambers P, Millar NS, Bingham RW, Emmerson PT (1986) Molecular cloning of complementary DNA to newcastle disease virus, and nucleotide sequence analysis of the junction between the genes encoding the haemmaglutinin-neuraminidase and the large protein J Gen Virol 67:475-86.

11. Samal SK (2011) Newcastle disease and related avian paramyxoviruses. The biology of paramyxoviruses 1:69-114.

12. Lamb R, Parks G (2007) Paramyxoviridae: The viruses and their replication. Fields virology. Williams y Wilkins 5.

13. Khulape SA, Gaikwad SS, Chellappa MM, Mishra BP, Dey S (2014) Genetic characterization and pathogenicity assessment of newcastle disease virus isolated from wild peacock. Virus Genes 49:449-55.

14. Cattoli G, Susta L, Terregino C, Brown C (2011) Newcastle disease: A review of field recognition and current methods of laboratory detection. J Vet Diagn Invest 23:637-56.

15. Nagai Y, Klenk H-D, Rott R (1976) Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of newcastle disease virus. Virology 72:494-508.

16. Alexander DJ (2011) Newcastle disease in the european union 2000 to 2009. Avian Pathol 40:547-58.

17. Lamb RA (2001) Paramyxoviridae: The viruses and their replication. Fields virology.

18. Nagai Y (1995) Virus activation by host proteinases. A pivotal role in the spread of infection, tissue tropism and pathogenicity. Microbiol Immunol 39:1-9.

19. Gotoh B, Ohnishi Y, Inocencio N, Esaki E, Nakayama K, et al (1992) Mammalian subtilisin-related proteinases in cleavage activation of the paramyxovirus fusion glycoprotein: Superiority of furin/pace to pc2 or pc1/pc3. J Virol 66:6391-97.

20. Nagai Y, Klenk H-D (1977) Activation of precursors to both glycoproteins of newcastle disease virus by proteolytic cleavage. Virology 77:125-34.

21. Peeters BP, de Leeuw OS, Koch G, Gielkens AL (1999) Rescue of newcastle disease virus from cloned cdna: Evidence that cleavability of the fusion protein is a major determinant for virulence. J Virol 73:5001-09.

22. Panda A, Huang Z, Elankumaran S, Rockemann DD, Samal SK (2004) Role of fusion protein cleavage site in the virulence of newcastle disease virus. Microb Pathog 36:1-10.

23. Wakamatsu N, King DJ, Seal BS, Peeters BP, Brown CC (2006) The effect on pathogenesis of newcastle disease virus lasota strain from a mutation of the fusion cleavage site to a virulent sequence. Avian Dis 50:483-88.

24. Römer-Oberdörfer A, Werner O, Veits J, Mebatsion T, Mettenleiter TC (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 84:3121-29.

25. de Leeuw OS, Koch G, Hartog L, Ravenshorst N, Peeters BPH (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 86:1759-69.

26. Estevez C, King D, Seal B, Yu Q (2007) Evaluation of newcastle disease virus chimeras expressing the hemagglutinin-neuraminidase protein of velogenic strains in the context of a mesogenic recombinant virus backbone. Virus Res 129:182-90.

27. Panda A, Elankumaran S, Krishnamurthy S, Huang Z, Samal SK (2004) Loss of n-linked glycosylation from the hemagglutinin-neuraminidase protein alters virulence of newcastle disease virus. J Virol 78:4965-75.

28. Wakamatsu N, King DJ, Seal BS, Samal SK, Brown CC (2006) The pathogenesis of newcastle disease: A comparison of selected newcastle disease virus wild-type strains and their infectious clones. Virology 353:333-43.

29. Diel DG, da Silva LH, Liu H, Wang Z, Miller PJ, et al (2012) Genetic diversity of avian paramyxovirus type 1: Proposal for a unified nomenclature and classification system of newcastle disease virus genotypes. Infect Genet Evol 12:1770-79.

30. Kim LM, King DJ, Suarez DL, Wong CW, Afonso CL (2007) Characterization of class i newcastle disease virus isolates from hong kong live bird markets and detection using real-time reverse transcription-pcr. J Clin Microbiol 45:1310-14.

31. Miller PJ, Decanini EL, Afonso CL (2010) Newcastle disease: Evolution of genotypes and the related diagnostic challenges. Infect Genet Evol 10:26-35.

32. Wise MG, Suarez DL, Seal BS, Pedersen JC, Senne DA, et al (2004) Development of a real-time reverse-transcription pcr for detection of newcastle disease virus rna in clinical samples. J Clinl Microbiol 42:329-38.

33. Dimitrov KM, Clavijo A, Sneed L (2014) Rna extraction for molecular detection of newcastle disease virus – comparative study of three methods. Review de Médicine Vétérinaire 165:172-75.

34. Kim LM, Suarez DL, Afonso CL (2008) Detection of a broad range of class i and ii newcastle disease viruses using a multiplex real-time reverse transcription polymerase chain reaction assay. J Vet Diagn Invest 20:414-25.

35. Fuller CM, Brodd L, Irvine RM, Alexander DJ, Aldous EW (2010) Development of an l gene real-time reverse-transcription pcr assay for the detection of avian paramyxovirus type 1 rna in clinical samples. Arch Virol 155:817-23.

36. Cattoli G, De Battisti C, Marciano S, Ormelli S, Monne I, et al (2009) False-negative results of a validated real-time pcr protocol for diagnosis of newcastle disease due to genetic variability of the matrix gene. J Clin Microbiol 47:3791-2.

37. Sabra M, Dimitrov KM, Goraichuk IV, Wajid A, Sharma P, et al (2017) Phylogenetic assessment reveals continuous evolution and circulation of pigeon-derived virulent avian avulaviruses 1 in eastern europe, asia, and africa. BMC Vet Res 13:291.

38. Kim LM, Afonso CL, Suarez DL (2006) Effect of probe-site mismatches on detection of virulent newcastle disease viruses using a fusion-gene real-time reverse transcription polymerase chain reaction test. J Vet Diagn Invest 18:519-28.

39. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, et al (2009) Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4:e7370.

40. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al (2012) A tale of three next generation sequencing platforms: Comparison of ion torrent, pacific biosciences and illumina miseq sequencers. BMC Genomics 13:341.

41. Knief C (2014) Analysis of plant microbe interactions in the era of next generation sequencing technologies. Frontiers in plant science 5:216.

42. Barzon L, Lavezzo E, Costanzi G, Franchin E, Toppo S, et al (2013) Next-generation sequencing technologies in diagnostic virology. J Clin Virol 58:346-50.

43. Kapgate S, Barbuddhe S, Kumanan K (2015) Next generation sequencing technologies: Tool to study avian virus diversity.

44. Day JM, Zsak L (2016) Molecular characterization of enteric picornaviruses in archived turkey and chicken samples from the united states. Avian Dis 60:500-05.

45. Zylberberg M, Van Hemert C, Dumbacher JP, Handel CM, Tihan T, et al (2016) Novel picornavirus associated with avian keratin disorder in alaskan birds. MBio 7:e00874-16.

46. Pankovics P, Boros Á, Kiss T, Delwart E, Reuter G (2015) Detection of a mammalian-like astrovirus in bird, european roller (coracias garrulus). Infect Genet Evol 34:114-21.

47. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, et al (2012) Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. MBio 3.

48. Dimitrov KM, Sharma P, Volkening JD, Goraichuk IV, Wajid A, et al (2017) A robust and cost-effective approach to sequence and analyze complete genomes of small rna viruses. Virol J 14:72.

49. Carrick DM, Mehaffey MG, Sachs MC, Altekruse S, Camalier C, et al (2015) Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. PLoS One 10:e0127353.

50. Bodewes R, van Run PR, Schurch AC, Koopmans MP, Osterhaus AD, et al (2015) Virus characterization and discovery in formalin-fixed paraffin-embedded tissues. J Virol Methods 214:54-9.

51. Mubemba B, Thompson PN, Odendaal L, Coetzee P, Venter EH (2017) Evaluation of positive rift valley fever virus formalin-fixed paraffin embedded samples as a source of sequence data for retrospective phylogenetic analysis. J Virol Methods 243:10-14.

52. Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, et al (2013) High-throughput rna sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. The Journal of Pathology 229:535-45.

53. He Y, Taylor TL, Dimitrov KM, Butt SL, Stanton JB, et al (2018) Whole-genome sequencing of genotype vi newcastle disease viruses from formalin-fixed paraffin-embedded tissues from wild pigeons reveals continuous evolution and previously unrecognized genetic diversity in the us. Virol J 15:9.

54. Mayahi V, Esmaelizad M (2017) Molecular evolution and epidemiological links study of newcastle disease virus isolates from 1995 to 2016 in iran. Arch Virol 162:3727-43.

55. Esmaelizad M, Mayahi V, Pashaei M, Goudarzi H (2017) Identification of novel newcastle disease virus sub-genotype vii-(j) based on the fusion protein. Arch Virol 162:971-78.

56. Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J (2016) High throughput sequencing: An overview of sequencing chemistry. Indian J Microbiol 56:394-404.

57. Rhoads A, Au KF (2015) Pacbio sequencing and its applications. Genomics Proteomics Bioinformatics 13:278-89.

58. Cruz-Rivera M, Forbi JC, Yamasaki LH, Vazquez-Chacon CA, Martinez-Guarneros A, et al (2013) Molecular epidemiology of viral diseases in the era of next generation sequencing. J Clin Virol 57:378-80.

59. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, et al (2013) Next generation sequencing of viral rna genomes. BMC Genomics 14:444.

60. Phan H, Stoesser N, Maciuca I, Toma F, Szekely E, et al (2017) Illumina short-read and minion long-read whole genome sequencing to characterise the molecular epidemiology of an ndm-1-serratia marcescens outbreak in romania. J Antimicrob Chemother 73 (3) 672–79.

61. Greninger A, Naccache S, Federman S, Yu G, Mbala P, et al (2015) Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med 7:99.

62. Jain M, Olsen HE, Paten B, Akeson M (2016) The oxford nanopore minion: Delivery of nanopore sequencing to the genomics community. Genome Biol 17:239.

63. Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, et al (2015) Bacterial and viral identification and differentiation by amplicon sequencing on the minion nanopore sequencer. Gigascience 4:12.

64. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, et al (2015) Minion nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol 33:296.

65. Lemon JK, Khil PP, Frank KM, Dekker JP (2017) Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. J Clin Microbiol 55:3530-43.

66. Wang J, Moore NE, Deng Y-M, Eccles DA, Hall RJ (2015) Minion nanopore sequencing of an influenza genome. Front Microbiol 6:766.

67. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, et al (2016) Real-time, portable genome sequencing for ebola surveillance. Nature 530:228.

68. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, et al (2017) Multiplex pcr method for minion and illumina sequencing of zika and other virus genomes directly from clinical samples. Nat Protoc 12:1261.

69. Scholz MB, Chien-Chi Lo, and Patrick SG Chain. (2012) Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. Curr Opin Biotechnol 23:9-15.

70. Crockett DK, Voelkerding KV, Brown AF, Stewart RL. Bioinformatics tools in clinical genomics. Genomic applications in pathology: Springer; 2019. pp. 209-34.

71. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK (2016) Sequencing 16s rrna gene fragments using the pacbio smrt DNA sequencing system. PeerJ 4:e1869.

72. Kim D, Song L, Breitwieser FP, Salzberg SL (2016) Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res.

73. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-10.

74. Huson DH, Weber N. Microbial community analysis using megan. Methods in enzymology. 531: Elsevier; 2013. pp. 465-85.

75. Wood DE, Salzberg SL (2014) Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46.

76. Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, et al (2015) Minion analysis and reference consortium: Phase 1 data release and analysis. F1000Research 4.

77. Jain M, Tyson JR, Loose M, Ip CL, Eccles DA, et al (2017) Minion analysis and reference consortium: Phase 2 data release and analysis of r9. 0 chemistry. F1000Research 6.

78. Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, et al (2016) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3-W10.

79. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, et al (2014) A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res 24:1180-92.

80. Wan Y, Renner DW, Albert I, Szpara ML (2015) Viramp: A galaxy-based viral genome assembly pipeline. Gigascience 4:19.

81. Kammon A, Heidari A, Dayhum A, Eldaghayes I, Sharif M, et al (2015) Characterization of avian influenza and newcastle disease viruses from poultry in libya. Avian Dis 59:422-30.

CHAPTER 2

ENHANCED PHYLOGENETIC RESOLUTION OF NEWCASTLE DISEASE OUTBREAKS USING COMPLETE VIRAL GENOME SEQUENCES FROM FORMALIN-FIXED PARAFFIN-EMBEDDED TISSUE SAMPLES[1]

[1] Salman Latif Butt, Kiril M. Dimitrov, Jian Zhang, Poonam Sharma, Abdul Wajid, Tasra Bibi, Asma Basharat, Corrie C. Brown, Shafqat F. Rehmani, Patti J. Miller, James B. Stanton, Claudio L. Afonso. *Virus Genes* 2019, 1-11, Reprinted here with permission of publisher.

Abstract

Highly virulent Newcastle disease virus (NDV) causes Newcastle disease (ND), which is a threat to poultry production worldwide. Effective disease management requires approaches to accurately determine sources of infection, which involves tracking of closely related viruses. Next-generation sequencing (NGS) has emerged as a research tool for thorough genetic characterization of infectious organisms. Previously formalin-fixed paraffin-embedded (FFPE) tissues have been used to conduct retrospective epidemiological studies of related but genetically distinct viruses. However, this study extends the applicability of NGS for complete genome analysis of viruses from FFPE tissues to track the evolution of closely related viruses. Total RNA was obtained from FFPE spleens, lungs, brains, and small intestines of chickens in 11 poultry flocks during disease outbreaks in Pakistan. The RNA was randomly sequenced on an Illumina MiSeq instrument and the raw data were analyzed using a custom data analysis pipeline that includes de novo assembly. Genomes of virulent NDV were detected in 10/11 birds: eight nearly complete (>95% coverage of concatenated coding sequence) and two partial genomes. Phylogeny of the NDV complete genome coding sequences was compared to current methods of analysis based on the full and partial fusion genes and determined that the approach provided a better phylogenetic resolution. Two distinct lineages of sub-genotype VIIi NDV were identified to be simultaneously circulating in Pakistani poultry. Non-targeted NGS of total RNA from FFPE tissues coupled with de novo assembly provided a reliable, safe, and affordable method to conduct epidemiological and evolutionary studies to facilitate management of ND in Pakistan.

.

Introduction

Newcastle disease (ND) is a significant worldwide disease of poultry caused by virulent strains of *Avian avulavirus* 1 (AAvV-1), commonly known as Newcastle disease virus (NDV) [1-3]. Endemicity of this virus in multiple countries is a major challenge to global poultry production and at least four panzootics have been recognized since it was first identified in the 1920s [4-6]. Reliable and affordable epidemiologic tools that can use tissues transported in a safe and convenient manner across international boundaries are needed to improve the management of ND and other infectious diseases.

Newcastle disease virus is a negative-sense, single-stranded, non-segmented, enveloped RNA virus of the *Paramyxoviridae* family [1]. The genome of NDV is approximately 15 kb and encodes six structural gene products: nucleocapsid protein (NP), phosphoprotein (P), matrix protein (M), fusion protein (F), hemagglutinin-neuraminidase (HN), and RNA-dependent-RNA polymerase (L) [7]. The amino acid sequence of the F protein cleavage site is a major molecular determinant of pathogenicity and virulence of NDV. Virulent NDVs have multiple (3 or more) basic amino acids (^{112}R/K-R-Q-R/K-R↓F^{117}) at the F protein cleavage site and phenylalanine at position 117 [8]. Sequencing of the F gene and other gene segments has been used to predict viral virulence and to study epidemiology and evolution; however, this and other targeted methods have limitations in comparison to random complete genome sequencing.

Genotypic characterization of NDV is commonly done by sequencing the complete F gene [9]; whereas, complete genome characterization may allow more reliable evolutionary and epidemiological studies. The commonly used Sanger sequencing with overlapping primers for complete genome sequencing has significant limitations such as high cost, reduced output, long turnaround time, and is dependent on pre-existing knowledge of the genetic makeup of the viruses [10-12] making this an impractical method for full genome sequencing.

Shipping and testing samples with live NDV is problematic because virulent NDV is classified as a select agent in many countries and working with live select agent demands extensive safety precautions (BSL-3 laboratories), expensive transportation, and special permissions. In contrast, formalin fixation inactivates infectious agents; thus, formalin-fixed paraffin-embedded (FFPE) tissues can be easily, safely, and affordably transported across boundaries and processed in a biosafety level 1 laboratory [13]. Additionally, formalin-fixation is the gold standard for pathologic preservation and FFPE tissues allow for a full pathologic analysis of the tissue, which will provide a better list of differential diagnoses. As such, FFPE tissues are universally collected clinical samples for routine histopathology [14]. Flinders Technology Associates filter papers (FTA® cards) have been used for hazard-free handling of sample and detection of NDV by RT-PCR [15], and while these show promise for full-genomic characterization of infecting viruses (https://www.ncbi.nlm.nih.gov/pubmed/28684566), FTA cards are limited to genetic analysis only, which prevents broader testing that is often required in a diagnostic setting when samples are collected prior to a definitive diagnosis. Immunohistochemistry (IHC), targeting nucleocapsid protein of NDV as an antigen within FFPE tissues samples, has been used in experimental studies to demonstrate tissue tropism of NDV [16]; however, it does not provide as complete, sensitive, or specific characterization of the NDV as nucleic acid-based methods [13].

The advancement of massively parallel sequencing (next-generation sequencing; NGS) has significantly improved the ability to sequence full-length genomes using a non-targeted approach [17]. The use of random primers in a non-targeted sequencing approach is ideal for RNA viruses, which rapidly mutate resulting in the problems listed above for PCR and Sanger sequencing. Furthermore, recent studies have shown that the cost of NDV complete genome sequencing by NGS can be reduced approximately 10-fold by multiplexing samples in a single sequencing run [10]. Previously, NDV detection and genome characterization by NGS have been done on egg-passaged live viruses [10]. There are limited reports on the application of NGS with FFPE samples for infectious disease studies [18-21]. Recently, a retrospective study describing the molecular evolution of pigeon-adapted NDV in the U.S. using archived FFPE tissue samples from wild pigeons was reported [22]. This approach, however, has not been tested to identify minor evolutionary changes or to trace the epidemiology of closely related isolates, as is often the case in endemic countries involved in intensive production of poultry. For example, recently reports showed that virulent NDVs have been circulating in Iran, a neighboring country of Pakistan, for last 21 years. A novel virulent strain of NDV has been recently reported and epidemiological information indicates its relatedness to the NDVs circulation in China which is also a neighboring country of Pakistan [6, 23, 24]. Due to the endemicity of NDV in Pakistan and the presence of constantly evolving genotypes VIIi in neighboring countries of Pakistan, clinical samples from Pakistan were chosen to test the ability of this method to identify minor evolutionary changes.

The aim of the current study was to sequence and characterize the genomes of NDV directly from clinical FFPE tissue samples by using NGS and to use the obtained data to track the epidemiology of closely related isolates. This method did not include any procedure to target NDV specifically (e.g., sequence-based capture) or to enrich for viral nucleotides generally (e.g., depletion of host ribosomal RNA). This study demonstrates the ability of NGS, assisted by a customized bioinformatics pipeline, to assemble nearly complete NDV genomes from FFPE tissues and that those genomes are suitable for improved phylogenetic differentiation between closely related NDV isolates.

Materials and methods

Sample collection and processing

Thirty-six FFPE field tissue samples (spleen, lung, brain, and small intestine) collected in 2015 from 11 chickens during disease outbreaks in 11 poultry flocks in five different regions of the Pakistani Punjab province were used in this study (Table 2.1). Collected tissues were fixed for 24 to 48 hours in neutral-buffered 10% formalin within 4 hours of collection. Fixed tissues were embedded in low-melt paraffin following routine procedures [25]. These paraffin blocks were stored at room temperature and subsequently shipped without refrigeration to the Southeast Poultry Research Laboratory of the USDA, Athens, Georgia, USA, for NDV characterization.

Immunohistochemistry assay

Immunohistochemistry was performed to detect NDV nucleocapsid protein in tissue samples (spleen, lung, brain, and small intestine). Briefly, sections of 3 µm thickness were cut and immunostained with monoclonal antibodies directed against NDV nucleocapsid protein using alkaline phosphatase method as described previously [26]. Immunostained sections were microscopically evaluated and the samples were scored as negative or positive.

RNA extraction

For each sample, six 10-µm thick tissue sections were cut and collected in one 1.5 ml centrifuge tube and immediately deparaffinized by CitriSolv^{TM} (VWR International, USA). Before and after collecting tissue sections from each sample, the microtome blade was decontaminated with RNase Away (Sigma, USA) to avoid cross-contamination between samples. Total RNA was extracted using the RNeasy FFPE Kit (Qiagen, USA) as per manufacturer’s instructions and quantified using the Qubit® RNA HS Assay Kit on a Qubit® fluorometer 3.0 (ThermoFisher Scientific, USA). The purity of RNA (a ratio of absorption at 260 nm and 280 nm wavelength, $A_{260/280}$) was measured on a NanoDrop Spectrophotometer 2000/2000c (ThermoFisher Scientific, USA). Fragment size of RNA was determined using the RNA 6000 Pico kit on an Agilent Bioanalyzer® instrument (Agilent Technologies, USA) as per manufacturer’s instructions.

NGS library preparation

DNA Libraries for NGS sequencing were prepared using the KAPA Stranded RNA-Seq Library Preparation Kit for Illumina platforms (Kapa Biosystems, USA) following manufacturer’s instructions. Briefly, the protocol involved synthesis of first strand and second strand of DNA from total extracted RNA by using random primers, marking and A-tailing of cDNA fragments and ligation of unique adapters allowing indexing of each sample for multiplexing purposes. Finally, bead-purified adapter-ligated libraries were amplified with PCR (12 cycles) using library amplification master mix provided with the kit. The Qubit® fluorometer 3.0 was used to quantify dsDNA concentration in libraries using the dsDNA High Sensitivity Assay kits (ThermoFisher Scientific, USA). The average DNA fragment size in each library was determined using the High Sensitivity DNA kit in the Agilent Bioanalyzer®. To facilitate uniform clustering during sequencing process in the MiSeq flow cell, libraries with an average DNA fragment size of 240 to 300 bp and a concentration greater than 3 ng/µl were used. The prepared libraries were diluted to 4 nM. Five microliters of each library were pooled and denatured with NaOH (0.2 N final concentration). After 5 minutes of incubation at room temperature, the pool was further diluted to 20 pM concentration with chilled HT1 hybridization buffer (Illumina, USA). Using the same buffer, the final concentration of the library pool was diluted to 10 pM. A control library (3% PhiX174, Illumina, USA) was added and the pool was chilled on ice. The paired-end sequencing was conducted on an Illumina MiSeq instrument using a 300 cycle (2 × 150) MiSeq Reagent Kit v2 (Illumina, USA). After automated cluster generation in MiSeq, the sequencing reads were demultiplexed based on their unique adapter.

Genome assembly

Pre-assembly processing from read quality assessment to digital normalization was conducted as previously described [10] using Fast QC [27], Cutadapt v1.6 (TruSeq LT index sequences were used as reference file the study herein) [28], BWA-MEM v0.2.1 [29], Filter sequencing by mapping v0.0.4 tool (https://github.com/peterjc/pico_galaxy/tree/master/tools/seq_filter_by_mapping), an in-house tool for forward and reverse read re-synchronization (https://github.com/jvolkening/b2b-utils), PEAR v0.9.6.0 [30], and Khmer package v1.1-1 [31]. De novo assembly of the filtered, trimmed and synchronized reads was performed with MIRA assembler v3.4.1 [32]; however, to account for the relatively low number of NDV reads, as compared to egg-grown NDV samples from the previous study [10], the following parameters for *de novo* assembly were changed: minimum number of reads required to build a contiguous sequence = 5, minimum overlap = 10, contig length cut off = 60 bp, default values were used for the rest of the settings. The assembled contigs were aligned to the NCBI nt database (accessed on 15 September 2017), using BLASTn (cut off *E* value = 0.001) and the best hit was further used as the reference genome. As previously described [10], the final consensus sequence was obtained by processing the trimmed and un-normalized reads (after read re-synchronization) through BWA-MEM [29] (using the BLASTn-defined genome as the reference) and an in-house tool for parsing (https://github.com/jvolkening/b2b-utils). In addition, when a bird had multiple tissues with NDV, the raw sequence data from those differing tissue samples were merged and re-analyzed again using the same bioinformatics workflow to see if there is any difference in depth and percentage of genome coverage on a per-tissue basis versus a per-bird basis. Tissues were considered NDV positive by NGS if at least one contig hit to NDV (see BLASTn step above)

Phylogenetic analyses

The final consensus sequences were aligned using ClustalW [33] and the concatenated complete genome coding sequences (CDS) were used for nucleotide (nt) distance estimation to closely related NDV isolates obtained from GenBank using MEGA6 [34] and the Maximum Composite Likelihood model [35]. Consensus sequences from six birds in this study (for which complete genome coding sequences were obtained), 32 sequences of sub-genotypes of genotype VII obtained from GenBank were used for final phylogenetic tree construction (See Table S2.1). Determination of the best-fit substitution model was performed using MEGA6, and the goodness-of-fit for each model was measured by corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC) [34]. The final tree was constructed using the maximum-likelihood method based on the General Time Reversible model as implemented in MEGA6, with 500 bootstrap replicates [36]. Additionally, two more datasets were parsed from the initial dataset – one containing the complete fusion gene coding sequences and one with the first 375 nucleotides of the fusion gene coding sequences (denoted as “partial fusion gene sequence”), both regions being commonly used in phylogenetic analyses and epidemiological studies [9, 37]. Nucleotide distance estimation was performed as described above and bootstrap maximum-likelihood phylogenetic trees based on the Tamura 3-parameter (Tamura, 1992) were constructed using these smaller datasets utilizing MEGA6.

Accession numbers

The sequences obtained in the current study were submitted to GenBank and are available under the accession numbers from MG200021 to MG200026.

Statistical analyses

Statistical analyses were performed in JMP Pro Version 13.2. Chi-square test (a contingency table 2 × 4) was performed to determine whether frequency of NDV identification by NGS is significantly different between different tissue types. In addition, a one-way-ANOVA by ranks Kruskal Wallis H test was performed to determine if mean percentage genome coverage is significantly different between different types of tissue.

Results

Immunohistochemistry

Thirty-six FFPE tissues (spleen, brain, lung, and small intestine) were examined by IHC with a primary monoclonal antibody to detect NDV nucleoprotein (Table 2.2, Fig. S2.1). At least one tissue per bird was positive by IHC. Of 36 tissue samples, 31 were IHC positive for NDV nucleocapsid protein and five samples were IHC negative.

RNA isolation and evaluation

Total RNA from thirty-six tissue samples was extracted with RNA concentrations of 15.8–84 ng/µl. The 260/280 ratios were 1.9–2.03, and the RNA integrity number (RIN) values were 1.7–6.5 (see Table S2.2 for details).

Next-generation sequencing and genome assembly

Raw read analysis

Five of the prepared thirty-six libraries did not meet the criteria (see materials and methods) set for library quality and were not submitted for sequencing. Suboptimal read generation was observed for sample 1163-spleen. Only 928 total reads (0.005% of the raw reads) were assigned to this sample and its reads were not included in downstream data analysis. Multiplexed NGS of the remaining 30 tissue samples generated a total of 19,078,363 raw reads (255,408 to 4,797,051 per sample). A total of 166,084 reads (237 to 47,742 per sample) remained after filtering out the host genome and PhiX control reads (Table 2.2). NDV reads were a small fraction of the total reads obtained (0.01%–1.9% per positive sample). Approximately 91–99% of the total raw reads mapped to the chicken genome. However, the chicken reads were filtered out for the purposes of this study. The reads that mapped to bacterial genome were also filtered out. These values are not presented in the results.

NDV contigs: Per tissue

Results were initially analyzed on a per tissue basis. By employing de novo assembly coupled with reference-based consensus re-calling, NDV contigs (NDV positive) were obtained from 64.5% (20/31) of the sequenced tissue samples (55.6% [20/36] of the total extracted), with NGS-positive samples having 179–44,525 NDV reads per positive sample. In 14 of the tissue samples, the NDV genome coverage was greater than 90% (91.75%–99.84%) (Table 2.2). Although NDV was identified by NGS in more lung tissues (72.7%) compared to other tissues (spleen = 62.5%, brain = 62.5%, and small intestines = 50%), these differences were not statistically significant ($\chi2$ chi-square test = 0.721; p = 0.8684, α = 0.05). In terms of genome recovery among different tissues types (Table 2.2), mean percentage genome coverage was not significantly different among different tissue types (Kruskal-Wallis statistic = 0.9916, p = 0.8033, α = 0.05). Additionally, no nucleotide differences were identified between sequences obtained from different tissues of a same bird (data not shown).

NDV contigs: Per bird

Since identical consensus sequences were identified among different tissues, the raw reads from different tissues of the same bird were merged so that the data could be analyzed on per-bird basis. NDV contigs were assembled from these data in 91% of the birds (10 out of 11 birds) with 191–54,086 NDV reads per bird When the raw sequence data obtained from different tissues of the same bird were merged, in 9 out of the 11 birds, the genome coverage was greater than 90% (91.87 to 99.83%). The mean depth of each assembled genome per bird ranged from 4 to 513 in the 9 out of 10 NGS-positive birds that had > 90% genome coverage. In one bird (#1171) that was NGS positive, the genome coverage was 45.02% with a max depth of 7× (See Table 2.3).

Phylogenetic analyses

Complete coding sequences analysis

The complete coding sequences of all six genes were obtained from six of the eleven birds and the complete fusion gene coding sequence was obtained from eight of the eleven birds. The deduced amino acid sequence of the fusion cleavage site for all sequences was specific for virulent viruses ($_{113}$RRQKR↓F$_{117}$, with the exception of sample #1168 that had $_{113}$RRQRR↓F$_{117}$) (See supplementary Table 2.3). The nucleotide distances between the studied sequences and GenBank sequences were estimated. Most of the studied sequences were very closely related to each other (99.7% mean identity within them), except one (#1168) that was 1.5% distant. This first group of sequences was most closely related (99.3–99.4%) to sequences from viruses isolated from a variety of species (chickens, pigeons, and peacock) from Pakistan in 2014 and 2015. The sequence that was more genetically distant (#1168) was closely related (99.2–99.4%) to sequences from viruses isolated from varying species (chickens, parakeets, pigeon, duck) in Pakistan in 2015 and 2016. Also, NDV isolates in this study had nucleotide distance of 1.2% from previously reported sub-genotypes VIIi and (See supplementary Table 2.4). To further confirm the evolutionary relationship between the NDV sequences studied here and sequences available in GenBank, phylogenetic analysis using the complete genome concatenated CDS was performed. In the created phylogenetic tree, the isolates studied here expectedly grouped together within sub-genotype VIIi with the viruses that showed highest nucleotide sequence identity to them (Fig. 2.1A). Taken together, the results from the distance analysis and the phylogenetic tree demonstrated that all sequences obtained from field FFPE tissue sample from diseased chicken in Pakistan in 2015 belong to NDV class II sub-genotype VIIi. In addition, two separate branches of isolates were identified in the tree. To assess the genetic diversity within sub-genotype VIIi, the nucleotide distance between the sequences in these two branches was estimated and they were found to be less closely related (1.2% nucleotide distance)

Complete and partial fusion gene coding sequences analysis

The evolutionary history was also inferred by phylogenetic analysis using the smaller datasets (same taxa as in the CDS analysis) comprising the complete F gene coding sequences (Fig. 2.1B) and the partial F gene coding sequence (Fig. 2.1C) and the results were compared to those of the complete genome sequences (Fig. 2.1A). The overall phylogenetic topology based on partial or complete F gene coding sequence analyses was consistent with phylogenetic grouping observed in analysis based on complete genome coding sequences. However, when the complete genome CDS were used for the analysis, higher bootstrap values were observed as compared to those in the partial and complete F gene coding sequences. In addition, sequences that appeared identical in the partial and complete fusion gene sequences analyses (MG200022/chicken/Pakistan/Kassur/1165/966/2015 and MG200026//chicken/Pakistan/Gujranwala/1174/978/2015 [bold and red], KX268688//parakeet/Pakistan/Rawalpindi/SFR-RP15/2015 and KX791183/parakeet/Pakistan/R-Pindi/SFR-16/2016 [bold font and blue], KX268689/parrot/Pakistan/Lahore/SFR-129/2015 and KX268691/parakeet/Pakistan/Lahore/SFR-148A/2015 [bold font and brown], KU885948/peacock/Pakistan/MZS-UVAS/2014 and KY076037/chicken/Pakistan/Buner/KPK/5A/1004/2015 [bold font and green]) (Fig. 1B and 1C) and had 100% nucleotide identity between them, were readily differentiated in the complete genome tree and had nucleotide distances ranging from 0.1 to 0.34%. To specify the genomic regions that readily differentiated NDV isolates that were 100% identical in the fusion protein gene based phylogeny, a pairwise comparison of nucleotide sequences of all six genes of these isolates (MG200022 vs MG200026, KX268691 vs. KX268689 and KY076037 vs. KU885948) was performed (See Table 2.4). It was observed that in MG200022 vs MG200026 comparison, phosphoprotein gene contributed highest variation (0.4%) followed by Hemagglutinin (HN), polymerase (L), matrix (M) and nucleoprotein (NP). The HN and L genes contributed most of the variation (0.5%) followed by P gene (0.3%) when a comparison of KX268691 vs. KX268689 was made.

Discussion

The feasibility of using FFPE tissues to sequence NDV complete genomes and to conduct evolutionary and epidemiological studies of closely related NDV-infected field tissue samples has been shown. The phylogenetic analyses confirmed that the detected viruses belonged to the virulent sub-genotype VIIi. Furthermore, this study demonstrates that the phylogenetic results from concatenated complete coding sequences obtained from FFPE tissues provide better resolution compared to individual or partial genes and is sufficient to demonstrate viral evolution that would otherwise remain unnoticed because of the limited genomic fragments analyzed. This is also the first use of NGS to characterize NDV genomes from FFPE tissue samples collected during recent disease outbreaks in commercial poultry. Additionally, some variants of NDV, including pigeon paramyxoviruses (PPMVs), are virulent without the polybasic amino acids at the Fusion protein cleavage site, as some of the PPMVs have shown increased virulence to chickens after only a few passages and without any change in the F gene coding region. In addition, replacing the F gene of the avirulent pigeon-adapted NDV with virulent NDV failed to produce virulent chimeric viruses. These observations underscore the importance of other genes in determining the virulence of NDV [39]. Therefore, these observations highlight the need and utility of complete genome analysis of field isolates.

As an example of the utility of this protocol, evolutionary and phylogenetic analyses were conducted. First, it was confirmed that the detected viruses belonged to the virulent sub-genotype VIIi, which is currently circulating in Pakistan [40-42]. Furthermore, two lineages of sub-genotype VIIi NDV were identified. The phylogenetic analysis showed clear separation of sequences from Pakistan isolated between 2014 and 2016 into two independent branches based on the complete coding sequences. While both groups of viruses were simultaneously circulating in geographically close areas, the genetic distance of 1.4% between them suggests at least 15 years of independent evolution of these two groups [43]. These findings demonstrate no direct, or very distant, epidemiological link between the viruses in these two groups. For example, they may represent two separate introductions events of NDV into Pakistan that evolved elsewhere or that they evolved locally from a common, unidentified ancestor introduced earlier in the region. Wajid and co-authors have recently reported the simultaneous circulation of sub-genotype VIIi virulent NDV in various poultry and non-poultry avian species in Pakistan based on complete F gene analyses [42]. While no particular interdependence among the hosts affected by ND was observed, the role of non-poultry species in the epidemiology and endemicity of NDV in Pakistan was confirmed.

The investigation of the epidemiologic link between highly related NDV viruses is of vital importance especially in closely located geographical regions where ND has acquired endemicity [6, 23, 24]. A phylogenetic analysis of the complete F gene assisted with evolutionary distances has been proposed previously for accurate classification of NDV genotypes [9]. In this study, the general separation of the sequences into major branches was consistent across the three different phylogenetic analyses. However, a better resolution in terms of higher bootstrap values of highly related NDV isolates were observed in the phylogenetic analysis based on complete genome coding sequences of NDV. In addition, the complete genome CDS analysis allowed the differentiation of viruses, which was otherwise impossible using only the complete or partial fusion gene sequences. These findings suggest the application of paraffin-embedded tissues from outbreaks to track epidemiologically very closely related viruses. Although, the utility of NGS to sequence complete genomes of NDV from clinical FFPE tissue samples was described here, the utility of this method in diagnostic settings would require further testing to more rigorously establish sensitivity, specificity and limit of detection, which is beyond the scope of the current study.

In the evaluation of per-tissue percentage genome assembly of NDV, all four tissue types showed different genome assembly percentages. While the sample size for any given tissue was relatively low, no statistically significant difference was identified in the frequency of positivity between the tissue types. This is consistent with, Barbezange and Jestin, which reported that RT-PCR showed no differences in the rate of NDV detection between lung, brain, trachea and spleen [44].

As described earlier, there was no difference between consensus sequences from different tissues of the same host; however, low sequencing depth and generation of short reads are limitations that currently prevent this approach from being used to study viral quasispecies diversity within clinical FFPE tissue samples from the same host. As the focus of this study was on consensus-level genome analysis, future studies are required to determine if altered protocols (e.g., targeted sequencing, increased depth of sequencing, etc.) could be used for the investigation of tissue effect on quasispecies diversity or similar question involving small genetic differences between sequences.

In this study, without any enrichment procedure for viral RNA, partial to complete AAvV-1 genomes coding sequences were assembled from archived FFPE tissues collected in 2015 from chickens during disease outbreaks. Genome assembly was up to 100% of the protein coding sequences and up to 99.81% of the complete genome of AAvV-1. These data suggest the feasibility of using FFPE tissue samples to sequence complete AAvV-1 genome for epidemiological studies of closely related virus isolates. Using FFPE tissues for direct sequencing of AAvV-1 is advantageous due to the fact that they can easily be transported as pathogens are inactivated make them available to transport across countries for research. As formalin fixation of tissues affect nucleic quality, DV200 values (estimated by Bioanalyzer) represent relative amounts of RNA fragments > 200 bp and may be an adequate predictor of RNA quality from FFPE tissue samples and may be evaluated in the future studies involving FFPE tissue samples. FTA® cards have also been used for hazard-free sample handling and evaluated for successful NDV detection [15], provide a method to inactive samples and are aimed at preserving nucleic acid integrity; however, the use of FTA cards by practicing clinicians, especially for animal pathogens is currently limited. Formalin-fixed tissues remain a standard sample as it allows for numerous diagnostic assays, which is valuable especially when there is no clinical diagnosis.

Conclusion

In conclusion, using FFPE tissues for direct sequencing of NDV genome is useful because FFPE tissues can be conveniently and affordably transported, due to pathogen inactivation, and because FFPE tissues are the primary means of preserving tissue for routine diagnostic and pathologic testing and for historical archival. This study demonstrates the capability of full-genome epidemiologic investigations in FFPE samples. The use of random sequencing coupled with absence of any virus enrichment procedure make this technique likely to be applicable to sequence virus genomes from clinical FFPE tissues in other viral infections. Additionally, the results demonstrate that sub-genotype VIIi viruses are still circulating and evolving in Pakistan after they were first identified in the country in 2011 and support that active epidemiologic surveillance for NDV is needed.

Acknowledgments The authors would like to thank Poonam Sharma, Jeremy D. Volkening and Dawn Williams-Coplin for their technical support during the study.

The mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer.

Funding: This work was supported by the U.S. Department of Agriculture ARS CRIS 6040-32000-072, the U.S. Department of State BEP/CRDF NDV 31063. Dr. Salman Latif Butt is conducting his PhD research program sponsored by the Fulbright U.S. Student Program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest

Human participants and/or animals No human subjects were used in this study. This article does not contain any studies with animals performed by any of the authors. Tissues were collected from dead birds being used for diagnostic purposes during disease outbreaks. Sampling was carried out by veterinarian, who took different samples as part of his routine work and under the permission of the farm owner. Since these are diagnostic specimens, sampling did not require the approval of the Ethics Committee.

References

1. Miller PJ, Koch G. Newcastle disease. In: Diseases of Poultry. 13th edn. Edited by Swayne DE, Glisson JR, McDougald LR, Nolan LK, Suarez DL, Nair V. Hoboken, New Jersey: Wiley-Blackwell; 2013; 89-138.
2. Afonso CL, Amarasinghe GK, Banyai K, Bao Y, Basler CF, Bavari S, Bejerman N, Blasdell KR, Briand FX, Briese T *et al*. Taxonomy of the order Mononegavirales: update 2016. Arch Virol 2016;161(8):2351-2360.
3. Amarasinghe GK, Ceballos NGA, Banyard AC, Basler CF, Bavari S, Bennett AJ, Blasdell KR, Briese T, Bukreyev A, Caì Y. Taxonomy of the order Mononegavirales: update 2018. Arch Virol 2018:1-12.
4. Dimitrov KM, Ramey AM, Qiu X, Bahl J, Afonso CL. Temporal, geographic, and host distribution of avian paramyxovirus 1 (Newcastle disease virus). Infect Genet Evol 2016;39(2016):22-34.
5. Sabouri F, Vasfi Marandi M, Bashashati M. Characterization of a novel VIIl sub-genotype of Newcastle disease virus circulating in Iran. Avian Pathol 2018;47(1):90-99.
6. Gowthaman V, Singh SD, Dhama K, Desingu PA, Kumar A, Malik YS, Munir M. Isolation and characterization of genotype XIII Newcastle disease virus from Emu in India. VirusDisease 2016;27(3):315-318.
7. Chambers P, Millar NS, Bingham RW, Emmerson PT. Molecular cloning of complementary DNA to Newcastle disease virus, and nucleotide sequence analysis of the

junction between the genes encoding the haemmaglutinin-neuraminidase and the large protein J Gen Virol 1986;67:475-486.

8. de Leeuw OS, Koch G, Hartog L, Ravenshorst N, Peeters BPH. 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 2005;86(Pt. 6):1759-1769.

9. Diel DG, da Silva LH, Liu H, Wang Z, Miller PJ, Afonso CL. Genetic diversity of avian paramyxovirus type 1: proposal for a unified nomenclature and classification system of Newcastle disease virus genotypes. Infect Genet Evol 2012;12(8):1770-1779.

10. Dimitrov KM, Sharma P, Volkening JD, Goraichuk IV, Wajid A, Rehmani SF, Basharat A, Shittu I, Joannis TM, Miller PJ *et al.* A robust and cost-effective approach to sequence and analyze complete genomes of small RNA viruses. Virol J 2017;14(1):72.

11. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, Spiro DJ, Sitz J, Koo H, Bolotov P *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 2005;437(7062):1162-1166.

12. Ansorge WJ. Next-generation DNA sequencing techniques. N Biotechnol 2009;25(4):195-203.

13. Wakamatsu N, King DJ, Seal BS, Brown CC. Detection of Newcastle disease virus RNA by reverse transcription-polymerase chain reaction using formalin-fixed, paraffin-embedded tissue and comparison with immunohistochemistry and in situ hybridization. J Vet Diagn Invest 2007;19(4):396-400.

14. Klopfleisch R, Weiss AT, Gruber AD. Excavation of a buried treasure--DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. Histol Histopathol 2011;26(6):797-810.

15. Perozo F, Villegas P, Estevez C, Alvarado I, Purvis LB. Use of FTA® filter paper for the molecular detection of Newcastle disease virus. Avian Pathol 2006;35(02):93-98.

16. Wakamatsu N, King D, Kapczynski D, Seal B, Brown C. Experimental pathogenesis for chickens, turkeys, and pigeons of exotic Newcastle disease virus from an outbreak in California during 2002-2003. Vet Pathol 2006;43(6):925-933.

17. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ *et al*. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. mBio 2012;3(6).

18. Carrick DM, Mehaffey MG, Sachs MC, Altekruse S, Camalier C, Chuaqui R, Cozen W, Das B, Hernandez BY, Lih CJ *et al*. Robustness of Next Generation Sequencing on Older Formalin-Fixed Paraffin-Embedded Tissue. PLoS One 2015;10(7):e0127353.

19. Bodewes R, van Run PR, Schurch AC, Koopmans MP, Osterhaus AD, Baumgartner W, Kuiken T, Smits SL. Virus characterization and discovery in formalin-fixed paraffin-embedded tissues. J Virol Methods 2015;214:54-59.

20. Mubemba B, Thompson PN, Odendaal L, Coetzee P, Venter EH. Evaluation of positive Rift Valley fever virus formalin-fixed paraffin embedded samples as a source of sequence data for retrospective phylogenetic analysis. J Virol Methods 2017;243:10-14.

21. Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK. High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. The Journal of pathology 2013;229(4):535-545.

22. He Y, Taylor TL, Dimitrov KM, Butt SL, Stanton JB, Goraichuk IV, Fenton H, Poulson R, Zhang J, Brown CC. Whole-genome sequencing of genotype VI Newcastle disease viruses from formalin-fixed paraffin-embedded tissues from wild pigeons reveals continuous evolution and previously unrecognized genetic diversity in the US. Virol J 2018;15(1):9.

23. Mayahi V, Esmaelizad M. Molecular evolution and epidemiological links study of Newcastle disease virus isolates from 1995 to 2016 in Iran. Arch Virol 2017;162(12):3727-3743.

24. Esmaelizad M, Mayahi V, Pashaei M, Goudarzi H. Identification of novel Newcastle disease virus sub-genotype VII-(j) based on the fusion protein. Arch Virol 2017;162(4):971-978.

25. Bancroft JD, Gamble M. Theory and practice of histological techniques: Elsevier Health Sciences; 2008.

26. Susta L, Miller PJ, Afonso CL, Brown CC. Clinicopathological characterization in poultry of three strains of Newcastle disease virus isolated from recent outbreaks. Vet Pathol 2011;48(2):349-360.

27. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

28. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 2011;17(1):pp. 10-12.

29. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997 2013.

30. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 2013;30(5):614-620.

31. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Research 2015;4.

32. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: German conference on bioinformatics: 1999: Hanover, Germany; 1999: 45-56.

33. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673-4680.

34. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 2013;30(12):2725-2729.

35. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 2004;101(30):11030-11035.

36. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993;10(3):512-526.

37. Courtney SC, Susta L, Gomez D, Hines NL, Pedersen JC, Brown CC, Miller PJ, Afonso CL. Highly divergent virulent isolates of Newcastle disease virus from the Dominican

Republic are members of a new genotype that may have evolved unnoticed for over 2 decades. J Clin Microbiol 2013;51(2):508-517.

38. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. Mol Biol Evol 1992;9(4):678-687.

39. Dortmans JC, Koch G, Rottier PJ, Peeters BP. Virulence of Newcastle disease virus: what is known so far? Vet Res 2011;42(1):122.

40. Rehmani SF, Wajid A, Bibi T, Nazir B, Mukhtar N, Hussain A, Lone NA, Yaqub T, Afonso CL. Presence of virulent Newcastle disease virus in vaccinated chickens in farms in Pakistan. J Clin Microbiol 2015;53(5):1715-1718.

41. Miller PJ, Haddas R, Simanov L, Lublin A, Rehmani SF, Wajid A, Bibi T, Khan TA, Yaqub T, Setiyaningsih S *et al*. Identification of new sub-genotypes of virulent Newcastle disease virus with potential panzootic features. Infect Genet Evol 2015;29:216-229.

42. Wajid A, Dimitrov KM, Wasim M, Rehmani SF, Basharat A, Bibi T, Arif S, Yaqub T, Tayyab M, Ababneh M *et al*. Repeated isolation of virulent Newcastle disease viruses in poultry and captive non-poultry avian species in Pakistan from 2011 to 2016. Prev Vet Med 2017;142:1-6.

43. Dimitrov KM, Lee D-H, Williams-Coplin D, Olivier TL, Miller PJ, Afonso CL. Newcastle Disease Viruses Causing Recent Outbreaks Worldwide Show Unexpectedly High Genetic Similarity to Historical Virulent Isolates from the 1940s. J Clin Microbiol 2016;54(5):1228-1235.

44. Barbezange C, Jestin V. Development of a RT-nested PCR test detecting pigeon Paramyxovirus-1 directly from organs of infected animals. J Virol Methods 2002;106(2):197-207.

Table 2.1. Background information of 36 field FFPE tissue samples collected from different regions of Pakistan during disease outbreaks in 2015.

Sample ID	**Collected organs**	**No. of samples**	**Breed of chicken**	**Location**
1162	spleen, lung, small intestine	3	Broiler chicken	Kassur
1163	spleen, lung, brain	3	Broiler chicken	Lahore
1164	spleen, lung, brain, small intestine	4	Desi chicken	Lahore
1165	spleen, lung, brain	3	Broiler chicken	Kassur
1166	spleen, lung, brain	3	Broiler chicken	Kamoki
1168	spleen, lung, brain, small intestine	4	Broiler chicken	Sheikhupura
1170	spleen, lung, brain, small intestine	4	Broiler chicken	Lahore
1171	spleen, lung, brain	3	Broiler chicken	Sheikhupura
1172	spleen, lung, brain	3	Broiler chicken	Lahore
1173	lung, brain, small intestine	3	Desi chicken	Gujranwala
1174	lung, brain, small intestine	3	Broiler chicken	Gujranwala

Table 2.2. Summary of sequencing and IHC data of 31 field FFPE tissue samples collected from different regions of Pakistan during disease outbreaks in 2015

Samples		IHC (+/-)	Raw read pairs	Filtered read pairs[a]	*De novo* NDV detection by contig (Yes/no)	Number of reads for contig	NDV reads (% of raw read pairs)	Percent genome coverage
SEPRL ID	Tissue							
1162-L	Lung	+	386682	31630	Yes	791	0.20	94.92
1162-SI	S. Intestine	-	372467	1349	No	NA	NA	NA
1163-SP[b]	Spleen	+	928	6	No	NA	NA	NA
1163-L	Lung	+	395061	808	No	NA	NA	NA
1163-B	Brain	+	385336	939	Yes	577	0.15	91.75
1164-SP	Spleen	+	441166	1194	Yes	346	0.08	77.71
1164-L	Lung	+	351767	6158	Yes	5383	1.53	99.72
1164-B[c]	Brain	+	366257	394	No	NA	NA	NA
1165-SP	Spleen	+	407295	3253	Yes	884	0.22	93.26
1165-L[c]	Lung	+	399209	1582	No	NA	NA	NA
1165-B	Brain	-	316718	13426	Yes	6033	1.90	99.72
1166-SP	Spleen	+	4797051	47742	Yes	44525	0.93	99.84
1166-L	Lung	+	1616222	14386	Yes	9358	0.58	99.81
1166-B	Brain	-	394885	5023	Yes	203	0.05	53.67
1168-SP	Spleen	+	445743	662	Yes	308	0.07	61.56
1168-L	Lung	+	398836	1676	Yes	255	0.06	55.27
1168-B	Brain	+	322405	5281	Yes	3801	1.18	99.77
1168-SI	S. Intestine	+	414401	3595	Yes	3100	0.75	99.47
1170-SP	Spleen	+	592168	1087	Yes	768	0.13	94.67
1170-L	Lung	+	545772	4250	Yes	2626	0.48	99.72
1170-B	Brain	+	411037	5430	Yes	761	0.19	96.29
1170-SI	S. Intestine	+	2006616	2589	Yes	197	0.01	51.17
1171-SP[c]	Spleen	+	530855	237	No	NA	NA	NA
1171-L	Lung	+	389248	1347	Yes	179	0.05	42.00
1172-SP[c]	Spleen	+	315083	380	No	NA	NA	NA
1172-L[c]	Lung	+	455865	1008	No	NA	NA	NA
1172-B	Brain	-	442780	1086	No	NA	NA	NA
1173-L	Lung	+	273342	2236	Yes	1057	0.39	98.06
1173-SI[c]	S. Intestine	+	373069	2256	No	NA	NA	NA
1174-L	Lung	-	255408	4670	Yes	2533	0.99	99.80
1174-B	Brain	+	275619	410	No	NA	NA	NA

[a] the number of paired reads after filtering out host (*Gallus gallus* 5.0) and internal control (PhiX) reads
[b] Suboptimal read generation was observed from #1163-spleen
[c] IHC-positive cells ranged from 10 to 200 per studied tissue section.
NA = not applicable

Table 2.3 Summary of genome assembly and sequencing data and from each of the 10 NDV positive birds.

<table>
<tr><th>SEPRL</th><th>Raw read pairs</th><th>Filtered read pairs[a]</th><th>Mean fragment length</th><th>Fragment length SD[b]</th><th>Forward read quality[c]</th><th>Reverse read quality[c]</th><th>Number of reads for consensus[d]</th><th>Final coverage depth[c]</th><th>Consensus nucleotide length</th><th>Percent genome coverage</th></tr>
<tr><td>1162</td><td>759149</td><td>43641</td><td>127</td><td>42</td><td>2|34|37|37|38</td><td>2|33|36|37|38</td><td>828</td><td>0|4|6|9|23</td><td>14469</td><td>95.24</td></tr>
<tr><td>1163</td><td>780861</td><td>7380</td><td>121</td><td>43</td><td>2|35|37|37|38</td><td>2|34|36|37|38</td><td>585</td><td>0|3|4|6|15</td><td>13958</td><td>91.87</td></tr>
<tr><td>1164</td><td>1146188</td><td>30829</td><td>133</td><td>48</td><td>2|34|37|37|38</td><td>2|33|36|37|38</td><td>5751</td><td>0|29|40|55|191</td><td>15149</td><td>99.71</td></tr>
<tr><td>1165</td><td>1123285</td><td>49586</td><td>121</td><td>36</td><td>2|34|37|37|38</td><td>2|33|36|37|38</td><td>6925</td><td>0|39|53|70|131</td><td>15149</td><td>99.71</td></tr>
<tr><td>1166</td><td>6808158</td><td>134809</td><td>148</td><td>50</td><td>2|35|37|38|38</td><td>2|33|36|37|38</td><td>54086</td><td>0|394|513|634|1057</td><td>15167</td><td>99.83</td></tr>
<tr><td>1168</td><td>1581385</td><td>135333</td><td>130</td><td>50</td><td>2|34|36|37|38</td><td>2|32|36|37|38</td><td>7464</td><td>0|47|59|78|161</td><td>15159</td><td>99.78</td></tr>
<tr><td>1170</td><td>3555593</td><td>42487</td><td>139</td><td>48</td><td>2|35|37|37|38</td><td>2|33|36|37|38</td><td>4352</td><td>0|30|37|47|96</td><td>15163</td><td>99.8</td></tr>
<tr><td>1171</td><td>920103</td><td>23922</td><td>106</td><td>42</td><td>2|34|37|37|38</td><td>2|32|36|37|38</td><td>191</td><td>0|0|1|2|7</td><td>6840</td><td>45.02</td></tr>
<tr><td>1173</td><td>646411</td><td>37909</td><td>126</td><td>43</td><td>2|34|37|37|38</td><td>2|33|36|37|38</td><td>1129</td><td>0|6|9|12|30</td><td>15013</td><td>98.82</td></tr>
<tr><td>1174</td><td>531027</td><td>16168</td><td>123</td><td>40</td><td>2|34|37|37|38</td><td>2|33|36|37|38</td><td>2536</td><td>0|14|20|26|58</td><td>15162</td><td>99.8</td></tr>
</table>

[a] the number of paired reads after filtering out host (*Gallus gallus* 5.0) and internal control (PhiX) reads

[b] SD = standard deviation

[c] the numbers represent read quality distribution (minimum | lower quartile | median | upper quartile | maximum).

[d] number of paired reads used for the final consensus sequence form each bird

Table 2.4 Evolutionary distance estimated using the complete coding sequences of individual genes of NDV genomes.

Gene ID		**MG200022 vs. MG200026**[a]	**KX268691 vs. KX268689**[b]	**KY076037 vs. KU885948**[c]
Fusion	(F)	0.000	0.000	0.000
Partial fusion*		0.000	0.000	0.000
Hemagglutinin	(HN)	0.002	0.001	0.005
Polymerase	(L)	0.001	0.002	0.005
Matrix	(M)	0.001	0.000	0.000
Nucleoprotein	(NP)	0.001	0.000	0.000
Phosphoprotein	(P)	0.004	0.000	0.003

[a] NDV isolates from current study (#1165 vs 1174)

[b, c] NDV genotype VIIi sequences from GenBank

*374 bp long sequence of Fusion gene that is frequently used for NDV classification and virulence prediction

Table S2.1 Complete genome coding sequences of NDV Genotype VIIi from Pakistan were used for constructing Maximum Likelihood phylogenetic trees.

Genotype	GenBank accession number	Host	Country	Isolate	Year
VIIi	MG200021*	Chicken	Pakistan	Lahore/1164	2015
VIIi	MG200022*	Chicken	Pakistan	Kassur/1165	2015
VIIi	MG200023*	Chicken	Pakistan	Kamoki/1166	2015
VIIi	MG200024*	Chicken	Pakistan	Sheikhupura/1168	2015
VIIi	MG200025*	Chicken	Pakistan	Lahore/1170	2015
VIIi	MG200026*	Chicken	Pakistan	Gurjanwala/1174	2015
VII i	KM670337	Chicken	Pakistan	SFR-611	2013
VII i	KP776462	Chicken	Pakistan	AW-14	2014
VII i	KU845252	Duck	Pakistan	AW-123	2015
VII i	KU885948	Peacock	Pakistan	MZS-UVAS	2014
VII i	KX268688	Parakeet	Pakistan	SFR-RP15	2015
VII i	KX268689	Parrot	Pakistan	Lahore/SFR-129	2015
VII i	KX268690	Parakeet	Pakistan	Lahore/SFR-148A	2015
VII i	KX268691	Parakeet	Pakistan	Lahore/SFR-148B	2015
VII i	KX496962	Pigeon	Pakistan	Lahore/20A 996	2015
VII i	KX496963	Pigeon	Pakistan	Lahore/22A 1001	2015
VII i	KX496964	Pigeon	Pakistan	Lahore/23A 997	2015
VII i	KX496965	Pigeon	Pakistan	Lahore/21A 1084	2015
VII i	KX791183	Parakeet	Pakistan	R-Pindi/SFR-16	2016
VII i	KX791184	Chicken	Pakistan	Lahore/SFR-144A	2016
VII i	KX791185	Chicken	Pakistan	Lahore/SFR-144B	2016
VII i	KX791186	Chicken	Pakistan	Lahore/SFR-144C	2016
VII i	KX791187	Chicken	Pakistan	Lahore/SFR-144D	2016
VII i	KX791188	Chicken	Pakistan	Lahore/SFR-144E	2016
VII i	KY076030	Chicken	Pakistan	Sheikhupura/12A/994	2015
VII i	KY076031	Chicken	Pakistan	Pakistan/995/15A	2015
VII i	KY076032	Chicken	Pakistan	Pakistan/998/26A	2011
VII i	KY076033	Chicken	Pakistan	Badhana/27A/999	2015
VII i	KY076034	Chicken	Pakistan	ChakShahzad/30A/1000	2015
VII i	KY076035	Chicken	Pakistan	1A/1002	2015
VII i	KY076036	Chicken	Pakistan	Buner/KPK/2A/1003	2015
VII i	KY076037	Chicken	Pakistan	Buner/KPK/5A/1004	2015
VII i	KY076038	Chicken	Pakistan	BhaiPhairu/6A/1007	2015
VII i	KY076039	Chicken	Pakistan	Gujranwala/13A/1009	2015
VII i	KY290560	Peacock	Pakistan	Lahore/AW-pck	2015
VII i	KY290561	Pheasant	Pakistan	Lahore/AW-pht	2015
VII i	KY967611.1	Duck	Pakistan	Pakistan/I/UVAS	2016
VII i	KY967612.1	Duck	Pakistan	Pakistan/II/UVAS	2016

*Isolates from current study

Table S2.2. 260nm/280nm, concentration, and RIN value of RNA extracted from formalin-fixed paraffin-embedded tissue samples of 31 field FFPE tissue samples collected from different regions of Pakistan during disease outbreaks in 2015.

SEPRL ID	Tissue	A260/280	RNA concentration (ng/µl)	RIN value
1162-L	Lungs	1.98	32.6	2.6
1162-SI	Small intestines	2.00	52.3	2.4
1163-L	Lungs	2.03	34.7	2.1
1163-B	Brain	2.02	72.0	2.3
1163-SP	Spleen	2.01	44.3	1.9
1164-SP	Spleen	1.99	61.0	2.1
1164-L	Lungs	1.99	68.0	2.1
1164-B	Brain	1.96	71.0	1.9
1165-SP	Spleen	1.96	67.0	2.9
1165-L	Lungs	1.94	51.0	6.5
1165-B	Brain	1.96	32.5	5.1
1166-SP	Spleen	1.99	84.0	2.4
1166-L	Lungs	1.94	58.0	2.9
1166-B	Brain	1.92	21.2	2.5
1168-SP	Spleen	1.98	61.0	2.4
1168-L	Lungs	1.98	20.8	7.7
1168-B	Brain	1.94	55.0	2.7
1168-SI	Small intestines	1.97	71.0	3.7
1170-SP	Spleen	1.99	66.0	2.4
1170-L	Lungs	1.98	70.0	2.8
1170-B	Brain	2.00	66.0	2.1
1170-SI	Small intestines	2.03	63.0	1.7
1171-SP	Spleen	1.90	71.0	2.1
1171-L	Lungs	1.96	15.8	1.8
1172-SP	Spleen	2.02	62.0	2.1
1172-L	Lungs	1.98	26.4	2.6
1172-B	Brain	1.98	43.3	3.6
1173-L	Lungs	1.99	26.8	4.9
1173-SI	Small intestines	1.97	52.0	2.5
1174-L	Lungs	1.90	29.8	2.5
1174-B	Brain	1.99	72.0	2.1

Table S2.3. Complete genome and fusion gene coding sequences coverage, and deduced amino acid cleavage site motif for the NDV sequences obtained from the merged data for each of the studied birds.

Sample ID	Complete genome		Fusion gene		Cleavage site 113-117
	Obtained coding sequence length	**Percentage coverage of coding sequence (13746 bp)**	**Obtained coding sequence length**	**Percentage coverage of coding sequence (1662 bp)**	
1162	13234	96.27	1662	100	RRQKR↓F
1163	12690	92.31	1319	79.36	No sequence
1164	13746	100	1662	100	RRQKR↓F
1165	13746	100	1662	100	RRQKR↓F
1166	13746	100	1662	100	RRQKR↓F
1168	13746	100	1662	100	RRQRR↓F
1170	13746	100	1662	100	RRQKR↓F
1171	6390	45.96	721	43.38	No sequence
1173	13669	99.43	1662	100	RRQKR↓F
1174	13746	100	1662	100	RRQKR↓F

Table S2.4. Evolutionary distance between class II Newcastle disease viruses of genotypes VII estimated using the complete genome coding sequences

Sub-genotypes (number of analyzed sequences)	VII b	VII d	VII e	VII f	VII h	[b]VII i	VII i
VII b (n = 4)							
VII d (n = 4)	0.036						
VII e (n = 4)	0.040	0.036					
VII f (n = 2)	0.054	0.050	0.043				
VII h (n = 4)	0.090	0.087	0.081	0.077			
[b]VII i (n = 6)	0.091	0.088	0.081	0.077	0.090		
VII i (n = 33)	0.090	0.086	0.079	0.075	0.089	0.012	
VII j (n = 4)	0.024	0.043	0.047	0.060	0.095	0.097	0.095

[a] The numbers of base substitutions per site from averaging over all sequence pairs between groups within genotype VII are shown. The analysis involved 77 nucleotide sequences. There were a total of 13746 positions in the final dataset

[b] Sub-genotype VIIi isolates from current study.

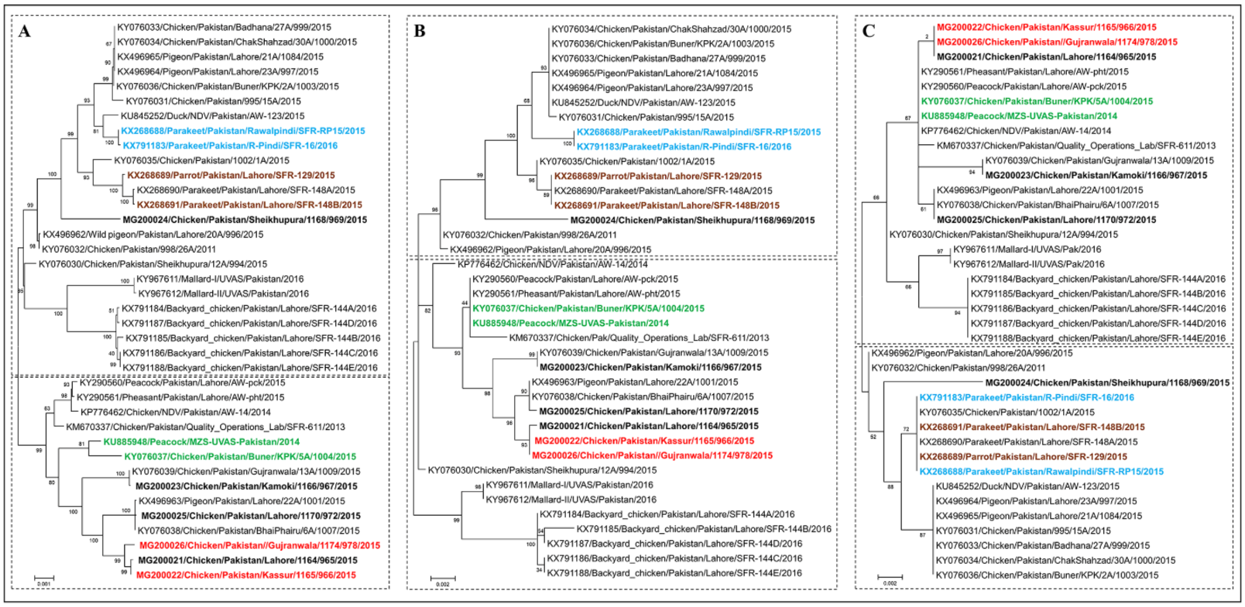

A
KY076033/Chicken/Pakistan/Badhana/27A/999/2015
KY076034/Chicken/Pakistan/ChakShahzad/30A/1000/2015
KX496965/Pigeon/Pakistan/Lahore/21A/1084/2015
KX496964/Pigeon/Pakistan/Lahore/23A/997/2015
KY076036/Chicken/Pakistan/Buner/KPK/2A/1003/2015
KY076031/Chicken/Pakistan/995/15A/2015
KU845252/Duck/NDV/Pakistan/AW-123/2015
KX268688/Parakeet/Pakistan/Rawalpindi/SFR-RP15/2015
KX791183/Parakeet/Pakistan/R-Pindi/SFR-16/2016
KY076035/Chicken/Pakistan/1002/1A/2015
KX268689/Parrot/Pakistan/Lahore/SFR-129/2015
KX268690/Parakeet/Pakistan/Lahore/SFR-148A/2015
KX268691/Parakeet/Pakistan/Lahore/SFR-148B/2015
MG200024/Chicken/Pakistan/Sheikhupura/1168/969/2015
KX496962/Wild pigeon/Pakistan/Lahore/20A/996/2015
KY076032/Chicken/Pakistan/998/26A/2011
KY076030/Chicken/Pakistan/Sheikhupura/12A/994/2015
KY967611/Mallard-I/UVAS/Pakistan/2016
KY967612/Mallard-II/UVAS/Pakistan/2016
KX791184/Backyard_chicken/Pakistan/Lahore/SFR-144A/2016
KX791187/Backyard_chicken/Pakistan/Lahore/SFR-144D/2016
KX791185/Backyard_chicken/Pakistan/Lahore/SFR-144B/2016
KX791186/Backyard_chicken/Pakistan/Lahore/SFR-144C/2016
KX791188/Backyard_chicken/Pakistan/Lahore/SFR-144E/2016
KY290560/Peacock/Pakistan/Lahore/AW-pck/2015
KY290561/Pheasant/Pakistan/Lahore/AW-pht/2015
KP776462/Chicken/NDV/Pakistan/AW-14/2014
KM670337/Chicken/Pakistan/Quality_Operations_Lab/SFR-611/2013
KU885948/Peacock/MZS-UVAS-Pakistan/2014
KY076037/Chicken/Pakistan/Buner/KPK/5A/1004/2015
KY076039/Chicken/Pakistan/Gujranwala/13A/1009/2015
MG200023/Chicken/Pakistan/Kamoki/1166/967/2015
KX496963/Pigeon/Pakistan/Lahore/22A/1001/2015
MG200025/Chicken/Pakistan/Lahore/1170/972/2015
KY076038/Chicken/Pakistan/BhaiPhairu/6A/1007/2015
MG200026/Chicken/Pakistan//Gujranwala/1174/978/2015
MG200021/Chicken/Pakistan/Lahore/1164/965/2015
MG200022/Chicken/Pakistan/Kassur/1165/966/2015
0.001
B
KY076034/Chicken/Pakistan/ChakShahzad/30A/1000/2015
KY076036/Chicken/Pakistan/Buner/KPK/2A/1003/2015
KY076033/Chicken/Pakistan/Badhana/27A/999/2015
KX496965/Pigeon/Pakistan/Lahore/21A/1084/2015
KX496964/Pigeon/Pakistan/Lahore/23A/997/2015
KU845252/Duck/NDV/Pakistan/AW-123/2015
KY076031/Chicken/Pakistan/995/15A/2015
KX268688/Parakeet/Pakistan/Rawalpindi/SFR-RP15/2015
KX791183/Parakeet/Pakistan/R-Pindi/SFR-16/2016
KY076035/Chicken/Pakistan/1002/1A/2015
KX268689/Parrot/Pakistan/Lahore/SFR-129/2015
KX268690/Parakeet/Pakistan/Lahore/SFR-148A/2015
KX268691/Parakeet/Pakistan/Lahore/SFR-148B/2015
MG200024/Chicken/Pakistan/Sheikhupura/1168/969/2015
KY076032/Chicken/Pakistan/998/26A/2011
KX496962/Pigeon/Pakistan/Lahore/20A/996/2015
KP776462/Chicken/NDV/Pakistan/AW-14/2014
KY290560/Peacock/Pakistan/Lahore/AW-pck/2015
KY290561/Pheasant/Pakistan/Lahore/AW-pht/2015
KY076037/Chicken/Pakistan/Buner/KPK/5A/1004/2015
KU885948/Peacock/MZS-UVAS-Pakistan/2014
KM670337/Chicken/Pak/Quality_Operations_Lab/SFR-611/2013
KY076039/Chicken/Pakistan/Gujranwala/13A/1009/2015
MG200023/Chicken/Pakistan/Kamoki/1166/967/2015
KX496963/Pigeon/Pakistan/Lahore/22A/1001/2015
KY076038/Chicken/Pakistan/BhaiPhairu/6A/1007/2015
MG200025/Chicken/Pakistan/Lahore/1170/972/2015
MG200021/Chicken/Pakistan/Lahore/1164/965/2015
MG200022/Chicken/Pakistan/Kassur/1165/966/2015
MG200026/Chicken/Pakistan//Gujranwala/1174/978/2015
KY076030/Chicken/Pakistan/Sheikhupura/12A/994/2015
KY967611/Mallard-I/UVAS/Pakistan/2016
KY967612/Mallard-II/UVAS/Pakistan/2016
KX791184/Backyard_chicken/Pakistan/Lahore/SFR-144A/2016
KX791185/Backyard_chicken/Pakistan/Lahore/SFR-144B/2016
KX791187/Backyard_chicken/Pakistan/Lahore/SFR-144D/2016
KX791186/Backyard_chicken/Pakistan/Lahore/SFR-144C/2016
KX791188/Backyard_chicken/Pakistan/Lahore/SFR-144E/2016
0.002
C
MG200022/Chicken/Pakistan/Kassur/1165/966/2015
MG200026/Chicken/Pakistan//Gujranwala/1174/978/2015
MG200021/Chicken/Pakistan/Lahore/1164/965/2015
KY290561/Pheasant/Pakistan/Lahore/AW-pht/2015
KY290560/Peacock/Pakistan/Lahore/AW-pck/2015
KY076037/Chicken/Pakistan/Buner/KPK/5A/1004/2015
KU885948/Peacock/MZS-UVAS-Pakistan/2014
KP776462/Chicken/NDV/Pakistan/AW-14/2014
KM670337/Chicken/Pakistan/Quality_Operations_Lab/SFR-611/2013
KY076039/Chicken/Pakistan/Gujranwala/13A/1009/2015
MG200023/Chicken/Pakistan/Kamoki/1166/967/2015
KX496963/Pigeon/Pakistan/Lahore/22A/1001/2015
KY076038/Chicken/Pakistan/BhaiPhairu/6A/1007/2015
MG200025/Chicken/Pakistan/Lahore/1170/972/2015
KY076030/Chicken/Pakistan/Sheikhupura/12A/994/2015
KY967611/Mallard-I/UVAS/Pak/2016
KY967612/Mallard-II/UVAS/Pak/2016
KX791184/Backyard_chicken/Pakistan/Lahore/SFR-144A/2016
KX791185/Backyard_chicken/Pakistan/Lahore/SFR-144B/2016
KX791186/Backyard_chicken/Pakistan/Lahore/SFR-144C/2016
KX791187/Backyard_chicken/Pakistan/Lahore/SFR-144D/2016
KX791188/Backyard_chicken/Pakistan/Lahore/SFR-144E/2016
KX496962/Pigeon/Pakistan/Lahore/20A/996/2015
KY076032/Chicken/Pakistan/998/26A/2011
MG200024/Chicken/Pakistan/Sheikhupura/1168/969/2015
KX791183/Parakeet/Pakistan/R-Pindi/SFR-16/2016
KY076035/Chicken/Pakistan/1002/1A/2015
KX268691/Parakeet/Pakistan/Lahore/SFR-148B/2015
KX268690/Parakeet/Pakistan/Lahore/SFR-148A/2015
KX268689/Parrot/Pakistan/Lahore/SFR-129/2015
KX268688/Parakeet/Pakistan/Rawalpindi/SFR-RP15/2015
KU845252/Duck/NDV/Pakistan/AW-123/2015
KX496964/Pigeon/Pakistan/Lahore/23A/997/2015
KX496965/Pigeon/Pakistan/Lahore/21A/1084/2015
KY076031/Chicken/Pakistan/995/15A/2015
KY076033/Chicken/Pakistan/Badhana/27A/999/2015
KY076034/Chicken/Pakistan/ChakShahzad/30A/1000/2015
KY076036/Chicken/Pakistan/Buner/KPK/2A/1003/2015
0.002

Fig. 2.1 Phylogenetic analyses based on the complete genome coding sequences (A), complete F gene coding sequences (B), and partial F gene coding sequences (C) of isolates representing class II sub-genotype VIIi Newcastle disease virus. The evolutionary histories were inferred by using the maximum-likelihood method based on General Time Reversible model with 1000 bootstrap replicates as implemented in MEGA 6 [36, 38]. The analyses involved 38 nucleotide sequences with a total of 13746 (A), 1662 (B), or 375 (C) positions in the final datasets. The sequences obtained in the current study are presented in bold font. Evolutionary analysis was conducted in MEGA6. For all analyses, the codon positions included were 1st+ 2nd+ 3rd+ noncoding, and all positions containing gaps and missing data were eliminated. The GenBank accession numbers are followed by host name, country of isolation, strain designation, and year of isolation. Isolates with bold font in red, brown, green and blue has 100% nucleotide identity between them based on F (Fig 1B) and partial F (Fig 1C) gene trees, were readily differentiated in the complete genome tree and had nucleotide distances ranging from 0.1 to 0.34%. Isolates in bold black font are from current study.

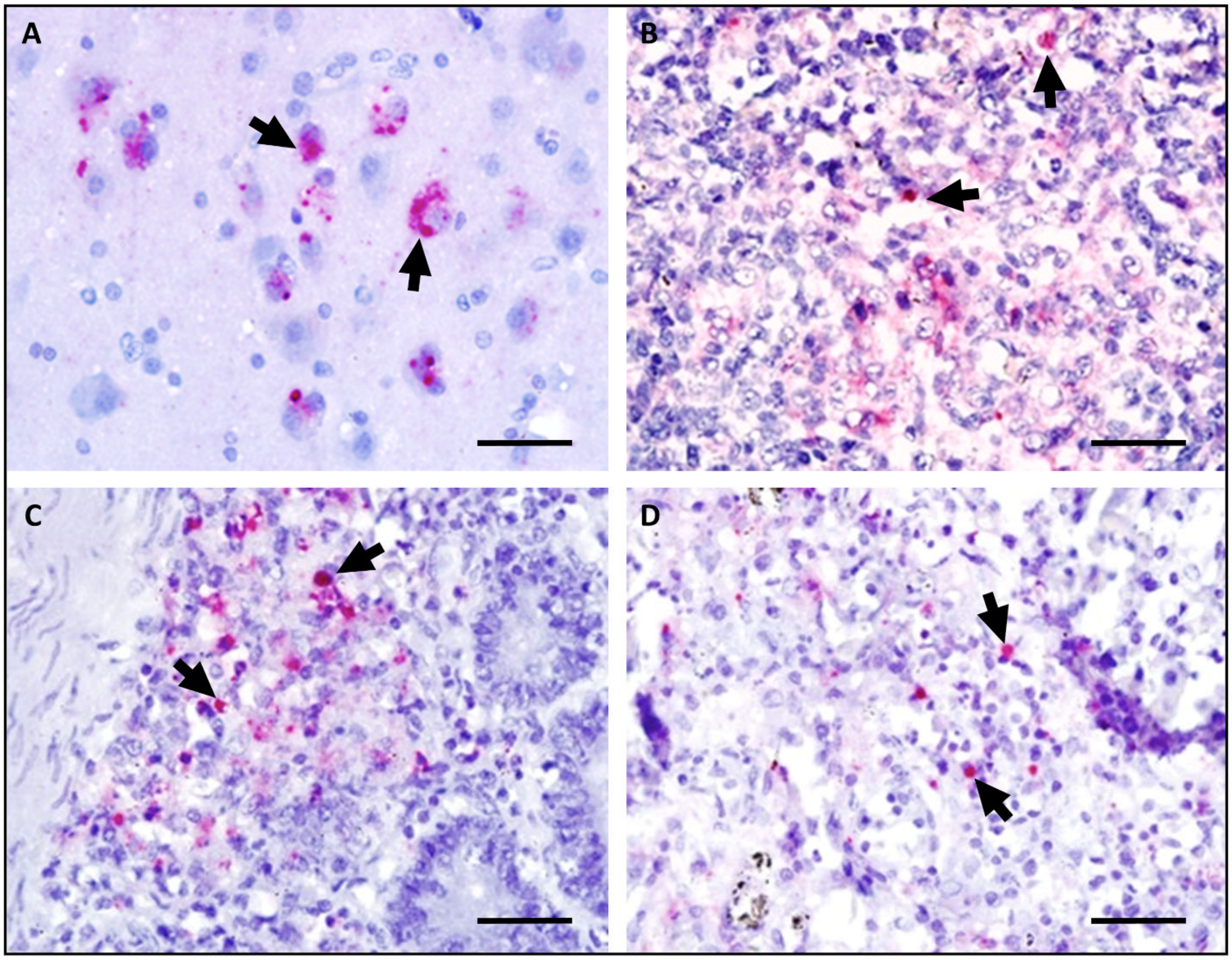


Fig. S2.1. A-D: Photomicrographs demonstrating NDV nucleoprotein immunostaining within formalin-fixed paraffin-embedded tissue sections from chicken. Black arrows showing granular staining of viral nucleoprotein in the cytoplasm of cells. A) Brain. B) Spleen. C) Small intestine. D) Lung. Bar = 50 µm.

CHAPTER 3

RAPID VIRULENCE PREDICTION AND IDENTIFICATION OF NEWCASTLE DISEASE VIRUS GENOTYPES USING THIRD-GENERATION SEQUENCING[2]

[2] Salman L. Butt, Tonya L. Taylor, Jeremy D. Volkening, Kiril M. Dimitrov, Dawn Williams-Coplin, Kevin K. Lahmers, Asif M. Rana, Patti J. Miller, David L. Suarez, Claudio L. Afonso, James B. Stanton. *Virology Journal* 2018 15:179, Reprinted here with permission of publisher.

Abstract

Background: Newcastle disease (ND) outbreaks are global challenges to the poultry industry. Effective management requires rapid identification and virulence prediction of the circulating Newcastle disease viruses (NDV), the causative agent of ND. However, these diagnostics are hindered by the genetic diversity and rapid evolution of NDVs.

Methods: An amplicon sequencing (AmpSeq) workflow for virulence and genotype prediction of NDV samples using a third-generation, real-time DNA sequencing platform is described here. 1D MinION sequencing of barcoded NDV amplicons was performed using 33 egg-grown isolates, (15 NDV genotypes), and 15 clinical swab samples collected from field outbreaks. Assembly-based data analysis was performed in a customized, Galaxy-based AmpSeq workflow. MinION-based results were compared to previously published sequences and to sequences obtained using a previously published Illumina MiSeq workflow.

Results: For all egg-grown isolates, NDV was detected and virulence and genotype were accurately predicted. For clinical samples, NDV was detected in ten of eleven NDV samples. Six of the clinical samples contained two mixed genotypes as determined by MiSeq, of which the MinION method detected both genotypes in four samples. Additionally, testing a dilution series of one NDV isolate resulted in NDV detection in a dilution as low as 101 50% egg infectious dose per milliliter. This was accomplished in as little as 7 minutes of sequencing time, with a 98.37% sequence identity compared to the expected consensus obtained by MiSeq.

Conclusion: The depth of sequencing, fast sequencing capabilities, accuracy of the consensus sequences, and the low cost of multiplexing allowed for effective virulence prediction and genotype identification of NDVs currently circulating worldwide. The sensitivity of this protocol

was preliminary tested using only one genotype. After more extensive evaluation of the sensitivity and specificity, this protocol will likely be applicable to the detection and characterization of NDV.

Background

Newcastle disease (ND) is one of the most important infectious diseases of poultry and is a major economic burden to the global poultry industry. Virulent strains of avian paramyxovirus 1 (APMV-1), commonly known as Newcastle disease virus (NDV) [1], are the cause of ND and have been recently reclassified as avian avulavirus-1 (AAvV-1) [2]. Newcastle disease viruses are a highly diverse group of viruses with two distinct classes, 19 accepted genotypes and a wide host range including domestic and wild bird species. In addition to the genotypic diversity of NDVs, these viruses are also diverse in their virulence. This includes low virulent viruses, whose replication is limited to the respiratory and digestive tracts and typically cause clinically inapparent infections, to highly virulent viruses that cause acute disease with high mortality rates [1, 3]. The global spread, constant evolution, varying virulence, and the wide host range of NDV are challenges to the control of ND [4].

Effective control of ND is dependent on specific diagnostic testing, which is typically oriented towards detection, genotyping, or prediction of virulence. Virulence of NDV is best assayed through *in vivo* pathogenicity studies [5], but due to the cost and time constraints associated with such methods, reverse transcriptase-quantitative PCR (RT-qPCR) and sequencing of the F gene cleavage site are used to predict NDV virulence [6, 7]. Genotyping of NDV is commonly achieved through sequencing of the coding sequence of the fusion gene [8], which also allows for prediction of virulence. Preliminary genotyping can be accomplished through partial fusion gene sequencing (i.e., variable region) [9]. For detection of NDV, PCR assays avoid the highly variable fusion gene and instead target more conservative regions of the genome (i.e., matrix and polymerase genes) [10-14]; however, while this increases the applicability of these assays across genotypes, these assays lack applicability for virulence and genotypic determination.

For example, while fusion-based assays can be used for detection [10, 15], the variability of this region, which makes it useful for genotyping, hinders the universal applicability of any single primer set [11, 12] and often requires screening samples with a different PCR assay prior to pathotyping [15]. Furthermore, for most current methods, detection, genotyping, and virulence prediction rely on Sanger sequencing; thus, they lack multiplexing capability and have limited sequencing depth, which complicates detection of mixed infections. In summary, there is a need to develop a method that will sensitively and rapidly detect NDV from multiple genotypes, while also providing genotype and virulence predictions.

Rapid advances in nucleic acid sequencing have led to different sequencing platforms [16, 17] being widely applied for identification of novel viruses [18], whole genome sequencing [19], transcriptomics, and metagenomics [20, 21]. However, high capital investments and relatively long turnaround times limit the widespread use of these next-generation sequencing (NGS) platforms, especially in developing countries [22]. Recent improvements in third-generation sequencing, including those introduced by Oxford Nanopore Technologies (ONT) [23], increase the utility of high-throughput sequencing as a useful tool for surveillance and pathogen characterization [24]. Among the transformative advantages of ONT's sequencing technology are the ability to perform real-time sequence analysis with a short turnaround time [25], the portability of the MinION device, the low startup cost compared to other high-throughput platforms, and the ability to sequence up to several thousand bases from individual RNA or DNA molecules. The MinION device has been successfully used to evaluate antibiotic resistance genes from several bacterial species [26, 27], obtain complete viral genome sequences of an influenza virus [28] and Ebola virus [29], and detect partial viral genome sequences (e.g., Zika virus [30] and poxviruses [25]) by sequencing PCR amplicons (AmpSeq). The MinION, therefore, represents an opportunity to take infectious disease diagnostics a step further and to perform rapid identification and genetic characterization of infectious agents at a lower cost.

As with any deep sequencing platform, the sequence analysis approach is integral for accurate interpretation. Primarily, two approaches for taxonomic profiling of microbial sequencing data have been employed: read-based and *de novo* assembly-based classifications. Read-based metagenomic classification software has been used for identification of microbial species from high-throughput sequencing data [23, 31-33]. Although the sequencing accuracy of the MinION is improving, the raw single-read error rate of nearly 10% [34] may limit the accuracy of this approach for Nanopore data [31], especially when attempting to subspecies level differentiation. *De novo* approaches that use quality-based filtering and clustering of reads [35], or use consensus-based error correction of Nanopore sequencing reads have been reported [36]; however, these are not optimized for amplicon sequencing data.

In this study, a specific and rapid protocol, using the MinION sequencer, was developed to detect representative isolates from all currently circulating (excluding the Madagascar-limited genotype XI) genotypes of NDV. This protocol was also tested on 15 clinical swab samples collected from chickens during disease outbreaks. Additionally, a Galaxy-based, *de novo* AmpSeq workflow is presented that results in accurate final consensus sequences allowing for accurate genotype and virulence prediction. This study represents the first step towards developing AmpSeq as a diagnostic tool for NDV.

Methods

Viruses and clinical samples

Thirty-three NDV isolates, representing 15 different genotypes of different virulence, and ten other avian avulaviruses (AAvV 2-10 and AAvV-13) from the Southeast Poultry Research Laboratory (SEPRL) repository, were propagated in 9–11-day-old specific pathogen free (SPF) eggs [36] and the harvested allantoic fluids were used in this study. Additionally, 15 oral and cloacal swab samples collected from chickens during disease outbreaks in Pakistan in 2015 were collected, and the resulting swab fluid was shipped on dry ice, and then stored at -80 °C. RNA was extracted as described below for both egg-grown and clinical swab samples. The background information of the egg-grown isolates and the clinical samples is summarized in Table S3.1 and Table S3.2, respectively.

RNA extraction

Total RNA from each sample was extracted from infectious allantoic fluids or directly from clinical swab media using TRIzol LS (Thermo Fisher Scientific, USA) following the manufacturer's instructions. RNA concentrations were determined by using Qubit® RNA HS Assay Kit on a Qubit® fluorometer 3.0 (Thermo Fisher Scientific, USA).

Amplicon synthesis and MinION library preparation

Approximately 20 ng (in 5 µl) of RNA was reverse transcribed, and cDNA was amplified with target-specific primers using the SuperScript™ III One-Step RT-PCR System (Thermo Fisher Scientific, USA). Previously published primers (4331F and 5090R) [9, 38] were used in this protocol to target NDV; however, the primers were tailed with universal adapter sequence of 22 nucleotides (in bold font) to allow PCR-based barcoding: 4331F Tailed: 5′-**TTTCTGTTGGTGCTGATATTGC**GAGGTTACCTCYACYAAGCTRGAGA-3′; 5090R Tailed: 5′-**ACTTGCCTGTCGCTCTATCTTC**TCATTAACAAAYTGCTGCATCTTCCCWAC-3′). The thermocycler conditions for the reaction were as follows: 50 °C for 30 minutes; 94 °C for 2 minutes; 40 cycles of 94 °C for 15 seconds, 56 °C for 30 seconds, and 68 °C for 60 seconds, followed by 68 °C for 5 minutes. The reaction amplified a 788 base pair (bp) NDV product (832 bp including primer tails) for all genotypes, which included 173 bp of the 3′ region of the end of the M gene and 615 bp of the 5′ end of the F gene (sizes and primer locations based on the Genotype V strain). Amplified DNA was purified by Agencourt AMPure XP beads (Beckman Coulter, USA) at 1.6:1 volumetric bead-to-DNA ratio and quantified using the dsDNA High Sensitivity Assay kit on a Qubit® fluorometer 3.0. MinION-compatible DNA libraries were prepared with approximately 1 µg of barcoded DNA in a total volume of 45 µL using nuclease-free water and using the 1D PCR Barcoding Amplicon Kit (Oxford Nanopore Technologies, UK) in conjunction with the Ligation Sequencing Kit 1D (SQK-LSK108) [23] as per manufacturer's instructions. Briefly, each of the amplicons were diluted to 0.5 nM for barcoding and amplified using LongAmp Taq 2X Master Mix (New England Biolabs, USA) with the following conditions; 95 °C for 3 min; 15 cycles of 95 °C for 15 seconds; 62 °C for 15 seconds, 65 °C for 50 seconds, followed by 65 °C for 50 seconds. The barcoded amplicons were bead purified, pooled into a single tube, end prepped, dA tailed, bead purified, and ligated to the sequencing adapters per manufacturer's instructions. Final DNA libraries were bead purified and stored frozen until used for sequencing.

Comparison of AmpSeq protocol to RT-qPCR assay

For comparison of this MinION-based protocol with the matrix gene reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) assay [10], both methods were used on a dilution series from a single isolate. NDV (LaSota strain) from the SEPRL repository was cultured in SPF 9–11-days-old eggs and the harvested allantoic fluids were diluted to titers ranging from 10^6 to 10^1 EID_{50}/mL in brain-heart infusion broth. RNA was extracted from dilutions, and DNA libraries were prepared following the same protocols as described above. Amplicons from each of the dilutions were barcoded separately. At the pooling step, equal concentrations of barcoded amplicons from different dilutions of LaSota were pooled together in single tube. Dilutions, extractions, library construction, and sequencing were performed twice (run 1 and run 2).

The same extracted RNA was also used as the input into the RT-qPCR using the AgPath-ID one-step RT-PCR Kit (Ambion, USA) on the ABI 7500 Fast Real-Time PCR system following the previously described protocols [10].

Sequencing by MinION

The libraries were sequenced with the MinION Nanopore sequencer [23]. A new FLO-MIN106 R9.4 flow cell, stored at 4°C prior to use, was allowed to equilibrate to room temperature for 10 minutes before priming it for sequencing. The flow cell was primed with running buffer as per manufacturer's instructions. The pooled DNA libraries were prepared by combining 12 µL of the libraries with 2.5 µL nuclease-free water, 35 µL RBF, and 25.5 µL library loading beads. After the MinION Platform QC run, the DNA library was loaded into the MinION flow cell via the SpotON port. The standard 48-h 1D sequencing protocol was initiated using the MinKNOW software v.5.12. Detailed information for all MinION runs in this study is provided in Table S3.3.

The complete steps from RNA isolation to MinION sequencing were performed twice for egg-grown viruses. One run consisted of six egg-grown isolates from different genotypes representative of vaccine and virulent NDV strains (run 3: 6-sample pool). The other run consisted of these same six viruses and an additional 27 egg-grown NDV isolates (run 4: 33-sample pool). The clinical samples (n = 15) were processed in runs 5, 6, and 7. A variable number of samples were pooled in these three sequencing runs to cluster libraries with similar concentrations.

To determine the accuracy of consensus sequences at different sequencing time points for accurate identification of the NDV genotypes, the raw data (FAST5 files) obtained from the 10-

fold serial dilution experiment (see above) were analyzed in subgroups based on time of acquisition and processed through the AmpSeq workflow as described below.

Development of MinION data analysis workflow

To analyze the Nanopore sequencing data, a custom, assembly-based AmpSeq workflow within the Galaxy platform interface [39] was developed, as diagrammed in Figure 3.1. The MinION raw reads in FAST5 format were archived (tar format) and uploaded into Galaxy workflow. The reads were base-called using the Albacore v2.02 (ONT). The NanoporeQC tool v0.001 (available in the Galaxy testing toolshed) was used to visualize read quality based on the summary table produced by Albacore. Porechop v0.2.2 [40] was used to demultiplex reads for each of the barcodes and trim the adapters at the ends of the reads by using default settings. Short reads (cutoff = 600 bp) were filtered out and the remaining reads were used as input to the in-house LAclust v0.002. LAclust performs single-linkage clustering of noisy reads based on alignment identity and length cutoffs from DALIGNER pairwise alignments [41] (minimum alignment coverage = 0.90, maximum identity difference = 0.35; minimum number of reads to save cluster = 5; maximum reads saved per cluster = 200, minimum read length = 600 bp; rank mode = number of intracluster linkages; randomized input read order = yes). Read clusters generated by LAclust were then aligned using the in-house Amplicon aligner v0.001 to generate a consensus sequence. This tool optionally subsamples reads (target depth used = 100), re-orients them as necessary, aligns them using Multiple Alignment using Fast Fourier Transform (MAFFT) [42] with highly relaxed gap opening and extension penalties, and calls a majority consensus. Next, each consensus was used as a reference sequence for mapping the full unfiltered read clusters from LAclust with BWA-MEM and ONT2D settings [43, 44]. The final consensus sequence for each sample was refined by using Nanopolish v0.8.5 [45], which calculates an improved consensus using the read alignments and raw signal information from the original FAST5 files. After manually trimming primer sequences from both 3′ (25 bp) and 5′ (29 bp) ends, the obtained consensus sequences (734 bp) were BLAST searched against NDV customized database, which consisted NCBI's nucleotide (nt) database and internal unpublished NDV sequences (NCBI database updated on May 23, 2018).

Sequencing by MiSeq

For comparison between nucleotide sequences obtained from MinION and MiSeq (a high accuracy sequencing platform), 24 NDV isolates from the SEPRL repository that were used for MinION sequencing (representing each currently circulating genotype except XI and multiple sub-genotypes of NDV) and 15 clinical swab samples (allantoic fluid of cultured swab samples) were processed for target-independent NGS sequencing. Briefly, paired-end random sequencing was conducted from cDNA libraries prepared from total RNA using KAPA Stranded RNA-Seq kit (KAPA Biosystems, USA) as per manufacturer's instructions and as previously described [46]. All libraries for NGS were loaded into the 300-cycle MiSeq Reagent Kit v2 (Illumina, USA) and pair-end sequencing (2 × 150 bp) was performed on the Illumina MiSeq instrument (Illumina, USA). Pre-processing and de-novo assembly of the raw sequencing data was completed within the Galaxy platform using a previously described approach [19].

Phylogenetic analysis

The assembled consensus sequences from different NDV genotypes and sub-genotypes (6 sequences from MinION run 3, 33 sequences from run 4 and 24 sequences from MiSeq; a total of 62 sequences) and selected (minimum of one sequence from each genotype/subgenotype) sequences from GenBank (n = 66) were aligned using ClustalW [47] in MEGA6 [48]. Determination of the best-fit substitution model was performed using MEGA6, and the goodness-of-fit for each model was measured by corrected Akaike information criterion (AICc) and

Bayesian information criterion (BIC) [48]. The final tree was constructed using the maximum-likelihood method based on the General Time Reversible model as implemented in MEGA6, with 500 bootstrap replicates [49]. The available GenBank accession number for each sequence in the phylogenetic tree is followed by the, host name, country of isolation, strain designation, and year of isolation.

Comparison of MinION and MiSeq sequence accuracy

To assess the accuracy of the MinION AmpSeq consensus sequences, 24 samples were sequenced by both deep-sequencing methods (MinION and MiSeq) described above. Pairwise nucleotide comparison between MinION and MiSeq was conducted using the Maximum Composite Likelihood model [50]. The variation rate among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 54 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 691 positions in the final dataset. The evolutionary distances were inferred by pairwise analysis using the MEGA6 [48].

Results

Comparison to the Matrix gene RT-qPCR assay

Six, sequential, 10-fold dilutions (from 10^6 EID_{50}/ml to 10^1 EID_{50}/ml) from one NDV isolate (LaSota) were used to compare the ability of AmpSeq and RT-qPCR to detect low quantities of NDV. In each of the six dilutions, AmpSeq and the matrix RT-qPCR detected NDV in all dilutions. AmpSeq resulted in 99.04–100.0% sequence identity to the LaSota isolate across all six dilutions in the first experiment (run 1) and 99.86–100.0% identity in the second experiment (run 2) (Table 3.1).

Time for data acquisition and analysis

To determine the minimal sequencing time needed for acquisition of accurate full-length amplicon consensus sequences at different serial dilutions, 28,000 reads, which were obtained within the first 19 minutes of sequencing in the first serial dilution experiment (run 1), were analyzed. For all concentrations, the first read that aligned to the reference LaSota sequence was obtained within 5 minutes after the sequencing run started. To obtain consensus sequences (5 reads required to build a consensus sequence) only 5 minutes of sequencing time were required for concentrations 10^6–10^3 EID_{50}/ml, which resulted in 99.18–100% sequence identity to the reference LaSota strain. Seven minutes were required to obtain NDV consensus sequences for the two lower concentrations: 10^1 EID50/ml = 8 reads, 98.77% identity and 10^2 EID_{50}/ml = 5 reads, 98.37% identity (Table 3.2). After as little as 10 minutes of sequencing, the identity to the reference sequence was above 99% for even in the most dilute sample.

PCR specificity and range of reactivity for NDV genotypes

To determine the utility of the primers for the currently circulating NDV genotypes and the potential cross-reactivity for other AAvVs, which are relatively nonpathogenic in poultry but can confound diagnosis of NDV [50], total RNA from 43 AAvVs, including 23 AAvV-1 genotypes and sub-genotypes (15 different NDV genotypes, 8 different subgenotypes), as well as AAvV-2–10 and -13 (n = 10) were tested. All AAvV-1 genotypes that are currently circulating globally were amplified with tailed primers; samples 19 and 36 had weak bands of the desired molecular weight (i.e., 832 bp) compared to other lanes. Two bands larger than 800 bp were visible on the electrophoresis of samples #19, #20, #21, #31 and #32 (see Figure S1 legend for interpretation of this result). All non-AAvV-1 viruses failed to produce bands visible by gel electrophoresis (Figure S3.1)

Quality metrics

The Nanopore QC tool was used to obtain quality metrics plots of all sequencing runs. For MinION runs 3 (6-sample pool) and 4 (33-sample pool), more than 70% of total reads had a quality score greater than ten (Q10 score = 90% accuracy) (Figure 3.2 A and B). The average overall mean read quality scores in both runs were comparable (run 3 = 10.7, run 4 = 11.0), and the mean quality scores of reads ≥ 10 (mean $Q_{\geq 10}$) were similar (11.8) for both runs (Figure 3.2 C and D). In addition, analysis of five consecutive batches of reads (each batch = 20,000 reads) obtained at different time intervals from run 4 indicated that the overall mean read quality for each 20,000 read batch remained above 10 (Table S3.4). Similarly, the mean $Q_{\geq 10}$ over time remained consistent in the clinical sample runs (runs 5–7), which had long (12 hrs) sequencing runs (Figure 3.3 A, B and C, blue lines).

Sub-genotypic resolution of AAvV-1 viruses with MinION sequencing

To determine the capability to effectively detect and differentiate viruses of different genotypes and sub-genotypes, PCR amplicons from 33 egg-grown isolates, which were representative of 23 different NDV genotypes and sub-genotypes, were barcoded, pooled, and sequenced in a single 12-hour MinION run (run = 4) generating a total of 2.076 million reads. The first 100,000 reads, which were obtained in 3 hours and 10 minutes, were analyzed for identification of all 33 NDV isolates used in the study. All 33 NDV isolates were correctly identified to the sub-genotype level (Table 3.3), with 97.82%–100% sequence identity. Thirty-one of thirty-three samples were greater than 99% identical to the expected genotype in each of the sample, with 22 of 33 having 100% sequence identity. Samples with higher sequence identity represent those isolates whose sequences (Sanger or MiSeq based) had already been deposited in GenBank, while those samples with lower sequence identity lacked replicate sequences from those particular isolates. For sample #37, MiSeq detected genotypes XIIIb and VIc, but the AmpSeq workflow only detected genotype XIIIb (e.g., see pairwise comparison section for further demonstration of this protocol's accuracy).

While 832 bp was the expected amplicon size, genotypes III, IV, and IX (all previously untested genotypes with this primer set) yielded an unexpected electrophoresis product of ~1000 bp (see above). The analysis of sequences obtained from these NDV isolates revealed that in addition to the 788 bp adapter-trimmed consensus sequence, an upstream region of NDV genome was amplified, resulting in an 1067 bp adapter-trimmed consensus sequence that contains the targeted NDV sequence.

Clinical swab samples from chicken

To assess the potential utility of this protocol on field samples from disease outbreaks, MinION libraries were generated directly from clinical swab samples. These swab samples were also propagated in eggs and the allantoic fluid was sequenced using a MiSeq-based workflow (runs 5, 6, and 7) to compare to the MinION results. Out of 11 NDV-positive samples with the MiSeq method, 10 samples were NDV positive by the MinION protocol (Sample #52 being the exception) (Table 3.4). In the six NDV-positive samples that contained one NDV genotype, as detected by the MiSeq method, the same NDV genotype was also detected with the MinION protocol. The MiSeq method detected two genotypes in samples #45, #46, #47, and #49; whereas, the MinION protocol only detected dual genotypes in samples #45 and #46. In sample #48, only one NDV genotype was detected by MiSeq but two NDV genotypes were detected by the MinION protocol. All 4 samples negative by MiSeq were also negative by the MinION protocol.

Pairwise comparison of replicated MinION sequences and MiSeq sequences

Pairwise nucleotide distance analysis was used to compare the consensus sequences in six samples across two separate MinION runs. There was no variation in the consensus sequence between the MinION runs across those six samples. Pairwise nucleotide distance analysis was also used to compare the MinION consensus sequence to the MiSeq consensus sequence in 24 isolates (one isolate representing each genotype and subgenotype; 24 samples with asterisks in Table 3 were used for pairwise nucleotide comparison). The MinION and MiSeq consensus sequences were 100% identical, except in four samples (#20, #25, #36, and #37), in which the percent identity was 99.18%–99.86% (nota bene: the samples in Table 3.3 without asterisks did not have a second sequence directly from that stock for comparison; thus, the percent identity may be low due to the exact isolate not having a representative sequence in GenBank, e.g., sample 21 with 99.05% similarity). In addition, there were no differences at the fusion gene cleavage site between AmpSeq and either previous Sanger or previous MiSeq results (Tables 3.3 and 3.4, last column). Collectively, these results demonstrate the repeatable high accuracy of the MinION-AmpSeq method.

Phylogeny of NDV genotypes

To confirm the ability of the MinION-acquired partial matrix and fusion gene sequences to be used for accurate analysis of evolutionary relatedness, phylogenetic analysis using consensus sequences (734 bp; trimmed of adapter and primer sequences) obtained from two independent MinION runs (run 3 and 4) was performed. Additionally, the 24 sequences from MiSeq were also included in the phylogenetic tree (Figure S3.2, to further illustrate the agreement between these two sequencing methods. In the phylogenetic tree, the isolates (n = 33; green font) grouped together with the viruses that showed highest nucleotide sequence identity to them, including those in which MiSeq sequences were available (red font). The six isolates that were sequenced twice (blue font) clustered together. Taken together, the results demonstrated that all sequences clustered to the expected genotype/sub-genotype branch of the phylogenetic tree.

Time and cost estimation

The time of sample processing and cost estimation of reagents to multiplex and sequence samples (n = 6; n = 33) from RNA extraction to obtain final consensus sequences is presented in Table S3.5. From RNA extraction to final consensus sequence calculations, the average time (including sequencing time) to process six samples was approximately 9–10 person-hours and for 33 samples approximately 26 person-hours. Assuming that flow cell can be used multiple times (twice when 33 samples pooled and five times when six samples pooled to prepare one cDNA library) for sequencing, cost per sequencing run and cost per sample were estimated. The cost per sample decreased from $53 (six samples multiplexed) to $31 (33 samples multiplexed).

Discussion

This study describes the development of a single protocol for rapid and accurate detection, virulence determination, and preliminary genotype identification (with sub-genotype resolution) of NDV utilizing the low-cost MinION sequencer. Additionally, an assembly-based sequence analysis workflow for MinION amplicon sequencing data was developed. This MinION AmpSeq workflow detected all currently circulating genotypes when using egg-grown viruses. Furthermore, clinical swab samples were used to demonstrate proof of concept that such samples contained sufficient NDV nucleic acid for detection, and interestingly AmpSeq detected 2 NDV genotypes (vaccine and virulent strains) in several clinical swab samples. These capabilities suggest this protocol may be useful for research and ancillary diagnostic procedures and indicates that further development and validation of NDV AmpSeq would be useful, especially in developing countries where NDV is endemic and there is a need for affordable epidemiological surveillance to track reservoirs and disease outbreaks.

The sequence heterogeneity among AAvV-1 genomes, which hinders the ability to develop a single test that sensitively detects NDV while also predicting the genotypic classification and virulence, is well known [4, 51, 52]. Currently, an RT-qPCR targeting the M gene [10] is most sensitive and is used for screening samples, but this assay only provides positive and negative results of the samples. RT-qPCRs that predict virulence based on the fusion gene are available [10, 54]; however, the lower sensitivity of these assays and the inability of at least one of these assays to detect viruses of all genotypes (e.g., genotypes Va and VI) [10, 12, 13] complicate diagnostic interpretation when the matrix and fusion tests have conflicting results. Thus, the only truly reliable option to detect a broad range of viruses and to determine virulence from some strains is to design multiple tests that include genotype-specific primers and probes [7, 12, 53]. Recently, Miller *et al* reported that the primer set used in this study detected Class I and all nine of the tested class II genotypes [38]; however, this primer set was not tested against other currently circulating genotypes. The current study includes six additional genotypes, collectively representing all currently circulating genotypes (excluding the Madagascar-limited genotype XI). Furthermore, the ability to use AmpSeq as the final measure of a PCR allows for larger amplicon sizes as compared to RT-qPCR. As such, there will be one less restriction on primer site design when trying to create a pan-NDV primer set. Work is in progress to utilize the ability of MinION to sequence longer amplicon fragments, which will provide more complete phylogenetic information. After optimizing pan-NDV primer design for AmpSeq, sensitivity of pan-NDV AmpSeq will need to be further evaluated.

Additionally, while the preliminary analytical sensitivity of this protocol was determined using only one NDV genotype, the sensitivity of the MinION AmpSeq was comparable to the matrix RT-qPCR test, which does not allow inference of virulence. Further testing of the NDV

AmpSeq sensitivity to current virulence-predicting RT-qPCR tests are warranted [14, 15]; however, even these tests lack the genotyping capability of AmpSeq. The ability of this AmpSeq method to detect different genotypes of NDV was further aided by barcoding PCR, which adds another round of PCR to the assay and the ability to adjust the concentration of samples during the library preparation phase. This latter step allows for more volume of low concentration (i.e., weak positives) samples to be added to the library pool. While the additional steps for library synthesis provide these advantages, they also add time to the assay (see below for further discussion of time efficiency). However, the benefit of implementing detection, genotype prediction, and virulence prediction into a single test adds value to this assay.

While the multifaceted nature of this MinION AmpSeq protocol is an advantage, time and cost efficiency must be maintained for it to be useful. MinION is inherently rapid due to the real-time nature of the sequencing. For example, this method identified the correct NDV genotype in all serial dilutions, with an accuracy of 98.37–100%, after only 7 minutes of sequencing. Because the MinION provides real-time sequence data, it is possible to monitor the sequencing run to determine the optimal run length for each library. Additionally, samples can be multiplexed into a single sequencing run, which reduces time and cost [55]. Recently, multiplexing and MinION sequencing of the PCR products from a panel of 5 samples was reported [55]. Here a panel of 33 samples was multiplexed while maintaining successful NDV genotyping from data collected within 3 hrs and 10 minutes of sequencing and without affecting mean read quality and percentage of high-quality reads. Thus, this protocol provides the flexibility to rapidly and economically obtain accurate sequence data for a preliminary genotyping and virulence prediction.

While Nanopore sequencing has numerous benefits, the high error rate poses unique challenges to data analysis. Thus, it is important to extract accurate consensus sequences from raw sequencing data [56]. As previously discussed, pathogen typing from sequencing data can be done with read count-based profiling or *de novo* assembly approaches [31]. However, there are a limited number of available tools suitable for handling the noisy reads currently produced by the MinION platform. The approach in this study takes advantage of the fact that single MinION reads often represent full-length amplicon sequences. By clustering full-length reads based on pairwise identity and subsequently performing consensus calling using standard multiple alignment software, this method quickly and reliably generates accurate (consistently greater than 99% sequence identity to paired MiSeq) *de novo* assemblies from amplicon datasets using as few as twenty reads per amplicon, and correct genotypic prediction with as few as five reads per amplicon. Thus, this approach overcomes the inherently high error rate (~90% accuracy) of Nanopore sequencing [57] and sequence identification and differentiation at the sub-genotype level can be highly reliable.

One known source of error in Nanopore sequencing is that 5-mers of A and T in the individual reads are difficult to identify accurately with MinION sequencing [58]. Importantly, a 5-mer run of a single base is not present in the cleavage site of the Fusion gene, however, two positions on the consensus sequence where there are 5-mers of A and C are present. Because of 2 instances in which only 4 nucleotides were read (one two nucleotide gap at 153–57 bp position and second on 233 bp) on 5-mer site, sample #37 had less than 100% identity to the respective Miseq data. It should be noted that this type of system error can be easily detected and manually corrected for a paramyxovirus (including NDV), which are viruses that do not tolerate single nucleotide deletions or insertions (rule of six) [59]. Because of the consensus-based approach, despite the relatively high, read-based error rate, the cleavage site was accurately determined.

Because this protocol relies on identity-based clustering prior to assembly, it maintains the ability to detect samples with mixed NDV genotypes. For example, in this study four clinical samples had two different genotypes as detected by MiSeq analysis, two of which were correctly identified by the MinION AmpSeq workflow. In a fifth case, a mixed sample was detected by MinION AmpSeq, but not by MiSeq. A potential explanation for these differences could be that the MiSeq sequencing was performed on egg-amplified samples, which may have altered the relative levels of the two genotypes, as compared to the direct clinical swab sample used for MinION sequencing. Additionally, the differences in molecular techniques (i.e., MinION: targeted; MiSeq: random) may have altered the relative abundance of the genotypes within the sequencing libraries. While further studies into the ability of this workflow to sensitively detect and differentiate NDV in samples with more than one genotype are ongoing, rapid NDV genotyping from clinical samples without culturing the virus in SPF eggs has the potential to facilitate disease diagnostics.

Conclusions

Taken together, this protocol reliably detected, genotyped, and predicted the virulence of NDV using laboratory stocks of all genetic variants currently circulating worldwide. Furthermore, preliminary testing of clinical-based samples suggests its feasibility using clinical swab samples. This assay can be used for research purposes and as an ancillary test in field investigations; however, further testing, including sensitivity validation on clinical samples and testing the effect of multiple isolates on sensitivity are warranted. Furthermore, the advantages of MinION AmpSeq allow for further optimization not possible with other techniques. For example, PCR product length is less of a restriction with MinION AmpSeq as compared to RT-qPCR. Overall, MinION AmpSeq improves the depth of information obtained from PCRs and allows for more flexibility in assay design, which can be broadly applied to the detection and characterization of numerous infectious agents.

Abbreviations

ND: Newcastle disease, NDV: Newcastle disease viruses, **AmpSeq:** amplicon sequencing, **EID50:** 50% egg infectious dose, **APMV-1**: avian paramyxovirus 1, **AAvV:** avian avulavirus, **RT-qPCR:** reverse transcriptase-quantitative polymerase chain reaction, **ONT:** Oxford Nanopore Technologies, **SEPRL:** Southeast Poultry Research Laboratory, **SPF:** specific pathogen free, **MAFFT:** Multiple Alignment using Fast Fourier Transform, **AICc:** Akaike information criterion, **BIC:** Bayesian information criterion.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Availability of data and materials

The sequences obtained in the current study were submitted to GenBank and are available under the accession numbers from MH392212 to MH392228.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was supported by USDA National Institute of Food and Agriculture Grant no. 2018-67015-28306, USDA CRIS 6040-32000-072, DTRA grant 58.0210.5.006 and partially funded by

BAA project [grant#685/FRCALL 12-6-2-0015] and University of Georgia Research Foundation, Inc., Faculty Research Grants Program.

Authors' contributions

S. L. Butt extracted RNA from egg-grown isolates and clinical samples, created the MinION libraries, analyzed the MinION data, conducted the phylogenetic analyses, and wrote the manuscript. T. L. Taylor helped with the RT-qPCR and in creation and sequencing of the MinION libraries. J. D. Volkening developed the MinION data analysis workflow and assisted with manuscript preparation. K. M. Dimitrov contributed to the preparation and analysis of sequencing data, the phylogenetic analyses, and manuscript preparation. D. Williams-Coplin prepared NGS libraries. K. K. Lahmers assisted in data analysis and manuscript preparation, A. M. Rana provided clinical swab samples. D. L. Suarez assisted in data interpretation and manuscript preparation. C. L. Afonso and J. B. Stanton were involved in the design of the study, data analysis, data interpretation, and writing of the manuscript. All authors were involved with editing the manuscript.

Acknowledgments

We gratefully acknowledge Tim Olivier for technical assistance. Special thanks to the **Fulbright U.S. Student Program** for sponsoring Dr. Salman's PhD research work.

References

1. Miller PJ, Koch G. Newcastle disease. Diseases of Poultry, 13th ed(Swayne, DE, Glisson, JR, McDougald, LR, Nolan, LK, Suarez, DL and Nair, VL eds), John Wilkey and Sons, Inc, Ames. 2013:89-107.
2. Amarasinghe GK, Ceballos NGA, Banyard AC, Basler CF, Bavari S, Bennett AJ, et al. Taxonomy of the order mononegavirales: Update 2018. Arch Virol. 2018:1-12.
3. Nagai Y, Klenk H-D, Rott R. Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of newcastle disease virus. Virology. 1976; 72:494-508.
4. Dimitrov KM, Ramey AM, Qiu X, Bahl J, Afonso CL. Temporal, geographic, and host distribution of avian paramyxovirus 1 (newcastle disease virus). Infect Genet Evol. 2016; 39:22-34.
5. Commission IOoEBS, Committee IOoEI. Manual of diagnostic tests and vaccines for terrestrial animals: Mammals, birds and bees: Office international des épizooties; 2008.
6. Aldous E, Mynn J, Banks J, Alexander D. 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. 2003; 32:237-55.
7. Kim LM, King DJ, Guzman H, Tesh RB, da Rosa APT, Bueno R, et al. Biological and phylogenetic characterization of pigeon paramyxovirus serotype 1 circulating in wild north american pigeons and doves. J Clin Microbiol. 2008; 46:3303-10.
8. Diel DG, da Silva LH, Liu H, Wang Z, Miller PJ, Afonso CL. Genetic diversity of avian paramyxovirus type 1: Proposal for a unified nomenclature and classification system of newcastle disease virus genotypes. Infect Genet Evol. 2012; 12:1770-79.

9. Kim LM, King DJ, Suarez DL, Wong CW, Afonso CL. Characterization of class i newcastle disease virus isolates from hong kong live bird markets and detection using real-time reverse transcription-pcr. J Clin Microbiol. 2007; 45:1310-14.

10. Wise MG, Suarez DL, Seal BS, Pedersen JC, Senne DA, King DJ, et al. Development of a real-time reverse-transcription pcr for detection of newcastle disease virus rna in clinical samples. J Clinl Microbiol. 2004; 42:329-38.

11. Kim LM, Afonso CL, Suarez DL. Effect of probe-site mismatches on detection of virulent newcastle disease viruses using a fusion-gene real-time reverse transcription polymerase chain reaction test. J Vet Diagn Invest. 2006; 18:519-28.

12. Sabra M, Dimitrov KM, Goraichuk IV, Wajid A, Sharma P, Williams-Coplin D, et al. Phylogenetic assessment reveals continuous evolution and circulation of pigeon-derived virulent avian avulaviruses 1 in eastern europe, asia, and africa. BMC Vet Res. 2017; 13:291.

13. Kim LM, Suarez DL, Afonso CL. Detection of a broad range of class i and ii newcastle disease viruses using a multiplex real-time reverse transcription polymerase chain reaction assay. J Vet Diagn Invest. 2008; 20:414-25.

14. Fuller CM, Brodd L, Irvine RM, Alexander DJ, Aldous EW. Development of an l gene real-time reverse-transcription pcr assay for the detection of avian paramyxovirus type 1 rna in clinical samples. Arch Virol. 2010; 155:817-23.

15. FLU-LAB-NET. https://science.vla.gov.uk/flu-lab-net/docs/pub-protocol-avian-avulavirus-mole-pathotyp.pdf. Accessed 24 September 2018.

16. Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. High throughput sequencing: An overview of sequencing chemistry. Indian J Microbiol. 2016; 56:394-404.

17. Rhoads A, Au KF. Pacbio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015; 13:278-89.

18. Chiu CY. Viral pathogen discovery. Curr Opin Microbiol. 2013; 16:468-78.

19. Dimitrov KM, Sharma P, Volkening JD, Goraichuk IV, Wajid A, Rehmani SF, et al. A robust and cost-effective approach to sequence and analyze complete genomes of small rna viruses. Virol J. 2017; 14:72.

20. Cruz-Rivera M, Forbi JC, Yamasaki L, Vazquez-Chacon CA, Martinez-Guarneros A, Carpio-Pedroza JC, et al. Molecular epidemiology of viral diseases in the era of next generation sequencing. J Clin Virol. 2013; 57:378-80.

21. Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, et al. Next generation sequencing of viral rna genomes. BMC Genomics. 2013; 14:444.

22. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. J Pathol Inform. 2012; 3:40.

23. Phan H, Stoesser N, Maciuca I, Toma F, Szekely E, Flonta M, et al. Illumina short-read and minion long-read whole genome sequencing to characterise the molecular epidemiology of an ndm-1-serratia marcescens outbreak in romania. J Antimicrob Chemother. 2017; 73 (3) 672–79.

24. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015; 7:99.

25. Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, et al. Bacterial and viral identification and differentiation by amplicon sequencing on the minion nanopore sequencer. Gigascience. 2015; 4:12.

26. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. Minion nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015; 33:296.

27. Lemon JK, Khil PP, Frank KM, Dekker JP. Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. J Clin Microbiol. 2017; 55:3530-43.

28. Wang J, Moore NE, Deng Y-M, Eccles DA, Hall RJ. Minion nanopore sequencing of an influenza genome. Front Microbiol. 2015; 6:766.

29. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for ebola surveillance. Nature. 2016; 530:228.

30. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex pcr method for minion and illumina sequencing of zika and other virus genomes directly from clinical samples. Nat Protoc. 2017; 12:1261.

31. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res. 2016; 26:1721-29.

32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. Qiime allows analysis of high-throughput community sequencing data. Nat Methods. 2010; 7:335.

33. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009; 75:7537-41.

34. Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. Minion analysis and reference consortium: Phase 1 data release and analysis. F1000Research. 2015; 4.

35. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. Sequencing 16s rrna gene fragments using the pacbio smrt DNA sequencing system. PeerJ. 2016; 4:e1869.

36. Li C, Chng KR, Boey EJH, Ng AHQ, Wilm A, Nagarajan N. Inc-seq: Accurate single molecule reads using nanopore sequencing. GigaScience. 2016; 5:34.

37. Alexander D, Swayne D. Newcastle disease virus and other avian paramyxoviruses, p 156–163. A laboratory manual for the isolation and identification of avian pathogens. 1998; 4.

38. Miller PJ, Dimitrov KM, Williams-Coplin D, Peterson MP, Pantin-Jackwood MJ, Swayne DE, et al. International biological engagement programs facilitate newcastle disease epidemiological studies. Front Public Health. 2015; 3:235.

39. Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016; 44:W3-W10.

40. Porechop- an adapter trimming tool https://github.com/rrwick/Porechop.

41. Myers G, editor Efficient local alignment discovery amongst noisy long reads2014; Berlin, Heidelberg: Springer Berlin Heidelberg.

42. Katoh K, Misawa K, Kuma Ki, Miyata T. Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res. 2002; 30:3059-66.

43. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 2009; 25:1754-60.

44. Li H. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arXiv preprint arXiv:13033997. 2013.

45. Nanopolish – a software package for signal-level analysis of Oxford Nanopore sequencing data to calculate an improved consensus sequence for a draft genome assembly. https://github.com/jts/nanopolish.

46. He Y, Taylor TL, Dimitrov KM, Butt SL, Stanton JB, Goraichuk IV, et al. Whole-genome sequencing of genotype vi newcastle disease viruses from formalin-fixed paraffin-embedded tissues from wild pigeons reveals continuous evolution and previously unrecognized genetic diversity in the us. Virol J. 2018; 15:9.

47. Thompson JD, Higgins DG, Gibson TJ. Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673-80.

48. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. Mega6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30:2725-29.

49. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993; 10:512-26.

50. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004; 101:11030-35.

51. Nayak B, Dias FM, Kumar S, Paldurai A, Collins PL, Samal SK. Avian paramyxovirus serotypes 2-9 (apmv-2-9) vary in the ability to induce protective immunity in chickens against challenge with virulent newcastle disease virus (apmv-1). Vaccine. 2012; 30:2220-27.

52. Seal BS, King DJ, Bennett JD. Characterization of newcastle disease virus isolates by reverse transcription pcr coupled to direct nucleotide sequencing and development of sequence

database for pathotype prediction and molecular epidemiological analysis. J Clin Microbiol. 1995; 33:2624-30.

53. Seal BS, King DJ, Locke DP, Senne DA, Jackwood MW. Phylogenetic relationships among highly virulent newcastle disease virus isolates obtained from exotic birds and poultry from 1989 to 1996. J Clin Microbiol. 1998; 36:1141-45.

54. Rue CA, Susta L, Brown CC, Pasick JM, Swafford SR, Wolf PC, et al. Evolutionary changes affecting rapid identification of 2008 newcastle disease viruses isolated from double-crested cormorants. J Clin Microbiol. 2010; 48:2440-48.

55. Wei S, Weiss ZR, Williams Z. Rapid multiplex small DNA sequencing on the minion nanopore sequencing platform. G3: Genes, Genomes, Genetics. 2018:g3. 200087.2018.

56. Li H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016; 32:2103-10.

57. Jain M, Tyson JR, Loose M, Ip CL, Eccles DA, O'Grady J, et al. Minion analysis and reference consortium: Phase 2 data release and analysis of r9. 0 chemistry. F1000Research. 2017; 6.

58. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nature methods. 2015; 12:733.

59. Phillips R, Samson A, Emmerson P. Nucleotide sequence of the 5′-terminus of newcastle disease virus and assembly of the complete genomic sequence: Agreement with the “rule of six”. Arch Virol. 1998; 143:1993-2002.

Table 3.1. Comparison of MinION sequencing to RT-qPCR for detection of NDV LaSota (runs 1 and 2)

Dilution (EID50/ml)	Total reads[a]	Total NDV reads[b]	Reads per consensus[c]	Percent identity[d]	Consensus length	RT-qPCR[e] (Ct)
	R1[f] | R2[g]	R1 | R2	R1 | R2	R1 | R2	R1 | R2	R1 | R2
10^6	6667 | 11366	6577 | 10861	200 | 200	100 | 100	734 | 734	21.8, 21.1 | 22.7, 22.7
10^5	4519 | 6801	4439 | 6540	200 | 200	100 | 100	734 | 734	26.3, 25.8 | 26.2, 26.4
10^4	3856 | 8289	3829 | 7890	200 | 200	100 | 100	734 | 734	28.9, 27.8 | 29.1, 29.3
10^3	164 | 9484	157 | 9061	157 | 200	100 | 100	734 | 734	31.4, 31.1 | 32.7, 32.8
10^2	94 | 4939	85 | 4725	85 | 200	100 | 100	734 | 734	34.2, 34.8 | 34.8, 35.3
10^1	133 | 2652	131 | 2520	131 | 200	99.04 | 99.86	729 | 734	34.9, 34.8 | 36.9, 36.7

[a] Obtained from output of Porechop

[b] Obtained from output of LAclust.

[c] Obtained from output of BWA-MEM. Input into BWA-MEM was limited to 200 reads based on LAclust options.

[d] Consensus sequence identity to the reference sequence of NDV LaSota sequenced with MiSeq (MH392212/chicken/USA/LaSota/1946)

[e] Each dilution performed in duplicate and threshold cycle (Ct) values from each well are shown here.

[e] Run 1

[f] Run 2

Note: All 60,000 reads obtained during 32 minutes of sequencing run (R1) were utilized for the analysis. All 98,916 reads from R2 were utilized for the analysis.

Table 3.2. Accuracy of consensus sequence from serial dilutions (EID_{50}/ml) of NDV LaSota during MinION sequencing run 1.

Sequencing run time (min)	Total Raw reads	10^6		10^5		10^4		10^3		10^2		10^1	
		NDV reads[a]	% Identity[b]	NDV reads	% Identity	NDV reads	% Identity	NDV reads	% Identity	NDV reads	% Identity	NDV reads	% Identity
5	4000	283	100	199	100	182	100	10	99.18	2	NA	3	NA
7	8000	644	100	406	100	368	100	14	99.32	8	98.77	5	98.37
10	12000	1029	100	661	100	571	100	23	99.32	16	99.32	8	99.18
12	16000	1397	100	901	100	759	100	29	99.86	17	99.32	14	99.45
14	20000	1772	100	1173	100	1014	100	36	99.73	20	99.45	17	99.59
16	24000	2183	100	1451	100	1231	100	43	99.73	23	99.45	21	99.59
19	28000	2643	100	1775	100	1498	100	56	99.86	29	99.73	25	99.45

[a] The numbers represent total number of NDV reads obtained from LAclust. A maximum of 200 reads (optional cut-off value) were used to generate full length consensus sequence. Minimum 5 reads were used as a cut-off to build consensus sequence.

[b] Consensus sequence identity to the reference sequence of NDV LaSota sequenced with MiSeq (MH392212/chicken/USA/LaSota/1946)

Consensus sequences were BLAST searched against NDV custom database

Note: Only 28,000 out of total 60,000 reads were utilized for the analysis

Table 3.3. Identification and virulence prediction of NDV genotypes from 33 egg-grown samples (runs 3 and 4).

Sample ID	Input genotype[a]	Output genotype[b]	BLAST search	Alignment length	Percent identity	Fusion protein cleavage site[c, ⸸]
1	**II***	**II**	**MH392212/chicken/USA/LaSota/1946**	**732**	**100**	**low virulent**
2	**II**	**II**	**KJ607167/LHLJ/2/goose/2006/China**	**734**	**100**	**low virulent**
3	**II**	**II**	**KJ607167/LHLJ/2/goose/2006/China**	**732**	**100**	**low virulent**
4	**II**	**II**	**EU289029/turkey/USA/VG/GA-clone_5/1987**	**734**	**99.86**	**low virulent**
5	**Ia***	**Ia**	**MH392213/chicken/Australia/Queensland/V-4/10/1966**	**734**	**100**	**low virulent**
15	**Ia**	**Ia**	**MH392213/chicken/Australia/Queensland/V-4/10/1966**	**734**	**99.86**	**low virulent**
16	VIId	VIId	KU295454/chicken/Ukraine/Lyubotyn/961/2003	735	99.46	virulent
17	II*	II	MH392228/poultry/Canada/Ontario/Berwick/853/1948	735	100	virulent
18	II	II	MH392228/poultry/Canada/Ontario/Berwick/853/1948	732	98.36	virulent
19	III*	III	*MH392214/chicken/India/Mukteswar/519/1940	734	100	virulent
20	IV*	IV	MH392215/chicken/Nigeria/Kano/1973/N52/899/1973	734	99.86	virulent
21	IV	IV	EU293914/Italy/Italien/1944	734	99.05	virulent
22	XIVb*	XIVb	KT948996/domestic_duck/Nigeria/NG-695/KG.LOM.11-16/2009	734	100	virulent
23	Va*	Va	MH392216/cormorant/USA/MN/92-40140/250/1992	734	100	virulent
24	Vb*	Vb	MH392217/turkey/Belize/4338-4/607/2008	734	100	virulent
25	Vc*	Vc	MH392218/chicken/Mexico/NC/23/686/2011	733	99.73	virulent
26	VIc*	VIc	KY042125/chicken/Bulgaria/Dolno_Linevo/1992	734	100	virulent
27	VIm*	VIm	KX236101/pigeon/Pakistan/Lahore/25A/2015	734	100	virulent
28	VIIj*	VIIj	MH392219/chicken/Egypt/Sohag/18/1020/2014	734	100	virulent
29	VIIe	VIIe	KJ782375/goose/China/GD-QY/1997	734	97.82	virulent
30	VIIi*	VIIi	KX496962/ wild_pigeon/Pakistan/Lahore/20A/996//2015	734	100	virulent
31	IX*	IX	MH392220/poultry/China/04-23/C12/647/2004	734	100	virulent
32	IX	IX	MH392220/poultry/China/04-23/C12/647/2004	734	99.86	virulent
33	**Xb***	**Xb**	**MH392221/mallard/USA/MN/99-376/163/1999**	**734**	**100**	**low virulent**
34	**Xa***	**Xa**	**GQ288378/ northern_pintail/USA/OH/87-486/1987**	**734**	**100**	**low virulent**
35	XIIa*	XIIa	JN800306/poultry/Peru/1918-03/2008	734	100	virulent
36	XIIIb*	XIIIb	MH392222/chicken/Pakistan/SPVC/Karachi/27/558/2007	734	99.18	virulent
37[d]	VIc/XIIIb*	XIIIb	MH392223/chicken/Pakistan/SPVC/Karachi/33/556-XIII/2007	734	99.46	virulent
38	XIVb*	XIVb	MH392225/chicken/Nigeria/KD/TW/03T/N45/720/2009	734	100	virulent
39	XVI*	XVI	MH392226/chicken/Dominican_Republic/FO/499-31/505/2008	734	100	virulent
40	XVIIa*	XVIIa	KY171995/VRD124/06/N11/867/chicken/2006/Nigeria	734	100	virulent
41	XVII*	XVII	KU058680/903/domestic_duck/Nigeria/KUDU-113/1992	734	100	virulent
42	XVIIIb*	XVIIIb	MH392227/chicken/Nigeria/OOT/4/1/N69/914/2009	734	100	virulent

[a] Input genotype was determined with MiSeq sequencing (*) or previous Sanger sequencing.

[b] Determined by MinION sequencing

[c] F protein cleavage sites of virulent NDV genotypes contains more than 3 basic amino acids [(112(R/K)-R-(Q/K/R)-(R/K)-R-F117)] and low virulent NDV genotypes has monobasic amino acids [(112(G/E)-(R/K)-Q-(G/E)-R-L117)]

[d] Illumina Miseq detected two NDV genotypes

*Matching MiSeq result from same isolate

⸸ The fusion protein cleavage sites did not vary between AmpSeq and either previous Sanger or previous MiSeq.

Note: Isolates known to have low virulence are highlighted in bold.

Table 3.4. Identification and virulence prediction of NDV genotypes in clinical samples collected during outbreaks in 2015 (run 5, 6, and 7).

Sample ID	Miseq genotypes	MinION genotypes	ID of the MinION hit	Reads/ cluster	Consensus length	Percent identity	Fusion protein cleavage site[⸸]
44	VIIi	VIIi	chicken/Pakistan/Wadana_Kasur/PNI_PF_(14F)/2015	200	734	100	virulent
45	VIIi	VIIi	chicken/Pakistan/Wadana_Kasur/PNI_PF_(14F)/2015	28	734	99.31	virulent
	II	**II**	**chicken/USA/LaSota/1946**	**5**	**733**	**96.44**	**low virulent**
46	VIIi	VIIi	chicken/Pakistan/Wadana_Kasur/PNI_PF_(14F)/2015	10	733	99.13	virulent
	II	**II**	**chicken/USA/LaSota/1946**	17	733	**98.51**	**low virulent**
47	VIIi	ND[d]	NA[e]	NA	NA	NA	NA
	II	**II**	**chicken/USA/LaSota/1946**	**139**	**732**	**99.32**	**low virulent**
48	**ND**	**II**	**chicken/USA/LaSota/1946**	**200**	**732**	**99.59**	**low virulent**
	VIIi	VIIi	chicken/Pakistan/Wadana_Kasur/PNI_PF_(14F)v/2015	21	733	99.13	virulent
49	VIIi	ND	NA	NA	NA	NA	NA
	II	**II**	**chicken/USA/LaSota/1946**	**200**	**732**	**99.32**	**low virulent**
50	VIIi	VIIi	chicken/Pakistan/Wadana_Kasur/PNI_PF_(14F)/2015	113	734	100	virulent
51	VIIi	VIIi	chicken/Pakistan/Mirpur_Khas/3EOS/2015	200	734	100	virulent
52[a]	VIIi	ND	NA	NA	NA	NA	NA
53	VIIi	VIIi	exotic Parakeets/Pakistan/Charah/Pk29/29A/2015	5	726	98.5	virulent
54	NO NDV	ND	NA	NA	NA	NA	NA
55	NO NDV	ND	NA	NA	NA	NA	NA
56	NO NDV	ND	NA	NA	NA	NA	NA
57	NO NDV	ND	NA	NA	NA	NA	NA
58	VIIi	VIIi	chicken/Pakistan/Gharoo/Three_star_PF_(7G)/2015	8	729	99.32	virulent
TN[b]	NA	ND	NA	NA	NA	NA	NA
EN[c]	NA	ND	NA	NA	NA	NA	NA

[a] After bead purification, the barcoded amplicon concentration of this sample was lowest in this pool.

[b] Template control negative

[c] Negative extraction control

[d] Not detected

[e] Not applicable

[⸸] The fusion protein cleavage sites did not vary between AmpSeq and previous MiSeq.

Note: Isolates known to have low virulence are highlighted in bold.

Table S3.1. The representative genotypes of AAvV-1 and other AAvVs used in this study (egg-grown viruses)

Sample ID	Isolate	Genotype/ serotype	MinION run[a]
1	chicken/USA/Lasota/1946	II	3, 4, MiSeq
2	chicken/USA/Hitchner/B1/1947	II	4
3	clone 30	II	4
4	turkey/USA/VG/GA/1989	II	4
5	chicken/Australia/Queensland/V-4/10/1966	Ia	3, 4, MiSeq
6	MN2000-495	APMV-2	NA
7	APMV-3/turkey/USA/WI/1968	APMV-3	NA
8	APMV-4/--/USA/MN/2000	APMV-4	NA
9	APMV-6/--/USA/MN/1999	APMV-6	NA
10	APMV-7/--/USA/TX02-12/	APMV-7	NA
11	APMV-8/goose/US/DE/1053/1976	APMV-8	NA
12	APMV-9/duck/USA/NY/22/1978	APMV-9	NA
13	APMV-10/penguin/Falkland Islands/324/2007	APMV-10	NA
14	APMV-13/white-fronted goose/Ukraine/Askania-Nova/48-15-02/2011	APMV-13	NA
15	Malaysia/5091/633/2009	Ia	4
16	APMV-p-S-221111/964	VIId	4
17	poultry/Canada/Ontario/Berwick/853/1948	II	4, MiSeq
18	poultry/USA/OH/Miller/778/1948	II	4
19	chicken/India/Mukteswar/519/1940s	III	4, MiSeq
20	chicken/Nigeria/Kano/1973/N52/899/1973	IV	4, MiSeq
21	Italy/Milano/1945	IV	4
22	duck/Nigeria/NG-695/KG.LOM.11-16/2009	XIVb	4, MiSeq
23	cormorant/USA/MN/92-40140/250/1992	Va	4, MiSeq
24	turkey/Belize/4338-4/607/2008	Vb	3, 4, MiSeq
25	chicken/Mexico/NC/23/686/2011	Vc	4, MiSeq
26	chicken/Bulgaria/Dolno_Linevo/1160/1992	VIc	4, MiSeq
27	pigeon/Pakistan/Lahore/25A/1011/2015	VIk	3, 4, MiSeq
28	chicken/Egypt/Sohag/18/1020/2014	VIIj	4, MiSeq
29	duck/Vietnam/Long Bien/78/2002	VIIe	4
30	pigeon/Pakistan/Lahore/20A/996/2015	VIIi	3, 4, MiSeq
31	poultry/China/04-23/C12/647/2004	IX	4, MiSeq
32	chicken/03-45/641/2003	IX	4
33	mallard/USA/MN/99-376/163/1999	Xb	4, MiSeq
34	northern_pintail/US(OH)/87-486/164/1987	Xa	4, MiSeq
35	poultry/Peru/1918-03/603/2008	XIIa	4, MiSeq
36	chicken/Pakistan/SPVC/Karachi/27/558/2007	XIIIb	4, MiSeq
37	chicken/Pakistan/SPVC/Karachi/33/556/2007	XIIIb, VIc	3, 4, MiSeq
38	pigeon/Nigeira/Katsina/KT/MSH/15C_(N2)/689/2009	XIVb	4, MiSeq
39	chicken/Dominican_Republic/FO/499-31/505/2008	XVI	4, MiSeq
40	chicken/Nigeria/VRD124/06/N11/867/2006	XVIIa	4, MiSeq
41	duck/Nigeria/KUDU-113/903/1992	XVII	4, MiSeq
42	chicken/Nigeria/OOT/4/1/N69/914/2009	XVIIIb	4, MiSeq
43	APMV-5/Japan/Tokyo/Kunitachi/1978	APMV-5	NA

[a] 3 = MinION run 3 (6 samples pooled), 4 = MinION run 4 (33 samples pooled), MiSeq = Isolates sequences with Illumina Miseq. For further information about MinION sequencing runs, see Table S3.

Table S3.2. Background information of clinical swab (oral and cloacal) samples collected from chicken during disease outbreaks in Pakistan in 2015

Sample ID	Isolate	Genotype[a]	Swab type	MinION run[b]
44	chicken/Pakistan/Punjab/1F/1062/2015	VII i	Oral	6
45	chicken/Pakistan/Punjab/2F//1063/2015	VII i and II	Oral	5
46	chicken/Pakistan/Punjab/4F/1065/2015	VII i and II	Oral	6
47	chicken/Pakistan/Punjab/5F/1066/2015	VII i and II	Oral	5
48	chicken/Pakistan/Punjab/7F/1067/2015	VII i	Oral	6
49	chicken/Pakistan/Punjab/8F/1068/2015	VII i and II	Oral	5
50	chicken/Pakista/Punjab/14F/1069/2015	VII i	Cloacal	5
51	chicken/Pakistan/Sindh/4E/1061/2015	VII i	Oral	5
52	chicken/Pakistan/Punjab/1H/1072/2015	VII i	Oral	7
53	chicken/Pakistan/Punjab/3H/1074/2015	VII i	Oral	5
54	chicken/Pakistan/Punjab/16H/1077/2015	NO NDV	Cloacal	5
55	chicken/Pakistan/Sindh/A3/1221/2015	NO NDV	Oral	6
56	chicken/Pakistan/Sindh/B1/1222/2015	NO NDV	Oral	5
57	chicken/Pakistan/Punjab/D2/1226/2015	NO NDV	Oral	5
58	chicken/Pakistan/Punjab/7G/1071/2015	VII i	Oral	7

[a] all isolates were identified as members of AAvV-1 with next-generations sequencing (MiSeq)

[b] 5 = MinION run 5, 6 = MinION run 6, 7 = MinION run 7. For further information about MinION sequencing runs, see Table S3.

Table S3.3. Detail of MinION sequencing runs

Run	Description of MinION sequencing run	Sample type	Sequencing run time	Total reads
1	LaSota serial dilution (n = 6) (R1)	Allantoic fluid	32 min	60,000
2	LaSota serial dilutions (n = 6) (R2)	Allantoic fluid	32 min	98916
3	NDV6; (n = 6)	Allantoic fluid	20 min	60,000
4	NDV33; (n = 33)	Allantoic fluid	12 hrs	2,084,000
5	*Clinical swab samples (n = 9)	Swab material	6 hrs	368,000
6	*Clinical swab samples (n = 4)	Swab material	7 hrs	224,000
7	*Clinical swab samples (n = 2)	Swab material	12 hrs	284,000

*Amplicons obtained from clinical samples were pooled together based on the variation in their concentration

Table S3.4. Time-based quality metrics of MinION sequencing run 4 (n = 33)

Sequencing run time for each batch of 20,000 reads (total run time)	**Mean read quality**	**Reads $Q \geq 10$**	**% of reads $Q \geq 10$**	**Mean read $Q_{\geq 10}$**
30 (30) min	11.3	16593	82.97	11.9
30 (60) min	11.2	16274	81.37	11.8
45 (105) min	11.0	15853	79.26	11.8
45 (150) min	11.0	15714	78.57	11.7
40 (190) min	10.8	15256	76.28	11.6

A Q score of 10 translates into 90% accuracy.

Table S3.5. Estimation of cost of reagents and sample processing time for MinION sequencing

Steps	**Time**		**Cost in USD**	
Multiplexed samples	*n = 6*	*n = 33*	*n = 6*	*n = 33*
RNA extraction	2 hrs	8 hrs	30	165
One step RT-PCR	2 hrs	2 hrs	48	264
Amplicon purification	20 min	2 hrs	18	99
Barcode kit	NA	NA	25	150
Single library preparation	4 hrs	8 hrs	99	99
Flow cell	NA	NA	500	500
Flow cell per sequencing run (n)	NA	NA	(n = 5) 100	(n = 2) 250
Sequencing run	32 min	3 hrs 10 min	320	1027
[a]Basecalling	1 hr	1 hr 40 min	NA	NA
Post basecalling data processing and consensus assembly	25 min	40 min	NA	NA
Total time	[b]**9-10 hrs**	**26 hrs**	NA	NA
Cost ($) per sample	NA	NA	**$53**	**$31**

[a]Basecalling time varies based on average length and total number of reads. For amplicon sizes in this study, basecalling took approximately 60 minutes for 60,000 reads. Longer times will be required for longer amplicons.

[b]Including the sequencing run time

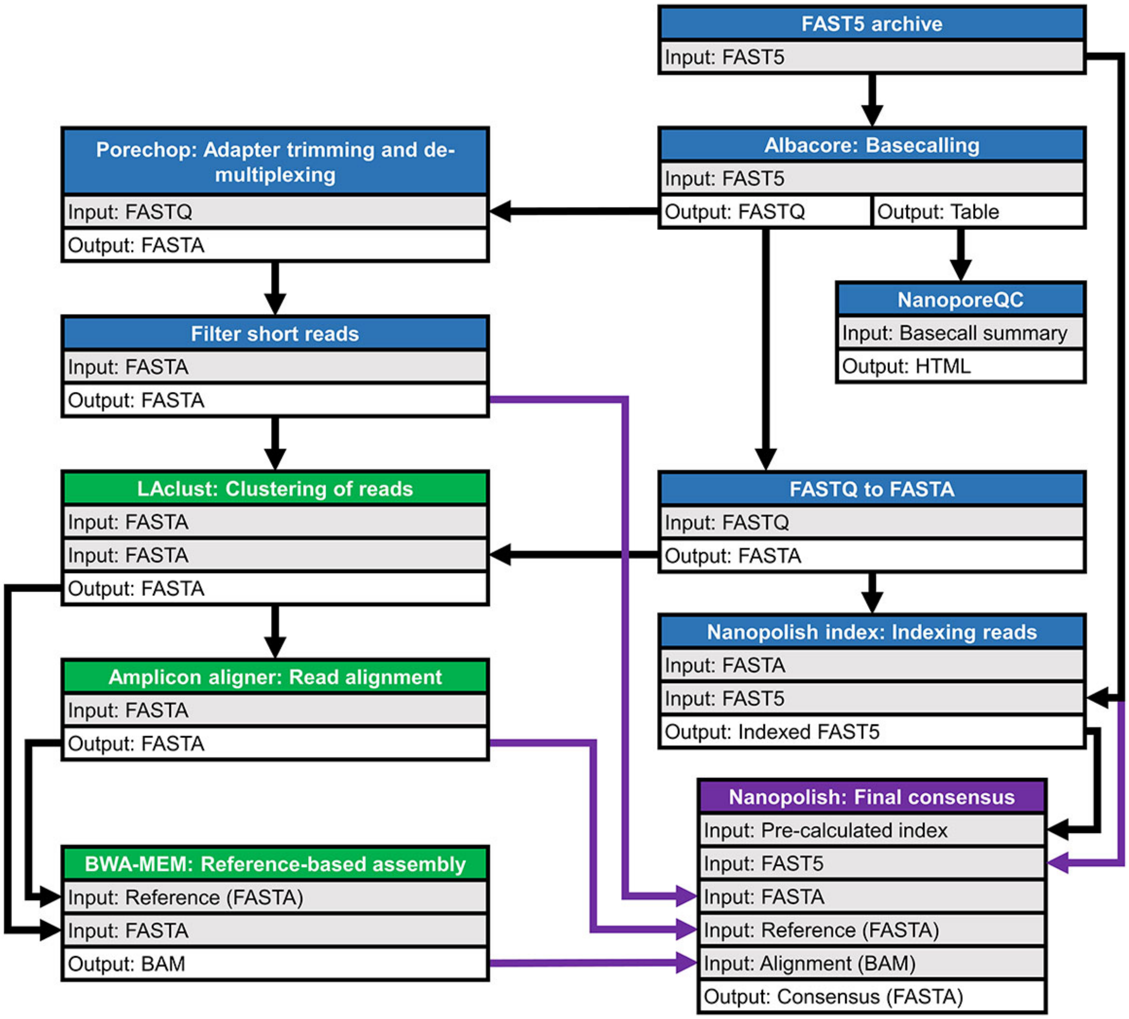


Figure 3.1. Schematic diagram of customized Galaxy workflow for MinION sequence data analysis. *Blue shading* indicates pre-processing steps. *Green shading* indicates post-processing steps; assembly/output is shaded purple. *Purple* arrows indicate different inputs for final consensus calculation.

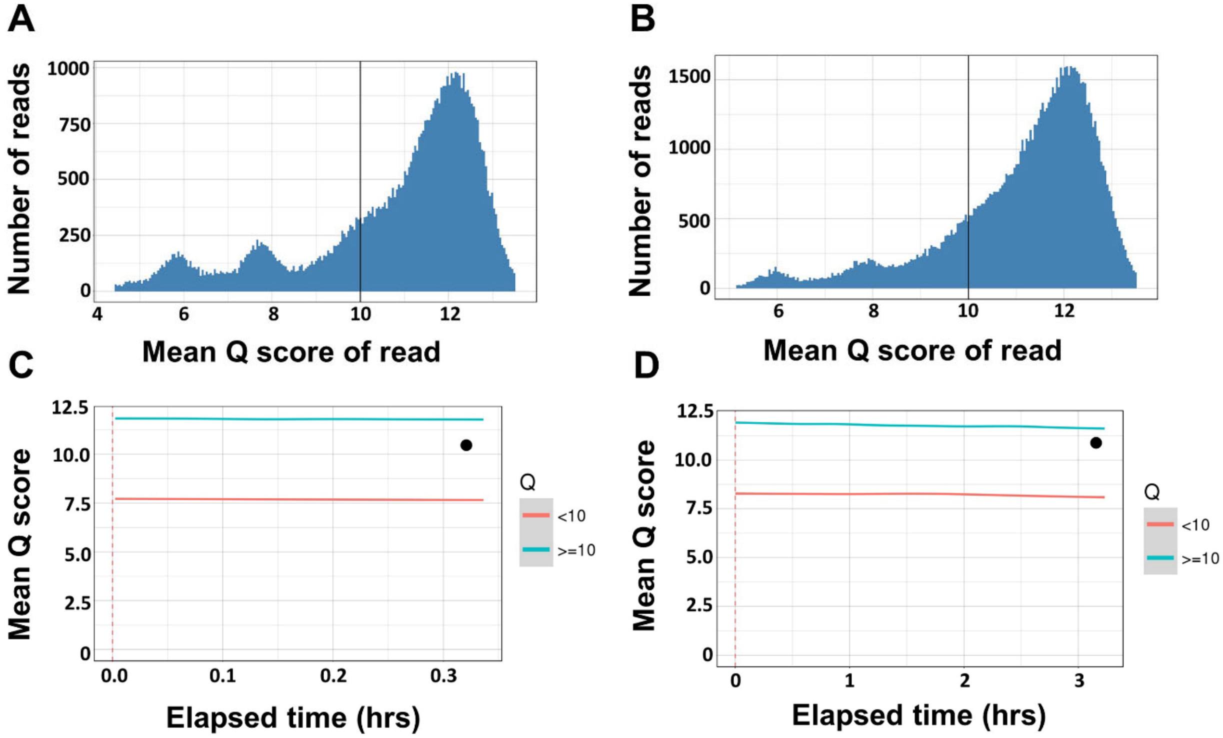


Figure 3.2: Quality metrics of two MinION sequencing runs. Mean read-based quality score distribution of 6 sample pooled run (run 3) (A) and 33 sample pooled run (run 4) (B). Mean run-based quality score over time of six sample pooled run (C) and 33 sample pooled run (D). The overall read quality average (●) remained above 10 in both runs.

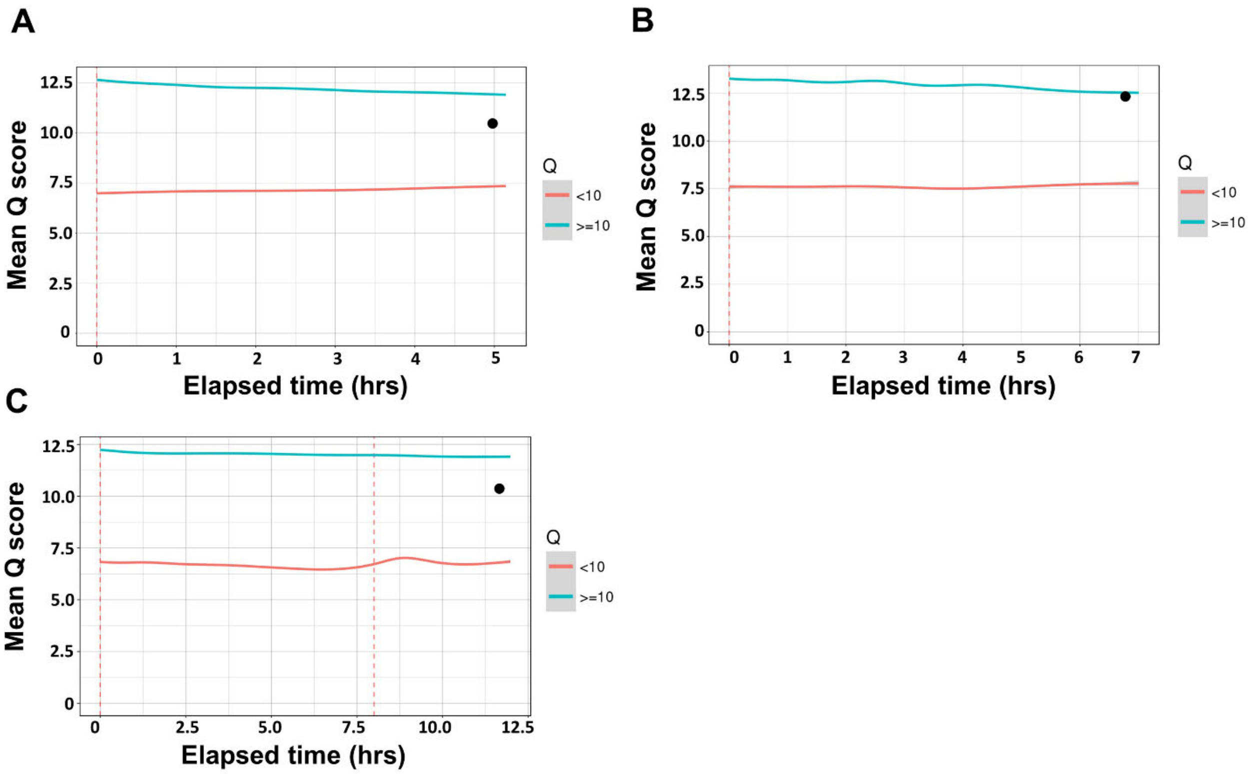


Figure 3.3: Mean Q≥10 (blue lines) and Q<10 (orange lines) of total reads over time during MinION sequencing runs using clinical swab samples. Additionally, the overall read quality average (●) for all three runs, remained above 10. A: MinION run 5, n = 9 samples, runtime = 6 hours. B: MinION run 6, n = 4 samples, runtime = 7 hours. C: MinION run 7, n = 2 samples, runtime = 12 hours.

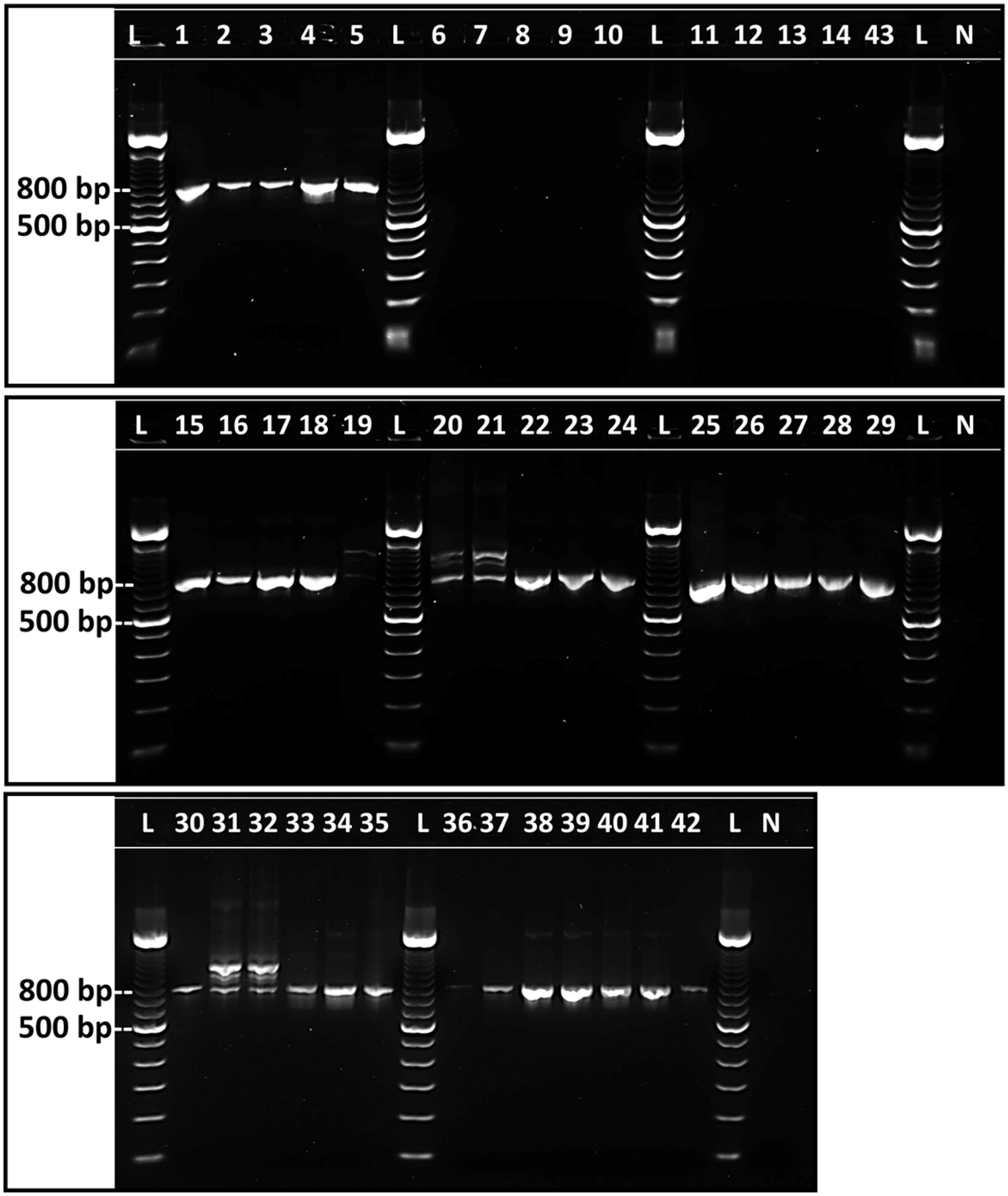
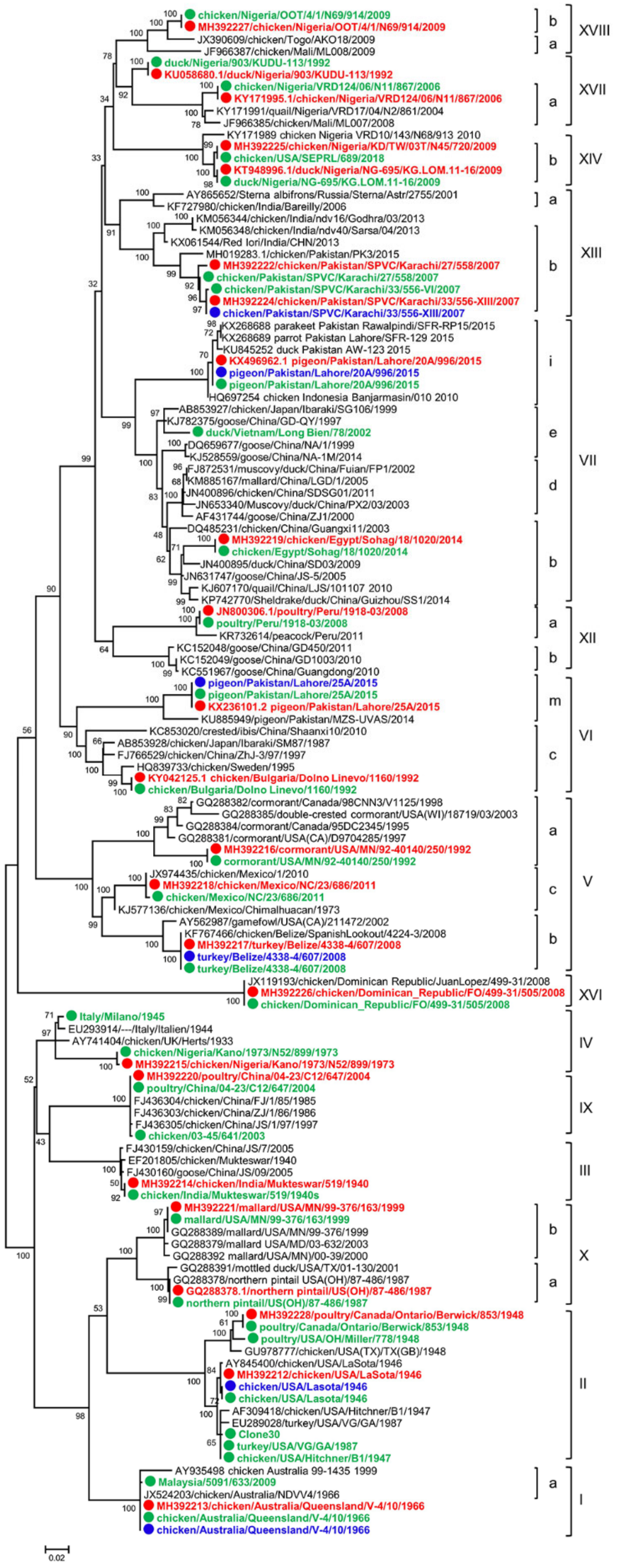

L 1 2 3 4 5 L 6 7 8 9 10 L 11 12 13 14 43 L N
800 bp
500 bp
L 15 16 17 18 19 L 20 21 22 23 24 L 25 26 27 28 29 L N
800 bp
500 bp
L 30 31 32 33 34 35 L 36 37 38 39 40 41 42 L N
800 bp
500 bp

Figure S3.1. Agarose gel electrophoresis of AAvVs. Sample 6–14 and 43 are AAvVs other than AAvV-1. A DNA ladder (100 bp) was loaded into lane L. A no-template control was loaded into lane N. Bright bands show the amplified target region of AAvA-1 genome (expected product size 788 bp). See Table S3.1 for key to lanes.



NDV Class II Genotypes

chicken/Nigeria/OOT/4/1/N69/914/2009
MH392227/chicken/Nigeria/OOT/4/1/N69/914/2009
JX390609/chicken/Togo/AKO18/2009
JF966387/chicken/Mali/ML008/2009
duck/Nigeria/903/KUDU-113/1992
KU058680.1/duck/Nigeria/903/KUDU-113/1992
chicken/Nigeria/VRD124/06/N11/867/2006
KY171995.1/chicken/Nigeria/VRD124/06/N11/867/2006
KY171991/quail/Nigeria/VRD17/04/N2/861/2004
JF966385/chicken/Mali/ML007/2008
KY171989 chicken Nigeria VRD10/143/N68/913 2010
MH392225/chicken/Nigeria/KD/TW/03T/N45/720/2009
chicken/USA/SEPRL/689/2018
KT948996.1/duck/Nigeria/NG-695/KG.LOM.11-16/2009
duck/Nigeria/NG-695/KG.LOM.11-16/2009
AY865652/Sterna albifrons/Russia/Sterna/Astr/2755/2001
KF727980/chicken/India/Bareilly/2006
KM056344/chicken/India/ndv16/Godhra/03/2013
KM056348/chicken/India/ndv40/Sarsa/04/2013
KX061544/Red lori/India/CHN/2013
MH019283.1/chicken/Pakistan/PK3/2015
MH392222/chicken/Pakistan/SPVC/Karachi/27/558/2007
chicken/Pakistan/SPVC/Karachi/27/558/2007
chicken/Pakistan/SPVC/Karachi/33/556-VI/2007
MH392224/chicken/Pakistan/SPVC/Karachi/33/556-XIII/2007
chicken/Pakistan/SPVC/Karachi/33/556-XIII/2007
KX268688 parakeet Pakistan Rawalpindi/SFR-RP15/2015
KX268689 parrot Pakistan Lahore/SFR-129 2015
KU845252 duck Pakistan AW-123 2015
KX496962.1 pigeon/Pakistan/Lahore/20A/996/2015
pigeon/Pakistan/Lahore/20A/996/2015
pigeon/Pakistan/Lahore/20A/996/2015
HQ697254 chicken Indonesia Banjarmasin/010 2010
AB853927/chicken/Japan/Ibaraki/SG106/1999
KJ782375/goose/China/GD-QY/1997
duck/Vietnam/Long Bien/78/2002
DQ659677/goose/China/NA/1/1999
KJ528559/goose/China/NA-1M/2014
FJ872531/muscovy/duck/China/Fuian/FP1/2002
KM885167/mallard/China/LGD/1/2005
JN400896/chicken/China/SDSG01/2011
JN653340/Muscovy/duck/China/PX2/03/2003
AF431744/goose/China/ZJ1/2000
DQ485231/chicken/China/Guangxi11/2003
MH392219/chicken/Egypt/Sohag/18/1020/2014
chicken/Egypt/Sohag/18/1020/2014
JN400895/duck/China/SD03/2009
JN631747/goose/China/JS-5/2005
KJ607170/quail/China/LJS/101107 2010
KP742770/Sheldrake/duck/China/Guizhou/SS1/2014
JN800306.1/poultry/Peru/1918-03/2008
poultry/Peru/1918-03/2008
KR732614/peacock/Peru/2011
KC152048/goose/China/GD450/2011
KC152049/goose/China/GD1003/2010
KC551967/goose/China/Guangdong/2010
pigeon/Pakistan/Lahore/25A/2015
pigeon/Pakistan/Lahore/25A/2015
KX236101.2 pigeon/Pakistan/Lahore/25A/2015
KU885949/pigeon/Pakistan/MZS-UVAS/2014
KC853020/crested/ibis/China/Shaanxi10/2010
AB853928/chicken/Japan/Ibaraki/SM87/1987
FJ766529/chicken/China/ZhJ-3/97/1997
HQ839733/chicken/Sweden/1995
KY042125.1 chicken/Bulgaria/Dolno Linevo/1160/1992
chicken/Bulgaria/Dolno Linevo/1160/1992
GQ288382/cormorant/Canada/98CNN3/V1125/1998
GQ288385/double-crested cormorant/USA(WI)/18719/03/2003
GQ288384/cormorant/Canada/95DC2345/1995
GQ288381/cormorant/USA(CA)/D9704285/1997
MH392216/cormorant/USA/MN/92-40140/250/1992
cormorant/USA/MN/92-40140/250/1992
JX974435/chicken/Mexico/1/2010
MH392218/chicken/Mexico/NC/23/686/2011
chicken/Mexico/NC/23/686/2011
KJ577136/chicken/Mexico/Chimalhuacan/1973
AY562987/gamefowl/USA(CA)/211472/2002
KF767466/chicken/Belize/SpanishLookout/4224-3/2008
MH392217/turkey/Belize/4338-4/607/2008
turkey/Belize/4338-4/607/2008
turkey/Belize/4338-4/607/2008
JX119193/chicken/Dominican Republic/JuanLopez/499-31/2008
MH392226/chicken/Dominican_Republic/FO/499-31/505/2008
chicken/Dominican_Republic/FO/499-31/505/2008
Italy/Milano/1945
EU293914/---/Italy/Italien/1944
AY741404/chicken/UK/Herts/1933
chicken/Nigeria/Kano/1973/N52/899/1973
MH392215/chicken/Nigeria/Kano/1973/N52/899/1973
MH392220/poultry/China/04-23/C12/647/2004
poultry/China/04-23/C12/647/2004
FJ436304/chicken/China/FJ/1/85/1985
FJ436303/chicken/China/ZJ/1/86/1986
FJ436305/chicken/China/JS/1/97/1997
chicken/03-45/641/2003
FJ430159/chicken/China/JS/7/2005
EF201805/chicken/Mukteswar/1940
FJ430160/goose/China/JS/09/2005
MH392214/chicken/India/Mukteswar/519/1940
chicken/India/Mukteswar/519/1940s
MH392221/mallard/USA/MN/99-376/163/1999
mallard/USA/MN/99-376/163/1999
GQ288389/mallard/USA/MN/99-376/1999
GQ288379/mallard USA/MD/03-632/2003
GQ288392 mallard/USA/MN)/00-39/2000
GQ288391/mottled duck/USA/TX/01-130/2001
GQ288378/northern pintail USA(OH)/87-486/1987
GQ288378.1/northern pintail/US(OH)/87-486/1987
northern pintail/US(OH)/87-486/1987
MH392228/poultry/Canada/Ontario/Berwick/853/1948
poultry/Canada/Ontario/Berwick/853/1948
poultry/USA/OH/Miller/778/1948
GU978777/chicken/USA(TX)/TX(GB)/1948
AY845400/chicken/USA/LaSota/1946
MH392212/chicken/USA/LaSota/1946
chicken/USA/Lasota/1946
chicken/USA/Lasota/1946
AF309418/chicken/USA/Hitchner/B1/1947
EU289028/turkey/USA/VG/GA/1987
Clone30
turkey/USA/VG/GA/1987
chicken/USA/Hitchner/B1/1947
AY935498 chicken Australia 99-1435 1999
Malaysia/5091/633/2009
JX524203/chicken/Australia/NDVV4/1966
MH392213/chicken/Australia/Queensland/V-4/10/1966
chicken/Australia/Queensland/V-4/10/1966
chicken/Australia/Queensland/V-4/10/1966
b
a
XVIII
a
XVII
b
XIV
a
b
XIII
i
e
d
b
VII
a
b
XII
m
c
VI
a
c
b
V
XVI
IV
IX
III
b
a
X
II
a
I
0.02

Figure S3.2. Phylogenetic tree constructed by using the nucleotide sequence (734 bp) of NDV isolates sequenced with MinION and MiSeq, with sequences of related NDV genotypes from GenBank. The evolutionary histories were inferred by using the maximum-likelihood method based on General Time Reversible model with 500 bootstrap replicates as implemented in MEGA 6. The tree with the highest log likelihood (-9347.8021) is shown. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories [+G, parameter = 0.9254]). The percentages of trees in which the associated sequences clustered together are shown below the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analyses involved 129 nucleotide sequences with a total of 725 positions in the final datasets. The sequences obtained in the current study are denoted with solid circles in front of the taxa name and bold font. Blue circles indicate isolates from MinION sequencing run 1, green circles indicate isolates from MinION sequencing run 2 and red circles indicate MiSeq sequencing

CHAPTER 4

SUMMARY AND CONCLUSIONS

Newcastle disease virus has a worldwide distribution and is endemic in many countries. Due to the impact of this disease on the global poultry industry, detection and characterization of NDV pathotypes is of utmost importance. Commonly, nucleic acid based Real Time PCR rapid diagnostic assays are being used for NDV diagnostics. However, rapid diagnostic assays have either little sequencing information or the obtained information is not enough to fully characterize the isolate. In current studies, two sequencing protocols (target independent using NGS and target dependent using MinION sequencing) were developed and optimized for detection and thorough characterization of currently circulating NDVs from different types of clinical samples.

Our first study tested the utility of FFPE tissues for direct sequencing of NDV genome using NGS. Different type of tissues collected during disease outbreaks were subjected to complete genome sequencing using random priming approach. This method is useful because FFPE tissues can be conveniently and affordably transported, due to pathogen inactivation, and because FFPE tissues are the primary means of archiving tissue, allowing for full-genome epidemiologic investigations in historical samples. The use of random sequencing coupled with absence of any virus enrichment procedure make this technique likely to be applicable to sequence virus genomes from clinical FFPE tissues in other viral infections. Additionally, the results demonstrate that sub-genotype VIIi viruses are still circulating and evolving in Pakistan after they were first identified in the country in 2011 and suggest that active epidemiologic surveillance for NDV is needed. This technique will help to investigate the epidemiologic link between highly related NDV viruses especially in closely located geographical regions where ND has acquired endemicity.

In our second study, we sought to develop and optimize a sequencing protocol that is time efficient and does not require expensive laboratory equipment. A MinION-based AmpSeq protocol was developed to detect all existing genotypes of APMV-1. Testing of this AmpSeq approach on cultured and clinical samples demonstrated that it provides high quality viral detection, preliminary genotyping, and predicts the virulence of NDV. It will be beneficial worldwide due to the rapid and comprehensive results, especially in developing countries where the endemicity of high-consequence diseases, such as NDV, and the lack of resources are additional challenges to monitoring and studying infectious diseases. MinION AmpSeq improves the depth of information obtained from PCRs and allows for more flexibility in assay design, which can be broadly applied to the detection and characterization of numerous infectious agents.

The development and optimization of sequencing protocols in the present study will further aid and supplement in detection and characterization of NDV isolates by current PCR-based rapid diagnostic assays. Additional testing of these protocols on a wide variety of clinical samples would be required to adopt these protocols to be used in diagnostic laboratories.