Genetic diversity of avian paramyxovirus type 1: Proposal for a unified nomenclature and classification system of Newcastle disease virus genotypes

Diego G. Diel a, Luciana H.A. da Silva a,b, Hualei Liu c, Zhiliang Wang c, Patti J. Miller a, Claudio L. Afonso a,⇑

a Southeast Poultry Research Laboratory, United States Department of Agriculture, Agricultural Research Services, 934 College Station Rd., Athens, GA 30605, United States
b Laboratory of Animal Virology, Institute of Biology, State University of Campinas, Campinas, SP, Brazil
c National Reference Laboratory for Newcastle Disease, China Animal Health and Epidemiology Center (CAHEC), 369 Nanjing Road, Qingdao 266032, China

Article history:
Received 20 April 2012
Received in revised form 11 July 2012
Accepted 12 July 2012
Available online 7 August 2012

Keywords:
APMV-1
NDV
Newcastle disease
Evolution
Epidemiology
Genotyping

Abstract

The avian paramyxovirus type 1 (APMV-1), or Newcastle disease virus (NDV), comprise a diverse group of viruses with a single-stranded, negative-sense RNA genome. Historically, two systems have been simultaneously used to classify NDV isolates into lineages or genotypes, generating confusion in the nomenclature and discrepancies in the assignment of genetic groups. In the present study we assessed the genetic diversity of the avian paramyxovirus type-1 (APMV-1) and propose a unified nomenclature and a classification system based on objective criteria to separate NDV into genotypes. Complete F gene sequences of class I \((n = 110)\) and class II \((n = 602)\) viruses were used for the phylogenetic reconstruction and to identify distinct taxonomic groups. The mean interpopulational evolutionary distance was estimated \((10\%)\) and set as the cutoff value to assign new genotypes. Results of our study revealed that class I viruses comprise a single genotype, while class II contains 15 genetic groups including 10 previously established \((I–IX, \text{ and XI})\) and five new genotypes \((X, XII, XIII, XIV \text{ and XV})\). Sub-genotypes were identified among class I and class II genotypes. Adoption of a unified nomenclature and of objective criteria to classify NDV isolates will facilitate studies on NDV epidemiology, evolution, disease control and diagnostics.

1. Introduction

Newcastle disease virus (NDV), or avian paramyxovirus type 1 (APMV-1), is a member of the genus Avulavirus of the family Paramyxoviridae (Alexander and Senne, 2008). The APMV type 1 comprise a diverse group of viruses with a single-stranded, negative-sense RNA genome (Alexander and Senne, 2008). As for other RNA viruses, the evolutionary dynamics of NDV is determined mainly by positive and negative selection and by the inherent error rate of the viral RNA dependent RNA polymerase (Holland et al., 1982; Charé et al., 2003; Miller et al., 2009b). While selection and polymerase errors are the main forces driving the evolution of NDV, it has been suggested that recombination may also play a role in shaping the genetic structure of the APMV type 1 (Han et al., 2008; Qin et al., 2008a,b; Miller et al., 2009b; Zhang et al., 2010).

Since the emergence of NDV in 1926, several genetic groups (lineages or genotypes) have been identified among the APMV-1 (Aldous et al., 2003; Cattoli et al., 2010; Miller et al., 2010), and phylogenetic studies have shown that viruses of distinct genetic groups undergo simultaneous evolutionary changes in different geographic locations of the world (Miller et al., 2010, 2009b). This evolutionary dynamics impose significant difficulties for disease control and diagnosis (Cattoli et al., 2010; Kim et al., 2007b; Miller et al., 2010; Rue et al., 2010). It has been shown, for example, that genotype-heterologous vaccines are less effective in preventing virus replication and shedding after challenge (Miller et al., 2007, 2009a). Therefore, genotype-homologous vaccines have been developed and are available in countries that experience significant economical burdens due to ND outbreaks (Hu et al., 2009, 2011; IASA – Investigación Aplicada S.A. de C.V., 2012). Evolutionary changes on the NDV genome have also been implicated in the failure of standard diagnostic tests to detect new genetic variants of the virus (Cattoli et al., 2010; Diel et al., 2011; Khan et al., 2010; Rue et al., 2010), requiring the development of new diagnostic tests to effectively detect variant viruses. These observations highlight the importance of constant epidemiologic and molecular surveillance for NDV.

Historically, NDV isolates have been classified into lineages or genotypes based on the phylogenetic analysis of the partial or complete nucleotide sequences of the fusion (F) gene (Aldous et al., 2003; Cattoli et al., 2010; Czeglédí et al., 2006; Kim et al., 2007a,b, 2008; Liu et al., 2003; Perozo et al., 2008; Snoeck et al., 2009). The lineage classification was defined by Aldous and collaborators, and initially grouped NDV isolates into six lineages \((1–6)\) and 13 sub-lineages (Aldous et al., 2003), with one additional lineage (lineage 7) and seven sub-lineages being proposed later.
The second system classifies NDV in two major groups designated class I and class II (Czeglédi et al., 2006; Miller et al., 2010; Seal et al., 1995). Class I has been divided in nine genotypes (1–9) (Kim et al., 2007a,b; Miller et al., 2010), while class II comprises eleven genotypes (I–XI) (Czeglédi et al., 2006; Miller et al., 2010; Mamininaia et al., 2010; Tsai et al., 2004), with genotypes I, II, VI and VII being further divided in sub-genotypes la and lb, Ila and IIa, Vla through Vlf, and Vlla through Vlhb (Kim et al., 2007a; Liu et al., 2003; Wehmann et al., 2003; Wu et al., 2010; Yu et al., 2001; Miller et al., 2009a, 2010). Both systems have been simultaneously used to classify NDV, generating confusion in the nomenclature and discrepancies in the assignment of multiple genetic groups. In the lineage system, for example, viruses of genotypes III, IV, V and VIII are classified into one large group named lineage 3 (Aldous et al., 2003). Other common problems of the current NDV classification systems include the large genetic diversity observed within sub-genotypes (sometimes larger than the genetic diversity between genotypes) (Miller et al., 2010), and the lack of objective criteria to define taxonomic groups. The genotype system is most widely used to classify NDV, as there is a stronger correlation among the phylogenetic relationship and the evolutionary distances between genetic groups. Rapid determination of the phylogenetic relationship between NDV isolates has, however, been hindered by the lack of uniformity in the nomenclature and by the lack of objective criteria to define major taxonomic groups. Therefore, the present study was conducted: (1). To obtain a clear representation of the genetic diversity of the APMV-1; (2). To develop a unified nomenclature for existing NDV taxonomic groups; (3). To define objective criteria to classify new NDV isolates; and (4). To provide a dataset with reference sequences for future classification of new NDV isolates.

Given the importance of the F gene for NDV virulence (de Leeuw et al., 2003, 2005), the large number of sequences available on GenBank, and the fact that current classification systems already use phylogenetic analysis of the F gene to classify NDV (Aldous et al., 2003; Kim et al., 2007b; Tsai et al., 2004), we have designed a classification system based on the phylogenetic analyses of all complete F gene sequences available on GenBank. Results of our analysis revealed that class I viruses contain a single genotype whereas class II NDV isolates can be effectively divided in 15 genotypes.

2. Materials and methods

2.1. Phylogenetic analyses and inference of the evolutionary distances

A preliminary phylogenetic analysis was performed with 1995 sequences (available on GenBank as of 12/12/2011) corresponding to the partial coding sequence of the F gene (374 nt) (data not shown). However, here we present data based on the analysis of the complete F gene sequences (class I, n = 110; and class II, n = 602) (Supplementary Table S1). The phylogenetic analysis was performed with the software MEGA5 (MEGA, version 5) (Tamura et al., 2011). The GTR + G + I model of nucleotide substitution was used for the analysis, with an estimated gamma shape parameter of 1.228. The robustness of the groups was assessed using the bootstrap approach with 100 replicates. The numbers presented in the phylogenetic trees represent the GenInfo Identifier (GI) sequence identification number in GenBank and the numbers in the tree nodes represent the bootstrap support.

The largest complete F gene dataset (n = 602), used for the phylogenetic reconstruction, was also used to infer the evolutionary distances within and between groups and to calculate the mean interpopulational evolutionary distance. The distances were inferred by pair-wise analysis using the MEGAS5 software (Tamura and Kumar, 2002, 2011) and the number of base substitutions per site was calculated by two different methods. The most simplified method (p-distance) was performed by averaging all sequence pairs between groups, while the second method used the Maximum Composite Likelihood model (Nei and Kumar, 2000). The variation rate among sites was modeled with a gamma distribution (shape parameter = 1, set based on a preliminary estimation from our dataset) and the changes in the composition bias among sequences were considered in the evolutionary comparisons (Tamura et al., 2011). Codon positions included in the analysis were the 1st, 2nd, 3rd, and non-coding. All positions containing gaps and missing data were eliminated.

2.2. Recombination analysis

Recombination analysis was performed with all class II complete genome (n = 103) and complete F gene (n = 602) sequences using the program RDP3 (Martin et al., 2010). Four statistical methods (RDP, Genecomv, Maxchi and Chimera) were used to identify putative recombinant sequences. Sequences with recombination events identified by at least two detection methods (p < 0.001) were considered as true recombinants.

2.3. Genotype classification criteria

Genotypes and sub-genotypes were assigned based on the phylogenetic topology and on the evolutionary distances between different taxonomic groups. Initially, a large phylogenetic tree was obtained with 1995 sequences corresponding to the partial coding sequence of the F gene (374 nt). As low bootstrap support was observed for most taxonomic groups, large phylogenetic trees were constructed with the complete nucleotide sequences of the F gene from class I (n = 118) and class II (n = 602) viruses and genotypes.
sub-genotypes were assigned according to the previous classification. New genotypes were assigned in numerical order and identified by the presence of a distinct common node shared by at least four isolates (bootstrap >60% at the defining edge). All complete F gene sequences of the genotypes and sub-genotypes identified in the phylogenetic analysis were used to infer the evolutionary distances between and within groups. To assign virus isolates to specific sub-genotypes, a separation into small clades was performed based on the Maximum Likelihood tree (complete F gene sequences; $n = 602$), and the distance between clades was inferred. The mean interpopulational evolutionary diversity (mean evolutionary distance between genotypes) was determined using the Maximum Composite Likelihood model (Tamura et al., 2004) (10%) and set as the cutoff value to confirm the classification based on the phylogenetic analysis. All taxonomic groups with less than four virus isolates were not classified into genotypes and were designated as “U” (unclassified). Similar classification criteria have been proposed by the World Health Organization, World Organization for Animal Health, Food and Agriculture Organization, and the H5N1 Evolution Working Group to classify highly pathogenic avian influenza virus isolates (WHO/OIE/FDA and Evolution Working Group, 2008).

2.4. New NDV isolates included in the study


2.5. RNA isolation and nucleotide sequencing

The NDV isolates Malaysia/1041-632/2008, Malaysia/5091-633/2009, China/NDV04-23-647/2004, China/NDV03-45-641/2003, chicken/MX/NC02-634/2010, chicken/MX/NC04-635/2010, goose/China-GD/12/2011, goose/China-GD/17/2011, goose/China-GD/20/2011, goose/China-GD/1003/2010, and goose/China-GD/450/2011, were propagated in specific pathogen free (SPF) embryonated chicken eggs. Total RNA was extracted from allantoic fluids using Trizol LS (Invitrogen, Carlsbad, CA) following the manufacturer’s instructions. The complete F gene nucleotide sequences were determined by using a RT-PCR/sequencing approach (Diel et al., 2011). Amplification reactions were performed with a one-step RT-PCR kit (Qiagen, Valencia, CA) and a set of F gene specific primers (fwd_upper-f1-ttgcttatagttagttgctgc, and rev_down-f2-acccgtcttgcttctg). The PCR amplicons were subjected to electrophoresis in 1% agarose gels and the DNA bands were excised from the gels and purified by using the QuickClean DNA gel extraction kit (Qiagen, Valencia, CA). The purified PCR products were cloned in the TOPO TA vector (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions and subjected to DNA sequencing. All sequencing reactions were performed with fluorescent dideoxy-nucleotide terminators in an ABI 3700 automated sequencer (Applied Biosystems Inc., Foster City, CA). Sequence editing and assembly were performed with LaserGene sequence analysis software package (LaserGene, version 5.07; DNASTar, Inc., Madison, WI).

2.6. Accession numbers


3. Results and discussion

In the present study, we assessed the genetic diversity of the APMV type 1 and developed a unified nomenclature and classification system to define NDV genotypes. Two previous classifications that grouped NDV isolates into lineages or genotypes were considered (Aldous et al., 2003; Miller et al., 2010; Pedersen et al., 2004; Seal et al., 1995). However, given that in the genotype classification there is a stronger correlation between the phylogenetic topology and the evolutionary distances within and between clades, we have used the genotype nomenclature and classification as the base to develop objective criteria to classify new NDV isolates. The classification criteria were determined based on the phylogenetic topology and on specific evolutionary distances that reflect the diversity of the APMV type 1. In the sections that follow, we will describe and discuss the results of our analysis and proposals for the future assignment of new NDV genotypes and sub-genotypes as they are discovered.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Criteria used to define Newcastle disease virus genotypes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>Description</td>
</tr>
<tr>
<td>1</td>
<td>Genotype and sub-genotype designations were maintained as previously described when possible (i.e. genotype I remains I and sub-genotype Ia remains Ia)</td>
</tr>
<tr>
<td>2</td>
<td>New genotypes and sub-genotypes were assigned based on the phylogenetic tree topology using the Maximum Likelihood method and the optimum nucleotide model (GTR +G +I) as determined by MEGAS. Phylodynamic tree topology was supported by the evolutionary distances between groups</td>
</tr>
<tr>
<td>3</td>
<td>New genotypes or sub-genotypes were designated only when the complete F gene sequence of at least four independent isolates without a direct epidemiologic link (i.e. distinct outbreaks) were available</td>
</tr>
<tr>
<td>4</td>
<td>The mean nucleotide distance (evolutionary distances) between and within groups was inferred as base substitutions per site by averaging all sequence pairs</td>
</tr>
<tr>
<td>5</td>
<td>The mean nucleotide distance (evolutionary distances) between and within groups was estimated using the Maximum Composite Likelihood model (0.106; standard error = 0.0061) and set as the cutoff value to assign new genotypes</td>
</tr>
<tr>
<td>6</td>
<td>Different genotypes should have an average distance per site &gt;10% (0.1) Different sub-genotypes should have an average distance per site between 3 (0.03) and 10% (0.1)</td>
</tr>
<tr>
<td>7</td>
<td>Bootstrap value at the genotype and sub-genotype defining node should be &gt;60%</td>
</tr>
</tbody>
</table>
3.1. Classification of NDV into genotypes

The rationale for classifying viruses into genotypes is based on the need to determine the association of viruses within a genotype with their geographic and temporal distribution, their pathogenicity, and the host species from which they were isolated (Czeglédi et al., 2006). NDV isolates have been classified into multiple genotypes and sub-genotypes based on the phylogenetic relationship between virus isolates (Czeglédi et al., 2006; Khan et al., 2010; Kim et al., 2007a, 2008; Liu et al., 2003; Tsai et al., 2004; Wu et al., 2010; Yu et al., 2001). However, the lack of objective criteria to define major taxonomic groups makes the classification of new NDV isolates difficult. Here we designed a system to classify NDV based on the phylogenetic topology and on the evolutionary distances between different taxonomic groups (Table 1).

The evolutionary history of NDV was inferred by phylogenetic analysis using the partial (374 nt) or complete nucleotide sequences of the F gene, and compared to the evolutionary history inferred from the complete genome sequences. Phylogenetic analysis based on the partial (374 nt; n = 1995) or complete (1649 nt; n = 602) F gene sequences clearly separated class II NDV isolates into 15 taxonomic groups (data not shown; Supplementary Fig. S1). However, when the partial F gene sequence (374 nt) was used for the analysis, inconsistencies in the phylogenetic tree topology and low bootstrap values (<60%) were observed for most taxonomic groups (data not shown). The overall phylogenetic topology based on the complete F gene sequences was consistent with the phylogenetic grouping obtained from the analysis of the complete genome sequences (congruence index [I(cong)] = 5.410, p = 2.24e-45). Additionally, when the alternative algorithm PhyML was used for the phylogenetic reconstruction (n = 118, complete F gene sequences), the grouping of all viruses was consistent with the grouping obtained in MEGAS, and all genotypes and sub-genotypes were supported by bootstrap values >60% (Supplementary Fig. S2, Fig. 3). Therefore, the complete F gene sequence was selected for all other analyses conducted here and is proposed for future classification of NDV isolates.

3.2. Classification criteria to define NDV genotypes

The specific classification criteria determined here by the analysis of all complete F gene sequences available on GenBank and used to define NDV genotypes are presented in Table 1. To avoid further inconsistencies in the classification of NDV isolates, previously assigned genotypes and sub-genotypes were maintained when possible. New genotypes were assigned when at least four isolates formed a distinct taxonomic group (bootstrap at the defining node >60%) with a mean evolutionary distance between other groups greater than 10%. The mean interpopulational evolutionary distance (10%) is proposed here as a cutoff to validate genotypes identified by phylogenetic analysis, because it represents the average diversity between distinct NDV taxonomic groups. The bootstrap value and the mean interpopulational evolutionary distance have also been proposed for the classification of highly pathogenic avian influenza viruses (WHO/OIE/FAA and Evolution Working Group, 2008). By using these specific criteria we identified one genotype in class I and 15 genotypes in class II (described in detail below). These genotypes were identified from the consensus topology of large phylogenetic trees (class I, n = 102; class II, n = 602) and were confirmed by the analysis of the evolutionary distances between and within groups (Supplementary Fig. S1; Figs. 1–3; Table 2, Supplementary Tables S2, S3, S5). Genotypes and sub-genotypes of class II, assigned based on large phylogenetic trees, consistently grouped (Fig. 3) when a smaller number of isolates (at least four viruses) were used for the phylogenetic reconstruction. Although genotypes were consistently identified between the large and the small datasets, slight differences in the topologic organization and in the bootstrap values were observed (Supplementary Fig. S1; Fig. 3). These observations highlight the importance of using the largest dataset possible to classify NDV isolates into genotypes. The phylogenetic topology of each genotype was similar, and all viruses classified within the same taxonomic group when the Maximum Likelihood and the Neighbor-Joining phylogenetic reconstruction methods were used (data not shown).

3.3. Analysis of class I viruses revealed the existence of a single genotype

Class I comprise viruses that have been isolated mainly from waterfowl and shorebirds, usually from samples collected in live bird markets worldwide (Kim et al., 2007a, b; Miller et al., 2010, 2009a). These viruses have been previously classified by our group and by others into nine genotypes (1 to 9), based on phylogenetic analysis performed with partial nucleotide sequences (374 nt) of the F gene (Kim et al., 2007a; Miller et al., 2010; Wu et al., 2010). Results here, however, indicate that according to the new classification criteria these viruses comprise a single phylogenetic group, which was designated as genotype 1. The phylogenetic relationship between viruses of class I suggested the existence of at least three distinct sub-groups within genotype 1 (Fig. 1). Analysis of the evolutionary distances between these sub-groups (Table 2) confirmed the phylogenetic topology, indicating that genotype 1 is divided in three sub-genotypes (1a, 1b, and 1c; distances between groups varying from 0.076 [7.6%] to 0.079 [7.9%]) (Fig. 1; Table 2). The differences between the previous classification of class I viruses and the one proposed here are most likely due to the use of more stringent criteria (larger distances and requirement of four independent isolates) to clearly define the genotypes. However, the possibility that additional class I genotypes exist and were not identified in our analysis cannot be formally excluded (e.g. there are two clearly distinct groups of viruses in Fig. 1 that were assigned as ‘U’ because there is only three sequences in GenBank). An analysis involving a larger number of sequences should be performed in the future as more complete F gene sequences of class I viruses become available.

3.4. Analysis of class II viruses revealed the existence of 15 genotypes

Class II contains viruses that have been isolated from multiple wild bird and poultry species (Supplementary Table S1; Fig. 3). Most viruses within this group are virulent (Miller et al., 2009a) and cause significant economic losses to the poultry industry worldwide. These viruses are also highly diverse and multiple phylogenetic studies have shown that they are continuously evolving (Czeglédi et al., 2006; Diel et al., 2011; Khan et al., 2010; Maminina et al., 2010; Miller et al., 2010; Pedersen et al., 2004; Perozo et al., 2008; Rue et al., 2010; Wu et al., 2010). Given the large genetic diversity observed within class II, these viruses have been classified in 11 genotypes (I–XI) in the past (Tsai et al., 2004). In our study, we confirmed the existence of 10 of these genotypes (I–IX, and XI), with genotype X, previously assigned by Tsai and collaborators (Tsai et al., 2004), not being included in our analysis due to the lack of complete F gene sequences on public databases. By using the genotyping criteria developed here (Table 1), we have also identified five new genotypes, which were designated as genotypes X, XII, XIII, XIV and XV (Figs. 2 and 3; Table 3). The classification of class II viruses into 15 genotypes (I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, and XV) is supported by the phylogenetic topology and by the evolutionary distances between (>10%) and within (<10%) groups (Figs. 2 and 3; Tables 3 and 4).
Fig. 1. Phylogenetic analysis based on the complete nucleotide sequence of the F gene of viruses representing Newcastle disease virus class I. The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood \((-8604.4794)\) is shown. Initial tree(s) for the heuristic search were obtained automatically as follows. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.6182). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 27.9736% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 110 nucleotide sequences. Codon positions included were the 1st, 2nd, 3rd, and non-coding. All positions containing gaps and missing data were eliminated. There were a total of 1662 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.
The evolutionary distances (inferred from the complete F gene sequences) between most class II genotypes was greater than 10% (Table 3), which is the mean interpopulational evolutionary distances between them (Table 3). Since the evolutionary distances between these three genotypes were close to 10% (the value set as cutoff for genotype assignment), and these groups represent mostly isolates obtained several years ago we decided to maintain their classification as distinct genotypes, to avoid inconsistencies with the existing literature. The phylogenetic tree topology (bootstrap >60%) (Supplementary Fig. S1), and the evolutionary distances inferred from the complete genome sequences (>10%) (Supplementary Table S2) further support their classification as distinct genotypes. The classification of class II viruses into 15 genotypes was also supported by the phylogenetic and evolutionary distances inferred from the analysis of the nucleotide sequences of the complete genome (data not shown; Supplementary Table S2).

Genotypes I, II, V, VI and VII comprise highly diverse viruses, therefore these groups have been further divided in sub-genotypes (Kim et al., 2007a; Liu et al., 2003; Miller et al., 2010; Qin et al., 2008a,b; Tan et al., 2010; Wang et al., 2006; Wu et al., 2010). The criteria used here and proposed for future sub-genotype assignment are presented in Table 1. Genotype I has been previously classified in sub-genotypes Ia and Ib (Kim et al., 2007), which was confirmed here by the phylogenetic relationship and the evolutionary distances between viruses of these sub-genotypes (Supplementary Fig. S1; Fig. 3; Table 4A). Genotype II has been previously divided into two sub-genotypes named sub-genotype II and IIa (Kim et al., 2007; Miller et al., 2010). Here, however, analysis of the evolutionary distances between these two groups indicated that viruses of sub-genotype IIa form a distinct genotype, which was designated genotype X (Fig. 3; Supplementary Fig. S1; Supplementary Table S3). Although genotype V comprises a highly diverse group of viruses, this genotype has not been previously classified in sub-genotypes. Analysis of the evolutionary distances between different taxonomic groups identified within genotype V indicated the existence of at least two sub-genotypes among these viruses, which were designated here as sub-genotypes Va and Vb (Fig. 3; Supplementary Fig. S1; Table 4B). Genotype VI has been classified into eight sub-genotypes (Wang et al., 2006), and here we confirmed the existence of four of these sub-genotypes consisting of sub-genotypes Via, Vlb, Vlc, and Vle (Fig. 3; Supplementary Fig. S1; Table 4A), according to the nomenclature previously used in the literature.

### 3.5. Recombination analysis

An increasing number of recombination events have been recently reported for NDV. (Han et al., 2008; Qin et al., 2008a,b; Miller et al., 2009b; Zhang et al., 2010). However, the contribution of recombination for NDV evolution and its occurrence in nature are still largely debated (Afonso, 2008; Collins et al., 2008; Han and Worobey, 2011; Song et al., 2011). The dataset used in the present study was subjected to a recombination analysis, and using the program RDP3 we identified 14/103 (13.5%) recombinant sequences in the complete genome dataset and 15/602 (2.5%) recombination events in the complete F gene dataset (Supplementary Tables S3 and S4). To assess the effect of recombination on the evolutionary distances between genotypes (Table 3 and Supplementary Table S2), recombinant sequences were removed from the dataset and the distances between genotypes were re-inferred (Supplementary Tables S5 and S6). As shown on Supplementary Tables S5 and S6 the presence of recombinant sequences had only minimal effects on the evolutionary distances between NDV genotypes (Supplementary Tables S2 and S3). Therefore, recombinant sequences were kept in our dataset and labeled in the phylogenetic trees as recombinant forms (RF) (Fig. 3; Supplementary Fig. S1). Notably, all sequences identified as recombinants in our analysis, lack a common recombination breakpoint (Supplementary Table S5), suggesting that these recombination events may not be of evolutionary significance.

### 3.6. Epidemiologic characteristics of class II NDV genotypes

Genotypes I, II, III, IV, and IX emerged between 1930s and 1960s and are considered “early” genotypes (Czeglédi et al., 2006). Among these, genotype II comprise mainly viruses of low virulence and strains of this genotype have been used as live vaccines against NDV for more than 40 years (Miller et al., 2010). Genotype IX includes the first virulent NDV strain obtained in China in 1948, and viruses of this genotype are still circulating in Asia with sporadic outbreaks being reported in chickens and domestic ducks (Qiu et al., 2011; Wang et al., 2006; Zhang et al., 2011). Genotypes V, VI, VII, VIII and XI emerged after 1960s and are considered “late” genotypes (Czeglédi et al., 2006). Genotypes V, VI, VII and VIII
Fig. 3. Phylogenetic analysis based on the complete nucleotide sequence of the F gene of viruses representing Newcastle disease virus class II. At least four sequences representing each NDV genotype or sub-genotype were used in the analysis. The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood (−23299.0541) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically as follows. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 1.2175). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 30.6963% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 118 nucleotide sequences. Codon positions included were 1st, 2nd, 3rd, and non-coding. All positions containing gaps and missing data were eliminated. There were a total of 1647 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. RF = recombinant form identified with the program RDP3.
contain only virulent viruses and are the predominant genotypes circulating worldwide (Miller et al., 2009a). Among these, geno-
type V viruses emerged in 1970s and are frequently isolated in South and Central America from poultry species and in North America from cormorants (Diel et al., 2011; Kinde et al., 2005; Pedersen et al., 2004; Perozo et al., 2008; Rue et al., 2010; Weh-
mann et al., 2003). Genotype VI, include viruses that have been iso-
lated from multiple avian species (Alexander, 2011; Czeglédi et al.,
2006; Kim et al., 2008; Liu et al., 2003; Wehmann et al., 2003).
Viruses of this genotype are particularly important because of their
frequent association with doves and pigeons and the consequent
risk for introduction into poultry flocks (Alexander, 2011; Kim et al.,
2008; Kammers et al., 2001, 2002). Currently, genotype VII is the
most frequently associated with outbreaks of ND in the Middle East and Asia (Miller et al., 2010), and outbreaks
caused by these viruses are of particular concern given that some
strains have shown increased virulence in poultry, while others
have expanded their host range and are now able to cause disease in
goose (Huang et al., 2004; Liu et al., 2003). Additionally, an out-
break of ND in South America has been recently attributed to a
genotype VII virus, indicating that these viruses are spreading to
other locations around the world (Perozo et al., 2012).

Genotypes X, XII, XIV and XV comprise viruses with differ-
ent epidemiologic characteristics and geographic origins (Fig. 3;
Supplementary Table S1). New genotype X, which contains viruses
that have been previously classified into genotype Ila (Kim et al.,
2007a), comprise viruses of low virulence that have been isolated from
waterfowl and shorebirds in North America between late 1980s and early 2000s. Genotype XII is a novel genotype that
contains virulent viruses recently isolated from poultry in South
America (Diel et al., 2012) and from goose in China (isolates goose/China-GD/12/2011, goose/China-GD/17/2011, goose/China-
450/2011 characterized in the present study). The South American
isolate of genotype XII (poultry/Peru/2011-03/2008) caused an
outbreak of ND in Peru in 2008 (Diel et al., 2012), and phylogenetic
related viruses have been recently isolated from healthy goose in
live bird markets in China. The epidemiologic link between the
Peruvian and the Chinese viruses is still unknown and further
investigation is needed to define the origin and source of these
viruses. Genotype XIII, previously classified as genotype VII (Khan
et al., 2010), comprises virulent viruses that were isolated in Rus-
sia, Iran and Pakistan between 1995 and 2008 (Fig. 3). Genotype XIV contains virulent viruses obtained in West and Central Africa
between 2006 and 2008 (Fig. 3). These viruses have been previ-
ously classified by using the lineage system into lineage 7 (Cattoli
et al., 2010). Genotype XV, comprises viruses obtained from chick-
gen and goose in China, which have been previously classified into
sub-genotype VId (isolates XJ-2/97 and FJ-2/99) or Vlle (isolate JX-
2/99) (Liu et al., 2003). The re-classification proposed here, con-
taining 15 genotypes, is supported by the phylogenetic relation-
ship and the evolutionary distances between genotypes (Figs. 2 and
3; Tables 3 and 4). A summary with the epidemiologic information of all viruses used in the present study is presented in
Supplementary Table S1.

4. Conclusions

Given the rapid evolutionary dynamics of NDV, its broad host
range (over 250 species of birds) and the use of mass vaccination
strategies to control the disease worldwide, additional genetic
variants will likely be discovered. To ensure their correct and con-
sistent classification it is essential to use a unified nomenclature

Table 2
Estimates of evolutionary distances between class I sub-genotypes.a

<table>
<thead>
<tr>
<th>Sub-genotype</th>
<th>No. of base substitutions/site in sub-genotypesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>1b 0.071 (0.007) 1c 0.079 (0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Inferred from the complete nucleotide F gene sequence.</td>
</tr>
</tbody>
</table>

b The number of base substitutions per site from averaging all sequence pairs between class I sub-genotypes is shown. All results are based on the pair wise analysis of 110 sequences and the number of sequences analyzed per groups was: 1a, n = 44; 1b, n = 22; 1c, n = 36; Unclassified, n = 8. Evolutionary distances were inferred by the analysis of 1662 nt positions of the F gene. The analysis was con-
ducted using the Maximum Composite Likelihood method in MEGAS. Codon posi-
tions included were 1st, 2nd, 3rd, and non-coding. All positions containing gaps and
missing data were eliminated. The rate variation among sites was modeled with a
gamma distribution (shape parameter = 1). The differences in the composition bias
among sequences were considered in evolutionary comparisons. Values in paren-
theses represent standard errors, obtained by a bootstrap procedure (500 replicates).

Table 3
Estimates of evolutionary distances between class II genotypes.a

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of base substitutions/site in genotypeb</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I 0.128 (0.012) II 0.116 (0.012) III 0.101 (0.012) IV 0.194 (0.013) V 0.177 (0.011) VI 0.179 (0.011) VII 0.169 (0.011) VIII 0.160 (0.011) IX 0.140 (0.011) X 0.137 (0.011) XI 0.139 (0.011) XII 0.127 (0.012) XIII 0.121 (0.012) XIV 0.113 (0.012) XV 0.103 (0.012)</td>
</tr>
</tbody>
</table>

a Inferred from the complete nucleotide F gene sequence.

b The number of base substitutions per site from averaging all sequence pairs between class II genotypes is shown. All results are based on the pair wise analysis of 602 sequences and the number of sequences analyzed per groups was: I, n = 70; II, n = 101; III, n = 9; IV, n = 6; V, n = 44; VI, n = 56; VII, n = 234; VIII, n = 4; IX, n = 18; X, n = 17; XI, n = 4; XII, n = 6; XIII, n = 8; XIV, n = 5; XV, n = 4; Unclassified, n = 12. Evolutionary distances were inferred by the analysis of 1662 nt positions of the F gene. The analysis was con-
ducted using the Maximum Composite Likelihood method in MEGAS. Codon posi-
tions included were 1st, 2nd, 3rd, and non-coding. All positions containing gaps and
missing data were eliminated. The rate variation among sites was modeled with a
gamma distribution (shape parameter = 1). The differences in the composition bias
among sequences were considered in evolutionary comparisons. Values in parentheses are standard errors, obtained by a bootstrap procedure (500 replicates).
Appendix A. Supplementary data

Table 4

<table>
<thead>
<tr>
<th>Genotype I sub-genotype</th>
<th>No. of base substitutions/site in sub-genotype</th>
<th>Genotype VII sub-genotype</th>
<th>No. of base substitutions/site in sub-genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b (0.006)</td>
<td>b</td>
<td>(0.003)</td>
</tr>
<tr>
<td>b</td>
<td>0.07</td>
<td>d</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.007)</td>
</tr>
<tr>
<td>b</td>
<td>0.089</td>
<td>0.084</td>
<td>b (0.008)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>c (0.007)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

a The percentage representing the number of base substitutions per site from averaging all sequence pairs between Class II sub-genotypes is shown. All results are based on the pair wise analysis of 602 sequences and the number of sequences analyzed per group was: Ia, n = 37; Ib, n = 32; Va, n = 24; Vb, n = 12; Vla, n = 27; Vlb, n = 16; Vlc, n = 10; Vle, n = 7; Vlh, n = 86; Vld, n = 116; Vlle, n = 9; Vlf, n = 5. Distances were inferred by the analysis of 1662 nt positions of the F gene. The analysis was conducted using the Maximum Composite Likelihood method in MEGA5. Codon positions were included as 1st, 2nd, 3rd, and non-coding. Codon positions included were 1st + 2nd + 3rd + non-coding. All positions containing gaps and missing data were eliminated. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The differences in the composition bias among sequences were considered in evolutionary comparisons. There were a total of 1662 positions in the final dataset. Values in parentheses are standard errors, obtained by a bootstrap procedure (500 replicates).

b Inferred from the complete F gene sequence.

References


