Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genomic Diversity and Evolutionary Insights of Newcastle Disease Virus(Ndv)

Eyerusalem Fetene

eyerusfetene@gmail.com

KU Leuven

Tesfaye Sisay Tessema

Addis Ababa University

Jan Paeshuyse

KU Leuven

Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11364678/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GENOMIC DIVERSITY AND EVOLUTIONARY INSIGHTS OF NEWCASTLE DISEASE VIRUS(NDV)

Eyerusalem Fetene[1*], Tesfaye Sisay Tessema[2], Jan Paeshuyse[1*]

1 KU Leuven, Department of Biosystems, Laboratory of Host Pathogen Interactions in Livestock, Leuven, Belgium

2 AAU, Biotechnology Research Center, Ethiopia

*Corresponding Author: Jan Paeshuyse, email: Jan.paeshuyse@kuleuven.be

*Corresponding Author: Eyerusalem Fetene, email: eyerusfetene@gmail.be

ABSTRACT

Newcastle disease virus (NDV) is a non-segmented, negative-sense single-stranded RNA virus of the family *Paramyxoviridae*. NDV is classified as a Class I virus and a Class II virus, with Class II viruses exhibiting substantially greater genetic diversity. Despite the increasing availability of complete NDV whole-genome sequences, comprehensive comparative analyses of genomic diversity and evolutionary dynamics across geographically diverse populations remain limited. Therefore, this study investigated NDV genomic diversity, recombination patterns, phylogenetic relationships, and evolutionary dynamics using complete whole-genome sequences from diverse geographic regions. We retrieved 935 whole-genome NDV sequences from NCBI. These sequences, reported from 48 countries, included 132 from class I, 798 from class II, and the remaining 5 were unclassified. Both class I and class II viruses show high nucleotide diversity (π) of 0.05754 and 0.10777 and average pairwise nucleotide differences (k) of 605.8 and 1273.9, respectively. Recombination detection analysis revealed 3 putative recombination events in class I viruses and 26 events in class II viruses. Bayesian evolutionary analyses estimated mean evolutionary rates of 2.04×10^{-3} and 1.133×10^{-3} substitutions/site/year for Class I and Class II viruses, respectively. The estimated times to the most recent common ancestor were 1878 and 1851 for Class I and Class II viruses, respectively. Overall, these findings demonstrate substantial genomic diversity and distinct evolutionary dynamics between NDV classes and provide insights into the temporal and genetic processes shaping NDV populations. The results contribute to a better understanding of NDV evolutionary history and provide a genomic framework to support molecular surveillance, risk assessment, and the development of effective disease-control strategies.

INTRODUCTION

Avian Paramyxovirus serotype 1 (APMV1), also known as Newcastle Disease Virus (NDV), belongs to the genus *Avulavirus* of the family *Paramyxoviridae*. It is the causative agent of Newcastle Disease (ND), which can affect around 236 species of wild birds and poultry. ND is reportable to the World Organization for Animal Health (WOAH). It can affect the respiratory, gastrointestinal, nervous, and reproductive systems, and cause up to 100% mortality in immunologically naive chickens[1]. Despite efforts to control ND through vaccination and strict biosecurity and quarantine measures, the disease remains endemic in many countries across Africa, the Middle East, and Asia[23].

NDV is a pleomorphic enveloped virus that contains a single-stranded, non-segmented RNA genome of ~15.2 kilobases[4]. Although genome length varies, the virus encodes six structural proteins: nucleocapsid protein (NP), phosphoprotein (P), matrix protein (M), fusion protein (F), hemagglutinin-neuraminidase protein (HN), and large RNA-dependent RNA polymerase protein (L) plus two non-structural proteins (V and W) encoded by RNA editing in the phosphoprotein (P) gene[5]. The F and HN proteins play a significant role in viral attachment, fusion, and entry into host cells, as observed in other paramyxoviruses. At the same time, NP, P, and L interact with viral RNA to form the ribonucleoprotein complex required for viral RNA replication.

NDV is divided into two groups, namely class I and class II. Class I isolates are classified as a single genotype with four sub-genotypes, while class II isolates were found to be much more diverse, divided into 18 genotypes, many of which are further differentiated into sub-genotypes[6]. Recent advancements have prompted the virus's reclassification, comprehensively reevaluated its genetic lineages, and provided fresh insights into virus diversity[7]. The extensive circulation of Newcastle disease virus (NDV) among domestic poultry and wild birds, together with its broad host range and worldwide distribution, has contributed to substantial genetic diversity and the emergence and dissemination of genetically distinct viral lineages[8].

Studying complete whole-genome sequences allows examination of phylogenetic relationships, genotype distribution, nucleotide diversity, recombination patterns, and temporal evolutionary dynamics, as highlighted[9–11]. For example, Miller et al. provided important evidence for the roles of mutation, selection, and recombination in NDV evolution[10]. Since

then, the number of publicly available sequences has expanded considerably, accompanied by the identification of additional genetic lineages and an internationally coordinated revision of NDV classification[7].

Recent studies have revealed substantial genetic diversity and complex evolutionary dynamics of NDV at regional and genotype-specific levels. Analyses of genotype VII Class II viruses and Class I viruses in China demonstrated multiple circulating lineages, geographic dissemination, and temporal changes in evolutionary and population dynamics[12,13]. Genome-wide analyses have further highlighted the roles of recombination, selection, and vaccination-associated pressures in shaping NDV evolution[14]. However, many of these studies have focused on specific regions, genotypes, or individual evolutionary processes, with limited integration of genome-wide diversity, phylogeny, recombination, and temporal dynamics at a global scale. Therefore, a comprehensive analysis of a large, carefully curated whole-genome dataset is warranted to characterize the global genomic diversity and evolutionary dynamics of both Class I and Class II NDV within an integrated framework.

This study used a large, geographically diverse collection of complete NDV genomes to characterize the phylogenetic structure, genetic diversity, genotype distribution, and genome-wide variation of Class I and Class II viruses. We further assessed recombination and evolutionary rates. By integrating these approaches, this study provides an updated assessment of NDV's global genomic diversity and evolutionary dynamics.

RESULTS

Comparative genome features

A total of 935 whole-genome sequences were reported from 48 countries, with China taking the lead by reporting 548 sequences (58.6%) (Table 1). Of the 935 sequences, 132 (14%) were Class I, 798 (85.3%) were Class II, and the remaining 5 (0.7%) were unidentified. Comparative genomic analysis revealed variable genome lengths ranging from 15060 to 15240 nucleotides. Most sequences had 15192 nucleotides, followed by 15186 and 15198 nucleotides (Supplementary Table S1).

The total length of genes NP, P and F varied across the whole genome, for the NP gene (1467-1563nt), encoding 488-520aa, the P gene (1188-1200nt), encoding 395-399aa, the F gene (1623-1683nt), encoding 540-564aa, and the HN gene (1716-1851nt), encoding 571-610aa.

The length was constant for the M gene (1095 nt), encoding 364 aa, and the L gene (6615 nt), encoding 2204aa.

Nucleotide Diversity

Analysis of the concatenated whole-genome sequences revealed substantial nucleotide polymorphism within the Class I viruses. A total of 13937 sites, 4734 polymorphic and 5795 monomorphic sites, were examined. The polymorphic sites within Class I generated 6451 total mutations, including 3978 parsimony-informative sites and 756 singleton variable sites. The nucleotide diversity (π) for Class I was 0,05754, with an average pairwise nucleotide difference (k) of 605,8. Analysis of haplotypic variation identified 124 distinct haplotypes with a haplotype diversity (Hd) of 0,9990. The mean Tajima's D value for Class I viruses was (-1.32), with a minimum of -1.53 and a maximum of -1.04.

For class II viruses, among 14,160 sites were examined, 7,636 were polymorphic, and 4,185 were monomorphic. The polymorphic sites showed 12,214 mutations in total, with 6,458 parsimony-informative sites and 1,178 singleton variable sites. Nucleotide diversity (π) was estimated at 0.10777, with an average pairwise nucleotide difference (k) of 1273.9. Analysis of sequence variation further identified 762 distinct haplotypes, with a haplotype diversity (Hd) of 0.9998, indicating extremely high haplotypic variation within the dataset. Tajima's D values in the sliding window ranged from -0.71 to 0.07, with a mean of -0.35.

Phylogenetic Tree

The maximum-likelihood phylogenetic analysis revealed substantial genetic diversity among the Newcastle disease virus (NDV) whole-genome sequences. The 132 Class I whole-genome sequences showed considerably less genotype-level diversity. All Class I viruses clustered within a single genotype, which was further subdivided into three subgenotypes (Figure 1). In contrast, among the 798 Class II whole-genome sequences, 18 distinct genotypes were identified based on clustering within the established NDV phylogenetic classification framework (Figure 2). Genotype representation varied, with genotype VII comprising the largest proportion of the dataset (269 sequences), followed by genotypes VI and II (224 and 52 sequences, respectively.

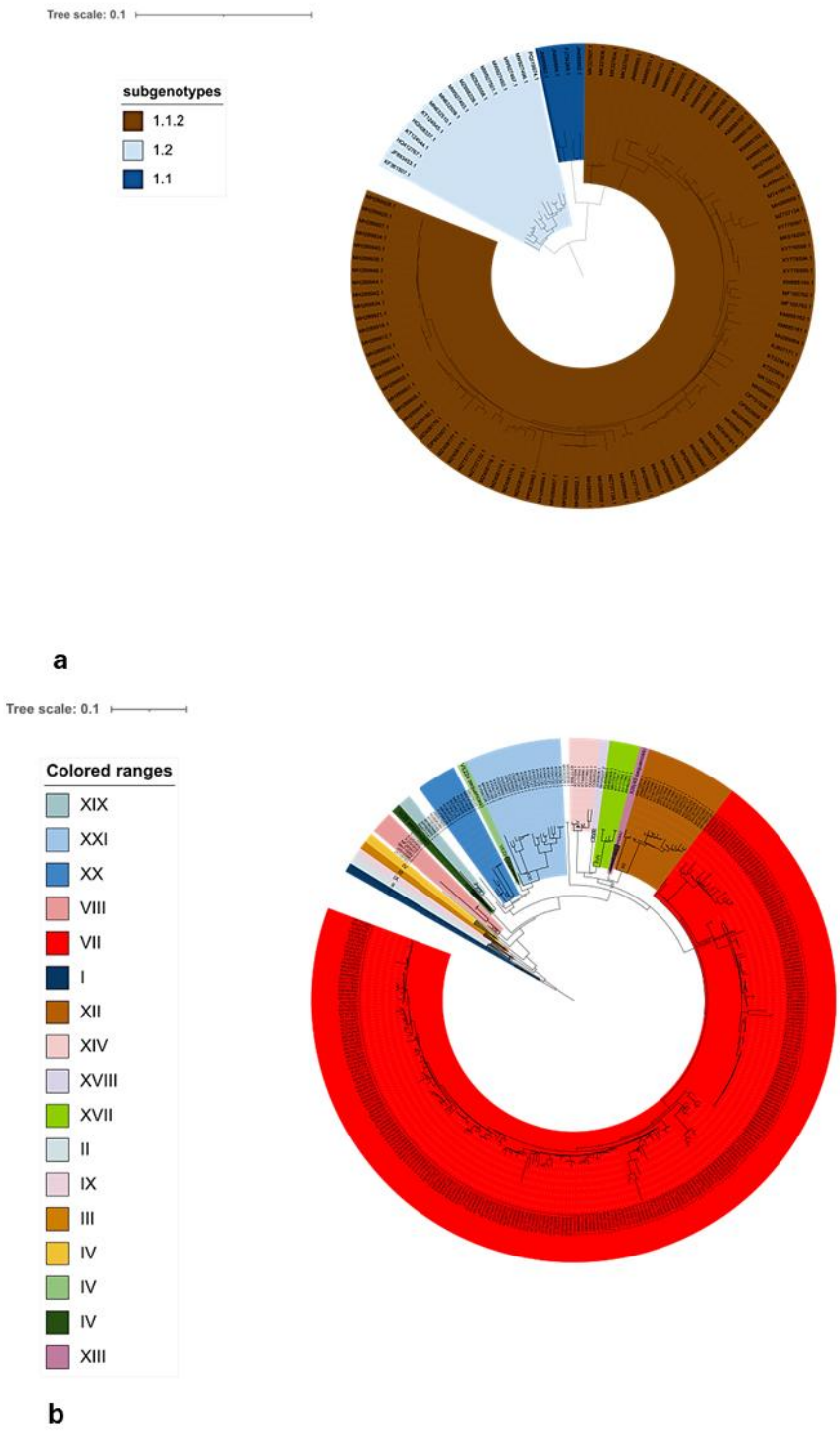


Figure 1. Maximum-likelihood phylogenies of Newcastle disease virus (NDV) Class I and Class II genomes**.** (a) Class I tree, with subgenotypes 1.1, 1.2 and 1.1.2 distinguished by colour. (b) Class II tree, with genotype clades distinguished by colour. Each tip represents a sequence, and the branches show the inferred evolutionary relationships among the sequences.

Geographical distribution of genotypes

Class I whole-genome sequences were reported from only three countries, with China representing 113 (85.6%), followed by Russia and Taiwan, with 11 (8.4%) and 8 (6%) sequences, respectively. All three subgenotypes (1.1.2, 1.2 and 1.1) were reported from China, whereas only subgenotype 1.2 was reported from Russia and Taiwan (Table 1).

Table 1: Geographic distribution, temporal representation, and genotype diversity of NDV class I sequences by country

No	Continent	Country	Number of sequences	Number of genotypes	First year	Last year	Genotypes reported
0	Asia	China	113	3	2006	2024	1.1, 1.1.2, 1.2
1	Europe	Russia	11	1	2014	2022	1.2
2	Asia	Taiwan	8	1	2013	2020	1.2

The geographic representation of class II whole-genome NDV sequences was highly uneven across countries. Sequences have been reported from 47 countries, with China contributing the most sequences (n = 431 (54.%)) and showing the broadest genotype representation, including genotypes I, II, III, IX, VI, VII, VIII, XII, and XX. India was the second most represented country (n = 56 (7%)), followed by Pakistan (n = 38(~5%)), Kenya (n = 30 (~4%)), Belgium (n = 24(3%)), and the USA (n = 21(~2.6%)) (Supplementary Table S2). Whole-genome sequences of genotype VII have been reported from 26 countries, followed by VI and II from 15 and 12 countries, respectively.

A total of 92 sequences were represented from African countries; Kenya contributed the most sequences (n = 30), followed by Tanzania (n = 19), Egypt (n = 17), Nigeria (n = 15), and Ethiopia (n = 4). Genotype composition varied among these countries: Tanzania included genotypes V, VII, and XIII; Egypt included II, VII, and XXI; Nigeria included XIV, XVII, and XVIII; while Ethiopia was represented by genotype VII in this dataset.

Temporal distribution of genotypes

Temporal distribution of class I genotype I showed that subgenotypes 1.1.2 and 1.2 whole-genome sequences were most represented throughout most of the study period from their first presentation in 2006 and 2007, respectively. Subgenotype 1.1.2 showed the greatest sequence representation, particularly around 2015, while 1.2 remained less abundant but consistently

detected over time. In contrast, subgenotype 1.1 was restricted to 2008-2010 and was represented by only a few sequences.

Temporal analysis revealed progressive changes in genotype representation across class II viruses (Figure 2). Whole-genome sequences of genotypes IX and II were the earliest in the dataset, observed in 1946 and 1948, respectively. Genotypes I, XX and V were subsequently represented in 1966, 1968 and 1971, followed by genotype VIII in 1979 and genotype VI in 1982. Six other genotypes, III (2000), XIX (2004), XXI (2007), XVI (2008), XII (2008), and XVIII and XIV (2009), first appeared in the dataset after 2000. Genotypes VI and VII exhibited the most pronounced increases in sequence representation during recent decades, particularly from the mid-2000s through the 2010s. Genotype VII showed especially high representation during 2005-2019, while genotype VI increased substantially after 2005 and remained well represented in the most recent period.

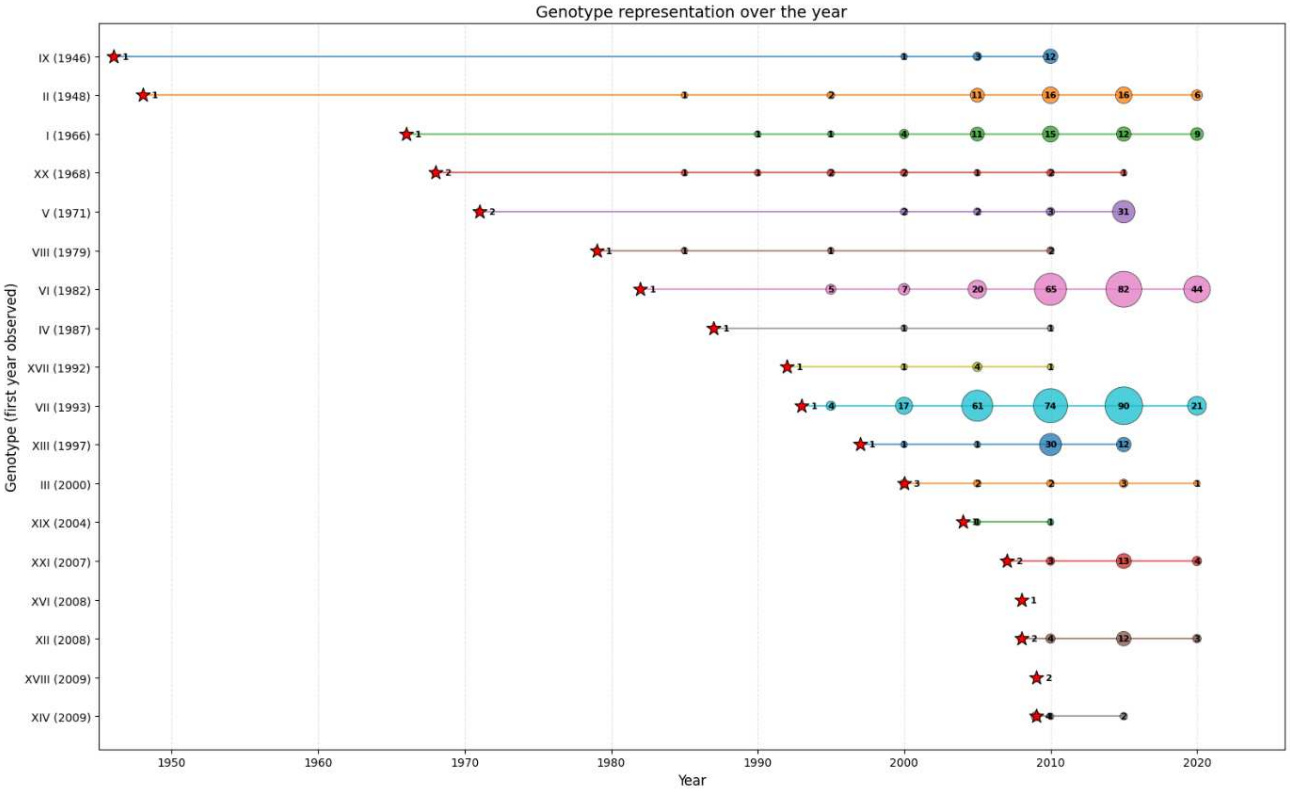


Figure 2: Temporal distribution and relative representation of NDV class II genotypes

Genetic distance between class II genotypes and vaccine reference strains

Genetic relatedness to the vaccine reference strains varied among class II genotypes. Genotype II showed the highest mean nucleotide identity with the LaSota vaccine strain (98.18%), the B1/Hitchner B1 vaccine strain (97.93%), and the VG/VA vaccine strain(97.89%),

corresponding to low mean divergence of approximately 1.8-2.1%. In contrast, its mean identity with I-2 vaccine strain was lower (88.85%). Genotype I showed its highest identity with I-2 (91.27%), while genotypes III and IX showed intermediate identities of approximately 89-90% across the vaccine references (Table 2). Most other genotypes, including the highly represented VI and VII, showed mean identities of approximately 84-86%, corresponding to about 14-16% nucleotide divergence. Across many of the more divergent genotypes, I-2 generally showed slightly higher mean nucleotide identity than the other vaccine references; for example, identities with genotypes VI and VII were 85.23% and 85.27%, respectively.

Recombination Events

Using the predefined criterion of support from at least four of the seven RDP4 algorithms at $P < 0.001$, RDP4 analysis identified three putative recombination events among Class I whole-genome sequences and 26 events among the 798 Class II sequences. These putative recombination events involved different putative major and minor parental sequences and occurred at distinct positions across the genome (Table 3). None of the Class I putative recombination events involved vaccine-associated strains. In Class II, four putative recombination events involved vaccine-associated sequences as putative major parental strains. KJ769262.1 (India, genotype II) was associated with JF950509.1 (Mukteswar vaccine strain), while JF930146.1 (China, genotype II) was associated with JF950510.1 (LaSota vaccine strain). Two Ethiopian genotype VII sequences, PV189286.1 and PV189288.1, were associated with AY562991.1 (Ulster 2C vaccine strain) as the putative major parent.

Evolutionary Rate

Bayesian coalescent model with a relaxed molecular clock was used to analyze the evolutionary dynamics of class I and class II whole-genome NDV viruses. For class I viruses, the mean evolutionary rate was 2.04×10^{-3} substitutions per site per year (95% HDP interval: 1.4922×10^{-3} and 2.9744×10^{-3}). The most recent common ancestor (TMRCA) was estimated to be 1878. Bayesian skyline plots (BSP) show that effective population size remained stable until the 1980s, then declined gradually, becoming more pronounced in the 2000s. Population size reached its lowest around 2011-2012, then gradually recovered from 2015 and stabilized in recent years (Figure 3a). The mean evolution rate of class II viruses was 1.04×10^{-3} substitutions per site per year(95% HDP interval: 9.444×10^{-4} and 1.153×10^{-3}), and the TMRCA for this group was 1851. BSP, used to infer changes in effective population size over time, indicates that effective population size grew slowly until 1960 and remained relatively stable

Table 2: Genotype-level nucleotide identity and genetic divergence between class II sequences and vaccine reference strains

		AF077761.1 Lasota				**AF375823.1 B1/Hitchner B1**				**AY935499.2 I2**				**EU289028.1 VG/VA**			
Genotype	no	Mean identity	Min identity	Max identity	Mean divergence	Mean identity	Min identity	Max identity	Mean divergence	Mean identity	Min identity	Max identity	Mean divergence	Mean identity	Min identity	Max identity	Mean divergence
I	57	89.16	73.94	96.93	10.83	89.27	73.97	96.68	10.72	91.27	74.34	96.03	8.7	89.24	73.96	96.63	10.7
II	53	98.18	84.32	99.85	1.81	97.93	84.41	99.88	2.08	88.85	84.89	92.02	11.14	97.89	84.38	99.83	2.1
III	11	88.82	88.72	88.87	11.17	88.94	88.84	88.99	11.06	89.73	89.62	89.8	10.26	88.92	88.82	88.96	11.08
IV	3	87.31	86.9	87.82	12.6	87.38	86.98	87.9	12.61	87.72	87.2	88.4	12.28	87.37	86.97	87.88	12.63
IX	17	89.04	88.91	89.11	10.9	89.12	88.99	89.19	10.88	89.35	89.26	89.42	10.64	89.10	88.96	89.17	10.90
V	40	84.15	83.63	86.51	15.8	84.26	83.76	86.55	15.74	84.60	84.03	87.13	15.39	84.19	83.68	86.51	15.81
VI	224	84.54	73.97	91.06	15.4	84.68	74.02	91.11	15.31	85.23	74.23	93.58	14.77	84.64	74.0	91.09	15.35
VII	268	84.49	83.58	87.01	15.5	84.64	83.69	87.1	15.35	85.27	84.14	86.42	14.73	84.59	83.65	87.05	15.4
VIII	5	85.73	83.95	87.04	14.25	85.80	84.0	87.15	14.19	86.48	84.31	88.04	13.51	85.78	83.97	87.1	14.21
XII	21	84.02	83.62	84.46	15.97	84.12	83.73	84.57	15.87	84.60	84.23	85.25	15.39	84.11	83.71	84.56	15.89
XIII	45	84.41	83.78	85.79	15.58	84.54	83.94	85.95	15.45	84.73	84.07	85.97	15.26	84.51	83.91	85.9	15.48
XIV	7	83.52	83.05	83.84	16.47	83.69	83.23	83.97	16.30	84.23	83.68	84.54	15.77	83.66	83.2	83.95	16.33
XIX	3	84.59	84.09	84.96	15.41	84.65	84.18	85.01	15.34	85.12	84.7	85.41	14.88	84.62	84.14	84.98	15.38
XVI	1	84.33	84.33	84.33	15.67	84.38	84.38	84.38	15.62	84.82	84.82	84.82	15.18	84.36	84.36	84.36	15.64
XVII	7	84.28	83.97	85.62	15.27	84.47	84.16	85.8	15.52	85.13	84.78	86.36	14.86	84.42	84.11	85.75	15.57
XVIII	2	84.40	84.33	84.48	15.59	84.56	84.5	84.61	15.44	84.90	84.86	84.95	15.09	84.50	84.45	84.54	15.5
XX	12	85.69	85.05	86.67	14.31	85.86	85.28	86.85	14.13	86.27	85.34	87.56	13.72	85.83	85.23	86.81	14.17
XXI	22	84.61	84.07	87.6	15.39	84.75	84.26	87.65	15.24	85.24	84.6	86.58	14.75	84.75	84.22	87.63	15.25

Table 3: Recombination events detected for the whole-genome sequences of NDVs

no	ID	Class	Major parent	Minor parent
1	MZ737134.1	I	MH289859.1	MZ737133.1
2	JN688863.1	I	MK327604.1	JN688864.1
3	KY776595.1	I	MK516205.1	JN688864.1
4	KT355595.1	II	MK764386.1	MN547991.1
5	JX193078.1	II	KJ499462.1	OR733341.1
6	AY562989.1	II	KX822746.1	GQ429293.1
7	HM357251.1	II	JX867334.1	PQ106774.1
8	KJ769262.1	II	JF950509.1	KJ636208.1
9	PV189286.1	II	AY562991.1	PV189285.1
10	MZ408157.1	II	PX060143.1	KC542913.1
11	JF930146.1	II	JF950510.1	HQ317394.1
12	JX532092.1	II	JN682211.1	KU845252.1
13	JX316216.1	II	EF201805.1	DQ060053.1
14	KF219497.1	II	EU167540.1	FJ436306.1
15	KU885390.1	II	JX193083.1	AB853926.2
16	JX193083.1	II	JX401403.1	KX765173.1
17	JX401404.1	II	MK796808.1	JX401403.1
18	EU167540.1	II	DQ060053.1	KC461214.1
19	KP089979.1	II	EU546165.2	KM056346.1
20	JX854452.1	II	MN609929.1	HQ697254.1
21	PV189288.1	II	AY562991.1	PV189285.1
22	JN682210.1	II	JX193082.1	KX061544.1
23	KP861633.1	II	FJ766531.1	JX486552.1
24	AY562985.1	II	GU978777.1	KM408752.1
25	OR860423.1	II	OR860422.1	MW147625.1
26	OP066059.1	II	OP066060.1	JX193076.1
27	JN618348.1	II	JF343538.1	DQ485231.1
28	KM056348.1	II	KJ670427.1	KP089979.1
29	MW412840.1	II	AJ880277.1	GQ338311.1

between 1960 and the late 1980s. The population size declined abruptly in the late 1990s. It recovered in 2000 and grew until the 2010s (Figure 3b).

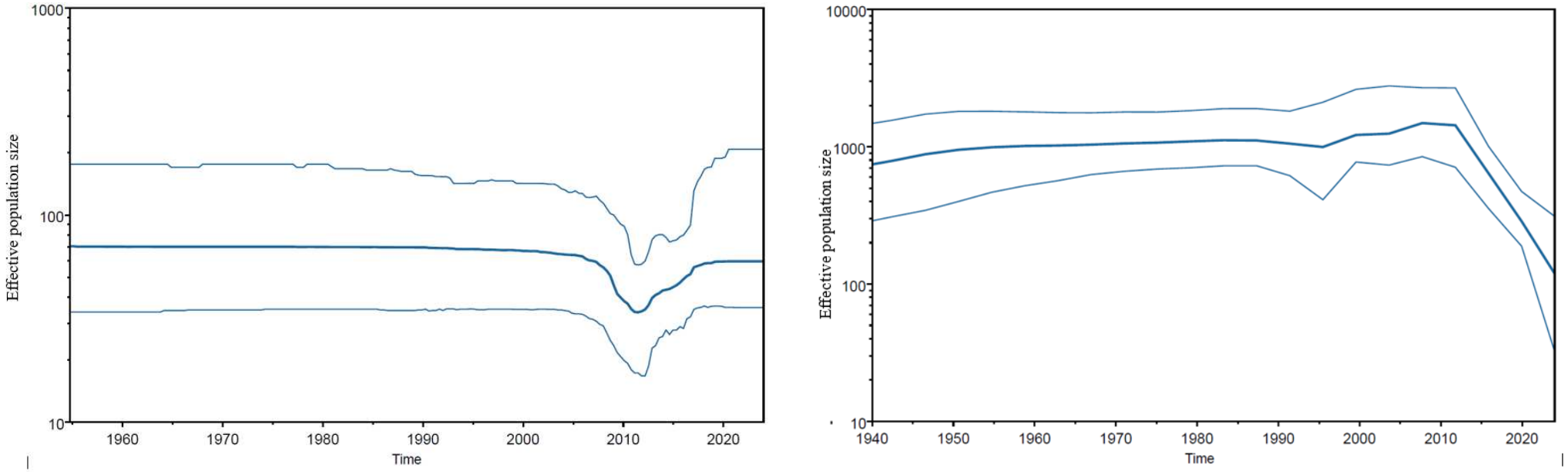


Figure 3: Bayesian skyline plots (BSP) depicting effective population size of whole-genome NDV sequences over time

Effective population size estimates are expressed on a logarithmic scale (*y-axis*). The bold lines denote the mean estimates, and the other two lines give the 95% HPD intervals of the estimates (a) BSP for class I sequences for the years between 1950 and 2025; (b) BSP for class II sequences for the years between 1940 and 2025

DISCUSSION

We presented a comparative genetic, phylogenomic, and evolutionary analysis of whole-genome NDV sequences reported in public databases to date. The 935 whole-genome sequences reported from 48 countries indicate a wide geographical distribution of NDV, where China contributed more than half of the sequences, suggesting a geographical bias in sequence availability. China has an active and intensive surveillance program, with a focus on molecular epidemiology studies[7,15].

The observed difference in genome length suggests distinct evolutionary lineages among NDVs. The three classical genome sizes (15192bp, 15186bp, and 15198bp) are associated with the emergence of new genotypes. The early genotypes (I-IV) have a genome length of 15186bp, whereas the recent genotypes have a genome length of 15192bp due to the insertion of 6 nt into the 5' non-coding region of the NP gene[16,17]. The conservation of M and L genes aligns with their role in viral transcription, replication, and assembly, making them subject to strong purifying selection [18]. The variability in F and HN genes is associated with their role as surface glycoproteins, as they are consequently under immune-mediated selective pressure and evolve more rapidly[19,20]. For gene P, the variability is associated with its RNA-editing mechanism, in which polymerase stuttering during transcription generates alternative P-gene mRNAs encoding the P, V, and W proteins[21,22].

Analysis of both class I and II viruses revealed high nucleotide polymorphism, indicating NDV genetic diversity. The high average pairwise nucleotide diversity (1273.9 for class II and 605.8 for class I) and the haplotype variance (0.9998 and 0.9990 for class II and class I viruses, respectively) reflect the nature of RNA viruses, which lack proofreading and accumulate mutations during replication[23].

Class I sequences cluster as a single genotype, while class II sequences diversify, aligning with the updated unified phylogenetic classification system for Newcastle disease virus (NDV)[7]. The diversity in class II viruses is consistent with its established status as the primary cause of ND outbreaks worldwide[24]. This is most likely associated with the evolution of the virus across different geographical locations and host populations[25,26]. In addition to host and geographical diversity, continuous mutations and natural selection play an important role in the diversity of class II viruses[12]. In contrast, class I viruses are less diverse and avirulent. The primary host of class I viruses is wild birds, which occasionally spill over to poultry[27,28].

The geographic imbalance in available whole-genome sequences likely reflects substantial variation in surveillance and sequencing intensity among countries. The predominance of Class I sequences from China, particularly subgenotype 1.1.2, agrees with surveillance studies showing widespread circulation and dominance of this subgenotype in Chinese poultry populations[29]. For Class II viruses, the broad geographic distribution of genotypes VII and VI is consistent with previous reports describing these lineages across multiple continents, with genotype VII particularly associated with extensive international dissemination. The African data similarly demonstrate considerable genotype heterogeneity despite comparatively limited whole-genome representation. Previous reviews have documented numerous NDV genotypes across Africa and identified genotype VII as one of the most geographically widespread, while emphasizing substantial gaps in molecular surveillance across the continent[15].

The high nucleotide identity of genotype II with the vaccine starins LaSota, B1, and VG/GA is expected because these commonly used vaccine strains belong to genotype II and are themselves highly genetically related[30]. In contrast, the lower identity of genotypes VI, VII, and several other contemporary genotypes indicates substantial genetic divergence from these conventional vaccine strains, consistent with previous reports describing marked differences between genotype II vaccines and circulating genotype VII viruses[31]. However, this genetic divergence should not be interpreted as direct evidence of vaccine failure, because NDV remains a single serotype and conventional vaccines can still protect against clinical disease caused by heterologous genotypes. Experimental studies nevertheless suggest that greater antigenic similarity can reduce virus shedding and improve control of transmission[32] Thus, the observed divergence of widely represented genotypes such as VI and VII from traditional vaccine strains highlights the value of continued monitoring of vaccine-field strain relatedness.

The identification of putative recombination events involving different genomic regions and parental lineages indicates that recombination may contribute to NDV genetic diversity. Notably, four Class II putative recombination events involved commonly used vaccine strains, including Mukteswar, LaSota, and Ulster 2C, as putative major parents, suggesting possible genetic interactions between vaccine-associated and circulating field viruses. Similar recombination signals involving vaccine-associated NDV strains have been reported previously[33], including potential recombination between the LaSota vaccine strain and circulating NDV strains. However, these findings should be interpreted cautiously, as RDP4 detects sequence patterns consistent with recombination but cannot confirm natural genetic exchange. Artificial recombination during PCR has also been demonstrated in NDV, and some

reported recombinants were not confirmed by resequencing[34]. Therefore, these events should be considered putative and require independent validation. If confirmed, they may indicate a role for recombination in NDV genetic diversity and warrant monitoring of interactions between vaccine-associated and field strains.

The evolutionary rates presented here for NDVs are consistent with the estimated rates of other RNA viruses, indicating that NDVs evolve rapidly[35]. The mean evolutionary rate of class I viruses aligns with Jia et al.[13]. The study described that class I viruses follow the same evolutionary path as class II viruses; however, current vaccine strategies targeting only class II viruses have allowed class I viruses to evolve relatively freely. The BSP also follows a similar pattern to the study mentioned above. In both findings, the effective population size of class I viruses declined sharply in the early 2010s; this may reflect limited sampling and high sequence similarity among isolates.

The evolutionary rate for class II viruses aligns with findings of several studies[14,36,37]. The apparent decline in population size for class II viruses identified in the BSP during the late 1990s aligned with the findings of Chong et al[14]. The article also discussed potential effects of El Niño in 1999 and the slaughter of millions of poultry during the H5N1 avian influenza outbreak as possible reasons for the abrupt decline. The TMRCA estimate of 1878 for class I and 1851 for class II viruses is in close agreement with the previous estimates of Chong et al[14] and Soñora et al[38]. This finding is also in line with studies done by Macpherson in 1956, which suggest that a disease outbreak in domestic birds in northwest Scotland between 1897 and 1898 was due to NDV[39].

MATERIALS AND METHODS

Complete genome dataset information

A total of 1312 complete whole-genome sequences of NDVs were accessed from NCBI (https://www.ncbi.nlm.nih.gov/) on July 25, 2025, and processed for subsequent bioinformatic analysis. 51 sequences that did not report the collection site were removed from the collection. Furthermore, given the “rule of six” genome atlas or polyhexameric genome length, 326 sequences were also excluded. All essential information related to the remaining 935 whole-genome sequences of the study-included strains is presented in Table 1.

Comparative genomic analysis and Nucleotide diversity

The complete genome dataset was aligned to a common length using MAFFT v7.490[40]. Based on variable sites for mutations and average numbers of pairwise nucleotide differences, the

nucleotide diversity among coding sequences (CDS) of each gene was assessed for genomic polymorphism by DnaSP version 6.12.03 (accessible at http://www.ub.es/dnasp). The departure from neutrality in all isolates' sequences was tested by Tajima's D statistical method[41].

Phylogenetic analysis

The whole-genome sequences of class I and class II NDV were aligned using IQ-TREE v2.0.7 [42]. The best-fit nucleotide substitution model was determined using ModelFinder implemented in IQ-TREE [43]. Model selection was evaluated using the Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc), and Bayesian Information Criterion (BIC), with the BIC-selected model preferred in cases of conflict. Maximum-likelihood phylogenies were inferred in IQ-TREE, with branch support assessed using the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT), approximate Bayes test, and ultrafast bootstrap approximation (1000 replicates)[44].

Genetic distance between class II genotypes and vaccine reference strains

Pairwise whole-genome nucleotide identity between class II genotypes and commonly used vaccines, Lasota, B1 (Hitcher B1), VG/VA, and I-2 (thermostable seed) was assessed using the Biopython package in Python 3.13. Identity values between each sequence and vaccine reference were extracted and grouped by genotype. Genetic divergence was calculated as 100-nucleotide identity (%), and values were summarized by genotype using the mean, minimum, and maximum distance.

Detection of putative recombination events

The sequences were analyzed to identify reliable putative breakpoints using RDP4 version 4.101[45]. RDP4 employs a combination of seven algorithms: RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, and 3Seq. To better isolate putative recombinants from parents, an event should be consistently identified by at least four algorithms at a probability threshold of 0.001.

Evolutionary Rate

Bayesian phylogenetic analyses of Class I and Class II viruses were conducted separately using BEAST v2.7.7[46] with BEAGLE v4.0.1[47]. For both groups of viruses, putative recombinant sequences were removed from evolutionary analysis[48]. Furthermore, as more than half the sequences were reported from China, a random down-sampling technique was implemented to avoid bias[49]. Sequences were randomly selected based on both their geographical location and year of collection. Each random down-sample contains at most 15 sequences from each country and year; ultimately, 641 class II and 120 class I sequences were included in the final analysis.

Sequence evolution was modeled using partition-specific GTR+Γ4 substitution models and an uncorrelated lognormal relaxed clock linked across partitions[50]. Population dynamics were estimated using the Coalescent Bayesian Skyline model[50,51]. MCMC chains were run for 500 million generations with sampling every 50,000 generations. Convergence was assessed in Tracer v1.7.2 after a 10% burn-in, ensuring ESS values exceeded 200[52]. Bayesian Skyline plots were generated in Tracer v1.7.2. The most recent common ancestor (TMRCA) was estimated based on Maximum clade credibility (MCC) trees using TreeAnnotator, and tree visualization and annotation were performed in FigTree v1.4.4[53].

DATA AVAILABILITY

All data used in this study are available within the paper and its Supplementary Information.

REFERENCES

1. Suarez, D. L., Miller, P. J., Koch, G., Mundt, E. & Rautenschlein, S. Newcastle disease, other avian paramyxoviruses, and avian metapneumovirus infections. in *Diseases of Poultry* 111–166 (John Wiley & Sons, Ltd, 2019). doi:10.1002/9781119371199.ch3.

2. Charkhkar, S. *et al.* The evaluation and importance of Newcastle disease's economic loss in commercial layer poultry. *J. Poult. Sci. Avian Dis.* **2**, 1–4 (2024).

3. Zegeye, A., Temesgen, W., Molla, W., Setotaw, H. & Lakew, M. Epidemiology of Newcastle disease in chickens of Ethiopia: a systematic review and meta-analysis. *Trop. Anim. Health Prod.* **54**, (2022).

4. Gogoi, P., Ganar, K. & Kumar, S. Avian Paramyxovirus: A Brief Review. *Transboundary and Emerging Diseases* vol. 64 53–67 at https://doi.org/10.1111/tbed.12355 (2017).

5. Malik, Y. S., Arun Prince Milton, A., Ghatak, S. & Ghosh, S. Newcastle Disease and Other Avian Paramyxoviruses. 79–91 (2021) doi:10.1007/978-981-16-4554-9_6.

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

7. Dimitrov, K. M. *et al.* Updated unified phylogenetic classification system and revised nomenclature for Newcastle disease virus. *Infect. Genet. Evol.* **74**, 103917 (2019).

8. de Almeida, R. S. *et al.* New Avian Paramyxoviruses Type I Strains Identified in

Africa Provide New Outcomes for Phylogeny Reconstruction and Genotype Classification. *PLoS One* **8**, e76413 (2013).

9. Ebrahimi, M. M., Shahsavandi, S., Moazenijula, G. & Shamsara, M. Phylogeny and evolution of Newcastle disease virus genotypes isolated in Asia during 2008-2011. *Virus Genes* **45**, 63–68 (2012).

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

11. Hicks, J. T., Dimitrov, K. M., Afonso, C. L., Ramey, A. M. & Bahl, J. Global phylodynamic analysis of avian paramyxovirus-1 provides evidence of inter-host transmission and intercontinental spatial diffusion. *BMC Evol. Biol.* **19**, 1–15 (2019).

12. Xiang, B. *et al.* Insights into Genomic Epidemiology, Evolution, and Transmission Dynamics of Genotype VII of Class II Newcastle Disease Virus in China. *Pathogens* **9**, 837 (2020).

13. Jia, L. *et al.* Circulation, genomic characteristics, and evolutionary dynamics of class I Newcastle disease virus in China. *Virulence* **13**, 414–427 (2022).

14. Chong, Y. L., Padhi, A., Hudson, P. J. & Poss, M. The Effect of Vaccination on the Evolution and Population Dynamics of Avian Paramyxovirus-1. *PLOS Pathog.* **6**, e1000872 (2010).

15. Amoia, C. F. *et al.* Genomic Diversity and Geographic Distribution of Newcastle Disease Virus Genotypes in Africa: Implications for Diagnosis, Vaccination, and Regional Collaboration. *Viruses* **16**, (2024).

16. Munir, M., Abbas, M., Khan, M., Zohari, S. & Berg, M. Genomic and biological characterization of a velogenic Newcastle disease virus isolated from a healthy backyard poultry flock in 2010. *Virol. J.* **9**, (2012).

17. Duan, Y., Leng, G., Liu, M. & Duan, Z. Advancements in the genomic feature of Newcastle disease virus and the multifaceted roles of non-structural proteins V/W in viral replication and pathogenesis. *Poult. Sci.* **104**, 105527 (2025).

18. Jadhav, A. *et al.* Genomic Diversity and Evolution of Quasispecies in Newcastle Disease Virus Infections. *Viruses 2020, Vol. 12,* **12**, (2020).

19. Kumar, S., Nayak, B., Collins, P. L. & Samal, S. K. Evaluation of the Newcastle Disease Virus F and HN Proteins in Protective Immunity by Using a Recombinant Avian Paramyxovirus Type 3 Vector in Chickens. *J. Virol.* **85**, 6521–6534 (2011).

20. Gu, M. *et al.* Positive selection in the hemagglutinin-neuraminidase gene of Newcastle disease virus and its effect on vaccine efficacy. *Virol. J. 2011 81* **8**, 150- (2011).

21. Chen, X. *et al.* Identification of Newcastle disease virus P-gene editing using next-generation sequencing. *J. Vet. Med. Sci.* **82**, 1231 (2020).

22. Jadhav, A. *et al.* Patterns of RNA Editing in Newcastle Disease Virus Infections. *Viruses* **12**, 1249 (2020).

23. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).

24. Kondakova, O. A. *et al.* Genetic Diversity of Newcastle Disease Virus and Its Implications for Vaccine Development. *Vet. Sci. 2025, Vol. 12, Page 858* **12**, 858 (2025).

25. Yan, W. *et al.* Emergence and characterization of historically extinct virulent genotype IV Newcastle disease virus in wild and domestic birds: genetic insights, pathogenicity, and vaccine efficacy. *J. Virol.* **99**, e01646-25 (2025).

26. Guseva, N. A. *et al.* Subgenotype VII.1.1 Newcastle Disease Virus Evolution and Spread in the Russian Federation in 2019–2023. *Viruses* **17**, 1319 (2025).

27. Sun, J. *et al.* Surveillance of Class I Newcastle Disease Virus at Live Bird Markets in China and Identification of Variants with Increased Virulence and Replication Capacity. *J. Virol.* **96**, e00241-22 (2022).

28. Snoeck, C. J. *et al.* Genetic Diversity of Newcastle Disease Virus in Wild Birds and Pigeons in West Africa. *Appl. Environ. Microbiol.* **79**, 7867–7874 (2013).

29. Wang, J. *et al.* Continuous surveillance revealing a wide distribution of class I Newcastle disease viruses in China from 2011 to 2020. *PLoS One* **17**, e0264936 (2022).

30. Dimitrov, K. M., Afonso, C. L., Yu, Q. & Miller, P. J. Newcastle disease vaccines—A solved problem or a continuous challenge? *Vet. Microbiol.* **206**, 126 (2016).

31. Hu, Z., He, X., Deng, J., Hu, J. & Liu, X. Current situation and future direction of Newcastle disease vaccines. *Vet. Res.* **53**, 99 (2022).

32. Liu, H. *et al.* Can genotype mismatch really affect the level of protection conferred by Newcastle disease vaccines against heterologous virulent strains? *Vaccine* **36**, 3917–3925 (2018).

33. Bahoussi, A. N. *et al.* Multiple potential recombination events among Newcastle disease virus genomes in China between 1946 and 2020. *Front. Bahoussi, PT Shah, JQ Zhao, PH Wang, YY Guo, C Wu, L XingFrontiers Vet. Sci. 2023•frontiersin.org* **10**, (2023).

34. Afonso, C. L. Not So Fast on Recombination Analysis of Newcastle Disease Virus. *J. Virol.* **82**, 9303 (2008).

35. Pomeroy, L. W., Bjørnstad, O. N. & Holmes, E. C. The Evolutionary and Epidemiological Dynamics of the Paramyxoviridae. *J. Mol. Evol. 2008 662* **66**, 98–106 (2008).

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

37. Du, J. *et al.* Evolutionary dynamics and transmission patterns of Newcastle disease virus in China through Bayesian phylogeographical analysis. *PLoS One* **15**, e0239809 (2020).

38. Soñora, M. *et al.* An evolutionary insight into Newcastle disease viruses isolated in Antarctica. *Arch. Virol. 2015 1608* **160**, 1893–1900 (2015).

39. Macpherson, L. W. Some Observations On The Epizootiology Of NewCastle Disease. *Can. J. Comp. Med. Vet. Sci.* **20**, 155 (1956).

40. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772 (2013).

41. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**, 585 (1989).

42. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

43. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods 2017 146* **14**, 587–589 (2017).

44. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

45. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

46. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, (2019).

47. Ayres, D. L. *et al.* BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012).

48. Anisimova, M., Nielsen, R. & Yang, Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229–1236 (2003).

49. Hall, M. D., Woolhouse, M. E. J. & Rambaut, A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. *Virus Evol.* **2**, vew003 (2016).

50. Douglas, J., Zhang, R. & Bouckaert, R. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLOS Comput. Biol.* **17**, e1008322 (2021).

51. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).

52. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

53. Rambaut, A. Rambaut, A. (2010) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. - References - Scientific Research Publishing. https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1661474 (2010).

ACKNOWLEDGEMENTS
We thank Prof. Phillipe Lemey for his input and guidance while conducting the BEAST analysis.

FUNDINGS
KU Leuven Global Minds Doctoral Scholarship Program (project number 4520196970).

AUTHORS CONTRIBUTION
Conceptualization, data collection, analysis, formal analysis, visualization, and writing the original draft were conducted by EF. Conceptualization, editing, funding acquisition, and supervision were handled by JP & TS.

AUTHORS AND AFFILIATIONS

1. **Laboratory of Host Pathogen Interaction in Livestock, Division of Animal and Human Health Engineering, Department of Biosystems, KU Leuven, Leuven, 3001, Belgium**

Eyerusalem Fetene & Jan Paeshuyse

2. **Biotechnology Research Center, Addis Ababa University, Ethiopia**

Tesfaye Sisay Tessema

COMPETING INTERESTS
The authors declare no competing interests.

ETHICAL APPROVAL AND INFORMED CONSENT
This research did not involve human or animal participants.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfilegenomicdiversityandevolutionaryinsightsof NDV.docx