frontiers | Research Topics

Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases

Edited by
Iryna Goraichuk, Christina Leyson and Moh A. Alkhamis

Published in
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-8325-3119-8
DOI 10.3389/978-2-8325-3119-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases

Topic editors

Iryna Goraichuk — Southeast Poultry Research Laboratory, U.S. National Poultry Research Center, Agricultural Research Service (USDA), United States

Christina Leyson — Southeast Poultry Research Laboratory, U.S. National Poultry Research Center, Agricultural Research Service (USDA), United States

Moh A. Alkhamis — Kuwait University, Kuwait

Citation

Goraichuk, I., Leyson, C., Alkhamis, M. A., eds. (2023). *Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3119-8

Table of contents

frontiers | Frontiers in Veterinary Science

TYPE Editorial
PUBLISHED 13 July 2023
DOI 10.3389/fvets.2023.1236155

OPEN ACCESS

EDITED AND REVIEWED BY
Ulises Garza-Ramos,
National Institute of Public Health, Mexico

*CORRESPONDENCE
Christina M. Leyson
✉ cmleyson@gmail.com
Moh A. Alkhamis
✉ m.alkhamis@ku.edu.kw
Iryna V. Goraichuk
✉ iryna.goraichuk@usda.gov

RECEIVED 07 June 2023
ACCEPTED 06 July 2023
PUBLISHED 13 July 2023

CITATION
Leyson CM, Alkhamis MA and Goraichuk IV (2023) Editorial: Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases.
Front. Vet. Sci. 10:1236155.
doi: 10.3389/fvets.2023.1236155

Editorial: Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases

Christina M. Leyson[1]*, Moh A. Alkhamis[2]* and Iryna V. Goraichuk[3,4]*

[1]Balik Scientist Program, Department of Science and Technology, Taguig, Philippines, [2]Department of Epidemiology and Biostatistics, College of Public Health, Health Sciences Center, Kuwait University, Kuwait City, Kuwait, [3]Exotic and Emerging Avian Viral Disease Research Unit, Southeast Poultry Research Laboratory, U.S. National Poultry Research Center, Agricultural Research Service, United States Department of Agriculture, Athens, GA, United States, [4]National Scientific Center Institute of Experimental and Clinical Veterinary Medicine, Kharkiv, Ukraine

KEYWORDS

genome sequencing, next-generation sequencing, sanger sequencing, phylogenetic analysis, metagenomics, molecular surveillance, molecular epidemiology

Editorial on the Research Topic
Sequencing and phylogenetic analysis as a tool in molecular epidemiology of veterinary infectious diseases

In the past few decades, the rapid development and decreasing costs of sequencing technologies dramatically changed the landscape of epidemiological studies and surveillance of infectious diseases. Subsequently, pathogen genomic studies have become the forefront tool for investigating emerging infectious disease epidemics and supporting decision-making processes related to the mobilization of intervention resources. Since the introduction of next-generation sequencing (NGS), which expanded the capacity for whole genome sequencing (WGS), and the revolutionary growth of computational resources, viral or bacterial complete genomes can now be sequenced and characterized within a few days or even hours. Therefore, high-throughput sequencing technologies resulted in the exponential growth of genomic databases, unveiling novel insights into the biology, pathophysiology, and molecular epidemiology of infectious disease pathogens. Such significant developments in genome sequence technologies also resulted in important advances in the field of phylogenetic analysis. Modern analytical methods in phylogenetics improved the tracking and understanding of pathogen transmission and evolution of human and animal diseases. Thus, phylogenetic analysis of big genomic databases can be used to clarify key questions related to infectious disease epidemiology, such as the initial detection and characterization of outbreaks, accurate tracing of transmission chains between hosts, and dispersal among and within geographical regions. The threat of emerging and re-emerging infectious diseases continues to be a challenge to global public and animal health, in which sequencing is not just a critical tool in surveillance but also has a major role in a pandemic or outbreak response. The main objective of this Research Topic is to explore the current status and future perspectives of sequencing technologies in the control and prevention of infectious diseases, including the elucidation of diagnosis, molecular evolution, and epidemiology of infectious disease pathogens.

NGS has become a primary diagnostic and characterization tool for genomic surveillance of pathogens that threaten biosecurity and food safety. Kariithi et al., Kim H.-J. et al., and Abbas et al. demonstrated multiple approaches in utilizing sequencing and phylogenetic analysis resources for improving genomic surveillance of avian infectious bronchitis virus in poultry by characterization of phylogenetic relationships, detection of critical recombination events, assessing vaccination effectiveness, and identifying evolutionary origins of endemic and emerging strains. Furthermore, Kim S.-W. et al. were able to confirm that avian reoviruses circulating in poultry flocks were originating from wild birds using straightforward traditional molecular characterization tools, similar to what Goraichuk et al., did with the Newcastle disease virus. However, Baek et al. went a step further by implementing Bayesian phylodynamic analysis to shed deeper insights into the evolutionary epidemiology of H5N8 avian influenza viruses. They demonstrated that wild birds were the ancestral host for multiple introductions of H5N8 viruses into poultry, but domesticated ducks more important in virus circulation and transmission among poultry flocks.

African swine fever continues to be the most important devastating pathogen to swine populations, causing unprecedented annual economic loss on a global scale. NGS and WGS also continue to be the most essential tools for providing critical genetic, epidemiologic, therapeutic, and vaccine development resources for African swine fever intervention efforts, as illustrated by Hyeon, Tseren-Ochir, et al., Gallardo et al., and Kim G. et al. Furthermore, the combination of WGS and phylodynamic analytical approaches was used by Pamornchainavakul et al. to unveil novel findings on how the rapid recombination events among porcine reproductive and respiratory syndrome virus 2 strains can accelerate their genetic mutations leading to the emergence of more virulent strains. Similarly, Wei et al. were able to identify evolutionary characteristics and geographical origins of the lumpy skin disease virus in cattle, using multiple gene segments of a strain isolated from an outbreak in China in 2009. Rossi et al. extended their genomic analytical pipeline by integrating ecological niche models to quantify the role of environmental and demographic risk factors in shaping the evolutionary epidemiology of *Mycobacterium bovis* in Cameroon, which is considered a novel and critical step in improving genomic surveillance of infectious diseases.

Genomic surveillance of infectious diseases in wildlife is the foundational pillar of implementing the One Health concept globally. Indeed, the rapid emergence and spread of West Nile and Rabies viruses from wild animal origins played an important role in building the foundation of the One Health concept, which necessitates integrating disease surveillance of humans, wildlife, and domestic animals. In this Research Topic, Chung et al. and Hyeon, Helal, et al. used NGS and WGS approaches to deeply characterize the genetic features of West Nile virus and lyssaviruses from wild animals in selected regions in the United States in order to shed important insights about their origins and transmission dynamics. Additionally, it is important to include genomic surveillance of pathogens in exotic pet animals, such as reptiles, to monitor evolutionary characteristics that may result in the emergence of novel pathogens, as demonstrated by Varga-Kugler et al..

Continuous development and revision of NGS and WGS pipelines are critical for ensuring the sustainable generation and accumulation of sound and reliable genetic data. In this Research Topic, Lorente-Leal et al. revised the performance and agreement of four commonly used pipelines for WGS data analysis of *M. bovis*, the causative agent of bovine tuberculosis, Butt et al. highlighted the feasibility and utility of long-read random sequencing approaches to identify pathogens in clinical samples, and Cho et al. developed a novel tiling amplicon PCR method for feasible and rapid sequencing of complete genomes in clinical samples.

Our Research Topic reinforces the importance of sequencing technologies in changing the landscape of infectious disease surveillance in animals and humans, with the aim of highlighting NGS and WGS approaches that continue to become more feasible and accessible globally. This Research Topic highlights the versatility of techniques for bench science and sequence analysis in its application to many problems in infectious diseases of veterinary importance. In this Research Topic alone, investigations on 13 infectious diseases of poultry, swine, bovine, and wildlife were presented. Furthermore, the authors of these investigations span 23 countries and five continents, highlighting the accessibility of NGS techniques and its continued lowering of costs. In this Research Topic, we aspire to further motivate physicians, veterinarians, epidemiologists, microbiologists, diagnosticians, and other scientists from related fields to make sequencing technologies the gold standard for the diagnosis and surveillance of infectious disease pathogens in order to improve their current and future intervention efforts.

Author contributions

All authors were involved in the writing of this editorial and editing contributions to this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fvets.2021.780018

Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization

Víctor Lorente-Leal[1,2]*, *Damien Farrell*[3], *Beatriz Romero*[1,2], *Julio Álvarez*[1,2], *Lucía de Juan*[1,2] *and Stephen V. Gordon*[3]

[1] VISAVET Health Surveillance Center, Universidad Complutense de Madrid, Madrid, Spain, [2] Animal Health Department, Faculty of Veterinary Medicine, Universidad Complutense de Madrid, Madrid, Spain, [3] UCD School of Veterinary Medicine, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:
Iryna Goraichuk,
Institute of Experimental and Clinical Veterinary Medicine, Ukraine

Reviewed by:
Lorraine Michelet,
de l'Environnement et du Travail (ANSES), France
Daniela Brites,
Swiss Tropical and Public Health Institute (Swiss TPH), Switzerland

***Correspondence:**
Víctor Lorente-Leal
vicloren@ucm.es

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 20 September 2021
Accepted: 25 November 2021
Published: 14 December 2021

Citation:
Lorente-Leal V, Farrell D, Romero B, Álvarez J, Juan Ld and Gordon SV (2021) Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. Front. Vet. Sci. 8:780018. doi: 10.3389/fvets.2021.780018

Whole genome sequencing (WGS) and allied variant calling pipelines are a valuable tool for the control and eradication of infectious diseases, since they allow the assessment of the genetic relatedness of strains of animal pathogens. In the context of the control of tuberculosis (TB) in livestock, mainly caused by *Mycobacterium bovis*, these tools offer a high-resolution alternative to traditional molecular methods in the study of herd breakdown events. However, despite the increased use and efforts in the standardization of WGS methods in human tuberculosis around the world, the application of these WGS-enabled approaches to control TB in livestock is still in early development. Our study pursued an initial evaluation of the performance and agreement of four publicly available pipelines for the analysis of *M. bovis* WGS data (vSNP, SNiPgenie, BovTB, and MTBseq) on a set of simulated Illumina reads generated from a real-world setting with high TB prevalence in cattle and wildlife in the Republic of Ireland. The overall performance of the evaluated pipelines was high, with recall and precision rates above 99% once repeat-rich and problematic regions were removed from the analyses. In addition, when the same filters were applied, distances between inferred phylogenetic trees were similar and pairwise comparison revealed that most of the differences were due to the positioning of polytomies. Hence, under the studied conditions, all pipelines offer similar performance for variant calling to underpin real-world studies of *M. bovis* transmission dynamics.

Keywords: whole genome sequencing (WGS), bioinformatics, variant calling pipeline, SNP analysis, genomic epidemiology, Bovine Tuberculosis (bTB), *Mycobacterium bovis*, *Mycobacterium tuberculosis* complex (MTBC)

INTRODUCTION

Animal tuberculosis (aTB) is a chronic infectious disease that affects a wide variety of mammalian species, which is caused by members of the *Mycobacterium tuberculosis* complex (MTBC) (1). The principal agent of TB in cattle (bovine TB, bTB) is *Mycobacterium bovis*. In this manuscript, we will use aTB to refer to TB across wild and domestic animals, and bTB to refer specifically to TB in cattle.

Bovine TB is subjected to control and eradication programmes in many countries, not only due to its economic impact, as a result of reduced yields and animal mortality, but also because of the risk of zoonotic transfer of infection from affected animals to humans (2). Eradication programmes are usually based on a test and slaughter strategy in which cattle that are positive to an official immunological test, such as the intradermal tuberculin test, are culled (3–5). In order to confirm the presence of MTBC species, tissues from the affected animals are cultured in the laboratory (6). In order to eradicate bTB, breakdown events not only need to be detected but also studied for epidemiological links, a process that is greatly facilitated by the application of molecular genetic methods. Due to the clonal structure and limited genetic variability of MTBC species, based on the observed genetic differences between the strains isolated from the breakdown herd and from other aTB episodes, authorities can establish if the outbreak originated from cattle movement, residual infection or contact with wild animal reservoirs (7).

Traditionally, molecular epidemiological studies of aTB are based on techniques that analyse small fragments of the microbial genome, such as spoligotyping or mycobacterial interspersed repeat unit-variable number of tandem repeats (MIRU-VNTR) (8, 9). Although useful in large-scale studies (10–12), some of these methods are laborious and the use of a limited number of loci entails a higher risk of homoplasies and a lack of resolution, limiting their use in the study of local transmission events (13, 14).

The advent of Whole Genome Sequencing (WGS) has revolutionized the study of microbial populations. When applied to epidemiological studies, the availability of the whole genome of the microorganism of interest allows for a much higher resolution than that obtained with previous molecular techniques (15). As a result, the use of WGS in human TB outbreak investigations has rapidly increased in the last decade (16–18).

Due to the limited genetic diversity in MTBC genomes, the standard workflow in MTBC studies is based on the alignment of genomic sequences to a reference genome followed by the detection of genomic variants, usually single nucleotide polymorphisms (SNPs) (19). The procedure starts with genomic DNA extraction, usually through phenol-chloroform or CTAB extraction, library preparation and sequencing using short read sequencing technologies, followed by short-read mapping to the reference genome and variant calling. Variants are then filtered according to certain thresholds and parameters such as proximity to other SNPs, mapping quality, base depth or strand bias. Remaining SNPs are generally concatenated into multi-FASTA files representing multiple sequence alignments and a phylogeny is reconstructed based on SNP differences.

There are several variant calling pipelines for human tuberculosis and, recently, several efforts have been made to assess their performance in human TB outbreak investigations (19–21). Regarding the veterinary field, there is a growing interest in the use of WGS for the analysis of bTB breakdowns, which has resulted in an increasing number of studies being published around the globe (22–26). Nevertheless, although several variant calling pipelines have been developed or are in the making, there are no tool-specific publications and there is a lack of information regarding their overall performance. The aim of this study was to evaluate similarities in design and performance of publicly available variant calling pipelines currently used in laboratories tasked with the application of WGS technologies for aTB eradication.

MATERIALS AND METHODS

Artificial Genome and Read Generation

In order to simulate a reference phylogeny, raw Variant Call Format (VCF) files were selected from an already published dataset from a bTB high prevalence setting in the Republic of Ireland (22). A total of 47 samples, including two outgroup isolates (isolates 161 and 182), were used to generate artificial mutant genomes by transferring the identified SNPs in the raw VCF files to the *M. bovis* AF2122/97 genome (NCBI RefSeq accession number: NC_002945.4) using simuG 1.0.0 (27) (**Figure 1**). ArtificialFASTQGenerator 1.0 was then used to generate artificial paired-end reads from the simulated genomes. Several parameters were tested to guarantee a full genome coverage and varying read depth across the whole sequence. Read length was set to 250 bp, template length mean to 650 bp (S.D. = 60), and peak coverage mean for a region was set to 250 (Standard Deviation or S.D. = 0.2) (28). Read qualities were obtained from real-life FASTQ files originating from other sequencing projects (unpublished) and sequencing errors were simulated based on these quality profiles.

Variant Calling

The artificially generated reads were analyzed with four variant calling pipelines: vSNP 2.03 (25), SNiPgenie 0.5.0, BovTB 20.4, and MTBseq 1.03 (29). The three first pipelines are used for analysis of aTB isolates in the USA (vSNP), Ireland (SNiPGenie), and UK (BovTB); MTBSeq was added as a comparator from the human TB field. Information regarding these pipelines is summarized in **Table 1** and described in Supplementary Materials and methods. All pipelines, with exception of SNiPgenie, were run using default settings in miniconda 4.9.2 in Ubuntu 18.04 and in Brigit, the HPC server of the Computer Services at Universidad Complutense of Madrid, using the default reference sequence of *M. bovis* AF2122/97 (NC_002945.4 or LT708304). In SNiPgenie, minimum mapping quality was set to 60 in variant calling and the minimum MQ was set to 30 in posterior filtering steps. Amended variant tables returned by MTBseq were converted to the VCF format using an in-house script for further comparisons. VCF files were generated from excel tables output by vSNP's step 2 script using an in-house script and including a zero coverage VCF of *M. bovis* AF2122/97 in order to include non-parsimonious SNPs.

Pipeline Performance Evaluation

As well as a FASTA file containing the artificial genome, SimuG generates a VCF file that contains all the variants included in the generated genome. These artificial VCF files were used as a reference standard to compare the VCF files output from the variant calling pipelines using the Haplotype Comparison Tools 0.3.12 (Som.py).

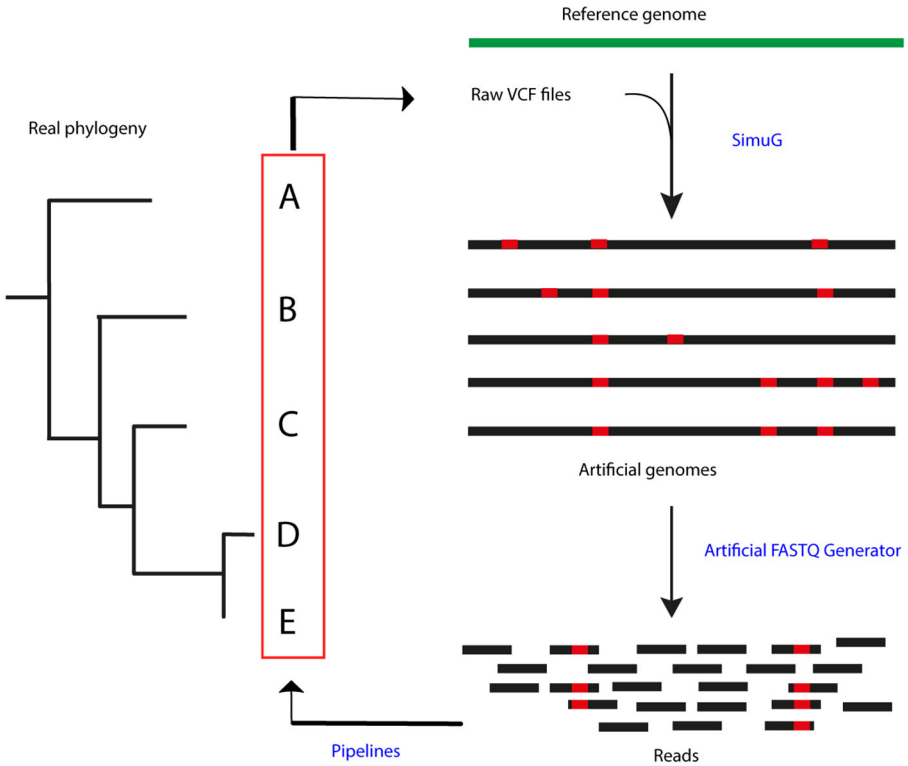


FIGURE 1 | Process summary. Raw VCF files obtained from a real-life phylogeny (22) were used to generate artificial genomes using SimuG and the *M. bovis* AF2122/97 reference genome. Artificial FASTQ Generator was then used to generate artificial reads and these were then input in the evaluated pipelines. Identified variants and output phylogenetic trees were compared between pipelines and the simulation.

Variants occurring in locations where no mutations existed in the simulated genome were considered "false positive SNPs," while mutations not detected by a given pipeline were considered "false negative SNPs".

The evaluated performance parameters were relative sensitivity or recall rate (true positives/true positives + false negatives) and relative specificity or precision (true positives/true positives + false positives). In addition, alternative (ALT) alleles were extracted from all sample VCF files obtained from each pipeline and combined to obtain the total amount of alleles identified per pipeline. The agreement between the different pipelines was then evaluated using Venn diagrams generated using VennDiagram v1.6.20 in R 3.6.3 (30).

In order to identify groups of genetic elements that usually give rise to false positive and negative calls, all VCF files were annotated using SnpEff 4.3t (31). These genetic elements were then divided into three categories: PE and PPE gene families, mobile genetic elements and other elements (including Direct Repeats and the *pks*12 gene), and their positions in the reference genome were extracted from the GFF3 annotation file available at the NCBI.

Performance was re-evaluated using different levels of hard filtering: (A) unfiltered, (B) a proximal window distance of 10 bp (22), (C) 10 bp window and pipeline default filters, (D) 10 bp window and PE/PPE family proteins, (E) 10 bp window, PE/PPE family proteins and mobile genetic elements, (F) 10 bp window, PE/PPE family proteins, mobile elements and others, and (G) PE/PPE family proteins, mobile elements and others. In order to assess the agreement between pipelines and the accuracy of these results with respect to the original simulated files, filtered positions were also removed from the simulated VCF files.

In addition, the effect of filtering on the number of identified homoplasies was assessed using HomoplasyFinder (32).

Evaluation of Phylogenetic Outputs and Epidemiological Conclusions

All pipelines, except for BovTB, generate a multi-FASTA alignment containing the concatenated variants. The SNPs in the alignment files obtained from vSNP and SNiPgenie only include polymorphic sites, whereas MTBseq alignments also include monomorphic sites. BovTB yields a consensus genome generated from the VCF files using the BCFtools consensus caller. In order to compare the different methods, core polymorphic SNPs were extracted from these consensus genomes using SNP-sites 2.5.1 (33). In addition, concatenated multi-FASTA files containing polymorphic SNPs were generated for the simulated VCFs using an in-house script.

Maximum-likelihood trees were reconstructed from the resulting multi-FASTA alignment files using RAxML 8.2.12 with

TABLE 1 | Pipeline properties of the different tools evaluated in this study.

	Pipeline			
	vSNP	**SNiPgenie**	**BovTB**	**MTBseq**
Institution	USDA-APHIS	UCD	APHA	LLI – RCB
Language	Python	Python	Nextflow	Perl
Reference	NC_002945.4	LT708304.1	LT708304.1	NC_002945.4
Parameter setup	No[a]	Yes	No	Yes
Pre-process				
Deduplication	Picard	No	FastUniq	Picard
Trimming	None	Yes[b]	Trimmomatic	None
Mapping and SNP calling				
Read aligner	BWA	BWA	BWA	BWA
SNP calling	FreeBayes	BCFtools	BCFtools	SAMtools + GATK
Phred base quality	20 (Step 1)	User defined	10	20
Normalize	No	No	Yes	Yes
SNP quality threshold	150	≥40 or User defined	None	None
Min. map quality	56	60	None	None
SNP coverage depth	None	30	5	4F and 4R
Region filter	Excel file (validated problematic positions)	BED file (PE/PPE genes)	TSV (95% similarity self-BLAST)	TSV file (repetitive sequences)
Proximality filter	None	Yes	None	Yes
Allele frequency/fraction	0.05	DP4>4	≥ 0.8	75%
Considers as diploid	Yes	No	No	No
Low coverage positions	Reference if QUAL < 50 N if 50 < QUAL < 150	Reference	Reference	Consensus base or ignore position if quality is below thresholds in >5% of samples
Alignment file	Core SNPs (polymorphic)	Core SNPs (polymorphic)	Consensus genome	Core SNPs (all)
Spoligotyping	Yes	Yes	No	No
Tree building	RAxML	RAxML	No	No
GUI	No	Yes	No	No
Other analyses	Lineage classification	INDEL analysis	Lineage classification	Lineage classification, antibiotic resistance annotation

[a] *Only allows for minor parameter settings, such as reference file or type of analysis in step 2.*
[b] *Deactivated by default.*

100 bootstraps and the GTRCATI model (34). The bipartitions and best trees obtained from each pipeline were evaluated using Robinson-Foulds (RF) distances and Ward's method for clustering through Treespace in R; briefly, RF pairwise distances between trees were decomposed into a low-dimensional space using a principal coordinate analysis (35). Trees obtained from hard filters that produced the best results in the performance evaluation were compared in a pairwise manner with the simulated phylogeny using Phytools 0.7.82 in R (36).

RESULTS

Artificial Read and Genome Simulation

An average of 2.5 x 10^6 reads (coefficient of variation or C.V. = 0.09%) were generated with an average depth of coverage of 145, with a minimum of 0 and a maximum of 310 reads per site.

Excluding the outgroup isolates, the average observed differences between isolates in the unfiltered simulation was 38.61, with an inter quartile range of 30–47, and a minimum and maximum number of 12 and 64 nucleotide differences, respectively.

Pipeline Performance

Recall rates were highest for SNiPgenie and BovTB when base parameters were employed, followed by vSNP and MTBseq (**Figure 2A**).

Sensitivity increased when increasing levels of hard filtering were applied for vSNP, SNiPgenie, and MTBseq, and remained similar for BovTB (**Figure 2A**). The positive effect was higher when pipeline-specific hard filters were used, in comparison to a proximal window alone (filter B) and a proximal window with additional PE/PPE filtering (filter D). However, recall rates of default filters (filter C) were slightly lower in comparison to the removal of combined proximal SNPs, loci encoding PE/PPE family proteins, mobile elements and other repetitive sequences (filter F) or PE/PPE family proteins, mobile elements and other repetitive sequences (filter G). This was specially the case for SNiPgenie, for which sensitivity increased to

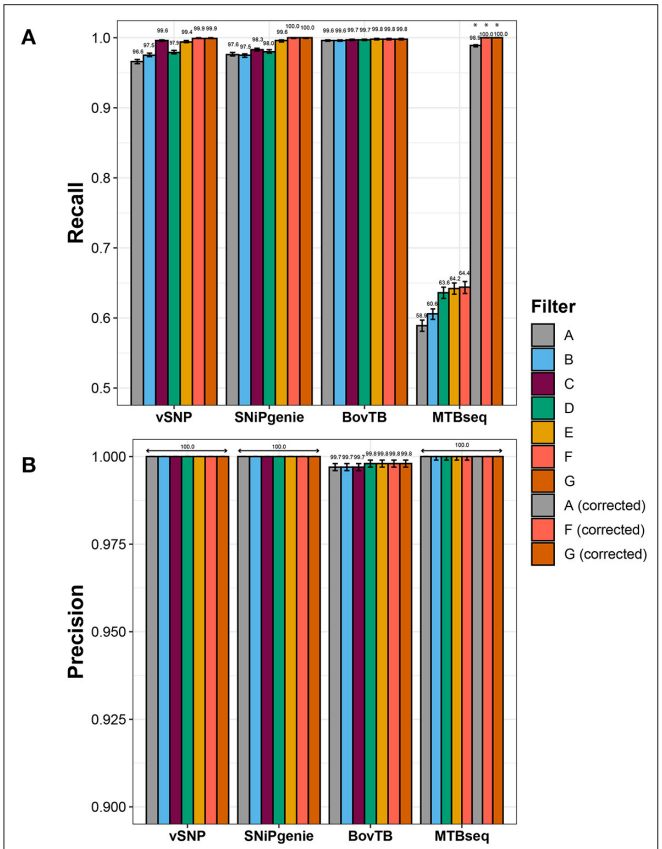


FIGURE 2 | Effect of hard filtering in the performance of the evaluated pipelines when compared to the simulated dataset as indicated in the **(A)** recall and **(B)** precision rates. Asterisks indicate filters in MTBseq for which the minimum coverage threshold was adjusted.

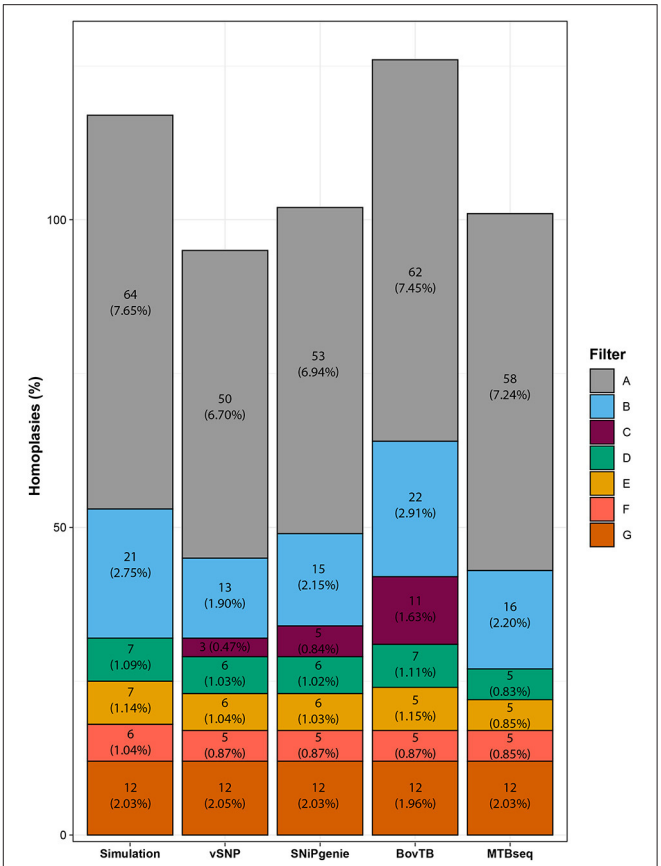


FIGURE 3 | Effect of filtering in the total number of homoplasic positions per pipeline identified by HomoplasyFinder. The percentage represents the proportion of homoplasic positions from the total number of identified positions using each specific filter.

levels similar to vSNP and BovTB when filters E, F, or G were applied.

Sensitivity remained below 65% for MTBseq despite the removal of problematic regions. Evaluation of the alignment files for this pipeline revealed that the increased amount of False Negative (FN) calls was produced by strand bias introduced by the artificial read generation, leading to forward or reverse read coverage being below the default minimum threshold ($n = 4$). Adjusting this threshold increased recall rates above 99% (**Figure 2A**).

After correction of MTBseq parameters, erroneous calls were further evaluated among the unfiltered pipeline results. FN calls were distributed unevenly among the simulated sequences (**Supplementary Figure 1**) and were mostly located within or near repetitive sequences (data not shown). More than half of the FN positions were shared by at least two of the pipelines, whereas 23 and 20% of the FN positions were identified only by BovTB and vSNP, respectively (**Supplementary Figure 2**). In addition, the majority of FN positions identified by BovTB in one sample were correctly detected as true SNPs in a varying number of samples (**Supplementary Figure 1C**).

A small proportion of false positive (FP) SNPs were identified by BovTB (43 SNPs across 37 positions) but, nevertheless, precision was high (>99%) for all of the evaluated pipelines (**Figure 2B**). Approximately 40% of FPs were located in repetitive regions and, although filtering improved precision in these cases, false positive SNPs were still detected (data not shown). Further analysis of the VCF files in BovTB revealed that the affected positions presented mixed calls caused by artificial sequencing errors. These positions were identified as both FNs and FPs by the Haplotype Caller and were appropriately removed by BovTB in later stages of the analysis. As a result, these mixed positions were ignored in the rest of the comparisons.

HomoplasyFinder identified 64 (7.65%) homoplasic positions among the generated sequences (**Figure 3**), mostly located within PE/PPE family proteins, intragenic regions or the *pks12* gene (data not shown). A similar proportion (6.70–7.45%) of homoplasies was identified in the alignments obtained from all of the evaluated pipelines. The removal of proximal SNPs reduced homoplasies to an average of 2%, similarly to what was

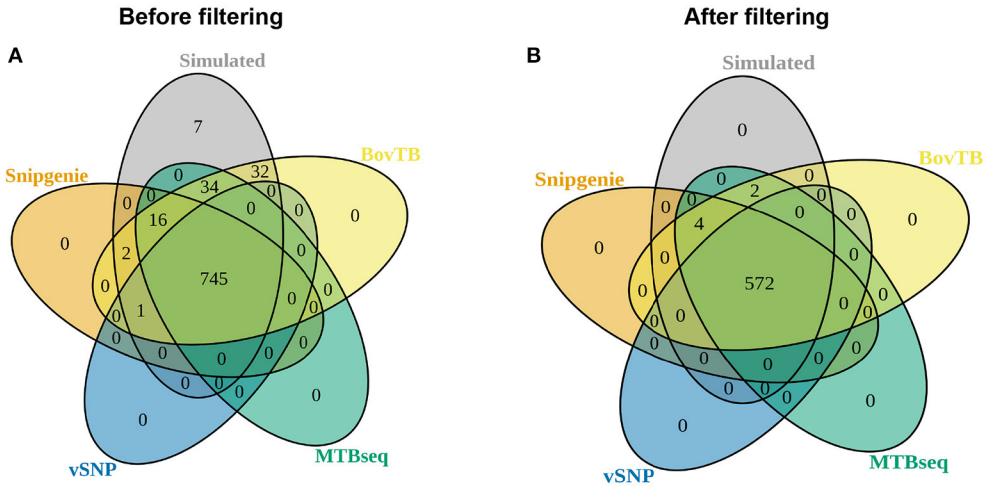


FIGURE 4 | Venn diagram showing the agreement between the positions identified by the different pipelines and the simulated dataset **(A)** before and **(B)** after hard filters (filter F) were applied.

observed for the removal of all of the problematic regions (filter G). Filtering of problematic regions with the proximity filter produced an additional reduction to 1%; most of the reduction was obtained with the removal of PE/PPE proteins alone and additional filters did not decrease the proportion significantly. Once all filters had been applied, all pipelines presented a reduced proportion of homoplasies compared to the ones present in the simulation. Finally, the use of default filters had a varying effect in the proportion of homoplasies, with vSNP and SNiPgenie obtaining the highest reduction in homoplasic positions.

Pipeline Agreement

There was a high agreement between the SNPs identified by the different pipelines and those in the simulated genomes, with the majority of simulated SNP positions being appropriately detected (**Figure 4A**). When proximal SNPs and repetitive sequences were filtered (filter F), there was an increase in the agreement between pipelines (**Figure 4B**). An identical agreement was observed when repetitive sequences were filtered without the proximity filter (filter G) (data not shown). SNiPgenie, BovTB and MTBseq were able to identify all of the SNPs from the simulation, while vSNP was not able to detect 7 SNPs (**Figure 4B**).

Tree Distance Comparison

The analysis of RF distances from best trees and bootstrap replicates revealed that trees output by the different pipelines clustered together with their simulated counterpart (**Figure 5**). In addition, cluster positioning was dependent on the type of hard filter used during the analysis. Trees obtained from the removal of problematic regions through filters D, E and F clustered together in one single group, whereas proximal filters (filter B) produced an intermediate clustering between unfiltered and filtered trees. The application of default filters (filter C) had an uneven effect in the different pipelines; BovTB trees did not separate considerably from proximally filtered trees (**Figure 5C**), whereas the trees produced by vSNP and SNiPgenie were closely related to other filtered trees.

Pairwise Phylogenetic Comparisons

In order to further evaluate topological differences among trees, a pairwise comparison of best trees obtained from each of the pipelines was carried out against their simulated counterpart (**Figure 6**; **Supplementary Figures 3, 4**). In general, there was a high level of agreement among trees and pipelines, with agreement being highest among filtered trees and, especially, among those obtained from BovTB (**Figure 6**). Among default filtered trees, those obtained from SNiPgenie and BovTB presented a higher agreement with the simulation than vSNP (**Supplementary Figure 4**).

Three major groups of taxa could be identified in all trees and no inter-cluster exchange was observed between pipelines. Among unfiltered trees, several isolates presented a small change in their relative location within the tree in the different pipelines (e.g., isolates 9, 10 and 11 in vSNP, 17 and 18 in SNiPgenie, or 28 in MTBseq and BovTB), sharing their Most Recent Common Ancestor (MRCA) with a different group of isolates to the one observed in the simulation. The main differences among filtered trees were produced by small topological variations among highly related taxa (e.g., isolates 12, 13, and 14) and the appearance of polytomies further contributed to the topological differences observed with the simulated tree. When compared against filter F, filter G resolved a small number of polytomies (**Supplementary Figure 5**, blue squares). In all cases, the filtered trees were highly congruent with the topology represented in the original publication (**Supplementary Figure 6**).

DISCUSSION

The application of WGS technologies in the study of aTB has increased in the last decade around the world. Despite its

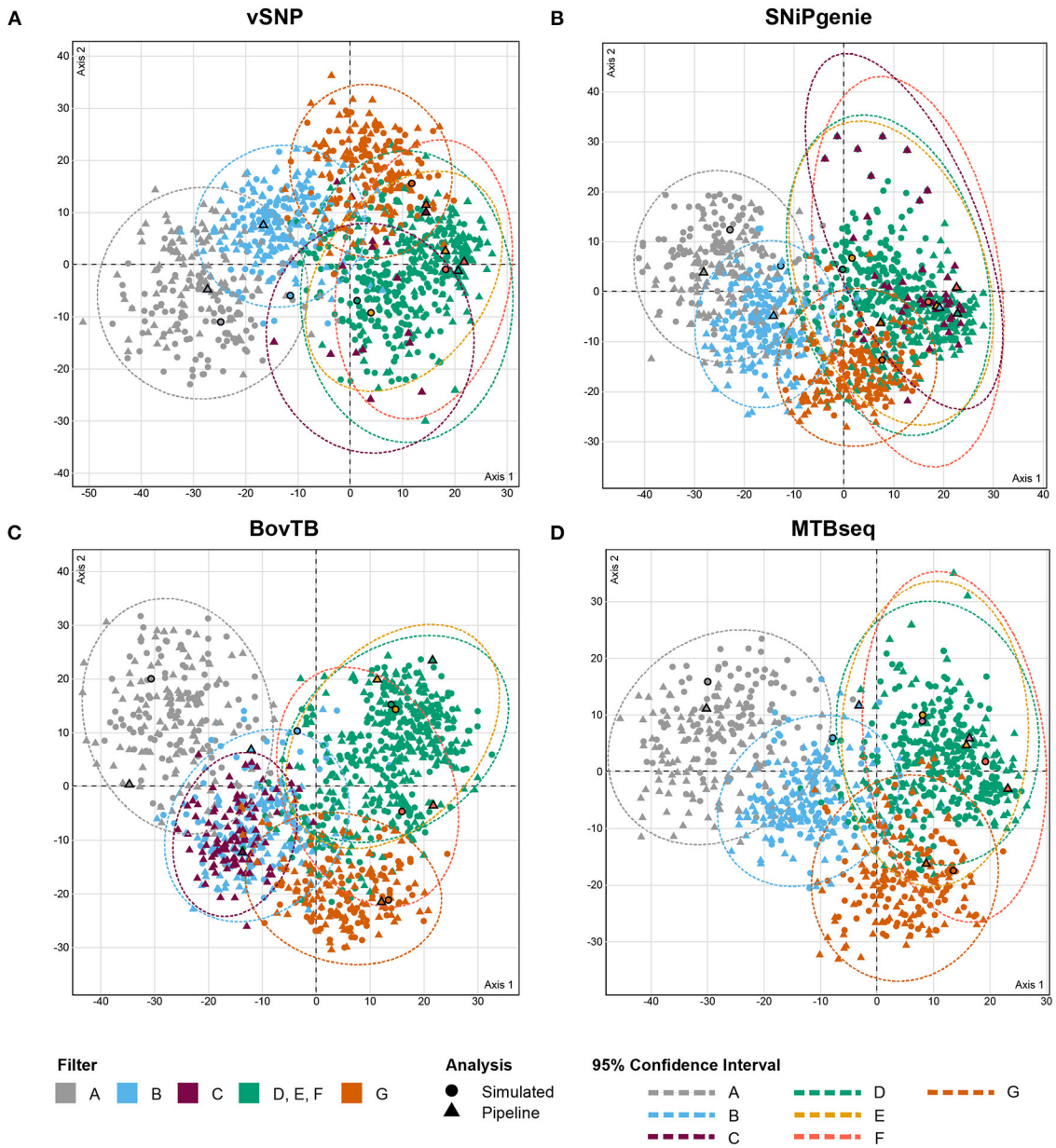


FIGURE 5 | MCA analysis of the RF distances (first two dimensions) between maximum likelihood (ML) trees obtained from core SNP multi-FASTA alignments produced by **(A)** vSNP, **(B)** SNiPgenie, **(C)** BovTB, and **(D)** MTBseq. Shapes without outlines correspond to bootstrap replicates whereas bold shapes correspond to the best ML trees output by RAxML. Color shading corresponds to the hard filtering approach used.

great promise as a higher resolution alternative to traditional molecular techniques in phylogenetic and epidemiological studies, its implementation in the eradication of bTB is still in development.

The digital nature of the data produced by WGS platforms and tools facilitates the exchange of information between laboratories, fostering collaboration between countries and organizations tasked with aTB control. However, the plethora of tools, parameters, protocols and types of analyses available may introduce variations that hamper this process of communication. Standardized procedures and parameters are needed in order to reduce the effect of these variations.

Prior to any standardization taking place, there is a need to evaluate the currently available techniques. None of the variant calling pipelines designed up to date for aTB have been benchmarked in the scientific literature, leading to uncertainty regarding the best method to implement in laboratories that are considering incorporating WGS analyses into their workflows. The aim of this study was to carry out an evaluation of the performance of the currently available aTB variant calling

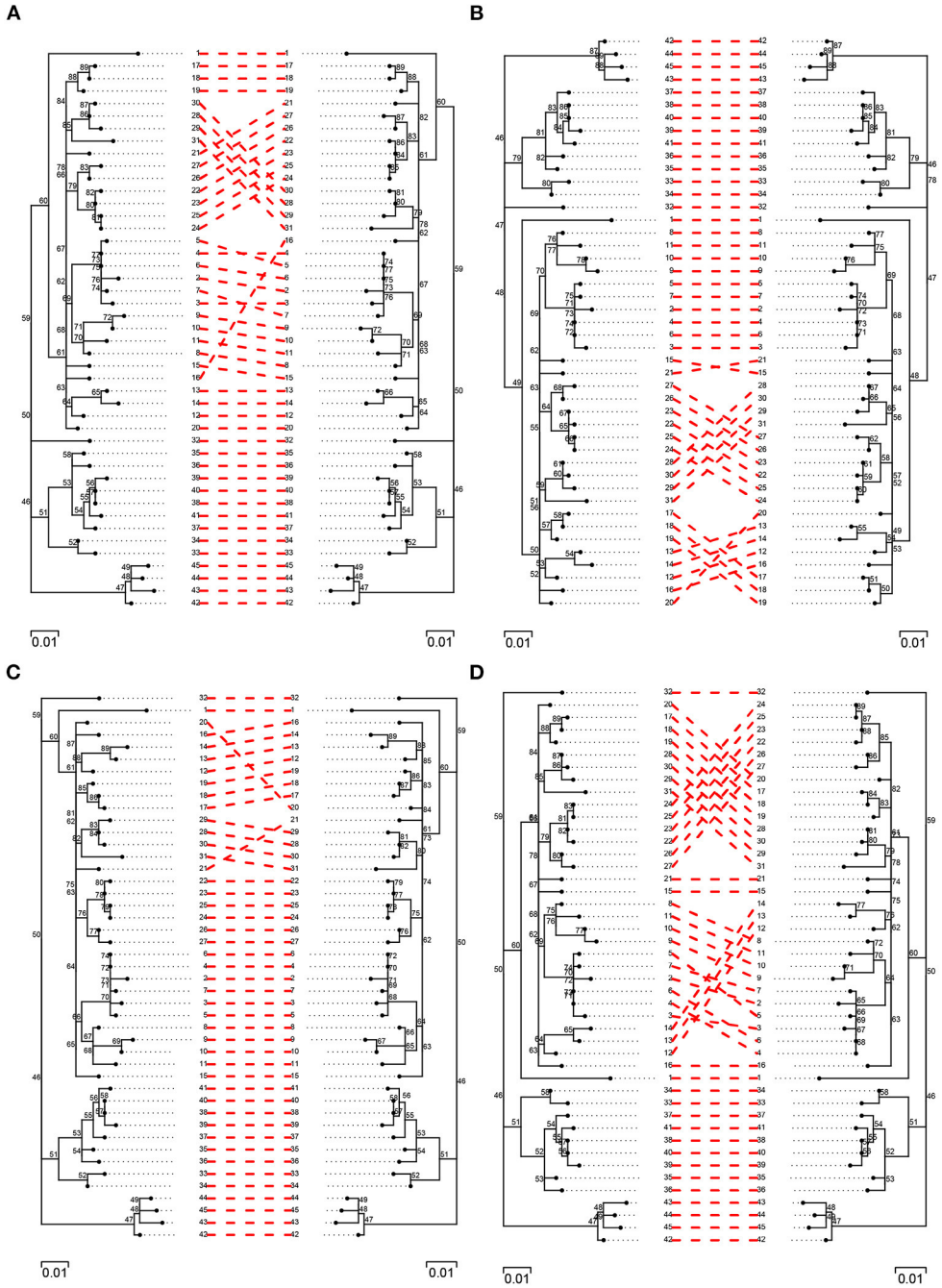


FIGURE 6 | Pairwise comparison of filtered (filter F) simulated trees (left) and trees obtained from the evaluated pipelines (right): **(A)** vSNP, **(B)** SNiPgenie, **(C)** BovTB, and **(D)** MTBseq.

pipelines and assess the degree of between-pipeline agreement in order to inform animal health authorities and laboratories.

The four pipelines evaluated in this study follow a similar procedure to other variant calling pipelines and employ bioinformatics tools widely used in the study of microorganisms. All pipelines use BWA as their sequence aligner (37), and the main procedural differences between them are related to the type of variant calling tools employed, and more importantly, the filtering process applied posteriorly.

The reduced sensitivity of vSNP, SNiPgenie, and MTBseq suggest that quality filters alone can negatively affect the performance of variant calling tools (21). The default minimum coverage settings of MTBseq, coupled with the minimum Allele Frequency of 75%, are important thresholds for the removal of

possible strand bias but were not well suited for the simulated dataset at hand in which SNPs were present in the sample but strand bias was present. This highlights the importance of evaluating and adapting these parameters to the data being evaluated, as suggested by the developers (29). Discrepancies in performance between pipelines were related to a varying proportion of erroneous calls (FP and FN SNPs). In general, a low number of FP and FN SNPs were found in this study, in contrast with previous results in which a high number of erroneous calls were identified by different caller combinations used along with BWA for the analysis of *M. tuberculosis* sequences (21). This could be due to the different approach used in this study for simulated genome and read generation and pipeline-specific filters.

All of the FNs produced by vSNP, SNiPgenie, and MTBseq, and a small proportion of those produced by BovTB, were located within or near repeat-rich regions. A small number of FN positions were identified by all four pipelines and only one of these positions was due to a reduced coverage (<10). This suggests that differences in FN calls could be due to how sample-specific filters deal with low quality regions.

BovTB was the only pipeline to identify FP SNPs and an evaluation of these positions revealed that they were a result of being identified as mixed positions. These were, in addition, partially responsible for a small proportion of the FN SNPs and were effectively removed from the analysis during the consensus calling. As a result, these positions did not have any effect in posterior analyses.

Identified variants are usually translated into phylogenetic trees as a visual aid to assess the genetic relatedness between strains, which can help to identify epidemiologically related isolates suggestive of transmission. In this study, the phylogenetic trees obtained from the different pipelines clustered together with their respective simulated trees and a pairwise inspection revealed a high level of agreement between simulated and pipeline-specific trees, especially on those obtained from BovTB. The phylogenetic trees inferred from the unfiltered SNP alignments obtained from each pipeline were compared against the reference phylogeny in order to assess the effect of pipeline performance in phylogenetic inference. Small divergences were identified in vSNP, SNiPgenie, and MTBseq with respect to the simulation, which are probably a consequence of a reduced sensitivity due to the application of stringent quality filters alone.

Despite the abovementioned effects of quality filters in performance, these are rarely the only parameter taken into consideration when carrying out variant calling in MTBC species. Repeat-rich regions, such as PE/PPE family proteins, mobile genetic elements or direct repeats, are generally considered low confidence regions either due to a higher error rate or mapping issues (19, 38), which complicates the variant calling process and could give rise to FN and FP SNPs. Indeed, the majority of erroneous calls in our simulation were identified in repeat-rich sequences, especially in *pe/ppe* genes and the *pks12* gene.

In addition to the technical constraints that repetitive regions pose to sequencing procedures and mapping algorithms, these can also have a negative impact in phylogenetic inference due to the occurrence of homoplasies. These are genetic traits that can arise independently in separate lineages due to different causes, mainly as a result of convergent evolution but also as a consequence of sequencing and mapping errors. Homoplasic events can add varying grades of background noise in phylogenetic signals and, therefore, must be taken into consideration (39). Due to the limited genetic variation in *M. bovis* strains, this could be especially relevant in closely related isolates and could potentially alter the epidemiological conclusions drawn from outbreak investigations (40). Although homoplasies can be identified anywhere in the genome, they are more frequent in repeat-rich regions such as *pe/ppe* genes (41).

In our study, a small proportion of homoplasies were identified in the simulation, probably due to the reduced number of variant positions in our dataset in comparison with published literature (32). Unfiltered alignments obtained from the different pipelines contained a similar number of homoplasies, although vSNP, and SNiPgenie presented a slightly lower proportion, probably due to their more stringent quality filters. A large reduction in homoplasies was observed when proximity filters were applied and, although the subsequent filtering of repetitive sequences decreased homoplasies further, filtering of these sequences alone (filter G) led to an increase in homoplasies. This indicates that the proximity filter could be an important feature to decrease homoplasies outside the standard repetitive sequences.

Quality filters are, therefore, usually coupled with the removal of problematic regions, an approach nowadays considered a standard procedure in WGS analyses of MTBC species (19). There is, however, no current consensus as to which of these regions should be included in the hard filtering process. In our study, filtering out a progressive amount of regions increased the sensitivity of vSNP and SNiPgenie to levels similar to those observed for BovTB or MTBseq. This positive effect in performance was especially evident for *pe/ppe* genes and mobile genetic elements, and is probably a result of the increased weight of these sequences in the overall composition of the *M. bovis* genome (7–10%) (42). Interestingly, the use of proximal filters had a strong effect in the clustering of phylogenetic trees with a clear separation of these from unfiltered trees, which in turn could be due to the large reduction in homoplasies. Agreement between pipelines also improved with filtering, indicating a difference in the stringency in which the evaluated pipelines deal with problematic regions, and their dependency on posterior masking for removal of low confidence regions.

Hard filtering also had a positive effect in the agreement between phylogenetic trees, as was reflected by the reduced differences among the best ML trees. Although topological differences were identified, these were limited to a reduced number of isolates and polytomies and did not alter the relationship between isolates as seen in the original publication (22). These topological variations are probably related to the overall low bootstrap support values of the identified clusters (43), which in turn could be due to the limited genetic diversity observed in the original dataset in which our simulation is based on. Indeed, *M. bovis* isolates in the original publication presented a maximum of 35 SNPs with respect to each other and a median distance of 14 SNPs once all filters were applied (22). Such a reduced diversity reflects a common drawback

encountered during *M. bovis* outbreak investigations, in which isolates from the same outbreak can accumulate a very small number of variants, hampering the definition of transmission events (26). Three different clusters were identified in our dataset in which genetic distances of *M. bovis* isolates were within 12 SNPs from each other; the maximum cut-off recommended for possible recent transmission of *M. tuberculosis* (44). In addition, polytomies can be resolved with increased isolate sampling, for example by including samples from wildlife or nearby breakdown events. However, this may not be a feasible option in many aTB outbreak investigations and, therefore, a removal of certain hard filters could be an interesting alternative to increase the amount of available informative SNPs. However, this alternative should be balanced to the risk of introducing possible biases or erroneous calls, such as FP SNPs. For example, although the removal of repetitive sequences without the proximity filter (filter G) increased the resolution of several polytomies, the increase in homoplasies could affect phylogenetic inference and needs to be considered.

There is little information as to how reliable low confidence regions are in phylogenetic inference, as their analysis has led to conflicting conclusions (45, 46). Nevertheless, there has been an increasing interest in the usefulness of filtering repeat-rich regions and recent data indicate that more than a half of the masked repetitive regions could be accurately identified using Illumina platforms (38). Even with the limitations of short-read sequencing platforms, the use of *de novo* assemblies or more refined masking filters may allow informative SNPs to be identified and retained (21, 38, 47). Furthermore, the introduction of long read sequencing could greatly improve the detection of variants within these regions of the genome (19). Improvements in the WGS analysis of problematic regions in MTBC species will surely benefit the field of aTB in the near future.

Pipeline choice may be based on other factors in addition to performance, and these have not been evaluated in this study. These include speed, use of disk space and memory, or ease of use, be it through the implementation of a GUI (SNiPgenie), limited command requirements (vSNP or BovTB) or by a straightforward data representation (vSNP and MTBseq) which could allow for more inexperienced users to access the bioinformatics analyses. In addition, the inclusion of additional analyses, such as antibiotic resistance profiling and cluster analysis (MTBseq), detection of INDELs and Regions of Difference (SNiPgenie), or lineage definition (vSNP, BovTB and MTBseq) could also be of interest for certain studies. However, in a similar manner to pipeline parameters, there is currently no standardized *M. bovis* lineage classification nor nomenclature based on WGS data. Although recent studies have suggested different lineages for *M. bovis* (48, 49), efforts toward this goal are still required. It is important to highlight that the results of our study are limited to simulated data and may not be representative of a real-life outbreak. The dataset used to generate our simulation does not correspond to an outbreak investigation but to a prevalence study. As a result, the capacity of each pipeline was approximated through their level of agreement with the simulation, rather than on their capacity to investigate true herd breakdown events. In addition, although this simulation partly mimics the negative impact of GC-rich sequences in genome coverage, it may be an underestimate in comparison to the actual sequencing of *M. bovis* isolates. Recent data highlight the existence of coverage blind spots in the *M. tuberculosis* reference genome which result from library preparation, sequencing as well as specific sequence attributes, such as homopolymers (50). Therefore, further work on a real-world dataset with a validated SNP profile and appropriate metadata is needed to evaluate these sources of bias.

Furthermore, the use of *M. bovis* AF2122/97 as a scaffold for the generation of simulated genomes meant that there were no sample-specific deletions, and therefore the capacity of these pipelines in calling SNPs near deletion events could not be evaluated. Furthermore, as is the case in human TB with *M. tuberculosis*, the choice of reference genome could also have an important effect in the WGS analysis of aTB due to differences in gene content between lineages, which could be masked by an inappropriate reference selection (19). This could be especially relevant when considering that traditional *M. bovis* lineages or clonal complexes are usually defined based on lineage-specific RDs, such as RDEu1 for the European 1 (Eu1) complex, or that certain genomic deletions may occur independently, such as the RD900 deletion. This study focused on the use of *M. bovis* AF2122/97, an Eu1 complex strain which is the default genome used by the evaluated aTB pipelines and the most extensively used *M. bovis* reference genome. However, the use of this reference genome in regions in which other clonal complexes are prevalent, such as the African 1 in western Africa, may lead to a loss of phylogenetic information. Therefore, other reference genomes may be better suited for different countries or regions and should be evaluated in the future.

Finally, it is important to note that, unlike other pipelines, manual and visual curation of SNPs is an important component of vSNP's design and functioning. As a result, a more detailed evaluation of this pipeline's results may have led to a reduced number of inconsistencies but would have added subjectivity to this comparison and was therefore avoided.

In conclusion, despite the above-mentioned limitations, the results of our comparison show that all evaluated pipelines perform well as long as similar hard filters are used, with minor differences amongst them with regard to performance and phylogenetic inference. This highlights the importance of standardizing and appropriately annotating filtering files when analyses are carried out between different laboratories or countries, and in particular in the context of aTB disease control.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.5281/zenodo.5179838, 5179838; https://github.com/Viloleal/bTB-pipeline-comparison-data-and-tools, None.

AUTHOR CONTRIBUTIONS

VL-L was involved in conceptualization, data curation, formal analysis, investigation, software, validation, and writing of the original draft. DF was involved in conceptualization, data curation, methodology, software, validation, and review and editing of the original draft. BR and LJ were involved in conceptualization, funding acquisition, and review and editing of the original draft. JÁ was involved in conceptualization and resources and review and editing of the original draft. SG was involved in conceptualization, funding acquisition, resources, validation, supervision, and review and editing of the original draft. All authors contributed to the article and approved the submitted version.

FUNDING

The author VL-L was funded with a predoctoral grant from the Complutense University of Madrid and Banco Santander 2017–2018 (CT17/17—CT18/17). This research project was made possible thanks to an international fellowship grant from the Complutense University of Madrid and Banco Santander 2020 (EB25/20). DF and SG acknowledge funding from the Department of Agriculture Food and the Marine (DAFM) award 2019R404 (BTBGenIE). This project was partially funded by the EU-RL for Bovine Tuberculosis.

ACKNOWLEDGMENTS

We would like to thank and acknowledge the valuable cooperation and feedback provided by Richard Ellis (APHA, BovTB), Jessica Hicks (USDA-APHIS, vSNP), and Tod Stuber (USDA-APHIS, vSNP) in the completion of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.780018/full#supplementary-material

REFERENCES

1. OIE. "Infection with *Mycobacterium tuberculosis* complex," In: *OIE—Terrestrial Animal Health Code*. (2021), 6.
2. OIE. "Bovine Tuberculosis," In: *OIE Terrestrial Manual*. (2018), 1058–1074.
3. European-Commission. "Regulation (EU) 2016/429 of the European Parliament and of the Council of 9 March 2016 on transmissible animal diseases and amending and repealing certain acts in the area of animal health ('Animal Health Law')," In: *EU 2016/429*. European Union Commision, (2016). EUR-Lex.
4. Livingstone PG, Hancox N, Nugent G, de Lisle GW. Toward eradication: the effect of *Mycobacterium bovis* infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. *N Z Vet J.* (2015) 63:4–18. doi: 10.1080/00480169.2014.971082
5. VerCauteren KC, Lavelle MJ, Campa H. Persistent spillback of bovine tuberculosis from white-tailed deer to cattle in Michigan, USA: status, strategies, and needs. *Front Vet Sci.* (2018) 5:301. doi: 10.3389/fvets.2018.00301
6. Gormley E, Corner LAL, Costello E, Rodriguez-Campos S. Bacteriological diagnosis and molecular strain typing of *Mycobacterium bovis* and *Mycobacterium caprae*. *Res Vet Sci.* (2014) 97:S30–43. doi: 10.1016/j.rvsc.2014.04.010
7. Durr PA, Clifton-Hadley RS, Hewinson GR. Molecular epidemiology of bovine tuberculosis - II. *Applicat Genotyp Revue Scientifique et Technique.* (2000) 19:689–701. doi: 10.20506/rst.19.3.1240
8. Frothingham, R., and Meeker-O'connell, W. A. (1998). Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology*. 144, 1189–1196. doi: 10.1099/00221287-144-5-1189
9. Kamerbeek, J., Schouls, L., Kolk, A., Agterveld, M., v., Soolingen, D., et al. Simultaneous Detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clinic Microbiol.* (1997) 35, 907–914. doi: 10.1128/jcm.35.4.907-914.1997
10. Rodriguez-Campos S, Navarro Y, Romero B, de Juan L, Bezos J, Mateos A, et al. Splitting of a prevalent *Mycobacterium bovis* spoligotype by variable-number tandem-repeat typing reveals high heterogeneity in an evolving clonal group. *J Clin Microbiol.* (2013) 51:3658–65. doi: 10.1128/JCM.01271-13
11. Rodriguez S, Romero B, Bezos J, de Juan L, Alvarez J, Castellanos E, et al. High spoligotype diversity within a *Mycobacterium bovis* population: clues to understanding the demography of the pathogen in Europe. *Vet Microbiol.* (2010) 141:89–95. doi: 10.1016/j.vetmic.2009.08.007
12. Smith NH. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infect Genet Evol.* (2012) 12:857–65. doi: 10.1016/j.meegid.2011.09.007
13. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE.* (2009) 4:e7815. doi: 10.1371/journal.pone.0007815
14. Rodriguez-Campos S, Aranaz A, de Juan L, Saez-Llorente JL, Romero B, Bezos J, et al. Limitations of spoligotyping and variable-number tandem-repeat typing for molecular tracing of *Mycobacterium bovis* in a high-diversity setting. *J Clin Microbiol.* (2011) 49:3361–4. doi: 10.1128/JCM.00301-11
15. Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev.* (2015) 28:541–563. doi: 10.1128/CMR.00075-13
16. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, et al. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol.* (2014) 52:2479–86. doi: 10.1128/JCM.00567-14
17. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential? *Clin Microbiol Infect.* (2018) 24:604–9. doi: 10.1016/j.cmi.2017.10.030
18. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis.* (2015) 15:1193–202. doi: 10.1016/S1473-3099(15)00062-6
19. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol.* (2019) 17:533–45. doi: 10.1038/s41579-019-0214-5
20. Jajou R, Kohl TA, Walker T, Norman A, Cirillo DM, Tagliani E, et al. Toward standardization: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Euro Surveill.* (2019) 24:1900130. doi: 10.2807/1560-7917.ES.2019.24.50.1900130
21. Walter KS, Colijn C, Cohen T, Mathema B, Liu Q, Bowers J, et al. Genomic variant-identification methods may alter *Mycobacterium*

tuberculosis transmission inferences. *Microb Genom.* (2020) 6:8. doi: 10.1099/mgen.0.000418
22. Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H, et al. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. *Microb Genom.* (2020) 6:20. doi: 10.1099/mgen.0.000388
23. Hauer A, Michelet L, Cochard T, Branger M, Nunez J, Boschiroli ML, et al. Accurate phylogenetic relationships among *Mycobacterium bovis* strains circulating in france based on whole genome sequencing and single nucleotide polymorphism analysis. *Front Microbiol.* (2019) 10:955. doi: 10.3389/fmicb.2019.00955
24. Kohl TA, Kranzer K, Andres S, Wirth T, Niemann S, Moser I. Population Structure of *Mycobacterium bovis* in Germany: a long-term study using whole-genome sequencing combined with conventional molecular typing methods. *J Clin Microbiol.* (2020) 58:e01573–20. doi: 10.1128/JCM.01573-20
25. Orloski K, Robbe-Austerman S, Stuber T, Hench B, Schoenbaum M. Whole genome sequencing of *Mycobacterium bovis* isolated from livestock in the United States, 1989–2018. *Front Vet Sci.* (2018) 5:253. doi: 10.3389/fvets.2018.00253
26. Price-Carter M, Brauning R, de Lisle GW, Livingstone P, Neill M, Sinclair J, et al. Whole genome sequencing for determining the source of *Mycobacterium bovis* infections in livestock herds and wildlife in New Zealand. *Front Vet Sci.* (2018) 5:272. doi: 10.3389/fvets.2018.00272
27. Yue JX, Liti G. simuG: a general-purpose genome simulator. *Bioinformatics.* (2019) 35, 4442–4. doi: 10.1093/bioinformatics/btz424
28. Frampton M, Houlston R. Generation of artificial FASTQ files to evaluate the performance of next-generation sequencing pipelines. *PLoS ONE.* (2012) 7:e49110. doi: 10.1371/journal.pone.0049110
29. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ.* (2018) 6:e5895. doi: 10.7717/peerj.5895
30. Chen, H. (2018). *Generate High-Resolution Venn and Euler Plots*. Available online at: https://cran.r-project.org/web/packages/VennDiagram/VennDiagram.pdf
31. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly.* 6:80–92. doi: 10.4161/fly.19695
32. Crispell J, Balaz D, Gordon SV. HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. *Microb Genom.* (2019) 5:1. doi: 10.1099/mgen.0.000245
33. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom.* (2016) 2:e000056. doi: 10.1099/mgen.0.000056
34. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
35. Jombart T, Kendall M, Almagro-Garcia J, Colijn C. treespace: Statistical exploration of landscapes of phylogenetic trees. *Mol Ecol Resour.* (2017) 17:1385–92. doi: 10.1111/1755-0998.12676
36. Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* (2012) 3:217–23. doi: 10.1111/j.2041-210X.2011.00169.x
37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
38. Marin M, Vargas R, Harris M, Jeffrey B, Epperson LE, Durbin D, et al. Genomic sequence characteristics and the empiric accuracy of short-read sequencing. *bioRxiv.* (2021) 2004:438862. doi: 10.1101/2021.04.08.438862
39. Brandley MC, Warren DL, Leache AD, McGuire JA. Homoplasy and clade support. *Syst Biol.* (2009) 58:184–98. doi: 10.1093/sysbio/syp019
40. Nakanishi N, Wada T, Arikawa K, Millet J, Rastogi N, Iwamoto T. Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages. *Infect Genet Evol.* (2013) 16:174–7. doi: 10.1016/j.meegid.2013.02.007
41. Tantivitayakul P, Ruangchai W, Juthayothin T, Smittipat N, Disratthakit A, Mahasirimongkol S, et al. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in *Mycobacterium tuberculosis*. *Sci Rep.* (2020) 10:1–10. doi: 10.1038/s41598-020-64895-4
42. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* (1998) 393:537–44. doi: 10.1038/31159
43. Pettengill JB, Luo Y, Davis S, Chen Y, Gonzalez-Escalona N, Ottesen A, et al. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. *PeerJ.* (2014) 2:e620. doi: 10.7717/peerj.620
44. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis.* (2013) 13:137–46. doi: 10.1016/S1473-3099(12)70277-3
45. Godfroid M, Dagan T, Kupczok A. Recombination signal in *Mycobacterium tuberculosis* stems from reference-guided assemblies and alignment artifacts. *Genome Biol Evol.* (2018) 10:1920–6. doi: 10.1093/gbe/evy143
46. Phelan JE, Coll F, Bergval I, Anthony RM, Warren R, Sampson SL, et al. Recombination in *pe/ppe* genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. *BMC Genom.* (2016) 17:151. doi: 10.1186/s12864-016-2467-y
47. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respirat Med.* (2013) 1:786–92. doi: 10.1016/S2213-2600(13)70231-5
48. Loiseau C, Menardo F, Aseffa A, Hailu E, Gumi B, Ameni G, et al. An African origin for *Mycobacterium bovis*. *Evol Med Public Health.* (2020) 2020:49–59. doi: 10.1093/emph/eoaa005
49. Zimpel CK, Patane JSL, Guedes ACP, de Souza RF, Silva-Pereira TT, Camargo NCS, et al. Global Distribution and Evolution of *Mycobacterium bovis* Lineages. *Front Microbiol.* (2020) 11:843. doi: 10.3389/fmicb.2020.00843
50. Modlin SJ, Robinhold C, Morrissey C, Mitchell SN, Ramirez-Busby SM, Shmaya T, et al. Exact mapping of Illumina blind spots in the *Mycobacterium tuberculosis* genome reveals platform-wide and workflow-specific biases. *Microb Genom.* (2021) 7:3. doi: 10.1099/mgen.0.000465

Conflict of Interest: SNiPgenie was developed by DF and SG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 16 December 2021
doi: 10.3389/fvets.2021.794228

Molecular Characterization of Infectious Bronchitis Virus Strain HH06 Isolated in a Poultry Farm in Northeastern China

Ghulam Abbas[1], *Yue Zhang*[1], *Xiaowei Sun*[1], *Huijie Chen*[2], *Yudong Ren*[3], *Xiurong Wang*[4], *Muhammad Zulfiqar Ahmad*[5], *Xiaodan Huang*[1] *and Guangxing Li*[1]*

[1] Heilongjiang Key Laboratory for Animal and Comparative Medicine, College of Veterinary Medicine, Northeast Agricultural University, Harbin, China, [2] College of Pharmaceutical Engineering, Jilin Agriculture Science and Technology University, Jilin, China, [3] Department of Computer Science and Technology, College of Electrical and Information Technology, Northeast Agricultural University, Harbin, China, [4] State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Science, Harbin, China, [5] Department of Plant Breeding and Genetics, Faculty of Agriculture, Gomal University, Dera Ismail Khan, Pakistan

OPEN ACCESS

Edited by:
Iryna Goraichuk,
Institute of Experimental and Clinical Veterinary Medicine, Ukraine

Reviewed by:
Tereza Cristina Cardoso,
Universidade Estadual de São Paulo, Brazil
Saima Ashraf,
University of Veterinary and Animal Sciences, Pakistan
Riaz Hussain,
Pir Mehr Ali Shah Arid Agriculture University, Pakistan

***Correspondence:**
Guangxing Li
ligx@neau.edu.cn

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 13 October 2021
Accepted: 25 November 2021
Published: 16 December 2021

Citation:
Abbas G, Zhang Y, Sun X, Chen H, Ren Y, Wang X, Ahmad MZ, Huang X and Li G (2021) Molecular Characterization of Infectious Bronchitis Virus Strain HH06 Isolated in a Poultry Farm in Northeastern China. Front. Vet. Sci. 8:794228. doi: 10.3389/fvets.2021.794228

Spike (S) glycoprotein is an important virulent factor for coronaviruses (CoVs), and variants of CoVs have been characterized based on S gene analysis. We present phylogenetic relationship of an isolated infectious bronchitis virus (IBV) strain with reference to the available genome and protein sequences based on network, multiple sequence, selection pressure, and evolutionary fingerprinting analysis in People's Republic of China. One hundred and elven strains of CoVs i.e., *Alphacoronaviruses* (Alpha-CoVs; $n = 12$), *Betacoronaviruses* (Beta-CoVs; $n = 37$), *Gammacoronaviruses* (Gamma-CoVs; $n = 46$), and *Deltacoronaviruses* (Delta-CoVs; $n = 16$) were selected for this purpose. Phylogenetically, SARS-CoV-2 and SARS-CoVs clustered together with Bat-CoVs and MERS-CoV of Beta-CoVs (C). The IBV HH06 of Avian-CoVs was closely related to Duck-CoV and partridge S14, LDT3 (teal and chicken host). Beluga whale-CoV (SW1) and Bottlenose dolphin-CoVs of mammalian origin branched distantly from other animal origin viruses, however, making group with Avian-CoVs altogether into Gamma-CoVs. The motif analysis indicated well-conserved domains on S protein, which were similar within the same phylogenetic class and but variable at different domains of different origins. Recombination network tree indicated SARS-CoV-2, SARS-CoV, and Bat-CoVs, although branched differently, shared common clades. The MERS-CoVs of camel and human origin spread branched into a different clade, however, was closely associated closely with SARS-CoV-2, SARS-CoV, and Bat-CoVs. Whereas, HCoV-OC43 has human origin and branched together with bovine CoVs with but significant distant from other CoVs like SARS CoV-2 and SARS-CoV of human origin. These findings explain that CoVs' constant genetic recombination and evolutionary process that might maintain them as a potential veterinary and human epidemic threat.

Keywords: molecular epidemiology, zoonosis, coronaviruses, infectious bronchitis virus, evolution

INTRODUCTION

Coronaviruses (CoVs) are a group of RNA viruses that mainly infect respiratory systems of domestic and wild birds as well as mammals including humans. These viruses belong to the subfamily *Orthocoronavirinae* of the family *Coronaviridae* (1, 2), further classified into *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* genera (3). The CoVs are enveloped viruses with a helical-symmetry nucleocapsid that projects club-shaped spikes. The genome is a positive-sense single-stranded RNA of 26-32 kilobase pairs that encodes main structural proteins i.e., spike (S) glycoprotein comprising 2 subunits (S1 and S2), envelop (E) protein, membrane (M) protein, and nucleocapsid (N) protein (4–6). The S glycoprotein is an important virulent factor i.e., plays role in viral adsorption and invasion into the host cells (7, 8). The evolution of S protein is more active and it often undergoes mutation. The changes of certain amino acids influence the conformation of antigenic determinants, resulting in the generation of new strains (9). Usually, difference of amino acids in S1 by 20∼50% is considered for different serotypes, however in some instances, only 2% or 10∼15 amino acids variation may lead to the emerging of different serotypes of infectious bronchitis virus (IBV) (10, 11). Hence, evolution process is considered important factor which plays major role in many emerging serotypes. Indicating the positions of amino acids evolutionary conservation is important for maintaining the protein structure and function (12, 13). Therefore, detection of selected sites may enlighten the selection forces and detects the functionally significant sites for CoVs S protein interaction.

IBV was the first coronavirus described, and was found by Schalk and Hawn (1931) in North Dakota of the United States of America (USA) (14). After that, the related CoVs have been isolated from other birds, mammals, and rodents (15); however, the first CoV in a human was identified in the 1960s and was associated with the common cold (16, 17). In the last couple of decades, disease pandemic viruses Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome (MERS) have caused a larger number of mortalities (18, 19). The last few days of the year 2019 bared the advent of a pathogenic disease caused by a novel epidemic of CoV (Severe Acute Respiratory Syndrome 2; SARS-CoV-2) (20). During the manifestation of disease, the CoVs has broader tissue tropism, mainly toward respiratory system, however can potentially infect other organ systems e.g., gastrointestinal and reproductive tract (21, 22). *Avian coronavirus* (Avian-CoV) from the genus *Gammacoronavirus* causes avian infectious bronchitis which is highly infectious, and affects respiratory, renal, and reproductive system. It causes significant drop in weight gain (in broilers) and egg production (in layers) (10, 23). Though, chickens (*Gallus gallus*) are considered natural hosts of IBV, these viruses have been reported to cause enteric diseases in turkeys (Turkey-CoV) (24), renal and respiratory disease in pheasants (Pheasant-CoV) (25). Duck coronavirus (Duck-CoV) (26), peafowl coronavirus (PeF-CoV) (27), pigeon coronavirus (Pi-CoV) (28), Canada goose coronavirus (GCoV) (29) seem to be less pathogenic. However, the host range might be even broader e.g., swans, mallards, geese, and gulls also exhibited IBV-like symptoms and yielded viruses that had gene fragments from M41, 793B, and QX lineages (25).

During replication, Avian-CoVs have high genetically recombination potential (30). Genetic techniques have played an important role in understanding the genetic relatedness of different microorganisms, pathogens, and the diseases caused by the pathogens as well as their evolutionary mechanism (31, 32). Recent pandemic of COVID-19 has drawn attention to the potential zoonotic threats of the CoVs (33, 34). In order to control and prevent the occurrence of such pandemics, it is important to understand the virus origin, genetic mechanics and its mode of transmission between intra-host species. In our study, we presented phylogenetic relationship of an isolated IBV strain with reference to the available genome and protein sequences based on phylogenetic, recombinant network, multiple sequence, selection pressure, and evolutionary fingerprinting analysis.

MATERIALS AND METHODS

Chicken Embryos

Specific-pathogen-free (SPF) chicken embryos were used for virus isolation and titration. 9–11-day-old SPF chicken embryos were purchased from the Experimental Animal Center of the Harbin Veterinary Research Institute (HVRI), Chinese Academy of Agricultural Sciences, People's Republic of China (PRC).

IBV HH06 Isolation and Identification

The IBV isolate HH06 (GenBank accession number MH181793.1) was isolated from Hy-Line chicken suspected of having infectious bronchitis infection in a farm at Northeastern China and kept in the Veterinary Pathology Laboratory of the College of Veterinary Medicine, Northeast Agricultural University as earlier described by Ren et al. (35). Briefly, the purification and propagation of the isolate was done by three times passaging in allantoic cavity of 9-day-old SPF embryonated chicken eggs (ECE) and distinct IBV characteristics e.g., embryo dwarfing, hemorrhages, curling or stunting of embryos were observed (36). The 50% embryo infectious dose (EID_{50}) was measured by inoculating 10-fold dilutions in groups of 9-day-old ECE as described previously (37).

Viral RNA Extraction and Reverse Transcription Polymerase Chain Reaction

RNA was extracted from the allantoic fluid using TRIzol reagent (TaKaRa, Dalian, China), according to the protocol of manufacturer. Reverse transcriptase reaction was performed according to procedures provided by Qiagen RT-PCR kit. Briefly, a total of 20 μl mix was prepared as follows; 8 μL DEPC, 4 μl 5×RT-buffer, 1 μl dNTP, 1 μl Oligo (dT), 5 μl RNA, 0.5 μl m-MLV, and 0.5 μl RNase. After preparation of cDNA, IBV-N primers (189 bp) [Sense: CAAGCTAGGTTTAAGCCAGGT; Antisense: TCTGAAAACCGTAGCGGATAT] (38) were used for RT-PCR IBV detection. PCR reactions included initial denaturation for 95°C for 5min, followed by 40 cycles of denaturation for 30 sec at 94°C, annealing for 30 sec at 55.7°C, and extension for 2 min, at 72°C and a final extension cycle at

72°C for 10 min with holding temperature of 4°C. PCR products were run on 1% agarose gel electrophoresis for confirmation and visualized by subsequent UV trans-illumination (Bio-Best 140E, SIM, USA) (39).

Cloning and Sequencing of Target Gene

Gel Extraction Mini Kit (Omega, USA) was used for DNA purification and recovery of the PCR products. Purified PCR products ligated with a TA cloning vector pMD18-T (TaKaRa, Japan) were transformed into competent *E. coli* cells strain JM109 (Beijing TransGen Biotech, PRC). Confirmation of clones containing recombinant plasmid was achieved by PCR and restriction enzyme digestion. The PCR conditions were the same as that for the above-mentioned PCR amplification. Three positive clones were randomly selected and cultured. Recombinant plasmids were sequenced at Shanghai Sanggong Biological Engineering Technology & Services Co., Ltd (Shanghai, China).

Genetic, Phylogenetic, Motif Analysis, and Comparative Sequence Alignment

A total of 111 corona viruses from the *Coronaviridae* family were selected to analyze phylogeny and genetic relatedness. The sequence of IBV strain HH06 (GenBank number MH181793.1) isolated in this study along with 110 viruses were aligned using ClustalW multiple alignment algorithm. The phylogenetic tree was constructed based on a maximum-likelihood method (JTT model) using the MEGA 7.0 version with bootstrap replicates (1,000) (https://www.megasoftware.net) (40). The sequences from GenBank (https://www.ncbi.nlm.nih.gov/genbank), which represent the well-established four genera of *Coronaviridae* are enlisted in **Supplementary Table S1**. The motif analysis was performed using the protein sequence of the S genes through the online database the MEME (https://meme-suite.org/meme/tools/meme). The ClustalW in MEGA 7.0 was used to align the amino acid sequence of S proteins of 31 CoVs representing different hosts, origins, and genotypes among all selected 111 CoVs of four genera of the *Coronaviridae* family (**Supplementary Table S1**). Aforementioned sequences were subjected to the GeneDoc program to shade the conserved amino acids in alignment (40).

Recombinant Network Tree

The spike protein sequence was analyzed to evaluate the degree of possible recombination. A network tree was assembled from protein sequences alignment of IBV strain HH06 and 110 reference CoV strains from different genera by using the SplitTree 4.13.1 (http://www.bio-soft.net/tree/SplitsTree.htm) (41).

Selection Pressure Analysis

Online database SELECTION (https://selecton.tau.ac.il/) was used to ratify codon sites under selection pressure. Aligned codon sequence of CoVs proteins was tested in the SELECTION that allows shifting the ω ratio between different codons within the aligned sequence and this was measured by maximum-likelihood test through Bayesian inference method (42). Moreover, the selection results are shown with color scales demonstrating various types of selection. The identification and accession numbers of protein coding sequence of gene (CDS) are presented in **Supplementary Table S1**.

Evolutionary Fingerprinting

We used the EFP model to represent evolutionary fingerprints as probability distributions and presented a methodology for comparing these distributions in a way that is robust against variations in data set size and divergence. The EFP was done by using an online Data Monkey classical tool (https://www.datamonkey.org/) (43) on the aligned CDS sequences of selected CoVs including SARS-CoV-2, SARS-CoV, MERS –CoV, and IBV.

Evolutionary Analysis of Diversifying Selection

Neutrality analysis was done based on maximum likelihood computation of dN-dS using the HyPhy software program implemented in MEGA 7.0, using the Nei-Gojobori method (44). All position gaps and missing data were eliminated. The evolutionary history was inferred using the maximum likelihood method based on the Kimura 2-parameter model and the phylogenetic tree of CoVs S gene was constructed using MEGA 7.0 software package based on maximum likelihood (45). ClustalW software was used for genetic sequence and the similarity analysis of S genes. For each codon, estimates of the numbers of inferred synonymous (s) and non-synonymous (n) substitutions are presented along with the numbers of sites that are estimated to be synonymous (S) and non-synonymous (N). These estimates were produced using the joint Maximum Likelihood reconstructions of ancestral states under a Muse-Gaut model of codon substitution and Tamura-Nei model of nucleotide substitution (43). For estimating ML values, a tree topology was automatically computed. The test statistic dN–dS is used for detecting codons that have undergone positive selection, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of non-synonymous substitutions per site (n/N). A positive value for the test statistic indicates an overabundance of non-synonymous substitutions. In this case, the probability of rejecting the null hypothesis of neutral evolution (p-value) was calculated (46). Values of $p < 0.05$ are considered significant at a 5% level and are highlighted. Normalized dN–dS for the test statistic is obtained using the total number of substitutions in the tree (measured in expected substitutions per site). It is useful for making comparisons across data sets. Maximum Likelihood computations of dN and dS were conducted using HyPhy software package. The analysis involved 110 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 2,625 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (40). Evolutionary analysis of diversifying selection was performed by various approaches to detect the episodic diversifying detection affecting individual codon sites. Mixed-effects model evolution (MEME) combines the fixed effects to identify instances of both episodic diversifying selection and pervasive positive selection at the individual branch site

level using Markov Chain Monte Carlo (MCMC) routine, which ensures the robustness against model misspecification over predefined sites through approximate Bayesian method (47). The fitting of MEME to alignment, MG94xREV codon model, was applied using parameter estimates $\omega = \beta/\alpha$ fitted to the data using the GTR nucleotide model as initial values. The selective pressure was measured with two parameters β: β- < α and β+ and the alternative model include four parameters for each site: β-, β+, and α estimating site to site substitution variability rates 42]. The values of $p < 0.05$ were considered as significant from the LRT based on χ2 asymptotic distribution (44).

RESULTS

Isolation, Identification, and Confirmation of IBV HH06

IBV strain HH06 was isolated using 9–11-day-old SPF ECE. The morphology and gross changes were observed, dwarfism, hemorrhage, and congestion were found (**Supplementary Figure S1**). The allantoic fluid was harvested. The presence of IBV HH06 was confirmed by RT-PCR using IBV-N specific product (189 bp) primers (**Supplementary Figure S2**).

Genetic, Phylogenetic, and MOTIF Analysis of the S Protein

A phylogenetic tree (**Figure 1**) was constructed based on amino acid CDS of S glycoprotein to assess the genetic relevancy and discrimination among the four main genera (*Alphacoronavirus, Betacoronavirus, Gammacoronavirus,* and *Deltacoronaviruse* of the *Coronaviridae* family) comprising different mammalian and avian coronaviruses. Among selected viruses were SARS-CoV-2, SARS-CoV, and MERS-CoV of mammalian origin along with IBV strain HH06 of avian origin isolated in this study. The IBV strain HH06 clustered into the GI-19 genotype (QX-type) of Avian-CoVs that belongs to *Gammacoronavirus* and was closely related to Duck-CoV DK/CH/HN/ZZ2004 (GenBank accession number AEO86768.1) and partridge S14 (GenBank accession number AAT70772.1), LDT3 (GenBank accession number AAU14248.1) (teal and chicken host) of GI-18 (LDT3-A). Ph-CoV strains ph/China/I0710 (GenBank accession number QDA76255.1) and PSH050513 (GenBank accession number AAZ85066.1) of Avian-CoVs also clustered closely in the same group. These indicate the intra-host evolution of Avian-CoVs from one genotype to another and from one host to another host. The S gene glycoprotein sequence of Beluga whale-CoV SW1 (GenBank accession number ABW87820.1) and Bottlenose dolphin-CoVs 37112-1 (GenBank accession numbers QII89019.1) and HKU22 (GenBank accession numbers 211 AHB63481.1) clustered at distant; however, making group along with Avian-CoVs altogether into *Gammacoronaviruses*. Another sequences set of delta-CoVs that comprises sparrow CoV, munia CoV, and Quail-CoV of Avian-CoVs along with porcine delta-CoV clustered together with Feline CoVs, Canine CoVs. PRCV, TGEV, and PEDV belonging to alpha CoVs, hence making a discrete cluster covering the coronaviruses from avian and mammalian species. A cluster of exactingly CoVs, branch off into the human CoVs (HCV-OC43), and a cluster containing murine CoVs, equine CoVs, and rodent CoVs. The MERS viruses of Beta-CoVs (C) were closely associated with SARS-CoVs-2 and SARS-CoVs. In the same manner, Bat-CoVs show close association with SARS-CoVs. Currently, the S gene mainly determines the serotype and tissue tropism of different virus strains. The conserved domains of spike proteins are determined by the MEME (**Figure 2**). In total, we found ten motifs in the S gene, and their annotations confirmed through the Pfam databases. All S protein motifs are generally well-conserved and similar within the same phylogenetic class; however, variation was also observed (**Supplementary Table S2**).

Recombinant Network Analysis

Further, a recombinant network tree was generated (**Figure 3A**) of 111 selected strains from four genera (Alpha, Beta, Gamma, and Delta) of CoVs comprising human, animals, and avian origin. The CoVs of human origin distributed in separate clades. The *Betacoronaviruses* B (SARS-CoV-2 and SARS-CoV) and *Betacoronaviruses* D (Bat-CoVs) shared common clades, although branched differently, however possible to be originated from bats. The MERS-CoVs of camel and human origin from *Betacoronaviruses* C distributed in individual clade, however, was closely associated with *Betacoronaviruses* B and D. Although human coronavirus (HCoV-OC43) has a human origin, it branched together with Bovine-CoVs and showed a significant distance from other human CoVs like SARS-CoV-2 and SARS-CoV. Feline- and Canine-CoVs showed common origin thus branched together, however, were significantly distant from other SARS-related viruses. The *Gammacoronaviruses* of avian origin indicated association at a distant place from *Betacoronaviruses* (SARS-CoV2, SARS-CoV, Bat-CoV and MERS-CoVs), however, Beluga whale-CoV strain SW1 and Bottlenose dolphin-CoVs strains 37112-1 and CF090325 of animal origin differentiated into separate clade from other CoVs of animal origin and clustered at distant however making group along with Avian-CoVs altogether into *Gammacoronaviruses*. Isolate HH06 from the current study was clustered into Avian-CoVs group, mainly of chicken origin, however, Duck-CoV and Pheasant-CoV showed a close relationship with the HH06 isolate from chicken. Similarly, Turkey-CoVs separated clade from other avian mainly chicken however 3 Turkey-CoVs differentiated into a separate clade.

Comparative Sequence Alignment

Alignment of interrelated amino acids sequences was performed in comparison to 4 genera of coronaviruses and these isolates expressed distinctive amino acids mutations in the HVRs compared to different species. The result of the alignment of the S gene of IBV isolate HH06 from this study shown the occurrence of numerous mutational sites (**Figure 3B**). The majority of mutation sites were located in HVRI in spike protein structure that was comparable to sequence alignment of amino acids of selected 31 CoVs of 4 genera of *Coronaviridae* family.

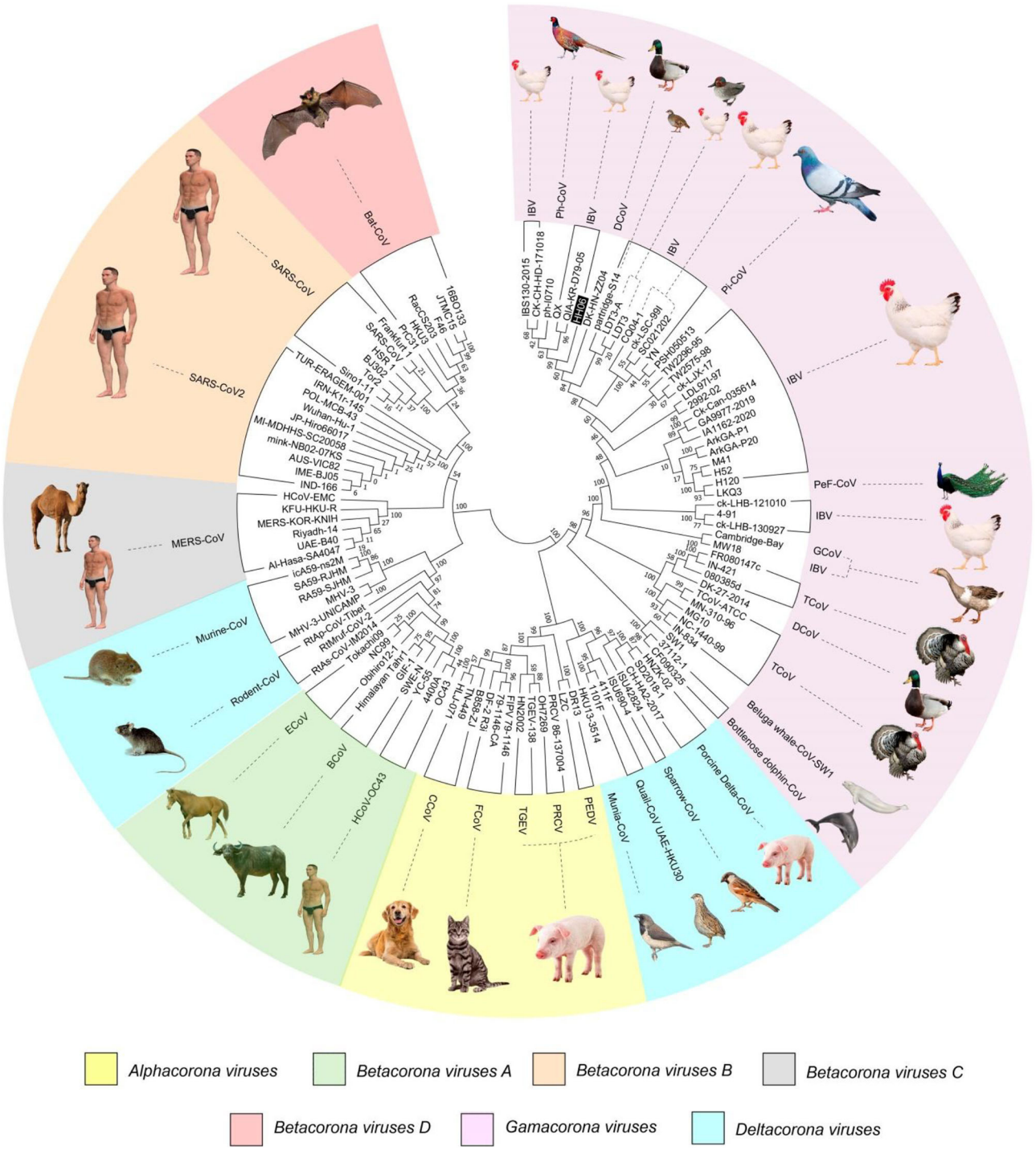


FIGURE 1 | A phylogenetic tree was constructed based on the Spike (S) gene from the four genera (Alpha, Beta, Gamma, and Delta coronaviruses) of *Coronaviridae* family comprising IBV HH06 and 110 reference CoV strains using the Maximum-Likelihood Method of MEGA7.0 version with bootstrap replicates of 1,000. The selected reference strains and their natural host are represented with a different color.

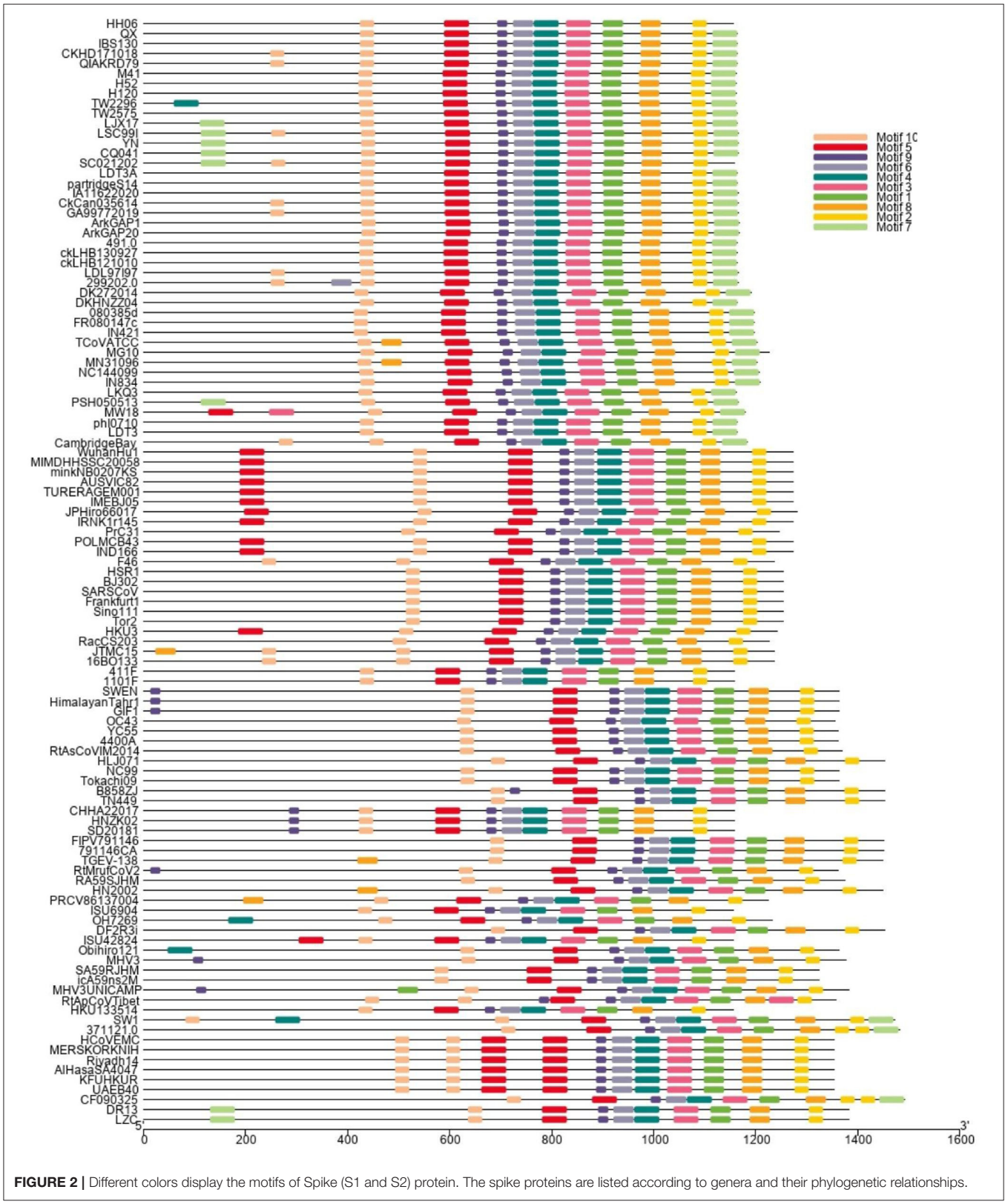


FIGURE 2 | Different colors display the motifs of Spike (S1 and S2) protein. The spike proteins are listed according to genera and their phylogenetic relationships.

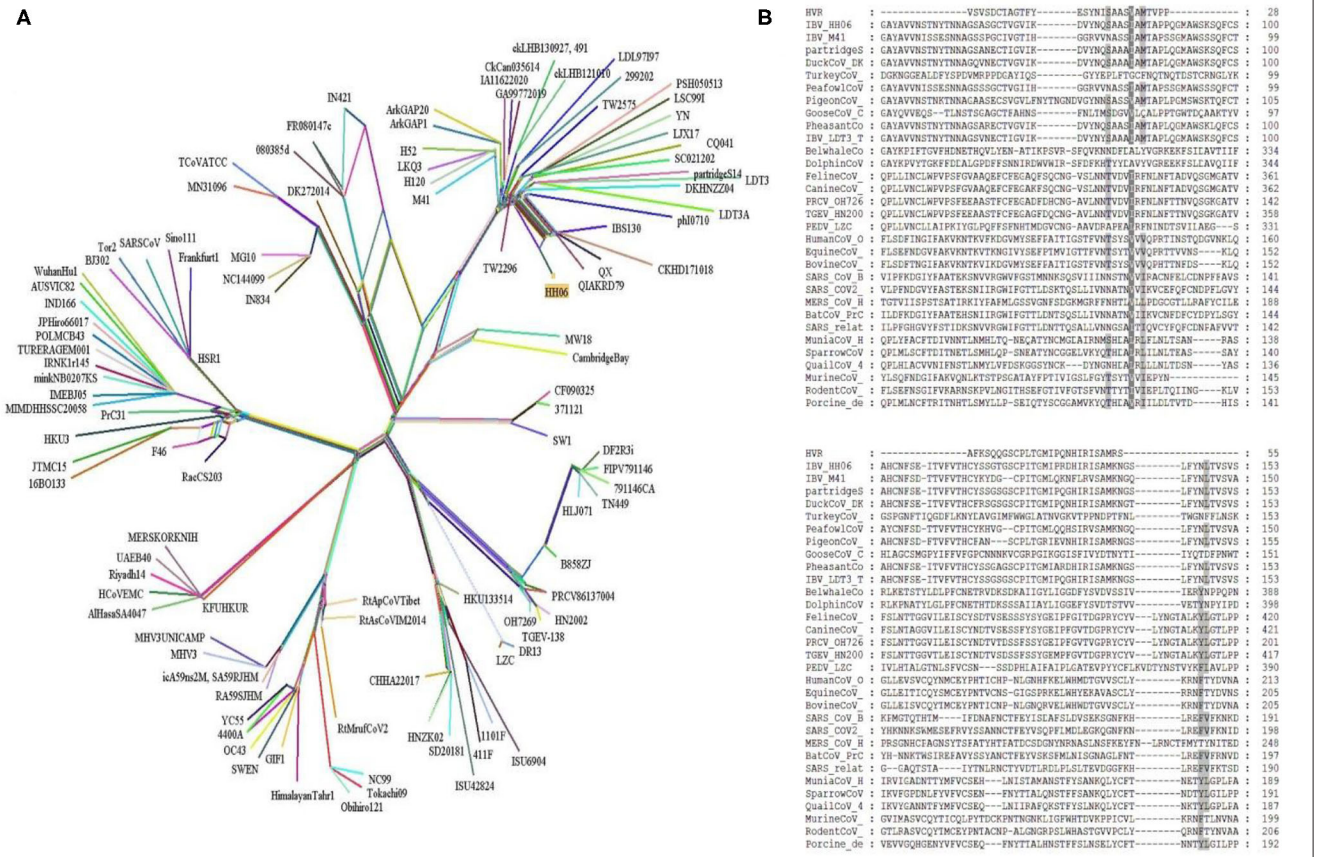


FIGURE 3 | A Network tree among 111 coronaviruses (see **Supplementary Table S1**) was constructed from protein sequence alignments of spike protein using the SplitsTree 4.14.5. **(A)** The multiple reticulate networks indicate the recombination events among different coronaviruses. The isolate (HH06) of this study is highlighted in yellow squire. The other main reference sequences of SARS-CoV2, SARS-CoV, and MERS-CoVs were also included. **(B)** Sequence alignment of S proteins from four genera of coronaviruses. Amino acid sequence alignment of 32 S proteins was done with the MEGA7. Further, the MEGA7 alignment was used in the GeneDoc program to shade the identical and similar amino acids among all alignments. Dark shad represents identical amino acids and the gray shade indicates similar amino acids.

Selection Pressure Analysis by the Position of Amino Acids

The selection pressure test was performed in the SELECTION server (http://selecton.tau.ac.il/) that uses the Mechanistic Empirical Combination (MEC) model for estimating the selection pressure at particular codons. The MEC model takes into account the variances between amino acid substitution rates. Adaptive selection pressure was found at various codons in S protein (**Figure 4A**), identified under positive selection. A total of 36.63% amino acids showed positive selection while the rest of the amino acids were under purifying process (**Figure 4A**). The result suggested that all gene pairs have evolved mainly under the influence of purifying selection.

Evolutionary Fingerprinting

Evolutionary fingerprinting of spike gene was inferred through codon model evolution based on synonymous and non-synonymous substitution rates using genetic algorithm. The probability of site-to-site distribution of synonymous and non-synonymous substitution rates exploited the ratio $\omega = \beta/\alpha$, which is estimated on the basis of the likelihood log and Akaike information criterion (AIC) using five rate classes identified neutral evolution at spike gene. The likelihood log was−129131.88712 for the S gene by using nine class rates with 49 parameters, and the AIC value was 258361.77424. The non-synonymous to synonymous ratio $\omega = \beta/\alpha$ values for all nine classes of CoVs were 0.004, 0.022, 0.117, 0.275, 0.077, 0.151, 0.262, 0.312, and 0.543, respectively (**Supplementary Table S3**). The graph showing the neutral evolution and only a few sites under the circle above diagonal has positive evolution at the S gene (**Figure 4B**).

Positive and Purifying Selection

There were 2,625 coding positions in the nucleotide sequences of the S gene and the coding positions were included 1st, 2nd, 3rd, and non-coding. Among 2,625 positions only 18 positions from codon 26 to 866 had undergone positive selection, as the values dN-dS > 1 indicated positive selection.

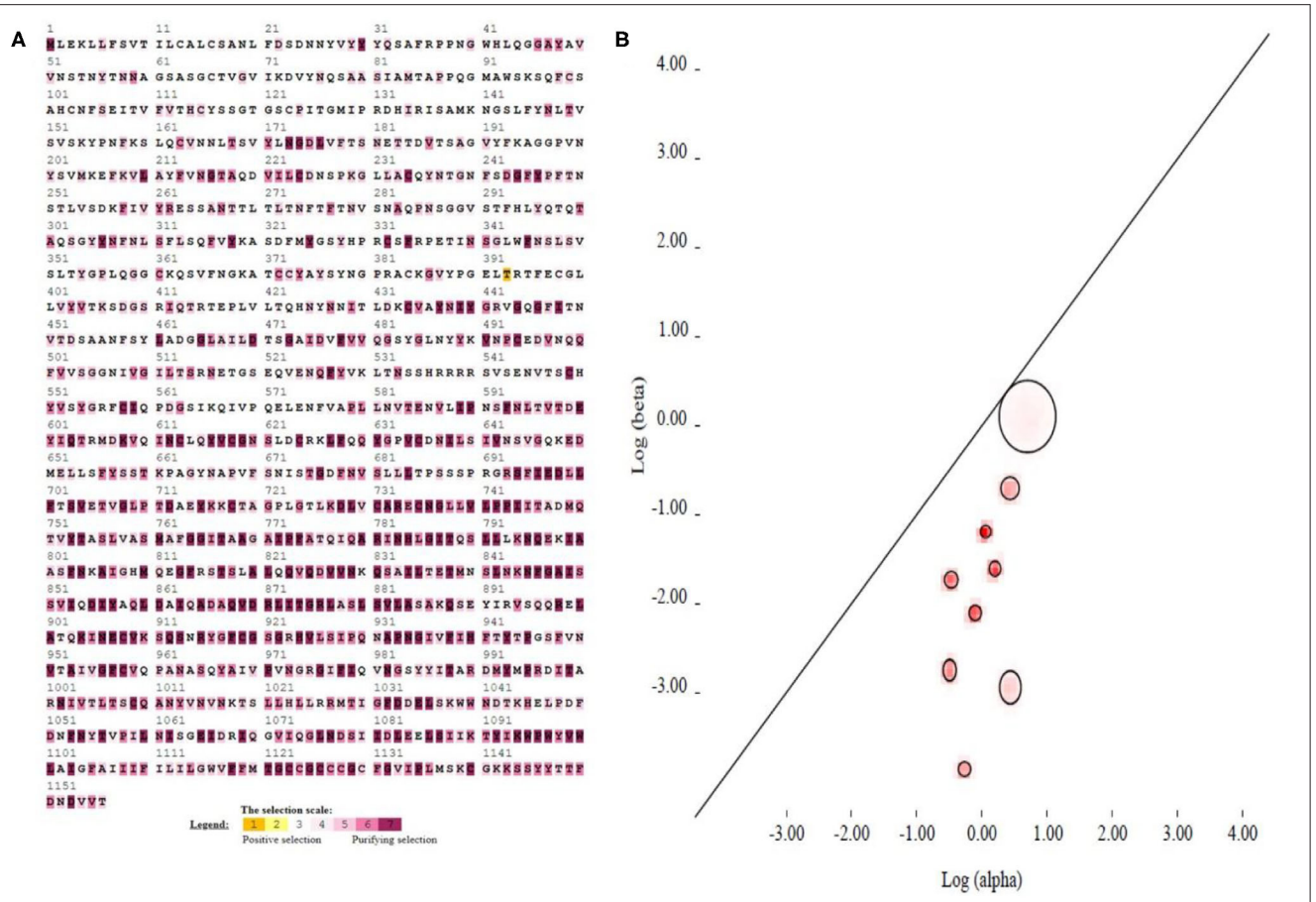


FIGURE 4 | Evolutionary mechanism of 111 coronaviruses. **(A)** The implementation of the MEC Model, which is found in the Selection Server, is used for selection pressure analysis of the viral spike protein sequences. The codons, which are highlighted with brown and yellow colors, represent positive selection, while white and gray colors are for the neutral codons. On the other hand, the codons, which are highlighted with purple, represent the purifying selection. **(B)** Evolutionary fingerprint (based on 1,000 distribution samples). The plot depicts the estimate of the distribution of synonymous and non-synonymous rates inferred from this alignment (on the log-log scale). The ellipses reflect a Gaussian-approximated variance in each individual rate estimate, and colored pixels show the density of the posterior sample of the distribution for a given rate. The diagonal line represents the idealized neutral evolution regime ($\omega = 1$), points above the line correspond to positive selection ($\omega > 1$), and points below the line–to negative selection ($\omega < 1$).

The number of synonymous and non-synonymous mutations along with synonymous and non-synonymous substitution sites was also given. The values of dN-dS were estimated for natural selection. The number of synonymous and non-synonymous codon changes, as well as the number of potential synonymous and non-synonymous changes for all sequences in pairwise comparison, was counted (**Supplementary Table S4**).

The MEME analysis detected 46 sites undergone episodic diversifying selection. All the 46 sites were detected under episodic diversifying having *p*-values <0.01 (**Table 1**). This model also estimates the synonymous (α) and non-synonymous (β) substitution rates, and the sites having values $\beta > \alpha$ were considered as significant and determined these sites under diversifying selection. The sites inferred to have experienced pervasive non-synonymous substitution throughout the evolutionary history with *p*-values < 0.01, and this site evolved with $\beta+ > \alpha$, as it is under positive selection in all analyses providing strong evidence of positive selection of S gene, whereas all other sites were conserved. The value of q was determined using the Simes' method to reduce the false discovery rate under a strict neutral null model (**Table 1**).

DISCUSSION

Genetic techniques have played an important role in understanding the genetic relatedness of different microorganisms, pathogens, and the diseases caused by the pathogens as well as their evolutionary relatedness. In our study, we presented phylogenetic relationship of an isolated IBV strain HH06 with reference to the available genome and protein sequences based on network, multiple sequence, selection pressure, and evolutionary fingerprinting analysis. Recent pandemic of COVID-19 has drawn attention to the potential zoonotic threats of the coronaviruses.

TABLE 1 | Mixed-effect model evolution (MEME) based episodic diversifying selection of S genes.

Codon	α	β^-	Pr[$\beta = \beta^-$]	β^+	Pr[$\beta = \beta^+$]	*p*-value	*q*-value
23	0.763537	0.0876926	0.791942	6.94636	0.208058	0.00599763	0.306802
32	0.0303846	0.0303846	0.845357	16.0443	0.154643	0.000109085	0.0217625
183	0	0	0.808421	1385.75	0.191579	4.86022e-05	0.0121202
292	2.25206	0.196196	0.983314	1017.51	0.0166862	0.000440945	0.0488714
298	1.42265	0.174859	0.939828	100.502	0.0601715	7.95141e-06	0.0158631
317	1.04444	0.137501	0.887011	18.7376	0.112989	0.00021124	0.0301017
325	1.13163	0.232676	0.92076	20.307	0.07924	0.00463552	0.249942
342	1.57608	0.159258	0.95851	79.7256	0.04149	0.00362301	0.24093
353	1.29453	0.352534	0.924912	28.1257	0.0750881	0.0069351	0.345888
386	0.753244	0.0399839	0.800754	8.28076	0.199246	0.00114942	0.0997
425	0.410353	0.02881	0.823356	11.616	0.176644	0.00245846	0.169125
433	0.593223	0.231261	0.93788	78.4249	0.06212	0.000153183	0.0254667
439	0.479128	0.113517	0.964491	72.4365	0.0355088	0.00124469	0.103465
552	0.99667	0.123464	0.905014	19.789	0.0949856	0.00458789	0.254245
665	0.438912	0.068045	0.916963	16.7519	0.0830372	0.000346139	0.0431592
800	0.697221	0.0904626	0.896715	28.2982	0.103285	5.45339e-05	0.0120883
801	1.19665	0.104413	0.94502	59.5819	0.0549798	3.53549e-05	0.0141066
812	0.435179	0.0407753	0.982768	209.337	0.0172323	2.63072e-05	0.0174943
945	0.311495	0.311495	0.931781	1385.76	0.0682191	0.00160388	0.114277
1010	0.666632	0.0706994	0.794554	6.50562	0.205446	4.24713e-05	0.0121043
1036	1.07674	0.149688	0.807791	11.1189	0.192209	0.00872274	0.395497
1063	1.34172	0.051046	0.772181	29.3489	0.227819	0.000724438	0.0722627
1065	0.23116	0.0690992	0.628878	4.16857	0.371122	0.00766241	0.372842
1070	1.27829	0.088108	0.950218	156.206	0.0497822	0.00159718	0.118014
1089	0.461102	0.0974815	0.955559	39.1503	0.0444409	3.42666e-05	0.0170905
1171	0.702561	0	0.919435	13.1505	0.0805655	0.00835174	0.387482
1176	1.23305	0.10632	0.992475	1292.18	0.00752452	0.000136251	0.0247109
1241	0.43565	0.142802	0.823178	7.65695	0.176822	0.00110546	0.100245
1282	2.62966	0.0713705	0.935338	41.2155	0.0646616	0.000524567	0.0550796
1285	0.384711	0.103926	0.889866	15.0998	0.110134	3.81763e-05	0.0126936
1381	0.443969	0.102546	0.956052	54.664	0.0439479	2.62908e-05	0.0262251
1446	1.01357	0.0717032	0.938069	22.6903	0.0619309	0.00362951	0.233577
1519	0.586696	0.0355755	0.970597	186.959	0.0294032	0.000404232	0.0474378
1563	0.141133	0.0203677	0.958524	192.671	0.0414764	0.000164024	0.0251713
1606	0.906054	0.0786118	0.902833	9.24085	0.0971675	0.00768589	0.36508
1658	1.94819	0.170368	0.985206	210.228	0.0147941	0.00967217	0.419478
1694	0.331159	0.0278589	0.986102	432.483	0.0138985	0.00125178	0.0998924
1727	0.453456	0.155761	0.932596	22.5383	0.0674039	0.00387794	0.241765
1819	0.480708	0.031874	0.987011	49.0222	0.0129895	0.00435959	0.255805
1840	0.515394	0.0643477	0.980805	250.885	0.0191948	0.000251536	0.0334543
1866	0.776281	0	0.887619	12.702	0.112381	0.0055095	0.289249
1939	0.232632	0	0.978126	33.5588	0.0218743	0.00875951	0.388338
1942	0.500544	0	0.962909	25.6181	0.0370913	0.00152886	0.117311
1972	0.115119	0.0412817	0.894805	4.66122	0.105195	0.00396383	0.239631
1973	0.310806	0.0140975	0.96477	21.6742	0.03523	0.000802131	0.0762024
1989	0	0	0.846684	1.51907	0.153316	0.00440476	0.251072

Genetic-recombinant network-analysis helps understanding complex genetic relationships and evolution including viruses (48). Network analysis showed clustering of IBV HH06 strain with Duck-CoV and Pheasant-CoV which indicated an evolutionary relationship between these two viruses. Previously, same results were reported for Pheasant-Cov which indicate the susceptibility of IBV to these both host species (25, 49, 50). To infer sequence homology, alignment of interrelated amino acids

sequences was done which indicated the majority of mutation sites in HVRI of the spike protein structure. Currently, CoVs are classified based on sequence comparison of structural protein, mainly spike protein gene (51). Spike protein facilitates the entry of virus to receptors present on the surface of host cells (52). These proteins have two cleaved forms called S1 and S2. The S1 plays a crucial role in the viral adsorption to the cellular glycoprotein receptor during the process of virus invasion into the host cell. The S2 play important role in fusion and forms stalk appearance to spike molecule. There is a very high conserved region in the C-terminal S2 mutation rate of S2 as it contains an extra cleavage site known as the furin S2' site, which directly influences the invasion mechanism of coronaviruses by host cells (53). The S protein is considered main determining factor in virus adaptation and tropism in the organs of host. Mutations that occurred in the S protein may influence the targeted host organ (54). We found 10 motifs in the S gene, and their annotations confirmed through the Pfam databases and assume their function in attachment and invasion/infection of the virus to the host cells.

A total of 36.63% positive selection on S protein amino acids indicated an influence of negative selection on all gene pairs. The selective removal of deleterious alleles that arise through random mutations can result in stabilizing selection. We selected Gaussian-approximate variance to evaluate the evolutionary fitness of our IBV strains based on S gene. It showed a majority of neutral evolution as compared to positive evolution on the S gene. To understand the dynamics of molecular sequence evolution, we estimated synonymous (α) and non-synonymous (β) substitution rates by MEME analysis, which indicated 46 episodic diversifying selection sites. In the present study, the phylogenic analysis indicated a broad mechanism, in which the MERS-CoVs of Beta-CoVs (C) were closely associated with SARS-CoVs-2 and SARS-CoVs of Beta-CoVs (B). In the same manner, Bat-CoVs show a close association with SARS-CoVs. Thus, similarly, in SARS-CoV-2 zoonotic transmission to humans, the viruses are considered to have originated in either bats or pangolins (31). Hence, the adaptation and recombination of SARS-CoV-2 have happened in another intermediate or reservoir host with the possibility of contact with pangolins or bats. Mutation and adaptation have determined the co-evolution of CoVs and their hosts (9). Gene recombination and mutation are both important means in producing multiple strains of CoVs, with multiple research reports on CoVs-recombination (55). There is a possibility of recombination among various strains, the recombination area and the antigenic profile of recombinant virus had important guiding significance for predicting the evolution and prevention of CoVs infection afterward (56, 57). Evolutionary conservation is critical for detecting amino acid positions and for sustaining the structural protein role. Consequently, during the selection pressure analysis, the sites that were detected as purifying and positive selection may instruct and clarify the spike protein gene function and evolution.

In the present analysis, purifying selection was observed during the selection pressure analysis. Li et al. (58) have reported purifying selection in different host species, along with frequent recombination among coronaviruses, proposes a common evolutionary mechanism that could lead to new emerging human coronaviruses. Similarly in our recombinant network tree analysis, the *Betacoronaviruses* B (SARS-CoV2 and SARS-CoV) and *Betacoronaviruses* D (Bat-CoVs) shared common clades, although branched differently, however possible to be originated from bats. Incidence of such intra-species transmission happenings in birds and mammals might be manifested by the predominant occurrence of CoVs at a large scale (59). In terms of Gamma-CoVs, thought provoking trends were observed during phylogenetic analysis. The Beluga whale- and Bottlenose dolphin-CoVs of mammals were closely associated with the branches of Avian-CoVs. Similarly, three avian-CoVs (sparrow, munia, and qual) were branched together with Porcine-CoV and this was suggestive of transmission among different species. Moreover, outcomes of the current analysis advocate that variation at the genomic and molecular level of the S gene sequence made viable CoVs adaptation to different host species. Additionally, the analysis of isolate from the present study IBV HH06 indicated that it belongs to QX-type G-19 of IBV, however, it was also closely associated with Duck-CoV and Pheasant-CoV. Abro et al. (60) carried bioinformatics of CoVs and found inter-species transmission of avian and mammalian CoVs that suggested inter-species transmission and were in agreement with our findings.

CONCLUSION

Based on the S glycoprotein, SARS-CoV-2, SARS-CoV, Bats-CoV, and MERS-CoV have close genetic relationship. IBV can potentially infect wider range of bird species beyond their natural hosts chicken e.g., ducks, teal, partridge, turkeys, and pheasants. Duck-CoV, and Pheasant-CoV grouped together with IBV strain HH06 and other QX-type viruses, hence there is an imperative need to work further on the transmission mechanism of these CoVs. Purifying selection contributed predominately to the evolutionary process in selection pressure analysis. Sufficient chances of recombination and evolution of CoVs result in the rise of novel CoVs, which have great ability of intra-species transmission and zoonosis. Thus, their continuous emerging worldwide needs to be controlled more efficiently.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

GA and YZ: conceptualization. GA and GL: data curation. GA and XS: formal analysis. GL: funding acquisition, project administration, resources, and supervision. HC and XH: investigation. GA: methodology and writing–original draft. GA and MA: software. YR and XW: validation. GA, HC, and GL: writing–review and editing. All authors agreed to the final version.

FUNDING

The National Natural Science Foundation of China under Grant (31172295 and 31272569) supported this research.

ACKNOWLEDGMENTS

China Scholarship Council was highly thanked for funding the stay of GA for conduction of his doctoral research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.794228/full#supplementary-material

Supplementary Table S1 | The information of Coronaviruses protein and gene sequences used in bioinformatics.

Supplementary Table S2 | Analysis and description of motif present in spike glycoprotein sequences of coronaviruses.

Supplementary Table S3 | Evolutionary fingerprinting analysis of Coronaviruses proteins.

Supplementary Table S4 | Maximum likelihood analysis of S protein for codon-by-codon positive selection.

Supplementary Figure S1 | Dwarfism and hemorrhage of embryos 72 h post-infection with IBV HH06 at the left, the control embryo at the right.

Supplementary Figure S2 | Confirmation of IBV Strain HH06 using IBV-N specific primers.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. China *Novel Coronavirus, Research T. A Novel Coronavirus from Patients with Pneumonia in China. N Engl J Med.* (2020) 382:727–33. doi: 10.1056/NEJMoa2001017
2. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. *Adv Virus Res.* (2011) 81:85–164. doi: 10.1016/B978-0-12-385885-6.00009-2
3. ICTV. *Coronaviridae*. Available online at: https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-senserna-viruses-2011/w/posrna_viruses/222/coronaviridae (accessed September 25, 2021).
4. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.* (2015) 1282:1–23. doi: 10.1007/978-1-4939-2438-7_1
5. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res.* (2006) 66:193–292. doi: 10.1016/S0065-3527(06)66005-3
6. Locker JK, Rose JK, Horzinek MC, Rottier PJ. Membrane assembly of the triple-spanning coronavirus M protein. Individual transmembrane domains show preferred orientation. *J Biol Chem.* (1992) 267:21911–8. doi: 10.1016/S0021-9258(19)36699-2
7. Lau SK, Lee P, Tsang AK, Yip CC, Tse H, Lee RA, et al. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol.* (2011) 85:11325–37. doi: 10.1128/JVI.05512-11
8. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* (2020) 181:271–80.e8. doi: 10.1016/j.cell.2020.02.052
9. I.Wickramasinghe NA, de Vries RP, Gröne A C.de Haan AM, Verheije MH. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J Virol.* (2011) 85:8903–12. doi: 10.1128/JVI.05112-11
10. Cavanagh D. Coronavirus avian infectious bronchitis virus. *Vet Res.* (2007) 38:281–97. doi: 10.1051/vetres:2006055
11. Hodgson T, Casais R, Dove B, Britton P, Cavanagh D. Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity. *J Virol.* (2004) 78:13804–11. doi: 10.1128/JVI.78.24.13804-13811
12. A Rao S, Shetty NP. Evolutionary selectivity of amino acid is inspired from the enhanced structural stability and flexibility of the folded protein. *Life Sci.* (2021) 281:119774. doi: 10.1016/j.lfs.2021.119774
13. Toro H, van Santen VL, Jackwood MW. Genetic diversity and selection regulates evolution of infectious bronchitis virus. *Avian Dis.* (2012) 56:449–55. doi: 10.1637/10072-020212-Review.1
14. Schalk AF, Hawn MC. An apparently new respiratory disease in baby chicks. *J Am Vet Med Assoc.* (1931) 78:413–22.
15. Beaudette F. Cultivation of the virus of infectious bronchitis. *J Am Vet Med Assoc.* (1937) 90:51–60.
16. Kendall E, Bynoe M, Tyrrell D. Virus isolations from common colds occurring in a residential school. *Br Med J.* (1962) 2:82. doi: 10.1136/bmj.2.5297.82
17. Bradburne AF, Bynoe ML, Tyrrell DA. New respiratory virus. *Br Med J.* (1967) 3:752. doi: 10.1136/bmj.3.5568.752
18. Varia M, Wilson S, Sarwal S. Hospital Outbreak Investigation Team. Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada. *CMAJ.* (2003) 169:285–92.
19. De Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, et al. Commentary: Middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. *J Virol.* (2013) 87:7790–2. doi: 10.1128/JVI.01244-13
20. Oude Munnink BB, Worp N, Nieuwenhuijse DF, Sikkema RS, Haagmans B, Fouchier RA, et al. The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology. *Nat Med.* (2021) 27:1–7. doi: 10.1038/s41591-021-01472-w
21. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* (2021) 19:141–54. doi: 10.1038/s41579-020-00459-7
22. Banerjee A. Virus hunters: discovering the evolutionary origins of SARS-CoV-2. *Cell Host Microbe.* (2021) 29:1031–3. doi: 10.1016/j.chom.2021.06.012
23. Cavanagh D, Naqi SA. Infectious bronchitis. In: Calnek BW, Barnes HJ, Beard CW, McDougald LR, Saif YM, editors, *Disease of Poultry*. Hoboken, NJ: Iowa State University Press (1997). p. 511–26
24. De Wit J. Detection of infectious bronchitis virus. *Avian Pathol.* (2000) 29:71–93. doi: 10.1080/03079450094108
25. Cavanagh D, Mawditt K D.d.Welchman B, Britton P, Gough R. Coronaviruses from pheasants (Phasianus colchicus) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys. *Avian Pathol.* (2002) 31:81–93. doi: 10.1080/03079450120106651
26. Wu X, Pan S, Zhou W, Wu Y, Huang Y, Wu B. The isolation and identification of infectious bronchitis virus PTFY strain in muscovy ducks. *Bing Du Xue Bao.* (2016) 32:203–9. doi: 10.13242/j.cnki.bingduxuebao.002908
27. Liu S, Chen J, Chen J, Kong X, Shao Y, Han Z, et al. Isolation of avian infectious bronchitis coronavirus from domestic peafowl (Pavo cristatus) and teal (Anas). *J Gen Virol.* (2005) 86:719–25. doi: 10.1099/vir.0.80546-0

28. Qian DH, Zhu GJ, Wu LZ, Hua GIsolation X, and characterization of a coronavirus from pigeons with pancreatitis. *Am J Vet Res.* (2006) 67:1575–9. doi: 10.2460/ajvr.67.9.1575
29. Papineau A, Berhane Y, Wylie TN, Wylie KM, Sharpe S, et al. Genome organization of canada goose coronavirus, a novel species identified in a mass die-off of Canada Geese. *Sci Rep.* (2019) 9:5954. doi: 10.1038/s41598-019-42355-y
30. Cook JK, Jackwood M, Jones RC. The long view: 40 years of infectious bronchitis research. *Avian Pathol.* (2012) 41:239–50. doi: 10.1080/03079457.2012.680432
31. Goraichuk IV, Arefiev V, Stegniy BT, Gerilovych AZoonotic P, and Reverse Zoonotic Transmissibility of SARS-CoV-2. *Virus Res.* (2021) 302:198473. doi: 10.1016/j.virusres.2021.198473
32. Alluwaimi AM, Alshubaith IH, Al-Ali AM, Abohelaika S. The coronaviruses of animals and birds: their zoonosis, vaccines, and models for SARS-CoV and SARS-CoV2. *Front Vet Sci.* (2020) 7:655. doi: 10.3389/fvets.2020.582287
33. Zhang G, Li B, Yoo D, Qin T, Zhang X, Jia Y, et al. Animal coronaviruses and SARS-CoV-2. *Transbound Emerg Dis.* (2021) 68:10971110. doi: 10.1111/tbed.13791
34. Rabaan AA, Al-Ahmed SH, Sah R, Alqumber MA, Haque S, Patel SK, et al. MERS-CoV: epidemiology, molecular dynamics, therapeutics, future challenges. *Ann Clin Microbiol Antimicrob.* (2021) 20:1–14. doi: 10.1186/s12941-020-00414-7
35. Ren X, Yin J, Ma D, Li Characterization G, and membrane gene-based phylogenetic analysis of avian infectious bronchitis virus Chinese strain HH06. *Virus Genes.* (2009) 38:39–45. doi: 10.1007/s11262-008-0280-7
36. Hewson K, Noormohammadi AH, Devlin JM, Mardani K, Ignjatovic J. Rapid detection and non-subjective characterisation of infectious bronchitis virus isolates using high-resolution melt curve analysis and a mathematical model. *Arch. Virol.* (2009) 154:649–60. doi: 10.1007/s00705-009-0357-1
37. Reed LJ, Muench H A. simple method of estimating fifty per cent endpoints12. *Am J Epidemiol.* (1938) 27:493–7. doi: 10.1093/oxfordjournals.aje.a118408
38. Sun X, Li L, Pan L, Wang Z, Chen H, Shao C, et al. Infectious bronchitis virus: identification of Gallus gallus APN high-affinity ligands with antiviral effects. *Antiviral Res.* (2021) 186:104998. doi: 10.1016/j.antiviral.2020.104998
39. Chen H, Muhammad I, Zhang Y, Ren Y, Zhang R, Huang X, et al. Antiviral Activity Against Infectious Bronchitis Virus and Bioactive Components of *Hypericum perforatum* L. *Front Pharmacol.* (2019) 10:1272. doi: 10.3389/fphar.2019.01272
40. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* (2016) 33:1870–4. doi: 10.1093/molbev/msw054
41. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* (2006) 23:254–67. doi: 10.1093/molbev/msj030
42. J-Yang R, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci U S A.* (2012) 109:E831–40. doi: 10.1073/pnas.1117408109
43. Kosakovsky Pond SL, Scheffler K, Gravenor MB, Poon AF, Frost SD. Evolutionary fingerprinting of genes. *Mol Biol Evol.* (2010) 27:520–36. doi: 10.1093/molbev/msp260
44. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* (2005) 22:1208–22. doi: 10.1093/molbev/msi105
45. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* (1980) 16:111–20. doi: 10.1007/BF01731581
46. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* (2013) 30:1196–205. doi: 10.1093/molbev/mst030
47. Newman M. *Networks*. Oxford: Oxford University Press (2018).
48. Wang Y, Zeng J, Zhang C, Chen C, Qiu Z, Pang J, et al. New framework for recombination and adaptive evolution analysis with application to the novel coronavirus SARS-CoV-2. *Brief Bioinform.* (2021) 22:bbab107. doi: 10.1093/bib/bbab107
49. Torres CA, Listorti V, Lupini C, Franzo G, Drigo M, Catelli E, et al. Gamma and Deltacoronaviruses in quail and pheasants from Northern Italy. *Poult Sci.* (2017) 96:717–22. doi: 10.3382/ps/pew332
50. Chen GQ, Zhuang QY, Wang KC, Liu S, Shao JZ, Jiang WM, et al. Identification and survey of a novel avian coronavirus in ducks. *PLoS ONE.* (2013) 8:e72918. doi: 10.1371/journal.pone.0072918
51. Gonzalez JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol.* (2003) 148:2207–35. doi: 10.1007/s00705-003-0162-1
52. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature.* (2020) 583:437–40. doi: 10.1038/s41586-020-2355-0
53. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* (2020) 26:450–2. doi: 10.1038/s41591-020-0820-9
54. Li W. Delving deep into the structural aspects of a furin cleavage site inserted into the spike protein of SARS-CoV-2: a structural biophysical perspective. *Biophys Chem.* (2020) 264:106420. doi: 10.1016/j.bpc.2020.106420
55. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. *Int J Biol Sci.* (2020) 16:1686. doi: 10.7150/ijbs.45472
56. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol.* (2010) 84:3134–46. doi: 10.1128/JVI.01394-09
57. Shafique L, Ihsan A, Liu Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens.* (2020) 9:240. doi: 10.3390/pathogens9030240
58. Li X, Giorgi EE, Marichannegowda MH, Foley B, Xiao C, Kong XP, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv.* (2020) 6:eabb9153. doi: 10.1126/sciadv.abb9153
59. Hulswit RJ, de Haan CA, Bosch BJ. Coronavirus spike protein and tropism changes. *Adv Virus Res.* (2016) 96:29–57. doi: 10.1016/bs.aivir.2016.08.004
60. Abro SH, Ullman K, Belak S, Baule C. Bioinformatics and evolutionary insight on the spike glycoprotein gene of QX-like and Massachusetts strains of infectious bronchitis virus. *Virol J.* (2012) 9:211. doi: 10.1186/1743-422X-9-211

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fvets.2022.794934

Isolation and Genomic Characterization of Avian Reovirus From Wild Birds in South Korea

Sang-Won Kim†, Yu-Ri Choi†, Jong-Yeol Park, Bai Wei, Ke Shang, Jun-Feng Zhang, Hyung-Kwan Jang, Se-Yeoun Cha and Min Kang**

Department of Veterinary Infectious Diseases and Avian Diseases, Center for Poultry Diseases Control, College of Veterinary Medicine, Jeonbuk National University, Iksan, South Korea

OPEN ACCESS

Edited by:
Christina Leyson, United States Department of Agriculture, United States

Reviewed by:
Rodrigo A. Gallardo, University of California, Davis, United States
Luiz Claudio Miletti, University of the State of Santa Catarina, Brazil

***Correspondence:**
Min Kang
vet.minkang@gmail.com
Se-Yeoun Cha
seyeouncha@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *14 October 2021*
Accepted: *06 January 2022*
Published: *28 January 2022*

Citation:
Kim S-W, Choi Y-R, Park J-Y, Wei B, Shang K, Zhang J-F, Jang H-K, Cha S-Y and Kang M (2022) Isolation and Genomic Characterization of Avian Reovirus From Wild Birds in South Korea. Front. Vet. Sci. 9:794934. doi: 10.3389/fvets.2022.794934

Avian reoviruses (ARVs) cause severe arthritis, tenosynovitis, pericarditis, and depressed growth in chickens, and these conditions have become increasingly frequent in recent years. Studies on the role of wild birds in the epidemiology of ARVs are insufficient. This study provides information about currently circulating ARVs in wild birds by gene detection using diagnostic RT-PCR, virus isolation, and genomic characterization. In this study, we isolated and identified 10 ARV isolates from 7,390 wild birds' fecal samples, including migratory bird species (bean goose, Eurasian teal, Indian spot-billed duck, and mallard duck) from 2015 to 2019 in South Korea. On comparing the amino acid sequences of the σC-encoding gene, most isolates, except A18-13, shared higher sequence similarity with the commercial vaccine isolate S1133 and Chinese isolates. However, the A18-13 isolate is similar to live attenuated vaccine av-S1133 and vaccine break isolates (SD09-1, LN09-1, and GX110116). For the p10- and p17-encoding genes, all isolates have identical fusion associated small transmembrane (FAST) protein and nuclear localization signal (SNL) motif to chicken-origin ARVs. Phylogenetic analysis of the amino acid sequences of the σC-encoding gene revealed that all isolates were belonged to genotypic cluster I. For the p10- and p17-encoding genes, the nucleotide sequences of all isolates indicated close relationship with commercial vaccine isolate S1133 and Chinese isolates. For the σNS-encoding gene, the nucleotide sequences of all isolates indicated close relationship with the Californian chicken-origin isolate K1600657 and belonged to chicken-origin ARV cluster. Our data indicates that wild birds ARVs were derived from the chicken farms. This finding suggests that wild birds serve as natural carriers of such viruses for domestic poultry.

Keywords: avian reoviruses, wild bird, chicken, South Korea, σC-encoding gene, σNS-encoding genes

INTRODUCTION

Avian reoviruses (ARVs), classified under the Orthoreovirus genus of *Reoviridae* family, have a non-enveloped icosahedral double capsid with 10 double-stranded RNA (dsRNA) genome segments which are further divided into three size classes based on their electrophoretic mobility: large (L1, L2, L3), medium (M1, M2, M3), and small (S1, S2, S3, S4) (1). ARV genome encodes eight structural (λA, λB, λC, μA, μB, σA, σB, and σC) and four non-structural viral proteins (μNS, σNS, p10, and p17) (2). Among the S-class segments of ARV, the segment S1 contains three open reading

frames (ORFs) that are translated into p10, p17, and σC proteins (2). The p10 protein induces cell fusion (3), while the p17 protein induces cellular protein translation shut-off and cell cycle arrest (4, 5). The σC protein encoded by the third ORF of the segment S1 is the cell attachment protein and elicits the production of ARV-specific neutralizing antibodies (6, 7). The σC sequence is used as a genetic marker to characterize and classify ARV isolates into different six or seven genotypic clusters (8, 9). The σNS protein encoded by the segment S4, has been reported for its single-stranded RNA (ssRNA) binding activity, and it is divided into diverse lineages that conserved by host origin of ARVs (10, 11); as well it has been used for diagnostic analysis (10, 12).

ARVs are widespread in nature and affect various commercial and wild avian species. ARVs have been isolated from poultry, such as chickens, ducks, turkeys, ostriches, and wild birds (13–17). Major damages from ARV infections are observed in young chickens, and ARV infections in broiler chickens are associated with up to 10% mortality and up to 40% morbidity. ARV infections in broilers have been associated with various clinical diseases, such as viral arthritis/tenosynovitis, malabsorption syndrome (MAS), runting–stunting syndrome (RSS), respiratory diseases, hepatitis, myocarditis, neurological signs (incoordination, tremors, and twisted necks), and immunosuppression (18, 19). ARV infections are responsible for significant economic losses in the poultry industry (20). There have been ARV infection outbreaks in the USA, Canada, Brazil, Europe, Israel, China, and Korea (8, 21–26).

ARVs have also been identified in wild birds, such as crows, magpies, partridges, black-capped chickadees, brown-eared bulbuls, psittacine bird species, and mallard ducks (27–35). ARV infections in wild birds have been associated with mortality and/or various clinical diseases, such as hepatitis, enteritidis, and neurological signs (29, 31, 33, 34, 36, 37). A previous study on the seroprevalence of ARVs in wild birds reported 25% seroprevalence in bean goose (*Anser fabalis*) and 34% seroprevalence in white-fronted goose *(Anser albifrons frontalis*) in Germany in 1998 (38). In addition, antigenic prevalence in wild birds representing 32 species was reportedly ∼30% in Poland from 2014 to 2016 (12). ARVs have been isolated from hooded crows in Finland, American crows in the USA, brown-eared bulbuls in Japan, psittacine bird species in Germany, and mallard ducks in China (27, 29, 30, 34, 35). Various hypotheses have been proposed about the role of wild birds in the dissemination and maintenance of ARVs and pathogenic potential of ARVs in wild birds that usually commingle around poultry farms. In previous studies, ARVs isolated from wild birds were somewhat distant from the those from chicken farms (29, 30, 33–35). However, recently research showed that the σA-encoding gene isolated from a healthy ostrich at a domestic farm in Japan showed great similarity to the chicken-origin ARVs, and the σC-encoding gene isolated from magpies was found to be genetically similar to the chicken-origin ARVs (13, 31). This genetic association between isolates from wild birds and poultry can be explained by the potential of wild birds to act as a carrier for the transmission of the ARVs (31). However, studies on the role of wild birds in the epidemiology of ARVs are insufficient.

This study provides information about currently circulating ARVs in wild birds by gene detection using diagnostic RT-PCR, virus isolation, and genomic characterization of S1 and S4 gene. Furthermore, it investigates the possibility of wild birds being potential sources of infection in poultry.

MATERIALS AND METHODS

Sample

A total of 7,390 fresh fecal samples were obtained from major migratory bird habitats, including near commercial/domestic chicken premises in South Korea by an avian influenza virus national monitoring program (2015–2019). All live migratory birds were trapped and collected feces by the South Korea Animal and Plant Quarantine Agency (QIA) and the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (39). Ten fecal samples obtained from pure or separately standing flocks of the same species were pooled into one test sample. Finally, the 739 resultant test samples were 10-fold diluted in sterile phosphate-buffered saline [PBS, pH7.4; supplemented with 100X antibiotic–antimycotic (Gibco, New York, USA)], thoroughly mixed by vortexing, and centrifugated at 600 x *g* for 10 min at 4°C. The supernatant was subsequently filtered (0.45-μm pore size; Minisart® NML, Sartorious, Germany). The filtered supernatant was then conserved in aliquots at −70°C for virus isolation and viral RNA extraction.

RNA Extraction and RT-PCR

Viral RNA was extracted from the clarified fecal samples using the MagMAX™ - 96 AI/ND Viral RNA isolation kit (Thermo Fisher Scientific, Vilnius, Lithuania) with KingFisher Duo Prime Purification system (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's protocol. Viral cDNA was generated from RNA samples using GoScript™ reverse transcriptase (Promega, Madison, WI USA) with random primers (9-mers; TaKaRa Bio. Inc., Otsu, Shiga, Japan). In the reverse transcription (RT) reaction, 8 μL of extracted RNA and 2 μL of dimethyl sulfoxide (DMSO, Tedia, USA) were heated at 100°C for 5 min and then placed in an ice bath for 5 min. Then, the following were added to this reaction mixture: 8 μL of GoScript™ 5X RT reaction buffer (Promega, Madison, WI USA), 10 μL of 2.0 mM of each dNTP (SolGent, Daejeon, Korea), 4 μL of $MgCl_2$ (Promega, Charbonnie're, France), 1 μL of 20 units Recombinant RNasin® Ribonuclease Inhibitor (Promega, Madison, WI, USA), 1 μL of GoScript™ reverse transcriptase, 1 μL of 50 pmol random primer, and 4 μL of diethylpyrocarbonate-treated water (DEPC water; Biosesang, Seoul, Korea); a final volume of 39 μL was obtained. The RT reaction mixture was incubated in this sequence: 25°C for 5 min, 42°C for 60 min, and 70°C for 15 min to inactivate the enzyme. The cDNA amplification was performed in 50 μL volume, containing 10X Taq buffer (Solgent, Daejeon, Korea), 5U Taq DNA polymerase (Solgent, Daejeon, Korea), 10 pmol of each primer (**Table 1**). For the detection of ARVs in wild bird feces, partial S1 gene was amplified. Thermal cycling protocols were as follows: Initial denaturation at 94°C for 5 min, 35 cycles (denaturation at 94°C for 1 min, annealing at 50°C for 1 min, extension at 72°C for 1 min) and one final extension

TABLE 1 | RT-PCR primer sequences and expected PCR products.

Designation	Sequence (5'-3')	Gene	Location	PCR products (bp)	Reference
MK87	5'-GGTGCGACTGCTGTATTTGGTAAC-3'	Partial S1	55–78	532	(40)
MK88	5'-AATGGAACGATAGCGTGTGGG-3'	Partial S1	568–588		
ARV-S1-1632-F	5'-CAATCCCTTGTTCGTCGATGYT-3'	Full S1	8–28	1,632	(2)
ARV-S1-1632-R	5'-AATAACCAATCCCMGTACGGCG-3'	Full S1	1,618–1,639		
SnS-F	5'-CTTTTTGAGTCCTTGTGCAGCCAT-3'	Full S4	2–26	1,185	(41)
SnS-R	5'-TAAGAGTCCAAGTCGCGGCAGAGG-3'	Full S4	1,163–1,186		

at 72°C for 10 min (40). For the sequence analysis, viral RNA was extracted from virus stocks using the Viral Gene-SpinTM Viral DNA/RNA Extraction kit (iNtRON, Daejeon, Korea). Viral cDNA was obtained from RNA samples by the above-mentioned method. For the amplification of the full S1 gene, thermal cycling protocols were as follows: Initial denaturation at 94°C for 5 min, 40 cycles (denaturation at 94°C for 1 min, annealing at 60°C for 1 min, extension at 72°C for 100 s) and one final extension at 72°C for 10 min (2). For the amplification of the full S4 gene, thermal cycling protocols were as follows: Initial denaturation at 94°C for 5 min, 35 cycles (denaturation at 94°C for 30 s, annealing at 58°C for 30 s, extension at 72°C for 90 s) and one final extension at 72°C for 10 min (41).

DNA Barcoding for Species Identification of Fecal Samples

The host species of the PCR-positive fecal samples were identified using the DNA barcoding technique as previously described (42). Briefly, DNA was extracted from fresh fecal samples using an Accuprep® Stool DNA Extraction kit (Bioneer, Daejeon, Korea) according to the manufacturer's protocol. The primers Aves-F (5′-GCATGAGCAGGAATAGTTGG-3′) and Aves-R (5′-AAGATGTAGACTTCTGGGTG-3′) were used to amplify the mitochondrial cytochrome oxidase gene subunit I present in host feces (43). PCR products were sequenced and identified using the information available at the Barcode of Life Data Systems website (Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada; http://www.barcodinglife.org/views/login.php) (44).

Virus Isolation

The PCR-positive samples were inoculated into the yolk sac of 6-day-old specific pathogen-free (SPF) embryos (SPAFAS Poultry Company, Jinan, China). Embryonated eggs were candled daily for 5 days. Chorioallantoic membranes (CAMs) of the infected embryos were collected by three cycles of freezing and thawing; then, 1.5 mL PBS was added, followed by centrifugation at 6,000 x *g* for 10 min. For cell culture passage, chicken embryo liver (CEL) cells were prepared from 14-day-old SPF chicken embryos as per the standard protocol and then dispensed into a 6-well-cell culture plate (SPL life sciences, Pocheon, Korea). The medium for CEL cell culture was Eagle's minimum essential medium supplemented with 8% fetal bovine serum (FBS) and 1% addition of 100X antibiotic–antimycotic. When the cell monolayers were ∼80% confluent, the medium was aspirated. Cell monolayers were infected with 0.2 mL of 10-fold diluted CAM fluids from the first passage of chicken embryos and incubated at 37°C for 60 min. Then, a maintenance medium containing 4% FBS was added. The cultures were incubated at 37°C under 5% CO_2 and were observed daily under a microscope to check for a cytopathic effect (CPE). Once 70–80% CPE was developed, the cultures were subjected to three cycles of freezing and thawing and then clarified with low centrifugation at 600 x *g* for 20 min (41).

Cloning and Sequencing

PCR products of full-length S1 and S4 genes were cloned and sequenced. The PCR products of the expected lengths were purified with the QIAquick gel extraction kit (Qiagen, Chatsworth, CA) and then cloned into the pGEM-T easy vector (Promega, Madison, WI, USA) according to the manufacturer's instructions, and nucleotide sequences were determined using an ABI 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA) by SolGent (Daegeon, Korea) (45).

Sequence Analysis and Phylogenetic Analysis

Based on the nucleotide and deduced amino acid sequences of the σC- and σNS-encoding genes, phylogenetic analysis was carried out. Sequence alignments and pairwise sequence comparisons were performed using the GENETYX software (Genetyx Corp., Tokyo, Japan) (46). Phylogenetic analysis was performed using the maximum likelihood method of MEGA-X software package (version 10.0.5) with bootstrap values calculated from 1,000 replicates (47). Recombination detection was accomplished by using Bootscan analysis within the Simplot program version 3.5.1, using the neighbor-joining method, with a Kimura 2-parameter applied and 100 replicates (48).

RESULTS

Virus Isolation and Propagation

A total of 14 PCR-positive samples from 739 samples (1.9%) were identified. Ten ARVs were isolated by using CEL cell from the 14 PCR-positive samples, and all isolates were confirmed by diagnostic RT-PCR as described above in **Table 1**. Five ARVs were detected from bean goose (*Anser fabalis*), one from mallard duck (*Anas platyrhtchos*), one from Eurasian teal (*Anas crecca*), one from oriental turtle dove (*Streptopelia orientalis*), and one from Indian spot-billed duck (*Anas poecilorhyncha*); the source could not be identified for one ARV (**Table 2**). CPE was

TABLE 2 | ARV obtained in the study, origin of sample, year of isolation, host species, sample type, and GenBank accession number.

No.	Isolate	Province	Year	Host	Sample type	GenBank accession number	
						S1	S4
1	A15-19	Jeonnam	2015	Not detected	Fecal	MW357863	MW357853
2	A15-48	Jeonnam	2015	Bean goose (*Anser fabalis*)	Fecal	MW357864	MW357854
3	A15-71	Jeonbuk	2015	Bean goose (*Anser fabalis*)	Fecal	MW357865	MW357855
4	A15-108	Jeonbuk	2015	Bean goose (*Anser fabalis*)	Fecal	MW357866	MW357856
5	A15-113	Jeonnam	2015	Bean goose (*Anser fabalis*)	Fecal	MW357867	MW357857
6	A15-157	Jeonbuk	2015	Oriental turtle dove (*Streptopelia orientalis*)	Fecal	MW357868	MW357858
7	A18-13	Gyeongnam	2018	Eurasian teal (*Anas crecca*)	Fecal	MW357869	MW357859
8	A18-19	Jeonnam	2018	Indian spot-billed duck (*Anas poecilorhyncha*)	Fecal	MW357870	MW357860
9	A18-205	Jeonnam	2018	Mallard (*Anas platyrhynchos*)	Fecal	MW357871	MW357861
10	A19-106	Gyeongnam	2019	Bean goose (*Anser fabalis*)	Fecal	MW357872	MW357862

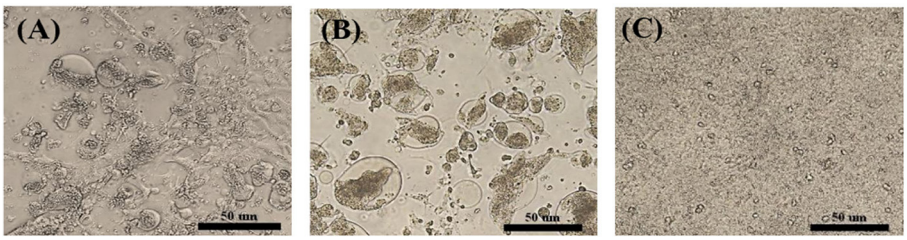


FIGURE 1 | Cytopathic effect (CPE) in ARV-infected CEL cells. **(A)** S1133/Chicken/USA/1971 infected CEL cells forming a syncytial cluster at 1 day post infection. **(B)** Representative wild bird isolate (A18-13/Wild bird/Korea/2018) infected CEL cells forming a syncytial cluster at 1 day post infection. **(C)** Non-infected CEL cell control.

observed in CEL cell infected with the 10 isolates. The CPE of isolates manifested as detachment of the monolayer or syncytium formation (**Figure 1**).

Sequence Analysis

The full-length S1 and S4 genes of all ARV isolates were sequenced, and the obtained sequences were submitted to the GenBank database under the accession numbers showed in **Table 2**. The σC- and σNS-encoding genes of all isolates were analyzed and compared with reference isolates (**Supplementary Table 1**). The alignment of nucleotide sequences of the σC-encoding gene of 10 isolates showed high similarity with commercial vaccine isolates (av-S1133, S1133, 2408, and 1733) (98.4–99.9%) and Chinese chicken-origin isolates (GuangxiR1, SD09-1, LN09-1, and GX110116) (98.6–99.6%) in nucleotide sequence. Conversely, the 10 isolates showed low similarity with previously identified Korean chicken-origin isolates (SNU0044, SNU0046, K738-14, and iREO0309) (50.2–74.5%) and Tvärminne avian virus (TVAV)-like wild bird isolates (Pycno-1, Corvus corone cornix 2002, and Tvärminne avian virus) (48.5–53.9%) (**Supplementary Table 2**). An alignment of nucleotide sequences of the σNS-encoding gene of the 10 isolates showed high similarity with the Californian chicken-origin isolate K1600657 (92.7%—92.9%) and low similarity with the TVAV-like wild bird isolates (Pycno-1, SD-12, Tvärminne avian virus, 71-03, Chickadee, and SRK) (58.2%—77.7%) (**Supplementary Table 3**).

We compared the amino acid sequences of the 10 isolates with those of 15 reference isolates. The results were found to be consistent with the results of pairwise comparisons of the σC-encoding gene (**Supplementary Table 4**). Compared at six conservative amino acid mutations (24, 71, 106, 113, 134, and 135), most newly isolated viruses, except A18-13 isolate, shared higher sequence similarity with commercial vaccine isolate S1133 and Chinese chicken-origin isolates (GuangxiR1 and GuangxiR2). Whereas 10 isolates shared lower similarity with Korean chicken-origin isolates (SNU0044, SNU0046, iREO0309, and K738-14). Nucleotide changes in the σC ORF unique to A18-13 isolate and the live attenuated vaccine av-S1133 and vaccine break isolates (SD09-1, LN09-1, and GX110116) were identified. A18-13 isolate, live attenuated vaccine, and vaccine break isolates carry four nonsynonymous nucleotide substitutions (A^{71}, G^{317}, T^{338}, and C^{405}), whereas commercial vaccine isolate S1133 have nucleotide substitutions at C^{71}, C^{317}, C^{338}, and A^{405} in the σC-encoding gene (**Supplementary Table 4**). These four non-synonymous nucleotide substitutions lead to deduced amino acid substitutions (T24N, T106R, T113I, and V135I) in the σC-encoding gene (**Table 3**). Nucleotide changes in the σNS ORF unique to all isolates and the Californian chicken-origin isolate K1600657 were identified. All isolates and K1600657 carry 10 non-synonymous nucleotide substitutions (G^{215}, T^{361}, A^{455}, A^{472}, G^{478}, G^{484}, T^{652}, A^{760}, G^{817}, and C^{1049}), whereas commercial vaccine isolates, including live attenuated vaccine have nucleotide substitutions at A^{215}, C^{361}, G^{455}, G^{472}, A^{478},

TABLE 3 | Deduced amino acid substitutions in the σC- and σNS-encoding genes of ARV isolates from wild bird.

Group	ORFs	σC (amino aicd)						σNS (amino acid)									
(Isolates)	Position	24	71	106	113	134	135	72	121	152	158	160	162	218	254	273	350
Commercial vaccine	S1133	T	T	T	T	D	V	D	P	R	A	S	P	A	L	T	M
	A15-19	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A15-48	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A15-71	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A15-108	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A15-113	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
Present study	A15-157	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A18-13	**N**	**T**	**R**	**I**	**D**	**I**	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A18-19	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A18-205	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	A19-106	T	T	T	T	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
Attenuated vaccine	av-S1133	**N**	**T**	**R**	**I**	**D**	**I**	D	P	R	A	S	P	A	L	T	M
	SD09-1	N	T	R	I	D	I	D	P	R	A	S	P	A	L	T	M
	LN09-1	N	T	R	I	D	I	D	P	R	A	S	P	A	L	T	M
China	GX110116	N	T	R	I	D	I	D	P	R	A	S	P	A	L	T	M
	GuangxiR1	T	I	T	T	D	V	D	P	R	A	S	P	A	L	T	M
	GuangxiR2	T	I	T	T	D	V	D	P	R	A	S	P	A	L	T	M
California	K1600657	T	A	R	S	D	V	**G**	**S**	**H**	**T**	**D**	**A**	**S**	**T**	**V**	**T**
	SNU0044	-	A	T	M	D	V	-	-	-	-	-	-	-	-	-	-
	SNU0046	-	A	T	M	D	V	-	-	-	-	-	-	-	-	-	-
Korea	iReo0309	-	S	N	S	D	V	-	-	-	-	-	-	-	-	-	-
	ADL112770-ARV	-	-	-	-	-	-	D	S	H	T	N	A	S	T	V	-
	ADL112782-ARV	-	-	-	-	-	-	D	S	H	T	N	A	S	T	V	-
	ADL121187-ARV	-	-	-	-	-	-	D	S	H	T	N	A	S	T	V	-
	K738-14	T	S	M	L	S	V	D	S	H	T	N	A	S	T	V	T

Deduced amino acid numbering corresponds to that of commercial vaccine isolate S1133, -: data is not available, specific deduced amino acid substitutions are showed in a bold.

C^{484}, G^{652}, C^{760}, A^{817}, and T^{1049} in the σNS-encoding gene (**Supplementary Table 4**). These 10 non-synonymous nucleotide substitutions lead to deduced amino acid substitutions (D72G, P121S, R152H, A158T, S160D, P162A, A218S, L254T, T273V, and M350T) in the σNS-encoding gene (**Table 3**). For the p10 protein, all isolates were not observed for deduced amino acid substitutions in putative transmembrane domains. For the p17 protein, all isolates were not observed for deduced amino acid substitutions in nuclear localization signal (NLS) motif (**Supplementary Table 5**).

Phylogenetic Analysis

Phylogenetic analysis of the σC- and σNS-encoding genes was demonstrated. For the σC-encoding gene, nucleotide sequences of the 10 isolates showed that they were closely related to each other and clustered with Chinese chicken-origin isolates and commercial vaccine isolates. Conversely, previously identified TVAV-like wild bird isolates (Pycno-1, SD-12, Tvärminne avian virus, and Corvus corone cornix 2002) were distinct from all isolates in this study (**Figure 2**). To classify the genotypes of our isolates, we constructed phylogenetic trees using deduced amino acid sequences of the σC-encoding gene. This analysis revealed that all isolates were in the same cluster (genotypic cluster I) as the commercial vaccine isolates (**Figure 3**). For the σNS-encoding gene, nucleotide sequences of the 10 isolates showed that they are grouped together in one branch and clustered with the Californian chicken-origin isolate K1600657. Conversely, Chinese chicken-origin isolates, TVAV-like wild bird isolates (Pycno-1, SD-12, Tvärminne avian virus, 71-03, Chickadee, and SRK) and commercial vaccine isolates were distinct from all isolates in this study (**Figure 4**). For the p10- and p17-encoding gene, nucleotide sequences of the 10 isolates showed that they are clustered with Chinese chicken-origin isolates and commercial vaccine isolates. Conversely, previously identified TVAV-like wild bird isolates (Pycno-1 and Tvärminne avian virus) were distinct from all isolates in this study (**Supplementary Figures 1**, **2**).

In Recombination analysis, representative 5 isolates (A15-113, A15-157, A18-13, A18-19, and A19-106) were selected by five independent recombination patterns in the σC-encoding gene of our 10 isolates. A15-113 has a similar pattern to A15-108, and A15-157 has a similar pattern to A15-19, A15-48, and A18-205. In addition, A18-13 has a similar pattern to A15-71, and A18-19 and A19-106 showed independent patterns. Whereas all isolates have identical recombination patterns in σNS-encoding gene. All reference sequences in **Supplementary Table 1** were analyzed. However, for the σC-encoding gene, meaningful

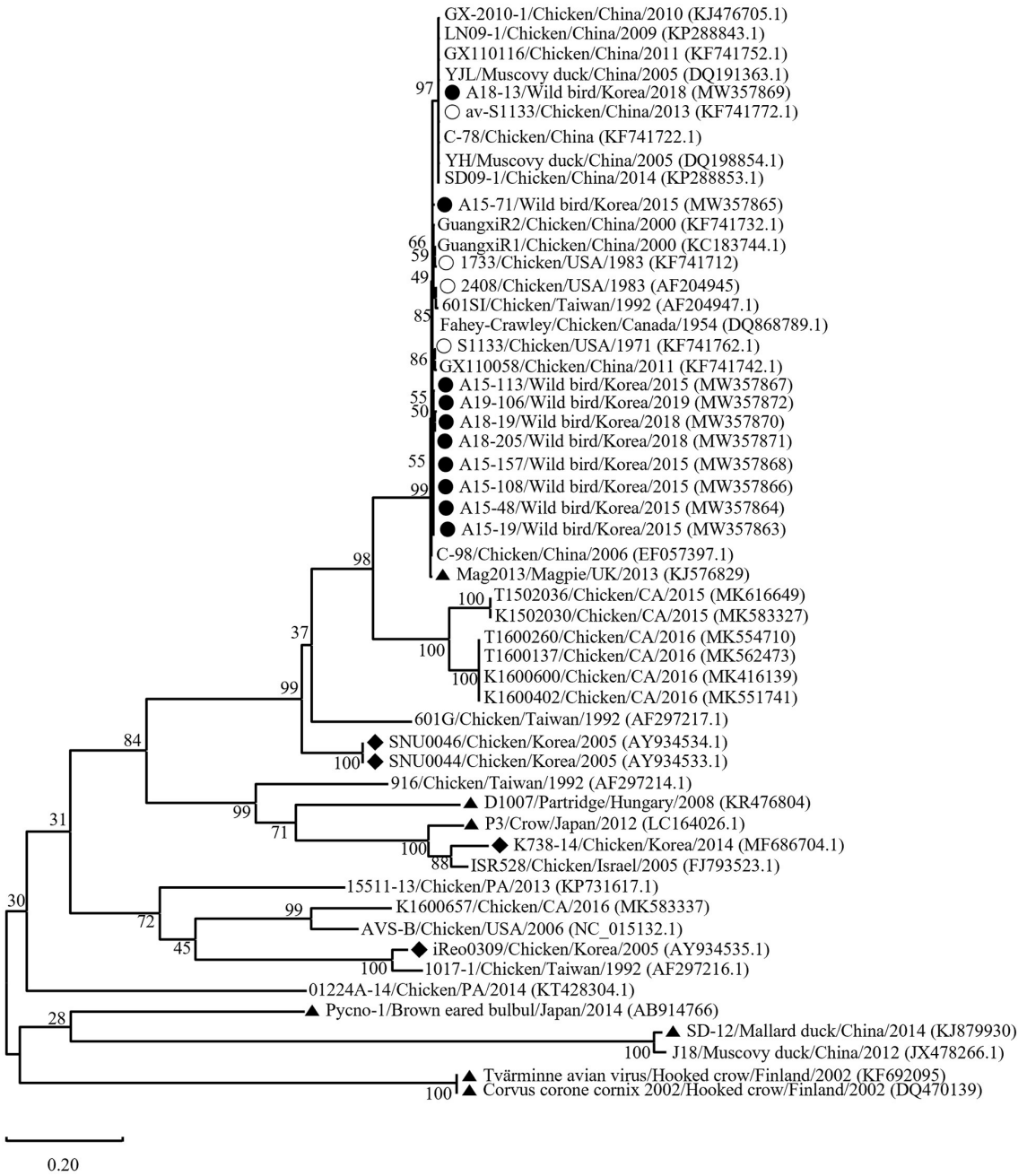


FIGURE 2 | Maximum likelihood phylogenetic tree for the σC-encoding gene based on nucleotide sequences (981 nt). Maximum likelihood phylogenetic analyses were conducted using MEGA-X software with the Kimura 2-parameter model and 1,000 bootstrap replicates. The 43 reference sequences were obtained from GenBank. The black circle (•) indicates our isolates, while the white circle (○) indicates the vaccine isolates. Additionally, the black diamond (◆) indicates field isolates in Korea, and the black triangle (▲) indicates previously isolated wild bird isolates. Each sequence on the tree is identified by the isolate name, host, country of origin, year of isolation, and GenBank accession number.

recombination with 3 isolates (GX110116, GuangxiR1, and K1600657) were detected. And for the σNS-encoding gene, meaningful recombination with 3 isolates (GX110116, AVS-B, and 01224A-14) were detected. According to a Bootscan analysis, it appears that the σC-encoding gene of A18-13 were recombined with Chinese vaccine break isolates GX110116 and Californian chicken-origin isolate K1600657. However, 4 isolates (A15-113, A15-157, A18-19, and A19-106) did not show clear evidence for recombination with these isolates (GX110116 and K1600657). For the σNS-encoding gene, there is evidence that 5 isolates may have recombined with Chinese vaccine break isolate GX110116 and American chicken-origin isolate AVS-B (**Figure 5**).

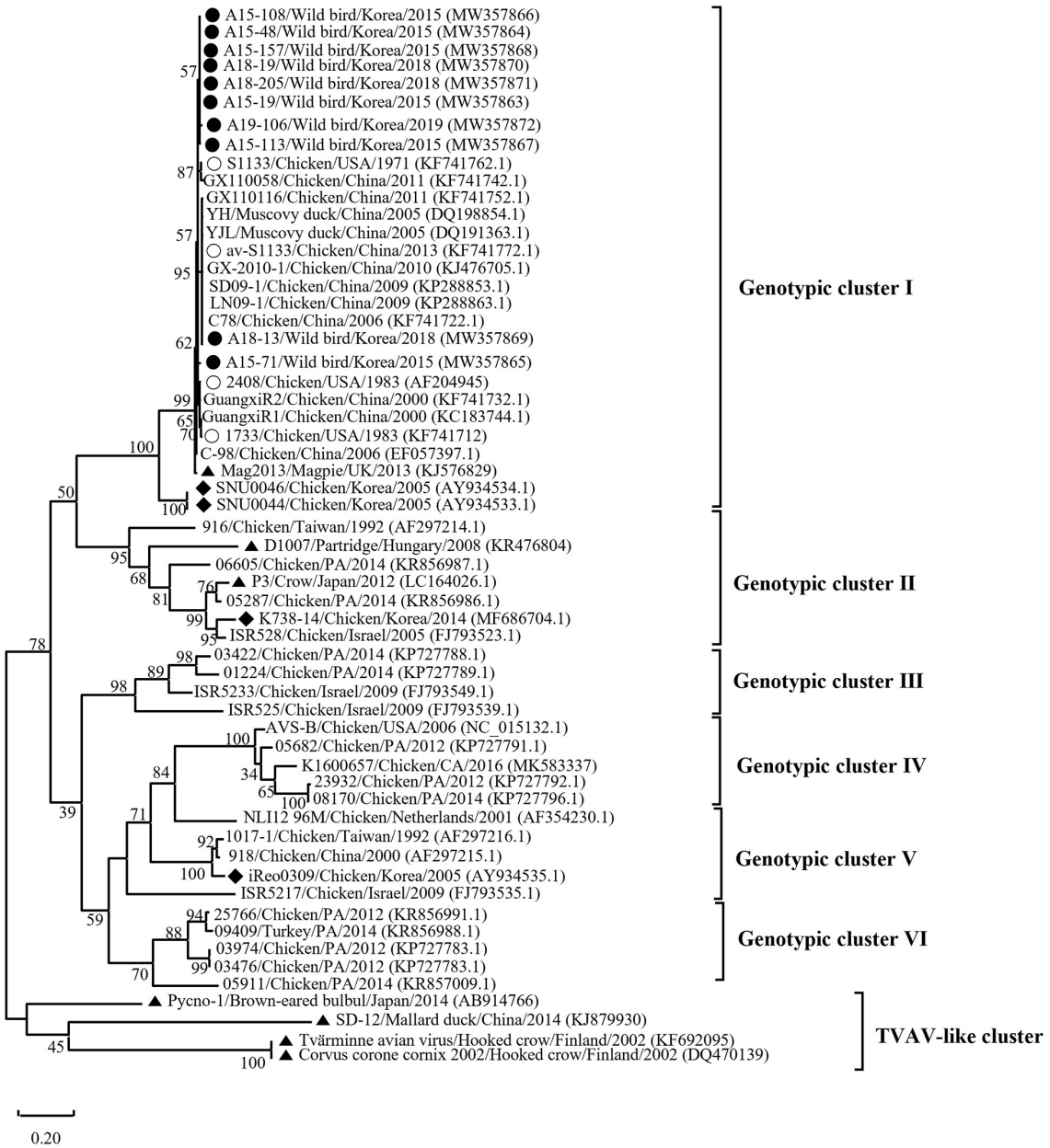


FIGURE 3 | Phylogenetic tree of ARV isolates based on the deduced amino acid sequences of the σC-encoding gene. Maximum likelihood phylogenetic analyses were conducted using MEGA-X software with the Jones–Taylor–Thornton (JTT) model and 1,000 bootstrap replicates. The tree shows genetic relationship between the σC protein sequences (326 amino acids) of our 10 isolates and 48 reference isolates that were isolated from around the world. The virus isolates are clustered into six genotypic clusters. The black circle (•) indicates our isolates, and the white circle (○) indicates the vaccine isolates. Additionally, the black diamond (♦) indicates field isolates in Korea, and the black triangle (▲) indicates previously isolated wild bird isolates. Each sequence on the tree is identified by the isolate name, host, country of origin, year of isolation, and GenBank accession number.

DISCUSSION

It has been hypothesized that wild birds have a role in the transmission and maintenance of ARVs and pathogenic potential of ARVs in wild birds that usually commingle around poultry farms. However, lack of evidence to support these hypotheses. Molecular surveillance is crucial to strategize control and prevention of endemic diseases. Contrary to the previous study results, recently detected reovirus isolates in ostrich (*Stuthio camelus*) and free-living magpie were genetically related to chicken-origin ARVs (13, 31). Despite of the need for a reliable and relevant source of genetic information of wild birds, available genetic information is rare. In this study, we conducted surveillance and molecular characterization of ARVs in wild birds, and ARVs were detected in 1.9% (14/739) of the wild bird feces. This result suggests a much lower prevalence of ARVs

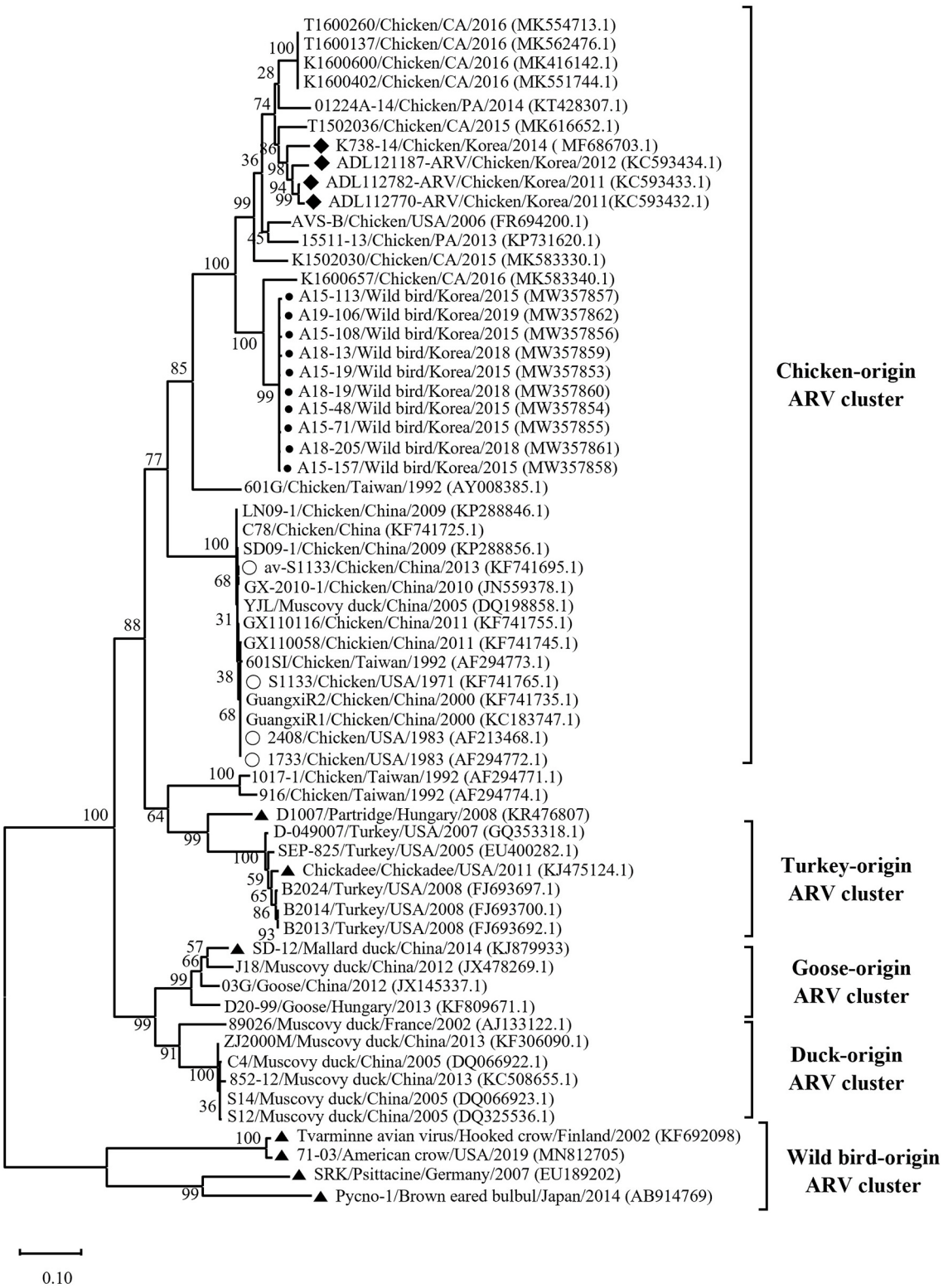


FIGURE 4 | Maximum likelihood phylogenetic tree for the σNS-encoding gene based on nucleotide sequences (1,104 nt). Maximum likelihood phylogenetic analyses were conducted using MEGA-X software with the Kimura 2-parameter model and 1,000 bootstrap replicates. The 52 reference sequences were obtained from GenBank. The black circle (•) indicates our isolates, and the white circle (○) indicates the vaccine isolates. Additionally, the black diamond (♦) indicates field isolates in Korea, and the black triangle (▲) indicates previously isolated wild bird isolates. Each sequence on the tree is identified by the isolate name, host, country of origin, year of isolation, and GenBank accession number.

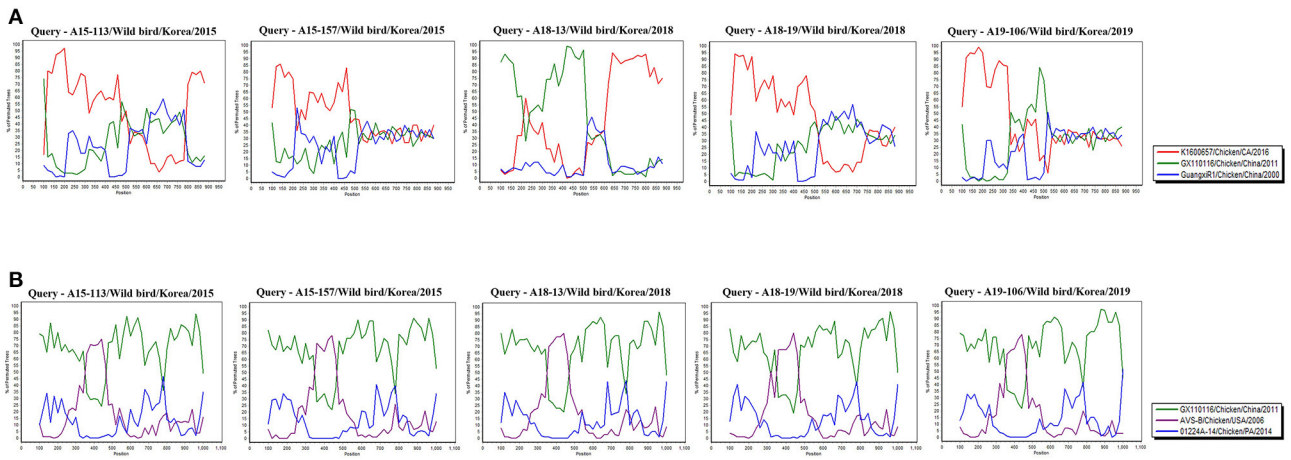


FIGURE 5 | Bootscan analysis of wild bird ARV for detecting recombination was performed using Simplot program version 3.5.1, the neighbor-joining method with a Kimura 2-parameter applied and 100 replicates. **(A)** σC-encoding gene, **(B)** σNS-encoding gene.

than reported previously in some studies in Poland (11.5%; 9/78 dead wild birds; intestine) (12) and the USA (25.0%; 4/20 dead American woodcock; intestine) (49). Both studies reported that ARVs in wild bird's intestine had high prevalence than ARVs in wild bird feces. These high prevalence in intestines are probably due to ARVs have enteric tropism (50) and low prevalence in feces may have influenced by several factors, such as environmental factors or feces can act as an anti-viral factor (51). A similar result to our study was reported in Brazil that conducted surveillance of ARVs in fecal samples from broiler chickens (1.9%; 7/378) (52). The differences in prevalence could be attributed to the sample type (feces *vs.* tissue) or regional differences.

We classified reovirus isolates in this study based on the σC protein, all isolates belonged to genotypic cluster I, which is the predominant cluster in chicken-origin ARVs and includes commercial vaccine isolates (**Figure 3**) (8, 53). Similar results in wild birds were described in the UK and ARV isolated from free-living magpie (*Pica pica*) belonged to genotypic cluster I (31). Conversely, in previous studies, wild bird-origin ARVs belonged to the TVAV-like cluster (27–30, 34, 35). Moreover, the phylogenetic analysis of the p10- and p17-encoding genes revealed that all isolates were closely clustered with Chinese chicken-origin isolates. For the σNS-encoding gene, our isolates were more closely clustered with the Californian chicken-origin isolate K1600657 that belonged to chicken-origin ARV cluster (**Figure 4**). These findings suggest that chicken-origin ARV can infect wild birds and can thus be a potential source of ARV infections that can spillover to chickens and vice versa.

Throughout this study, the 10 ARVs of the σC protein showed high degree of antigenic homogeneity to commercial vaccine isolates (inactivated vaccine) and Chinese chicken-origin isolates. However, these isolates were distinct from Korean chicken-origin isolates. In addition, analysis of the σNS protein showed high similarity to Californian chicken-origin isolate K1600657. Our results indicated that the deduced amino acid substitution patterns of all isolates are identical to those of chicken-origin ARVs in the σC and σNS proteins. Particularly, the A18-13 isolate had identical patterns with avirulent ARV (av-S1133) and virulent ARV (GX110116) in terms of the σC protein, which contributed to ARV virulence (54, 55). For the σNS protein, deduced amino acid substitutions at positions 72, 121, 152, 158, 160, 162, 218, 254, 273, and 350 were observed for all isolates with respect to S1133 (**Table 3**). These substitutions were also observed for the Californian chicken-origin isolate K1600657 that was associated with severe clinical signs (53). These findings suggested that pathotyping based on the σC protein has limitations. More analyses of polymorphisms and deduced amino acid substitutions of other segments are required for pathotyping of ARVs. For the p10 protein, all isolates had a putative transmembrane domains, which are members of the fusion associated small transmembrane (FAST) protein family that conserved in chicken-origin ARVS (56). For the p17 protein, all isolates had identical nuclear localization signal (NLS) motif with chicken-origin isolates, which is conserved in chicken-origin ARVS (**Supplementary Table 5**) (57). These finding suggested that 10 isolates have identical FAST protein and NLS to chicken-origin ARVs.

According to recombination analysis, the σC-encoding gene of A18-13 isolates might have recombined with Chinese vaccine break isolate GX110116 at ∼300–500 nucleotide sequence positions that include four non-synonymous nucleotide substitutions (G^{317}, T^{338}, A^{403}, and C^{405}) and Californian chicken-origin isolate K1600657 at approximate 700–900 nucleotide sequence positions with no non-synonymous nucleotide substitutions. Whereas other four isolates showed independent patterns in recombination analysis, although these isolates did not show clear evidence of recombination. The σNS-encoding gene of five isolates (A15-113, A15-157, A18-13, A18-19, and A19-106) might have recombined with American chicken-origin isolate AVS-B at ∼350–470 nucleotide sequence positions that include three non-synonymous nucleotide

substitutions (T^{361}, A^{455}, and A^{472}) and Chinese vaccine break isolate GX110116 at 100–350 and 470–1,000 nucleotide sequence positions that include seven non-synonymous nucleotide substitutions (G^{215}, G^{478}, G^{484}, T^{652}, A^{760}, G^{817}, and C^{1049}). In addition, all isolates showed identical recombination patterns in the σNS-encoding gene (**Figure 5** and **Supplementary Table 4**). These findings suggest that recombination patterns are seems to be associated with the characteristics of the σC and σNS protein. In terms of the σC protein, which is the most variable region in the viral genome (58). For the σNS protein, which is relatively conserved among other genome segments (59). Moreover, our 10 isolates genome reflects genomic reassortment events and intra-segmental recombination between Chinese vaccine break isolate and American chicken-origin isolate in the σNS-encoding gene (**Figure 5**). These finding indicated that our isolates are intracontinental recombined ARVs and can be potential risk in commercial chicken farms.

Wild bird ARVs have been isolated in sedentary bird species, including magpie, crow, partridge, black-capped chickadee, and psittacine birds (27, 29, 31–33). A previous study reported the isolation of wild bird ARV from a migratory bird species (mallard duck) (35). Eight isolates identified in this study (A15-48, A15-71, A15-108, A15-113, A18-13, A18-19, A18-205, and A19-106) were isolated from migratory bird species (bean goose, Eurasian teal, Indian spot-billed duck, and mallard duck) (**Table 2**). In the flyways of these birds, South Korea is an important wintering site for wild migratory birds, such as waterfowl which flies across the East Asia–Australian flyway, with Korea and China located along this flyway (60). The A15-157 isolate was from an oriental turtle dove. Notably, the oriental turtle dove is a sedentary bird species commonly found around chicken farms in Korea. These sedentary bird species may play a role in ARV transmission between poultry and wild birds.

CONCLUSION

We describe genetic characterization of 10 ARVs isolated from wild birds' feces and showed that all isolates were closely related to chicken-origin ARVs. It is possible that wild birds may play a potential role in the epidemiology of chicken-origin ARVs as a carrier. For further study to evaluate risk factors, it is essential to conduct whole-genome sequencing and investigate the potential pathogenicity of ARVs isolated from wild birds for domestic poultry.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

S-YC and MK: contributed to conception and design of experiments. H-KJ, S-WK, Y-RC, KS, BW, J-YP, and J-FZ: contributed to acquisition, analysis, and interpretation of data. S-WK, Y-RC, S-YC, and MK: drafted and/or revised the article. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Agriculture, Food and Rural Affairs Convergence Technologies Program for Educating Creative Global Leader (716002-7, 320005-4) funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.794934/full#supplementary-material

Supplementary Figure 1 | Maximum likely phylogenetic tree for p10-encoding gene based on nucleotide sequences (297 nt). Maximum likelihood phylogenetic analyses were conducted using MEGA X software with the Kimura 2-parameter model and 1,000 bootstrap replicates. The 31 reference sequences were obtained from GenBank. The black circle (•) indicates our isolates while the white circle (○) indicates the vaccine isolates. Additionally, the black diamond (♦) indicates field isolates in Korea while the black triangle (▲) indicates previously isolated wild bird isolates. Each sequence on the tree is identified by isolated name, host and country of origin, year of isolation, GenBank accession number.

Supplementary Figure 2 | Maximum likely phylogenetic tree for p17-encoding gene based on nucleotide sequences (441 nt). Maximum likelihood phylogenetic analyses were conducted using MEGA X software with the Kimura 2-parameter model and 1,000 bootstrap replicates. The 31 reference sequences were obtained from GenBank. The black circle (•) indicates our isolates while the white circle (○) indicates the vaccine isolates. Additionally, the black diamond (♦) indicates field isolates in Korea while the black triangle (▲) indicates previously isolated wild bird isolates. Each sequence on the tree is identified by isolated name, host and country of origin, year of isolation, GenBank accession number.

Supplementary Table 1 | Information of new isolated ARV and reference isolates used in this study.

Supplementary Table 2 | Sequence alignment of the σC-encoding gene (981 nt).

Supplementary Table 3 | Sequence alignment of the σNS-encoding gene (1,104 nt).

Supplementary Table 4 | Single nucleotide polymorphisms in the σC- and σNS-encoding genes of ARV isolates from wild bird.

Supplementary Table 5 | Deduced amino acid substitutions in the putative transmembrane domains of the p10 and nuclear localization signal motif in the p17.

REFERENCES

1. Spandidos Da, Graham A. Physical and chemical characterization of an avian reovirus. *J Virol.* (1976) 19:968–76. doi: 10.1128/jvi.19.3.968-976.1976
2. Bodelón G, Labrada L, Martinez-Costas J, Benavente J. The avian reovirus genome segment S1 is a functionally tricistronic gene that expresses one structural and two nonstructural proteins in infected cells. *Virology.* (2001) 290:181–91. doi: 10.1006/viro.2001.1159
3. Liu H-J, Lin P-Y, Wang L-R, Hsu H-Y, Liao M-H, Shih W-L. Activation of small GTPases RhoA and Rac1 is required for avian reovirus p10-induced syncytium formation. *Mol Cells.* (2008) 26:396–403.
4. Liu H-J, Lin P-Y, Lee J-W, Hsu H-Y, Shih W-L. Retardation of cell growth by avian reovirus p17 through the activation of p53 pathway. *Biochem Biophys Res Commun.* (2005) 336:709–15. doi: 10.1016/j.bbrc.2005.08.149
5. Chulu JL, Huang WR, Wang L, Shih WL, Liu HJ. Avian reovirus nonstructural protein p17-induced G2/M cell cycle arrest and host cellular protein translation shutoff involve activation of p53-dependent pathways. *J Virol.* (2010) 84:7683–94. doi: 10.1128/JVI.02604-09
6. Calvo PG, Fox GC, Parrado XLH, Llamas-Saiz AL, Costas C, Martínez-Costas J, et al. Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J Mol Biol.* (2005) 354:137–49. doi: 10.1016/j.jmb.2005.09.034
7. Wickramasinghe R, Meanger J, Enriquez CE, Wilcox GE. Avian reovirus proteins associated with neutralization of virus infectivity. *Virology.* (1993) 194:688–96. doi: 10.1006/viro.1993.1309
8. Lu H, Tang Y, Dunn PA, Wallner-Pendleton EA, Lin L, Knoll EA. Isolation and molecular characterization of newly emerging avian reovirus variants and novel strains in Pennsylvania, USA, 2011–2014. *Sci Rep.* (2015) 5:14727. doi: 10.1038/srep14727
9. De la Torre D, Astolfi-Ferreira CS, Chacón RD, Puga B, Piantino Ferreira A. Emerging new avian reovirus variants from cases of enteric disorders and arthritis/tenosynovitis in Brazilian poultry flocks. *Br Poult Sci.* (2021) 62:361–72. doi: 10.1080/00071668.2020.1864808
10. Pantin-Jackwood MJ, Day JM, Jackwood MW, Spackman E. Enteric viruses detected by molecular methods in commercial chicken and turkey flocks in the United States between 2005 and 2006. *Avian Dis.* (2008) 52:235–44. doi: 10.1637/8174-111507-Reg.1
11. Wang D, Shi J, Yuan Y, Zheng L, Zhang D. Complete sequence of a reovirus associated with necrotic focus formation in the liver and spleen of muscovy ducklings. *Vet Microbiol.* (2013) 166:109–22. doi: 10.1016/j.vetmic.2013.05.022
12. Styś-Fijoł N, Kozdruń W, Czekaj H. Detection of avian reoviruses in wild birds in Poland. *J Vet Res.* (2017) 61:239. doi: 10.1515/jvetres-2017-0033
13. Sakai K, Ueno Y, Ueda S, Yada K, Fukushi S, Saijo M, et al. Novel reovirus isolation from an Ostrich (Struthio camelus) in Japan. *Vet Microbiol.* (2009) 134:227–32. doi: 10.1016/j.vetmic.2008.08.022
14. Yan T, Guo L, Jiang X, Wang H, Yao Z, Zhu S, et al. Discovery of a novel recombinant avian orthoreovirus in China. *Vet Microbiol.* (2021) 260:109094. doi: 10.1016/j.vetmic.2021.109094
15. De Gussem J, Swam H, Lievens K, De Herdt P. Reovirus tenosynovitis in a flock of layer breeders. *Avian Pathology.* (2010) 39:169–70. doi: 10.1080/03079451003717597
16. Wang H, Gao B, Liu X, Zhang S, Diao Y, Tang Y. Pathogenicity of a variant duck orthoreovirus strain in Cherry Valley Ducklings. *Vet Microbiol.* (2020) 242:108546. doi: 10.1016/j.vetmic.2019.108546
17. Mor SK, Sharafeldin TA, Porter RE, Ziegler A, Patnayak DP, Goyal SM. Isolation and characterization of a turkey arthritis reovirus. *Avian Dis.* (2013) 57:97–103. doi: 10.1637/10353-090712-Reg.1
18. Neelima S, Ram G, Kataria J, Goswami T. Avian reovirus induces an inhibitory effect on lymphoproliferation in chickens. *Vet Res Commun.* (2003) 27:73–85. doi: 10.1023/A:1022014825451
19. Van de Zande S, Kuhn E-M. Central nervous system signs in chickens caused by a new avian reovirus strain: a pathogenesis study. *Vet Microbiol.* (2007) 120:42–9. doi: 10.1016/j.vetmic.2006.10.024
20. Troxler S, Rigomier P, Bilic I, Liebhart D, Prokofieva I, Robineau B, et al. Identification of a new reovirus causing substantial losses in broiler production in France, despite routine vaccination of breeders. *Vet Rec.* (2013) 172:556. doi: 10.1136/vr.101262
21. Palomino-Tapia V, Mitevski D, Inglis T, van der Meer F, Abdul-Careem MF. Molecular characterization of emerging avian reovirus variants isolated from viral arthritis cases in Western Canada 2012–2017 based on partial sigma (σ) C gene. *Virology.* (2018) 522:138–46. doi: 10.1016/j.virol.2018.06.006
22. De Carli S, Wolf JM, Gräf T, Lehmann FK, Fonseca AS, Canal CW, et al. Genotypic characterization and molecular evolution of avian reovirus in poultry flocks from Brazil. *Avian Pathology.* (2020) 49:611–20. doi: 10.1080/03079457.2020.1804528
23. Kant A, Balk F, Born L, van Roozelaar D, Heijmans J, Gielkens A, et al. Classification of Dutch and German avian reoviruses by sequencing the $\sigma $ C protein. *Vet Res.* (2003) 34:203–12. doi: 10.1051/vetres:2002067
24. Goldenberg D, Pasmanik-Chor M, Pirak M, Kass N, Lublin A, Yeheskel A, et al. Genetic and antigenic characterization of sigma C protein from avian reovirus. *Avian Pathology.* (2010) 39:189–99. doi: 10.1080/03079457.2010.480969
25. Teng L, Xie Z, Xie L, Liu J, Pang Y, Deng X, et al. Sequencing and phylogenetic analysis of an avian reovirus genome. *Virus Genes.* (2014) 48:381–6. doi: 10.1007/s11262-013-1029-5
26. Noh JY, Lee DH, Lim TH, Lee JH, Day JM, Song CS. Isolation and genomic characterization of a novel avian orthoreovirus strain in Korea, 2014. *Arch Virol.* (2018) 163:1307–16. doi: 10.1007/s00705-017-3667-8
27. de Kloet SR. Sequence analysis of four double-stranded RNA genomic segments reveals an orthoreovirus with a unique genotype infecting psittaciformes. *Avian Dis.* (2008) 52:480–6. doi: 10.1637/8212-011908-Reg.1
28. Huhtamo E, Uzcátegui NY, Manni T, Munsterhjelm R, Brummer-Korvenkontio M, Vaheri A, et al. Novel orthoreovirus from diseased crow, Finland. *Emerg Infect Dis.* (2007) 13:1967. doi: 10.3201/eid1312.070394
29. Forzán MJ, Renshaw RW, Bunting EM, Buckles E, Okoniewski J, Hynes K, et al. A novel orthoreovirus associated with epizootic necrotizing enteritis and splenic necrosis in American crows (Corvus brachyrhynchos). *J Wildl Dis.* (2019) 55:812–22. doi: 10.7589/2019-01-015
30. Dandar E, Huhtamo E, Farkas SL, Oldal M, Jakab F, Vapalahti O, et al. Complete genome analysis identifies Tvärminne avian virus as a candidate new species within the genus Orthoreovirus. *J Gen Virol.* (2014) 95:898–904. doi: 10.1099/vir.0.060699-0
31. Lawson B, Dastjerdi A, Shah S, Everest D, Núñez A, Pocknell A, et al. Mortality associated with avian reovirus infection in a free-living magpie (Pica pica) in Great Britain. *BMC Vet Res.* (2015) 11:20. doi: 10.1186/s12917-015-0329-5
32. Kugler R, Dandár E, Fehér E, Jakab F, Mató T, Palya V, et al. Phylogenetic analysis of a novel reassortant orthoreovirus strain detected in partridge (Perdix perdix). *Virus Res.* (2016) 215:99–103. doi: 10.1016/j.virusres.2015.11.018
33. Mor SK, Armién AG, Reed L, Schott R, Goyal SM. Detection and molecular characterization of a reovirus in black-capped chickadees (Poecile atricapillus) from Minnesota, USA. *J Wildl Dis.* (2014) 50:928–32. doi: 10.7589/2014-02-045
34. Ogasawara Y, Ueda H, Kikuchi N, Kirisawa R. Isolation and genomic characterization of a novel orthoreovirus from a brown-eared bulbul (Hypsipetes amaurotis) in Japan. *J Gen Virol.* (2015) 96:1777–86. doi: 10.1099/vir.0.000110
35. Yu K, Li Y, Han H, Song M, Ma X, Liu C, et al. Complete genome sequence of an avian reovirus isolated from wild mallard ducks in China. *Genome Announc.* (2014) 2:e00813–14. doi: 10.1128/genomeA.00813-14
36. Gough R, Alexander D, Collins M, Lister S, Cox W. Routine virus isolation or detection in the diagnosis of diseases in birds. *Avian Pathology.* (1988) 17:893–907. doi: 10.1080/03079458808436511
37. Vindevogel H, Meulemans G, Pastoret PP, Schwers A, Calberg-Bacq CM. Reovirus infection in the pigeon. In: *Annales de Recherches Vétérinaires.* (1982) 13:149–52.
38. Hlinak A, Müller T, Kramer M, Mühle R-U, Liebherr H, Ziedler K. Serological survey of viral pathogens in bean and white-fronted geese from Germany. *J Wildl Dis.* (1998) 34:479–86. doi: 10.7589/0090-3558-34.3.479
39. Lee Y-N, Cheon S-H, Lee E-K, Heo G-B, Bae Y-C, Joh S-J, et al. Pathogenesis and genetic characteristics of novel reassortant low-pathogenic avian influenza H7 viruses isolated from migratory birds in the Republic of Korea in the winter of 2016–2017. *Emerg Microbes Infect.* (2018) 7:182. doi: 10.1038/s41426-018-0181-3

40. Caterina KM, Frasca S Jr, Girshick T, Khan MI. Development of a multiplex PCR for detection of avian adenovirus, avian reovirus, infectious bursal disease virus, and chicken anemia virus. *Mol Cell Probes*. (2004) 18:293–8. doi: 10.1016/j.mcp.2004.04.003
41. Liu Q, Zhang G, Huang Y, Ren G, Chen L, Gao J, et al. Isolation and characterization of a reovirus causing spleen necrosis in Pekin ducklings. *Vet Microbiol*. (2011) 148:200–6. doi: 10.1016/j.vetmic.2010.09.016
42. Lee D-H, Lee H-J, Lee Y-J, Kang H-M, Jeong O-M, Kim M-C, et al. DNA barcoding techniques for avian influenza virus surveillance in migratory bird habitats. *J Wildl Dis*. (2010) 46:649–54. doi: 10.7589/0090-3558-46.2.649
43. Jeong S, Lee D-H, Kwon J-H, Kim Y-J, Lee S-H, Cho AY, et al. Highly pathogenic avian influenza clade 2.3. 4.4 b subtype H5N8 virus isolated from mandarin duck in South Korea, 2020. *Viruses*. (2020) 2:1389. doi: 10.3390/v12121389
44. Ratnasingham S, Hebert PD. BOLD The Barcode of Life Data System (http://www.barcodinglifeorg). *Mol Ecol Notes*. (2007) 7:355–64. doi: 10.1111/j.1471-8286.2007.01678.x
45. Reck C, Menin Á, Canever MF, Miletti LC. Rapid detection of mycoplasma synoviae and avian reovirus in clinical samples of poultry using multiplex PCR. *Avian Dis*. (2013) 57:220–4. doi: 10.1637/10425-101712-Reg.1
46. Belkhir K. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations (2004). Available online at: https://kimura.univ-montp2.fr/genetix/
47. Kumar S, Stecher G, Li M, Knyaz C, Tamura K, MEGA X. molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. (2018) 35:1547. doi: 10.1093/molbev/msy096
48. Strimmer K, Forslund K, Holland B, Moulton V. A novel exploratory method for visual recombination detection. *Genome Biol*. (2003) 4:R33. doi: 10.1186/gb-2003-4-5-r33
49. Docherty D, Converse K, Hansen W, Norman G. American woodcock (Scolopax minor) mortality associated with a reovirus. *Avian Dis*. (1994):899–904. doi: 10.2307/1592132
50. Sharafeldin TA, Mor SK, Sobhy NM, Xing Z, Reed KM, Goyal SM, et al. A newly emergent Turkey arthritis reovirus shows dominant enteric tropism and induces significantly elevated innate antiviral and t Helper-1 cytokine responses. *PLoS ONE*. (2015) 10:e0144085. doi: 10.1371/journal.pone.0144085
51. Takehara K, Chinen O, Jahangir A, Miyoshi Y, Ueno Y, Ueda S, et al. Ceramic powder made from chicken feces: anti-viral effects against avian influenza viruses. *Avian Dis*. (2009) 53:34–8. doi: 10.1637/8382-062008-Reg.1
52. Tamehiro CY, Alfieri AF, Médici C, Alfieri AA. Segmented double-stranded genomic RNA viruses in fecal samples from broiler chicken. *Braz J Microbiol*. (2003) 34:349–53. doi: 10.1590/S1517-83822003000400013
53. Egaña-Labrin S, Hauck R, Figueroa A, Stoute S, Shivaprasad H, Crispo M, et al. Genotypic characterization of emerging avian reovirus genetic variants in California. *Sci Rep*. (2019) 9:9351. doi: 10.1038/s41598-019-45494-4
54. Lin P-Y, Liu H-J, Chang C-D, Chang C-I, Hsu J-L, Liao M-H, et al. Avian reovirus S1133-induced DNA damage signaling and subsequent apoptosis in cultured cells and in chickens. *Arch Virol*. (2011) 156:1917. doi: 10.1007/s00705-011-1063-3
55. Zhang Z, Lin W, Li X, Cao H, Wang Y, Zheng SJ. Critical role of eukaryotic elongation factor 1 alpha 1 (EEF1A1) in avian reovirus sigma-C-induced apoptosis and inhibition of viral growth. *Arch Virol*. (2015) 160:1449–61. doi: 10.1007/s00705-015-2403-5
56. Shmulevitz M, Duncan R. A new class of fusion-associated small transmembrane (FAST) proteins encoded by the non-enveloped fusogenic reoviruses. *EMBO J*. (2000) 19:902–12. doi: 10.1093/emboj/19.5.902
57. Costas C, Martínez-Costas J, Bodelón G, Benavente J. The second open reading frame of the avian reovirus S1 gene encodes a transcription-dependent and CRM1-independent nucleocytoplasmic shuttling protein. *J Virol*. (2005) 79:2141–50. doi: 10.1128/JVI.79.4.2141-2150.2005
58. Ayalew LE, Gupta A, Fricke J, Ahmed KA, Popowich S, Lockerbie B, et al. Phenotypic, genotypic and antigenic characterization of emerging avian reoviruses isolated from clinical cases of arthritis in broilers in Saskatchewan, Canada. *Sci Rep*. (2017) 7:3565. doi: 10.1038/s41598-017-02743-8
59. Guo K, Dormitorio TV, Ou S-C, Giambrone JJ. Development of TaqMan real-time RT-PCR for detection of avian reoviruses. *J Virol Methods*. (2011) 177:75–9. doi: 10.1016/j.jviromet.2011.06.022
60. Lycett SJ, Duchatel F, Digard P. A brief history of bird flu. *Philosophical Transactions of the Royal Society B*. (2019) 374:20180257. doi: 10.1098/rstb.2018.0257

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/fvets.2022.846904

Measuring How Recombination Re-shapes the Evolutionary History of PRRSV-2: A Genome-Based Phylodynamic Analysis of the Emergence of a Novel PRRSV-2 Variant

*Nakarin Pamornchainavakul, Mariana Kikuti, Igor A. D. Paploski, Dennis N. Makau, Albert Rovira, Cesar A. Corzo and Kimberly VanderWaal**

Department of Veterinary Population Medicine, University of Minnesota, Saint Paul, MN, United States

OPEN ACCESS

Edited by:
Christina Leyson, Agricultural Research Service (USDA), United States

Reviewed by:
Jie Li, Changshu Institute of Technology, China
Sergei Raev, The Ohio State University, United States

***Correspondence:**
Kimberly VanderWaal
kvw@umn.edu

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *31 December 2021*
Accepted: *25 February 2022*
Published: *25 March 2022*

Citation:
Pamornchainavakul N, Kikuti M, Paploski IAD, Makau DN, Rovira A, Corzo CA and VanderWaal K (2022) Measuring How Recombination Re-shapes the Evolutionary History of PRRSV-2: A Genome-Based Phylodynamic Analysis of the Emergence of a Novel PRRSV-2 Variant. Front. Vet. Sci. 9:846904. doi: 10.3389/fvets.2022.846904

While the widespread and endemic circulation of porcine reproductive and respiratory syndrome virus type 2 (PRRSV-2) causes persistent economic losses to the U.S. swine industry, unusual increases of severe cases associated with the emergence of new genetic variants are a major source of concern for pork producers. Between 2020 and 2021, such an event occurred across pig production sites in the Midwestern U.S. The emerging viral clade is referred to as the novel sub-lineage 1C (L1C) 1-4-4 variant. This genetic classification is based on the open reading frame 5 (ORF5) gene. However, although whole genome sequence (WGS) suggested that this variant represented the emergence of a new strain, the true evolutionary history of this variant remains unclear. To better elucidate the variant's evolutionary history, we conducted a recombination detection analysis, time-scaled phylogenetic estimation, and discrete trait analysis on a set of L1C-1-4-4 WGSs (n = 19) alongside other publicly published WGSs (n = 232) collected over a 26-year period (1995–2021). Results from various methodologies consistently suggest that the novel L1C variant was a descendant of a recombinant ancestor characterized by recombination at the ORF1a gene between two segments that would be otherwise classified as L1C and L1A in the ORF5 gene. Based on analysis of different WGS fragments, the L1C-1-4-4 variant descended from an ancestor that existed around late 2018 to early 2019, with relatively high substitution rates in the proximal ORF1a as well as ORF5 regions. Two viruses from 2018 were found to be the closest relatives to the 2020-21 outbreak strain but had different recombination profiles, suggesting that these viruses were not direct ancestors. We also assessed the overall frequency of putative recombination amongst ORF5 and other parts of the genome and found that recombination events which leave detectable numbers of descendants are not common. However, the rapid spread and high virulence of the L1C-1-4-4 recombinant variant demonstrates that inter-sub-lineage recombination occasionally

found amongst the U.S. PRRSV-2 might be an evolutionary mechanisms that contributed to this emergence. More generally, recombination amongst PRRSV-2 accelerates genetic change and increases the chance of the emergence of high fitness variants.

Keywords: porcine reproductive and respiratory syndrome virus 2, viral recombination, whole genome sequencing, variant emergence, epidemics

INTRODUCTION

Over the three decades since its initial report, porcine reproductive and respiratory syndrome (PRRS) has undermined the stability of U.S. swine production both in terms of herd health and economics (1). The disease devastates pig production by causing reproductive failure in breeding herds and respiratory associated morbidity/mortality in growing pigs leading to poor production performance (2). In the U.S., outbreaks are mostly classified as porcine reproductive and respiratory syndrome virus 2 (PRRSV-2). The causative agent of PRRSV-2 is *Betaarterivirus suid 2*, an enveloped RNA virus belonging to *Arteriviridae* family in the *Nidovirales* order whose virulence varies by strain (3, 4). Due to lack of RNA proofreading during replication, PRRSV-2 has a high mutation rate even amongst RNA viruses (5, 6). Additionally, recombination–a process where genomic parts are exchanged between viral variants co-infecting a cell–can occasionally contribute to PRRSV-2 genetic variation. RNA viruses including PRRSV-2, are prone to recombination processes such as template switching in sub-genomic RNA synthesis (7–9).

Recently, pig producers in the Midwestern U.S. experienced atypical production losses caused by a fast-spreading variant of PRRSV-2 (10). From early 2020 to September 2021, 355 genetically similar viruses were detected (i.e., > 98% nucleotide identity based on the open reading frame 5–ORF5–gene). Based on data from the Morrison Swine Health Monitoring Project (MSHMP), which tracks the infection status of ∼50% of the U.S. breeding herd (cite Perez–Voluntary data sharing), 294 pig sites belonging to 15 different production systems, and ∼12% of breeding farms in the region had been impacted (11). The virus involved in this outbreak is referred to as a novel L1C-1-4-4 variant, as it falls within the L1C sub-lineage based on phylogenetic relatedness (12) and mostly possesses a 1-4-4 cut-pattern based on conventional restricted fragment length polymorphism (RFLP)-based classification. Both classifications are based on ORF5 sequences (13). While the exact case definition used is based on >98% nucleotide identity on ORF5 (10), here we refer to this variant simply as L1C-1-4-4. Although ORF5 has been widely used for virus classification or epidemiological assessment since it is highly variable and immunologically relevant (14), it only represents ∼4% of the genome and may not represent the full evolutionary history of the virus, particularly when recombination is involved. For example, when phylogenetic trees are constructed using whole genome sequences, the L1C-1-4-4 variant nests within a clade of viruses that are classified as sub-lineage L1A based on the ORF5 gene (10, 15), suggesting that recombination may be confounding the virus' genealogical tree topology. Because there is no WGS-based nomenclature for classifying PRRSV-2 genetic sub-types, here we refer to WGSs according to their ORF5 lineage classification and recognizing the limitation that genetic relatedness might not hold true when looking at other regions of the genome.

The rapid spread and high production impact of the newly emerged L1C-1-4-4 variant has drawn concern from the industry as similar events occurred in the past and the contributing factors to emergence of these strains are poorly understood. Given that PRRSV-2 in the U.S. is characterized by the cyclic emergence of new strains, and turnover in the dominant sub-lineage every few years (16), this emerging variant may continue increasing in prevalence, bringing further issues to the U.S. swine industry. However, previous emergence events have largely been documented based on ORF5 sequence data, which limits our ability to fully discern evolutionary processes associated with strain emergence. For example, PRRSV-2 virulence and antigenic determinants are multigenic, meaning that clinical presentation characteristics are influenced by a variety of genes throughout the viral genome (4). Thus, understanding the origin of this variant from a whole genome perspective is a crucial step in response to this ongoing outbreak, and may help elucidate evolutionary processes associated with strain emergence more broadly and lead to potential preventive interventions at the farm level. Here, we estimate divergence times, mutation rates, and parental strains of the novel L1C-1-4-4 variant using genomics-based approaches. More generally, to better understand the role of recombination in shaping PRRSV-2 phylogenies, we also quantify the frequency of potential inter- and intra-lineage recombination events.

METHODS

Data

A convenience sample of PRRSV-2 L1C-1-4-4 variant whole genome sequences (WGS) were obtained from the Veterinary Diagnostic Laboratory at the University of Minnesota [see (10) for details]. These samples were from multiple production systems in the Midwestern U.S. participating in the Morrison Swine Health Monitoring Project. Sample selection criteria for WGS included having ORF5 sequences within the emerging variant's phylogenetic clade (<2% genetic distance to at least one other sequence classified as L1C-1-4-4) and cycle threshold (Ct) value ≤25 for reverse transcription polymerase chain reaction (RT-PCR) using VetMAX™ NA and EU PRRSV Reagents (Thermo Fisher Scientific, MA, USA). Oral fluids and processing fluids samples were excluded due to the low success rate for whole genome sequencing (17, 18). At least one ORF5 sequence from each participating system was whole-genome sequenced. For systems that had two or more ORF5 L1C-1-4-4 sequences

identified during this period, the earliest and the most recent samples were selected for WGS sequencing. As described in Kikuti et al. (10), the selected samples were sequenced using Clontech SMARTer RNA Pico v2 kit on illumina MiSeq v3 (Illumina, CA, USA). Of the total 19 WGSs (GenBank accession numbers OL963961-OL963979), one isolate classified as the novel L1C-1-4-4 variant based on the above criteria was from an outbreak limited to a single production system in 2018, while the others were collected from multiple systems during the current epidemic (i.e., 2020–2021). We aligned the WGSs with all available PRRSV-2 WGSs from the U.S. that were publicly available and included date meta-data from NCBI GenBank ($n = 232$, **Supplementary Table 1**), ranging between 1995 and 2018 using MAFFT (19) and manual curation. Genetic distances were calculated between all sequences using seqcombo (20).

Recombination Detection

As a first step for screening sequences for recombination, the alignment was imported to RDP5 (21) for recombination detection. A putative recombination event was flagged when it was detected by at least four of seven methods: RDP (22), GENECONV (23), MaxChi (24), BootScan (25), SiScan (26), Chimaera (27), and 3Seq (28). We performed the analysis as a two-pronged approach. First, we specifically explored recombination in the novel L1C-1-4-4 variant group, which was set as a query against all GenBank WGS references. Second, the alignment was fully scanned (with no reference and query groups defined) to estimate the location of recombination hotspots within the genome. A recombination hotspot is defined as a genomic position in which the frequency of putative recombination exceeds neutral expectations (>99% confidence interval of the local density plot created by a permutation test) (29); genomic regions between hotspots are inferred to have low rates of recombination. Thus, the locations of hotspots can be used to subdivide the genome into fragments, where each fragment is relatively free of frequent within-fragment recombination and thus can be used for further phylogenetic analysis (30, 31).

Maximum likelihood phylogenies were built from each WGS fragment using W-IQ-TREE (32) with automated substitution model selection and 1,000 bootstraps. The consensus trees were assessed to: (1) check the temporal signal under a molecular clock assumption using TempEst (33), and any fragment whose phylogenetic reconstruction did not show a sufficient temporal signal was excluded from further time-scaled analyses. (2) Down-sample the dataset based on pairwise distances from the novel L1C-1-4-4 variant using *ape* version 5.5 (34) applied in R (35). Only the 50 most closely related sequences to each distinct fragment of the novel L1C-1-4-4 variant were retained, yielding a total of 142 sequences for further analysis (**Supplementary Table 1**).

Time-Scaled Phylogenetic Reconstruction

The time to the most recent common ancestor (tMRCA) and substitution rate of each fragment were estimated by Bayesian inference with Markov chain Monte Carlo (MCMC) applied in BEAST v.1.10.4 (36). According to IQ-TREE's substitution model test, we chose the general time reversible (GTR) with empirical base frequencies and gamma plus invariant site (G + I) heterogeneity model for all fragments. An uncorrelated relaxed clock (37) with log-normal distribution and the Gaussian Markov random field (GMRF) skyride (38) were specified as molecular clock model and coalescent prior, respectively. The alignments with these model settings were run with 500 million generations of MCMC. Maximum clade credibility (MCC) trees of each fragment were built using TreeAnnotator v.1.10.4 and visualized on the Nextstrain platform (39).

Discrete Trait Analysis

The frequency of inter- and intra-lineage recombination between ORF5 and other fragments was approximated through WGS-fragment phylogenies using discrete trait analysis in BEAST. For each WGS fragment, the ORF5-based lineage (14) or sub-lineage (12) of each sample was assigned as a discrete trait, and the ancestral trait of each internal node was inferred. Ancestral transitions between traits (i.e., the label the sequence received based on its ORF5 lineage) in the WGS-fragment phylogenies can be interpreted as putative recombination between the ORF5 gene and other WGS regions (i.e., instances where sequences are no longer clustered with other sequences that share the same ORF5-lineage label). Potential recombination in the WGS-fragment phylogenies were estimated from the number of trait (lineage) transitions with Bayes factors (BF) support obtained from an asymmetric substitution model with Bayesian stochastic search variable selection (BSSVS). Other parameters were set as the software default. The analyses were run with MCMC length of 500 million each. Ancestral states annotated on MCC trees were visualized using FigTree v.1.4.4 (40). Lineage and sub-lineage transitions were reported with BF computed by SpreaD3 (41). High numbers of transitions between inferred ORF5-lineage in the phylogenies of other WGS-fragments would provide support that recombination is more common, and that shared phylogenetic ancestry based on ORF5 lineage identity is scrambled on the whole genome due to putative recombination. Low transitions suggest that recombination events that leave descendants detected by surveillance activities are relatively rare, and that shared ancestry based on ORF5 lineages are relatively stable across the genome.

RESULTS

The 18 novel L1C-1-4-4 WGSs associated with the 2020-21 outbreak displayed a 98.2 to 99.9% nucleotide identity. The 2018 virus, which was included for whole genome sequencing based on its high similarity on ORF5, showed a 96.5 to 97.4 % pairwise similarity to the 2020-21 L1C-1-4-4 whole genomes. The greatest difference (<90 % similarity) between the 2020-21 group and the 2018 virus was in nsp9 to nsp10 in the ORF1b region (**Supplementary Figure 1**).

Recombination Profile

All 19 WGSs had a relatively similar recombination profile, with at least six putative recombinant regions in common across the viral genome. The minor parents (i.e., the parent

TABLE 1 | Ancestral date and evolutionary rate estimates of the novel L1C-1-4-4 variants and other PRRSV-2.

WGS fragment	Overall ($n = 161$)		2020–2021 novel L1C-1-4-4 ($n = 18$)		2018–2021 L1C-1-4-4 ($n = 19$)	
	tMRCA* (95% HPD)	mean rate** (95% HPD)	tMRCA* (95% HPD)	ancestral branch rate** (95% HPD)	tMRCA* (95% HPD)	ancestral branch rate** (95% HPD)
ORF1a-1	Oct 1988 (Mar 1983, April 1992)	3.81×10^{-3} (2.69×10^{-3}, 4.98×10^{-3})	Nov 2018 (Feb 2018, Sep 2019)	2.15×10^{-2} (2.00×10^{-3}, 6.79×10^{-2})	Nov 2017 (Jan 2016, Jun 2018)	1.22×10^{-2} (7.00×10^{-4}, 4.60×10^{-3})
ORF1a-2#	Aug 1546# (Jan 1194, Jan 1782)	4.07×10^{-4} (3.29×10^{-4}, 4.86×10^{-4})	May 2003# (Jun 1993, Mar 2014)	8.96×10^{-4} (1.22×10^{-4}, 2.39×10^{-3})	Jun 1997# (May 1987, Jan 2009)	2.92×10^{-3} (4.61×10^{-4}, 6.78×10^{-3})
ORF1b	Oct 1985 (Feb 1979, Jun 1991)	2.40×10^{-3} (1.67×10^{-3}, 3.07×10^{-3})	Jan 2019 (Apr 2018, Nov 2019)	8.82×10^{-3} (3.05×10^{-3}, 1.52×10^{-2})	NA (the 2018 taxon does not group with others)	NA (the 2018 taxon does not group with others)
3'ORFs	Jul 1987 (Apr 1981, May 1992)	2.55×10^{-3} (1.91×10^{-3}, 3.23×10^{-3})	Dec 2018 (Feb 2018, Sep 2019)	5.56×10^{-3} (6.64×10^{-4}, 1.17×10^{-2})	May 2017 (July 2014, May 2018)	1.60×10^{-3} (7.94×10^{-4}, 2.51×10^{-3})
ORF5	Nov 1989 (Oct 1984, May 1994)	3.20×10^{-3} (2.34×10^{-3}, 4.09×10^{-3})	Dec 2018 (Mar 2018, Nov 2019)	5.15×10^{-3} (2.65×10^{-4}, 1.27×10^{-2})	Sep 2017 (Aug 2015, Jun 2018)	2.04×10^{-3} (3.42×10^{-4}, 3.99×10^{-3})

Time to the most recent common ancestor.

**Evolutionary rate (substitutions/nucleotide site/year).*

#Estimates may be anomalous due to relatively poor temporal signal in this fragment.*

contributing the shorter part of the overall sequence) of several of these recombinant regions were viruses that clustered with other viruses classified as sub-lineage L1C in their ORF5 gene. For all 19 WGSs, a large recombinant region was identified in nsp2. The recombination detection algorithms implemented using RDP5 were not able to identify a feasible minor parent in the alignment of the 232 GenBank sequences for the nsp2 recombinant region and for a short recombinant region in ORF2 genes. The detectable minor parents of other events were identified as viruses belonging to the L1H (in nsp1) and L1C (in nsps2-9, and ORFs5-6) sub-lineages based on their ORF5 gene variation (**Figure 1B**, **Supplementary Table 2**). Major differences in the recombination pattern of the 2020-21 ($n = 18$) and the 2018 ($n = 1$) samples were found in the nsp9 to nsp12 of ORF1b gene, where the parents of the 2020–2021 sequences were other L1C, while the 2018's parents were mostly unknown (**Figure 1B**, **Supplementary Table 2**). In agreement with the estimated location of recombinant genomic regions, recombination breakpoints of this variant are located in the following genomic regions: nsp1 flanking regions, insertion and deletion (indel) sites of nsp2 (42), inside nsp9, ORF1ab-ORF2 junction, and ORF2 and ORF5 flanking regions (**Figures 1A,C**).

Several recombination hotspots were detected when employing an all-to-all approach for identifying recombination events (**Figure 1D**). We used these hotspots to extract four fragments within which there was a low frequency of recombination, namely ORF1a-1 [nucleotide position in the alignment (nt) 609 to 3,455], ORF1a-2 (nt 3,459 to 6,773), ORF1b (nt 6,895 to 10,225), and 3' ORFs (nt 11,916 to 14,657) after their genomic annotation (**Figure 1E**). A significant temporal signal was found in the best-fitting rooted maximum likelihood trees for all fragments, though unlike the other fragments, the temporal signal of ORF1a-2 was only significant when using the correlation and R-squared-based rooting methods (**Supplementary Table 3**).

Evolutionary Rate and Ancestral Date

Amongst time-scaled phylogenies of PRRSV-2 genomic fragments, mean evolutionary rates ranged from 2.40 to 3.81 $\times 10^{-3}$ substitutions/site/year, with the exception of ORF1a-2, which had the mean evolutionary rate 10 times lower than that of the rest of the genome (**Table 1**). ORF1a-2's temporal signal, as estimated by Tempest, was more uncertain which may be caused by the low evolutionary rate, ultimately resulting in an eccentric ancestral date estimation that may not be reliable. We thus excluded this fragment from the interpretation of time-scaled trees. The 161 viruses included in this analysis (L1C-1-4-4 variant and the most closely related GenBank sequences across each fragment) had median tMRCAs ranging from 1985 to 1989. The 2020–2021 novel L1C-1-4-4 samples ($n = 18$) form a monophyletic clade sharing a common ancestor in all WGS fragments' trees. Their tMRCA was dated from late 2018 to early 2019. If the 2018 sequence whose ORF5 gene had high nucleotide identity to the 2020–21 L1C-1-4-4 samples was included, tMRCA of the complete set of novel L1C-1-4-4 variants ($n = 19$) was estimated to be in 2017. The 2018 sequence was a basal taxon to L1C-1-4-4 clade in most fragments except the ORF1b tree, where the 2018 taxon was separated and embedded in a clade consisting of viruses labeled as the L1A sub-lineage, suggesting that the 2018 virus experienced a separate recombination event that did not occur in the 2020-21 sequences. The nucleotide substitution rate at the ancestral branch of the novel L1C (inclusive of the 2018 sequence) was lower in some fragments than the overall mean rate, whereas the ancestral branch of the more recent

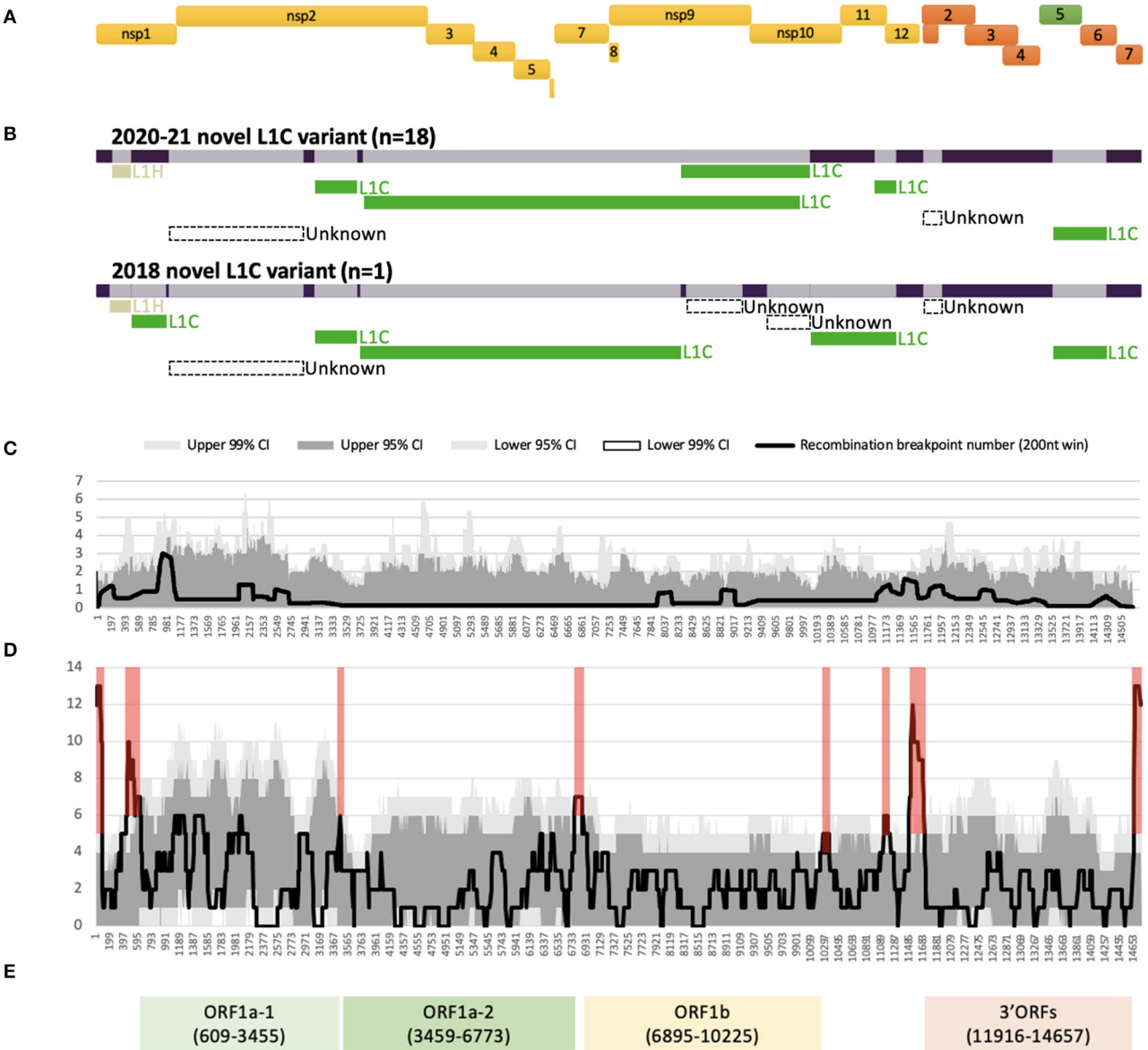


FIGURE 1 | Recombination profile of the novel L1C-1-4-4 viruses in relation to PRRSV-2 genomic organization. **(A)** PRRSV-2 genomic organization. **(B)** Putative recombinant regions and minor parents of the 2020–21 ($n = 18$) and the 2018 ($n = 1$) L1C-1-4-4 variants. The long bar across the top represents the viral genomic backbone. The smaller bars below represent putative minor parents labeled according to the ORF5-based sub-lineages. **(C)** Recombination breakpoint distribution of the novel L1C-1-4-4 WGSs as queries against other PRRSV-2 WGSs. **(D)** Overall recombination breakpoint distribution of the 251 PRRSV-2 WGSs. Recombination hotspots defined by the local density plot are highlighted in red. **(E)** Genomic fragments with low within-fragment recombination rates used for phylogenetic analyses. Nucleotide positions in the alignment are shown in the parenthesis.

2020–2021 epidemic samples had a higher rate than the mean (**Table 1**, trees in Nextstrain: "https://nextstrain.org/community/NakarinP/prrsv/").

Inter- and Intra-lineage Recombination

Phylogenetic clustering of samples on each WGS-fragment tree, labeled according to ORF5 lineages, are visually well-aligned with ORF5-based lineage classification. However, there were instances where clustering of sequences by ORF5 lineage did not translate perfectly to other WGS-fragments, which suggested the possibility of genomic recombination outside ORF5 gene. In the ORF5 tree, there was no significant mixing between lineages/sub-lineages except between sub-lineage L1G ancestors and L1B descendants; sub-lineage L1G is thought to have descended from L1B (16), so L1B-L1G mixing in the tree might be due to some misclassification of closely

related sequences. This pattern was also apparent in the 3' ORFs fragment (ORF5 is embed in this larger fragment), though the clade containing the novel L1C-1-4-4 group became the closest sister to a clade containing the majority of L1A in the 3' ORFs tree. Intermixture of lineage groupings was more apparent in the three ORF1 fragments, suggesting some level of recombination between these genomic regions and ORF5. Although most taxa remained grouped by their ORF5 classification, numerous ancestral recombination were observed between lineage one sub-lineages. This observation was supported by a high number of transitions between traits (i.e., ORF5 lineage label), Bayes factors (**Figure 2A**), and ORF1ab tree topology (**Figure 2B**). The novel L1C-1-4-4 variant's evolutionary history was part of that phenomenon since it was a descendant of the major L1A clade in ORF1a-1 tree. An L1A virus collected in early 2018 (MN073102) was its closest related taxon in all ORF1 trees regardless of whether the L1C-1-4-4 clade was embedded in a larger L1C or L1A clade (**Figure 2B**), suggesting that this virus had a similar evolutionary and recombination history throughout this genomic region.

DISCUSSION

Exploratory analysis of the genome and evolutionary history of viruses causing atypical outbreaks is a key step to understanding their origin. Here, we analyze a set of whole genome sequences (WGSs) from an emerging PRRSV-2 variant and contextualize its evolution using publicly available WGSs from the U.S. swine industry. Our results suggest that the 2020–21 epidemic associated with the novel L1C-1-4-4 viruses arose from a recombinant ancestor of which most genomic parts derived from viruses whose ORF5 genes were classified as sub-lineage L1C. An ancestor of those viruses was estimated to have emerged around late 2018 to early 2019 with a slightly higher mutation rate than the average rate. Two samples from 2018, classified as the L1C-1-4-4 variant and L1A (MN073102) based on ORF5 gene, are the closest relatives of the 2020–21 epidemic variants, with phylogenetic placement varied according to which genomic was examined. The observed shift in phylogenetic clustering of the L1C-1-4-4 variant from L1C in the ORF5-based tree to L1A in ORF1A-based tree, combined with the inferred frequency of recombination estimated from the discrete trait analysis, highlights the role of recombination within L1 sub-lineages in shaping PRRSV-2 genetic diversity.

Interpretations from our analysis should consider some key limitations. First, samples from the NCBI database may not represent the diversity of the U.S. PRRSV-2 population since whole genome sequencing (WGS) is not a routine practice for disease surveillance because of its cost and availability. In addition, viruses with atypical clinical presentations in the field

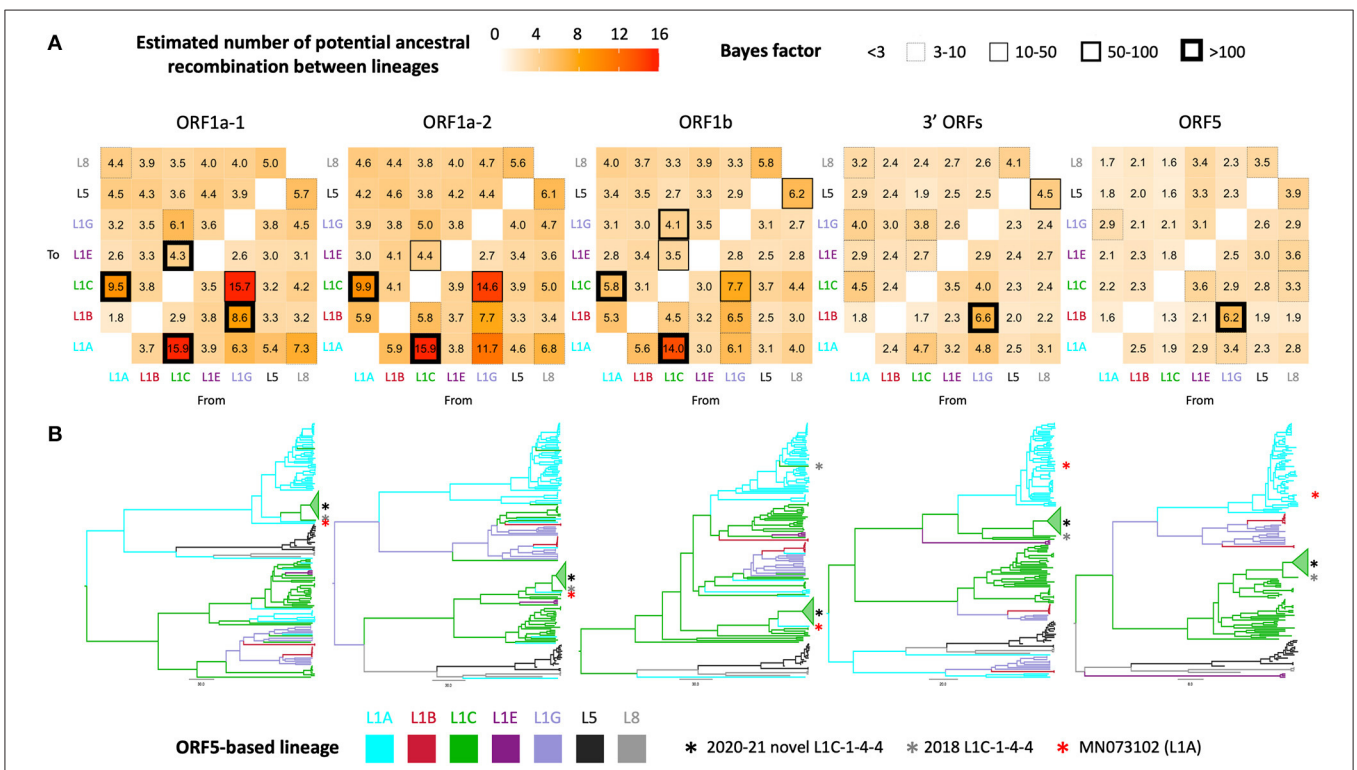


FIGURE 2 | Discrete trait analysis of PRRSV-2 lineage/sub-lineage recombination. **(A)** Heat map showing number of potential ancestral recombination between lineages/sub-lineages of each genomic fragment estimated from the trait transitions. Cell border thickness represents Bayes factor (BF) support for each recombination. **(B)** Bayesian MCC trees colored by ancestral ORF5-based lineage or sub-lineage. Asterisks locate the phylogenetic position of taxa of interest.

are more likely to undergo WGS. Thus, our recombination analysis only suggests the most likely parents or close relatives of the novel L1C-1-4-4 from amongst published sequences, which itself may be biased. In fact, the recombination detection was affected by data availability, as evidenced by several unknown parents of the novel L1C recombinant. Second, a fully recombination-free fragment, which is an ideal input for phylogenetic analysis, does not exist in the alignment because breakpoints are distributed across the genome. We alternatively used WGS-fragments with low frequencies of recombination to avoid recombination that may confound the genealogical tree. Genomic positions of such fragments nicely fit with three main protein coding regions of PRRSV-2 and other nidoviruses: ORF1a, ORF1b, and the nested set of multiple ORFs at the 3'-terminal (3' ORFs) (43). Last, the novel L1C-1-4-4 variant is defined by ORF5 genetic relatedness rather than clinical manifestation, and comparable data quantifying clinical aspects of disease were not available across data sources. Thus, an association between L1C-1-4-4's virulence and its evolution/recombination cannot be concluded from our study.

The inferred number of putative recombination events (trait transitions) from the discrete trait analysis reflect inter- and intra-lineage recombination between ORF5 and other genomic regions (i.e., ORF5 lineage was used as the discrete trait). From this analysis, we observe that recombination between lineages was rare, though this may be an artifact of the fact that the majority of included sequences belonged to a single lineage. However, recombination between sub-lineages within lineage one are more frequent, though still relatively uncommon. This corresponds to the mechanism of RNA recombination whereby the RNA polymerase is prone to switch from one RNA template to another that has a similar nucleotide sequence (7).

Additionally, recombination requires co-infection of the same cell, and viral prevalence will influence the likelihood that an animal is co-infected with two distinct viruses simultaneously. The prevalence of sub-lineages is temporally variable (12), which likely shapes opportunities for co-infection. Sub-lineages L1A, 1C, and 1H had the highest effective viral population sizes at the approximate tMRCA of the novel variant (12). Thus, the ancestor of the novel L1C-1-4-4 variant appears to have acquired each genomic portion from divergent viral subpopulations that were prevalent at the time. Recombination scanning along with phylogenetic tree analysis suggests that the majority of the 2020–2021 novel L1C-1-4-4 genomic fragments still derived from L1C viruses, while the proximal part of ORF1a, mostly nsp2, is genetically closer to viruses whose ORF5 is classified as L1A rather than L1C. This evidence coupled with the fact that nsp2 is the most variable gene in PRRSV-2 genome (44) explains why the novel L1C viruses clustered with viruses classified as L1A at the ORF5 level in the WGS tree in previous studies (10, 15).

The 2018 L1C-1-4-4 sample was included in this study because it carries an ORF5 gene closely related to the sequences associated with the 2020–2021 epidemic and was recovered from the same geographic area. However, some genomic parts as well as real world outbreak circumstances differ from the 2020–2021 epidemic. To our knowledge, there was no widespread PRRS outbreaks or heightened concern across the industry in connection with the 2018 virus, though anecdotally, field veterinarians noted that this particular virus transmitted readily between farms belonging to the same company and was challenging to control. The differential recombination profiles between the 2018 and 2020-21 L1C-1-4-4 viruses suggested by the RDP5 analysis were consistent with more robust phylogenetic analyses, which indicate that recombinant parents and phylogenetic position of the 2018 virus's ORF1b are different from the 2020–2021 sequences. Altogether, we hypothesize that both diverged in 2017 from the same recombinant ancestor that had a L1A-like ORF1a-1 fragment. The 2018 virus appears to be a result of an additional recombination event that appeared to leave very few progenies in our dataset. Other descendants kept evolving with or without recombination until they reached optimal fitness or a tipping point for exponential growth and became the 2020–2021 variant that is associated with the current outbreak.

An assessment of whether the acquisition of different WGS-fragments through recombination had a viral fitness benefit that allowed this variant to spread widely is beyond our limited understanding of the genetic determinants of pathogenicity and antigenicity. Therefore, we do not know the extent to which recombination contributed to the emergence or atypical clinical presentation of this virus. A study on SARS-CoV-2, a distant relative to PRRSV-2 in the same Nidovirales order, suggests the possibility that multi-strain recombination strengthens virulence (45). For PRRSV-2, all four genomic fragments we analyzed harbor at least one virulence-related gene. Mutations in nsp2, a part of both ORF1a-1 and ORF1a-2 fragments, are associated with target cell tropism of PRRSV-2 (46) and high fever in the host (47). RNA-dependent RNA polymerase (RdRp), a crucial component determining virus replication efficiency and pathogenicity (48), is encoded by nsp9 in ORF1b. Most of the 3'-terminal ORFs are transcribed and translated into the virus structural glycoprotein that directly interacts with either the target cell or host immune response (49, 50). Hypothetically, being able to rapidly shift antigenic phenotype through recombination may potentially confer a fitness advantage if it allows the virus to better evade population immunity. Genetic change in one of these genomic parts might be a key success of the novel L1C-1-4-4 variant but would need to be investigated by experimental studies such as targeted mutagenesis. However, our analysis better quantifies the contribution of recombination to PRRSV-2 genetic diversity and evolution, and points to the role of co-circulation of multiple variants within the same farm that may create conditions for recombination and selection for traits beneficial to the virus.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

NP and KV conceived and designed the study. NP performed data analysis and wrote the first draft of the manuscript. KV supervised the plan and findings of this work. CC and AR provided data for the analysis. NP, KV, MK, IP, and DM interpreted the results. All authors contributed to manuscript revision and approved the submitted version.

FUNDING

Funding was provided by the joint NIFA-NSF-NIH Ecology and Evolution of Infectious Disease award 2019-67015-29918. This work was also supported by the University of Minnesota College of Veterinary Medicine Signature Programs, Grant Number MIN-62-133. NP was supported by the Royal Thai Government Scholarship. This project was partially funded by the University of Minnesota Swine Disease Eradication Center (SDEC) and the Swine Health Information Center (SHIC) as the funding agency for MSHMP.

ACKNOWLEDGMENTS

Authors would like to thank the University of Minnesota Veterinary Diagnostic Laboratory, the Morrison Swine Health Monitoring Project (MSHMP) participating pig production systems and swine practitioners sharing samples, sequences and providing farm specific information. We would also like to thank (list the EEID team here that are not part of this manuscript) for constructive discussions related to this analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.846904/full#supplementary-material

REFERENCES

1. Holtkamp DJ, Kliebenstein JB, Neumann EJ, Zimmerman JJ, Rotto HF, Yoder TK, et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J Swine Heal Prod.* (2013) 21:72–84. doi: 10.31274/ans_air-180814-28
2. The OIE AD HOC group on porcine reproductive respiratory syndrome. PRRS: The disease, its diagnosis, prevention and control. Group (2008). Available online at: https://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/PRRS_guide_web_bulletin.pdf (accessed October 15, 2021).
3. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, et al. Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2021). *Arch Virol.* (2021) 166:2633–48. doi: 10.1007/s00705-021-05156-1
4. Ruedas-Torres I, Rodríguez-Gómez IM, Sánchez-Carvajal JM, Larenas-Muñoz F, Pallarés FJ, Carrasco L, et al. The jigsaw of PRRSV virulence. *Vet Microbiol.* (2021) 260:109168. doi: 10.1016/j.vetmic.2021.109168
5. Hanada K, Suzuki Y, Nakane T, Hirose O, Gojobori T. The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol Biol Evol.* (2005) 22:1024–31. doi: 10.1093/molbev/msi089
6. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* (2018) 16:1–6. doi: 10.1371/journal.pbio.3000003
7. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol.* (2011) 9:617–26. doi: 10.1038/nrmicro2614
8. Zhao H, Han Q, Zhang L, Zhang Z, Wu Y, Shen H, et al. Emergence of mosaic recombinant strains potentially associated with vaccine JXA1-R and predominant circulating strains of porcine reproductive and respiratory syndrome virus in different provinces of China. *Virol J.* (2017) 14:67. doi: 10.1186/s12985-017-0735-3
9. Anping Wang, Qi Chen LW, Darin Madson, Karen Harmon, Phillip Gauger, Jianqiang Zhang GL. Recombination between vaccine and field strains of porcine reproductive and respiratory syndrome virus. *Emerg Infect Dis.* (2019) 25:2335–7. doi: 10.3201/eid2512.191111
10. Kikuti M, Paploski IAD, Pamornchainavakul N, Picasso-Risso C, Schwartz M, Yeske P, et al. Emergence of a new lineage 1C variant of porcine reproductive and respiratory syndrome virus 2 in the United States. *Front Vet Sci.* (2021) 8:752938. doi: 10.3389/fvets.2021.752938
11. Kikuti M, Geary E, Picasso-Risso C, Medrano M, Corzo C. Updated epidemiological curve of cases associated with the new Lineage 1C RFLP1-4-4 variant. Saint Paul, MN: Morrison Swine Health Monitoring Project (2021). Available online at: https://vetmed.umn.edu/sites/vetmed.umn.edu/files/shmp_2021l22.02_updated_epi_curve_for_1-4-4_variant-_science_page.pdf (accessed November 12, 2021).
12. Paploski IAD, Corzo C, Rovira A, Murtaugh MP, Sanhueza JM, Vilalta C, et al. Temporal dynamics of co-circulating lineages of porcine reproductive and respiratory syndrome virus. *Front Microbiol.* (2019) 10:2486. doi: 10.3389/fmicb.2019.02486
13. Wesley RD, Mengeling WL, Lager KM, Clouser DF, Landgraf JG, Frey ML. Differentiation of a porcine reproductive and respiratory syndrome virus vaccine strain from North American field strains by restriction fragment length polymorphism analysis of ORF 5. *J Vet Diagnostic Investig.* (1998) 10:140–4. doi: 10.1177/104063879801000204
14. Shi M, Lam TT-Y, Hon C-C, Murtaugh MP, Davies PR, Hui RK-H, et al. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. *J Virol.* (2010) 84:8700–11. doi: 10.1128/JVI.02551-09
15. Schroeder DC, Odogwu NM, Kevill J, Yang M, Krishna VD, Kikuti M, et al. Phylogenetically distinct near-complete genome sequences of porcine reproductive and respiratory syndrome virus type 2 variants from four distinct disease outbreaks at U.S. swine farms over the past 6 years. *Microbiol Resour Announc.* (2021) 10:e0026021. doi: 10.1128/MRA.00260-21
16. Paploski IAD, Pamornchainavakul N, Makau DN, Rovira A, Corzo CA, Schroeder DC, et al. Phylogenetic structure and sequential dominance of sub-lineages of prrsv type-2 lineage 1 in the United States. *Vaccines.* (2021) 9:608. doi: 10.3390/vaccines9060608
17. Zhang J, Zheng Y, Xia XQ, Chen Q, Bade SA, Yoon KJ, et al. High-throughput whole genome sequencing of Porcine reproductive and respiratory syndrome virus from cell culture materials and clinical specimens using next-generation sequencing technology. *J Vet Diagnostic Investig.* (2017) 29:41–50. doi: 10.1177/1040638716673404
18. Gagnon CA, Lalonde C, Provost C. Porcine reproductive and respiratory syndrome virus whole-genome sequencing efficacy with field clinical samples using a poly(A)-tail viral genome purification method. *J Vet Diagnostic Investig.* (2021) 33:216–26. doi: 10.1177/1040638720952411
19. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* (2002) 30:3059–66. doi: 10.1093/nar/gkf436
20. Yu G. seqcombo: Visualization Tool for Sequence Recombination and Reassortment. R package version 1.4.1 (2021). Available online at: 10.18129/B9.bioc.seqcombo

21. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* (2021) 7:veaa087. doi: 10.1093/ve/veaa087
22. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics.* (2000) 16:562–3. doi: 10.1093/bioinformatics/16.6.562
23. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology.* (1999) 265:218–25. doi: 10.1006/viro.1999.0056
24. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol.* (1992) 34:126–9. doi: 10.1007/BF00182389
25. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses.* (2005) 21:98–102. doi: 10.1089/aid.2005.21.98
26. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in rebombinant sequences. *Bioinformatics.* (2000) 16:573–82. doi: 10.1093/bioinformatics/16.7.573
27. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A.* (2001) 98:13757–62. doi: 10.1073/pnas.241370698
28. Lam HM, Ratmann O, Boni MF. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. *Mol Biol Evol.* (2018) 35:247–51. doi: 10.1093/molbev/msx263
29. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* (2015) 1:vev003. doi: 10.1093/ve/vev003
30. Brito B, Pauszek SJ, Hartwig EJ, Smoliga GR, Vu LT, Dong PV, et al. A traditional evolutionary history of foot-and-mouth disease viruses in Southeast Asia challenged by analyses of non-structural protein coding sequences. *Sci Rep*. (2018) 8:6472. doi: 10.1038/s41598-018-24870-6
31. Aiewsakun P, Pamornchainavakul N, Inchaisri C. Early origin and global colonisation of foot-and-mouth disease virus. *Sci Rep*. (2020) 10:15268. doi: 10.1038/s41598-020-72246-6
32. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* (2016) 44:W232–5. doi: 10.1093/nar/gkw256
33. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. (2016) 2:vew007. doi: 10.1093/ve/vew007
34. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* (2019) 35:526–8. doi: 10.1093/bioinformatics/bty633
35. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (2019). Available online at: https://www.r-project.org/
36. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. (2018) 4:vey016. doi: 10.1093/ve/vey016
37. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* (2006) 4:e88. doi: 10.1371/journal.pbio.0040088
38. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* (2008) 25:1459–71. doi: 10.1093/molbev/msn090
39. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: real-time tracking of pathogen evolution. *Bioinformatics.* (2018) 34:4121–3. doi: 10.1093/bioinformatics/bty407
40. Rambaut A. FigTree v. 1.4.4. (2018). Available online at: http://tree.bio.ed.ac.uk/software/figtree/
41. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol.* (2016) 33:2167–9. doi: 10.1093/molbev/msw082
42. Yu F, Yan Y, Shi M, Liu H-Z, Zhang H-L, Yang Y-B, et al. Phylogenetics, genomic recombination, and NSP2 polymorphic patterns of porcine reproductive and respiratory syndrome virus in China and the United States in 2014–2018. *J Virol.* (2020) 94:e01813–19. doi: 10.1128/JVI.01813-19
43. Saberi A, Gulyaeva AA, Brubacher JL, Newmark PA, Gorbalenya AE. A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog.* (2018) 14:e1007314. doi: 10.1371/journal.ppat.1007314
44. Yoshii M, Okinaga T, Miyazaki A, Kato K, Ikeda H, Tsunemitsu H. Genetic polymorphism of the nsp2 gene in North American type-Porcine reproductive and respiratory syndrome virus. *Arch Virol.* (2008) 153:1323–34. doi: 10.1007/s00705-008-0098-6
45. Haddad D, John SE, Mohammad A, Hammad MM, Hebbar P, Channanath A, et al. SARS-CoV-2: possible recombination and emergence of potentially more virulent strains. *PLoS ONE.* (2021) 16:e0251368. doi: 10.1371/journal.pone.0251368
46. Song J, Gao P, Kong C, Zhou L, Ge X, Guo X, et al. The nsp2 hypervariable region of porcine reproductive and respiratory syndrome virus strain JXwn06 is associated with viral cellular tropism to primary porcine alveolar macrophages. *J Virol.* (2019) 93:e01436–19. doi: 10.1128/JVI.01436-19
47. Du L, Wang H, Liu F, Wei Z, Weng C, Tang J, et al. NSP2 is important for highly pathogenic porcine reproductive and respiratory syndrome virus to trigger high fever-related COX-2-PGE2 pathway in pigs. *Front Immunol.* (2021) 12:657071. doi: 10.3389/fimmu.2021.657071
48. Zhao K, Gao J-C, Xiong J-Y, Guo J-C, Yang Y-B, Jiang C-G, et al. Two residues in NSP9 contribute to the enhanced replication and pathogenicity of highly pathogenic porcine reproductive and respiratory syndrome virus. *J Virol.* (2018) 92:e02209–17. doi: 10.1128/JVI.02209-17
49. Das PB, Vu HLX, Dinh PX, Cooney JL, Kwon B, Osorio FA, et al. Glycosylation of minor envelope glycoproteins of porcine reproductive and respiratory syndrome virus in infectious virus recovery, receptor interaction, and immune response. *Virology.* (2011) 410:385–94. doi: 10.1016/j.virol.2010.12.002
50. Van Breedam W, Van Gorp H, Zhang JQ, Crocker PR, Delputte PL, Nauwynck HJ. The M/GP5 glycoprotein complex of porcine reproductive and respiratory syndrome virus binds the sialoadhesin receptor in a sialic acid-dependent manner. *PLoS Pathog.* (2010) 6:e1000730. doi: 10.1371/journal.ppat.1000730

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fvets.2022.906944

Evolution, Transmission, and Pathogenicity of High Pathogenicity Avian Influenza Virus A (H5N8) Clade 2.3.4.4, South Korea, 2014–2016

Yoon-Gi Baek[1], Yu-Na Lee[1], Yu-Ri Park[1], David H. Chung[2], Jung-Hoon Kwon[3], Young-Jae Si[1], Gyeong-Beom Heo[1], Youn-Jeong Lee[1], Dong-Hun Lee[4] and Eun-Kyoung Lee[1]**

[1] Avian Influenza Research & Diagnostic Division, Animal and Plant Quarantine Agency, Gimcheon-si, South Korea, [2] Department of Pathobiology and Veterinary Science, University of Connecticut, Storrs, CT, United States, [3] College of Veterinary Medicine, Kyungpook National University, Daegu, South Korea, [4] College of Veterinary Medicine, Konkuk University, Seoul, South Korea

OPEN ACCESS

Edited by:
Iryna Goraichuk,
National Scientific Center, Institute of Experimental and Clinical Veterinary Medicine, Ukraine

Reviewed by:
Ruyi Gao,
Yangzhou University, China
Daxin Peng,
Yangzhou University, China

***Correspondence:**
Dong-Hun Lee
donghunlee@konkuk.ac.kr
Eun-Kyoung Lee
ensenble@korea.kr

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 29 March 2022
Accepted: 26 May 2022
Published: 21 June 2022

Citation:
Baek Y-G, Lee Y-N, Park Y-R, Chung DH, Kwon J-H, Si Y-J, Heo G-B, Lee Y-J, Lee D-H and Lee E-K (2022) Evolution, Transmission, and Pathogenicity of High Pathogenicity Avian Influenza Virus A (H5N8) Clade 2.3.4.4, South Korea, 2014–2016. Front. Vet. Sci. 9:906944. doi: 10.3389/fvets.2022.906944

During 2014–2016, clade 2.3.4.4 H5N8 high pathogenicity avian influenza virus (HPAIV) caused the largest known avian influenza epidemic in South Korea. Based on data from earlier H5N8 outbreaks, primitive H5N8 virus in South Korea was classified into five subgroups: C1, C2, C3, C4, and C5. The present study investigated the pathogenic and molecular epidemiologic characteristics of H5N8 viruses obtained from 388 cases of poultry farms and 85 cases of wild bird infections in South Korea during 2014–2016. Representative viruses of subgroups C1, C2, and C4 showed significant pathobiological differences in specific pathogen-free (SPF) chickens, with the H1731 (C1) virus showing substantially lower infectivity, transmissibility, and pathogenicity than the H2102 (C2) and H1924 (C4) viruses. Full genome sequence analysis showed the number of mutations that significantly increased in domestic duck-origin H5N8 HPAIVs compared to the viruses from gallinaceous poultry. These differences may have been due to the long-term circulation of viruses in domestic duck farms. The same mutations, at positions 219 and 757 of PB1, independently evolving in the C0, C1, and C2 subgroups may have been positively selected, resulting in convergent evolution at the amino acid level. Bayesian discrete trait phylodynamic analysis (DTA) indicated multiple introductions of H5N8 HPAIV from wild birds into domestic poultry in various regions in South Korea. Following initial viral introduction into domestic duck farms in the western part of Korea, domestic ducks played a major role in viral transmission and maintenance. These findings highlight the need for continued genomic surveillance and pathobiological characterization of HPAIV in birds. Enhanced biosecurity in poultry farms should be implemented to prevent the introduction, maintenance, and spread of HPAIV.

Keywords: HPAI, H5N8, pathogenesis, selection pressure, phylodynamic analysis

INTRODUCTION

The H5 subtype of high pathogenicity avian influenza viruses (HPAIVs) belonging to the Goose/Guangdong (Gs/Gd) lineage have been circulating since 1996 and evolving into genetically diverse clades and subclades (1–3). The H5N8 subtype of Gs/Gd HPAIV belonging to the subclade 2.3.4.4 was initially identified in domestic ducks from eastern China in 2010 (4) and novel reassortant H5N8 HPAIVs were detected in live poultry markets in eastern China in late 2013. Subsequently, reassortant H5N8 HPAIVs were introduced into South Korea and Japan in early 2014 (5–7). In late 2014, several countries in Europe and North America experienced an invasion of HPAI H5Nx viruses (8–11). The transcontinental spread of these viruses was associated with dissemination in multiple directions by migratory wild birds (12).

In South Korea, four outbreaks of H5N1 viruses have been recorded, in 2003–2004 (clade 2.5), 2006–2007 (clade 2.2), 2008 (clade 2.3.2), and 2010–2011 (clade 2.3.2.1), with the lengths of these outbreaks ranging from 42 (2008) to 139 (2010–2011) days (13). The fifth HPAI outbreak started in January 2014 and lasted for 28 months (14). During this outbreak, South Korea experienced four waves of H5N8 outbreaks with 393 cases of poultry farms and 58 cases of wild bird infections. Moreover, two genetically distinct groups of clade 2.3.4.4 H5N8 viruses were identified: Buan2-like and Gochang1-like. The Buan2-like H5N8 viruses predominated during the epidemic and further diverged into five distinct subgroups: C1, C2, C3, C4, and C5 (8, 14). After the first wave, from January to July 2014, subgroups C1 and C5 were detected mainly in poultry farms in South Korea, whereas the other three subgroups were detected in wild birds and spread over long distances via wild birds including subgroups C2 in South Korea and Japan, C3 in North America and Japan, and C4 in Europe, Japan, and South Korea (8–11). The C2 and C4 subgroups were maintained in wild birds and re-introduced into South Korea during the fall migration season of wild birds in 2014.

Little is currently understood about the evolution and spread of H5N8 HPAIV during the 2014–2016 outbreak in South Korea. Previous studies investigating the origin and transmission of H5N8 HPAIV in South Korea included limited datasets, such as partial genome sequences of early isolates or selected viruses (8, 14, 15). The present study analyzed the complete genome sequences of H5N8 HPAIVs identified in 388 (of 393 total cases) poultry farms and 53 (of 58 total cases) wild bird cases during the 2014–2016 outbreak in South Korea. The pathogenic and molecular epidemiologic characteristics of the viruses were determined to identify the underlying causes of the prolonged outbreak and to reconstruct the evolutionary and transmission dynamics of H5N8 HPAIVs in South Korea.

MATERIALS AND METHODS

Samples

Fecal samples, bird carcasses, and oropharyngeal and cloacal swabs were collected from poultry farms and wild bird habitats in South Korea through active and passive surveillance from January 2014 to June 2016. Viruses were isolated by inoculating samples into 9–11-day-old specific-pathogen-free (SPF) embryonated chicken eggs after incubation for 4 days at 37°C.

Genome Sequencing

Viral RNA was extracted from the infected allantoic fluids of SPF chicken eggs using Maxwell® RSC simplyRNA Tissue Kits (Promega, Madison, WI, USA). Complementary DNA was synthesized by reverse transcription reaction using Superscript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) and eight segments of each virus were amplified by PCR using AccuPrime Pfx DNA Polymerase (Invitrogen, Carlsbad, CA, USA), as described (16). The complete genome sequences of 441 H5N8 HPAIVs isolated from 388 domestic poultry farms and 53 wild bird habitats were determined by next-generation sequencing with the Illumina Miseq platform using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Data were analyzed by CLC Genomics Workbench (Qiagen, Valencia, CA, USA). Nucleotide sequences were deposited under accession nos. EPI_ISL_157609-EPI_ISL_410213.

Genetic Subgrouping by Maximum-Likelihood Phylogenetic Analysis

The complete genomes of the 441 H5N8 HPAIVs sequenced in this study were subjected to phylogenetic analyses, along with 32 previously determined whole genome sequences of clade 2.3.4.4A H5N8 HPAIV identified in wild waterfowl in South Korea that had been deposited in the GISAID EpiFlu database (http://www.gisaid.org). Furthermore, our data set was added 20 sequences of the H5N8 viruses isolated from poultry and wild waterfowl in Japan. Maximum-likelihood (ML) phylogenies of the 493 concatenated complete genome sequences were generated using the Randomized Axelerated Maximum Likelihood (RAxML 8.2.12) (17) method, involving the general time reversible model of nucleotide substitution and the Gamma model of among-site rate heterogeneity model with 1,000 bootstrap iterations. A genetic subgroup was defined as a monophyletic cluster of sequences with high bootstrap support (>70%).

Animal Experiments

The objective of this study was to evaluate the infectivity, transmissibility, and pathogenicity of the first detected viruses (index virus) of three subgroups (C1, C2, and C4 except for C5 with only four cases) of H5N8 HPAIVs in chickens. Viruses evaluated in chickens included A/broiler duck/Korea/H1731/2014 (subgroup C1, abbreviated H1731), A/mallard/Korea/H2102/2015 (subgroup C2, abbreviated H2102), and A/mallard/Korea/H1924/2014 (subgroup C4, abbreviated H1924). To determine the mean bird lethal dose (BLD_{50}), five 4-week-old SPF white leghorn chickens each were inoculated intranasally with serial 10-fold dilutions (10^3 to 10^6 mean egg infectious dose [EID_{50}]) of each virus. Viral transmissibility was evaluated by housing three virus-naïve chickens with the chickens challenged with 10^6 EID_{50}/0.1 ml of each virus 8 h after infection. The chickens were monitored

daily for clinical signs for 14 days. Oropharyngeal and cloacal swabs were collected at 1–7, 10, and 14 days post-infection (dpi). Serum was collected from surviving chickens at 14 dpi to determine seroconversion by hemagglutination inhibition test (HI) (18). The tissue tropism and dissemination of the viruses were evaluated by intranasally inoculating three chickens with 10^6 EID_{50}/0.1 mL of each virus. These chickens were necropsied at 3 dpi and 12 organs (trachea, thymus, liver, pectoral major muscle, spleen, proventriculus, pancreas, cecal tonsil, lung, kidney, heart, and brain) were collected for viral titration. Virus titers were measured in DF-1 cells incubated with 10% tissue homogenates. Experiments in animals were reviewed and approved by the Institutional Animal Care and Use Committee of the Animal and Plant Quarantine Agency (APQA) (approval no: 2018-398 and 2018-412). All procedures were carried out in a biosafety level three facility at the APQA.

Shannon Entropy and Selective Pressure Analysis

Genomic variability between H5N8 HPAIVs derived from ducks and other hosts was evaluated by genome-wide Shannon entropy analysis using the entropy tool in the HIV sequence database (https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html). Shannon entropy differences (Hdiff) were calculated for each gene segment, with differences in variability considered significant when $p < 0.05$.

Gene- and site-specific selection pressures for all segments of the H5N8 viruses isolated in South Korea during 2014–2016 were measured as the ratio of non-synonymous substitutions (dN) to synonymous substitutions (dS) per site. These analyses were performed using a combination of four methods: single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), fast unconstrained Bayesian approximation (FUBAR), and a mixed effects model of evolution (MEME) available at the Datamonkey online version of the Hy-Phy package (19–21). The selective pressures of each gene of the C0, C1, C2, and C4 subtypes were compared and statistically significant positive selection of amino acid sites was estimated (SLAC, FEL, MEME, p-value < 0.1; FUBAR, posterior probability >0.9).

Bayesian Discrete Trait Phylodynamic Analysis

Hemagglutinin (HA) gene sequences were subjected to Bayesian discrete trait phylodynamic analysis (DTA) to investigate virus transmission history among regions and host species. These analyses included all 62 available complete HA genome sequences of clade 2.3.4.4A H5N8 HPAIVs identified in wild waterfowl in East Asia during 2013–2016 that had been downloaded from the GISAID EpiFlu database (http://www.gisaid.org). An ancestral state reconstruction approach and Bayesian stochastic search variable selection (BSSVS) were utilized to determine the most probable transmission history. Geographically discrete nominal categories were assigned at the province level for viruses collected in Korea; these included "Wild Bird" (WB) (ntax = 115 sequences), "Jeollanam-do" (JN) (ntax = 115 sequences), "Jeollabuk-do" (JB) (ntax = 63 sequences), "Gyeonggi-do" (GG) (ntax = 77 sequences), 'Gyeongsang-do' (GS) (ntax = 22 sequences), "Chungcheongnam-do" (CN) (ntax = 43 sequences), "Chungcheongbuk-do" (CB) (ntax = 69 sequences), "Gangwon-do" (GW) (ntax = 2 sequences), and "Jeju-do" (JJ) (ntax= 1 sequence). To account for potential sampling biases, sequences were subsampled to preserve diversity relative to geographic locations and host species. The WB, JN, JB, GG, and CB datasets were subsampled based on nucleotide sequence identities of 99.98, 99.9, 99.95, 99.95, and 99.95%, respectively, using CD-HIT (14). The GW and JJ datasets were excluded from the analysis because of their small sample sizes. The subsampled DTA dataset consisted of 328 sequences, including 62 WB, 57 JN, 45 JB, 55 GG, 22 GS, 43 CN, and 44 CB sequences.

DTA was also performed to determine the transmission dynamics between host species, including "Wild Bird" (ntax = 115 sequences), "Duck" (ntax = 294 sequences), "Chicken" (ntax = 71 sequences), and "minor poultry" (ntax =27 sequences). The Wild Bird, Duck, and Chicken datasets were subsampled using CD-HIT, based on nucleotide identities of 99.95, 99.87, and 100%, respectively (22). This subsampled DTA dataset consisted of 204 sequences, including 65 wild bird, 61 duck, 51 chicken, and 27 minor poultry sequences. For both analyses, Bayesian relaxed clock phylogenies were reconstructed using BEAST version 1.10.4 (23). The Hasegawa, Kishino, and Yano nucleotide substitution model with an uncorrelated log-normal distribution relaxed-clock method was used, along with a Gaussian Markov Random Field (GMRF) Bayesian skyride coalescent prior (24). The best supported viral transitions between discrete categories was identified by calculating the Bayes factor (BF) using SPREAD version 1.0.7 (25). A transition was defined as significant when BF >3 and posterior probability >0.5. The rate and number of viral transitions between discrete categories (Markov jump) and the time spent in a state between transitions (Markov reward) were estimated using stochastic mapping (26). The Markov Chain Monte Carlo (MCMC) was run in parallel for three chains, each with 100 million steps and samples across chains combined after 10% burn-in. The parameters, each of which had effective sample sizes >200, were analyzed with TRACER v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). A maximum clade credibility (MCC) tree was generated using TreeAnnotator and visualized using FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The number of transition events and the discrete state proportion over time in monthly intervals were calculated based on analyses of posterior trees using the program PACT (http://www.trevorbedford.com/pact).

The impact of viral source/sink sample sizes on Bayesian discrete inference datasets was determined using generalized linear model (GLM) analysis as an extension of DTA (23). GLM calculates the virus migration rates as a linear combination of coefficients and coefficient indicators, and predictors. The sample size of each discrete state was used to predict the viral transition rates of two separate DTAs. The coefficient indicator and coefficient describe if and to what degree each predictor contributes to the migration rate.

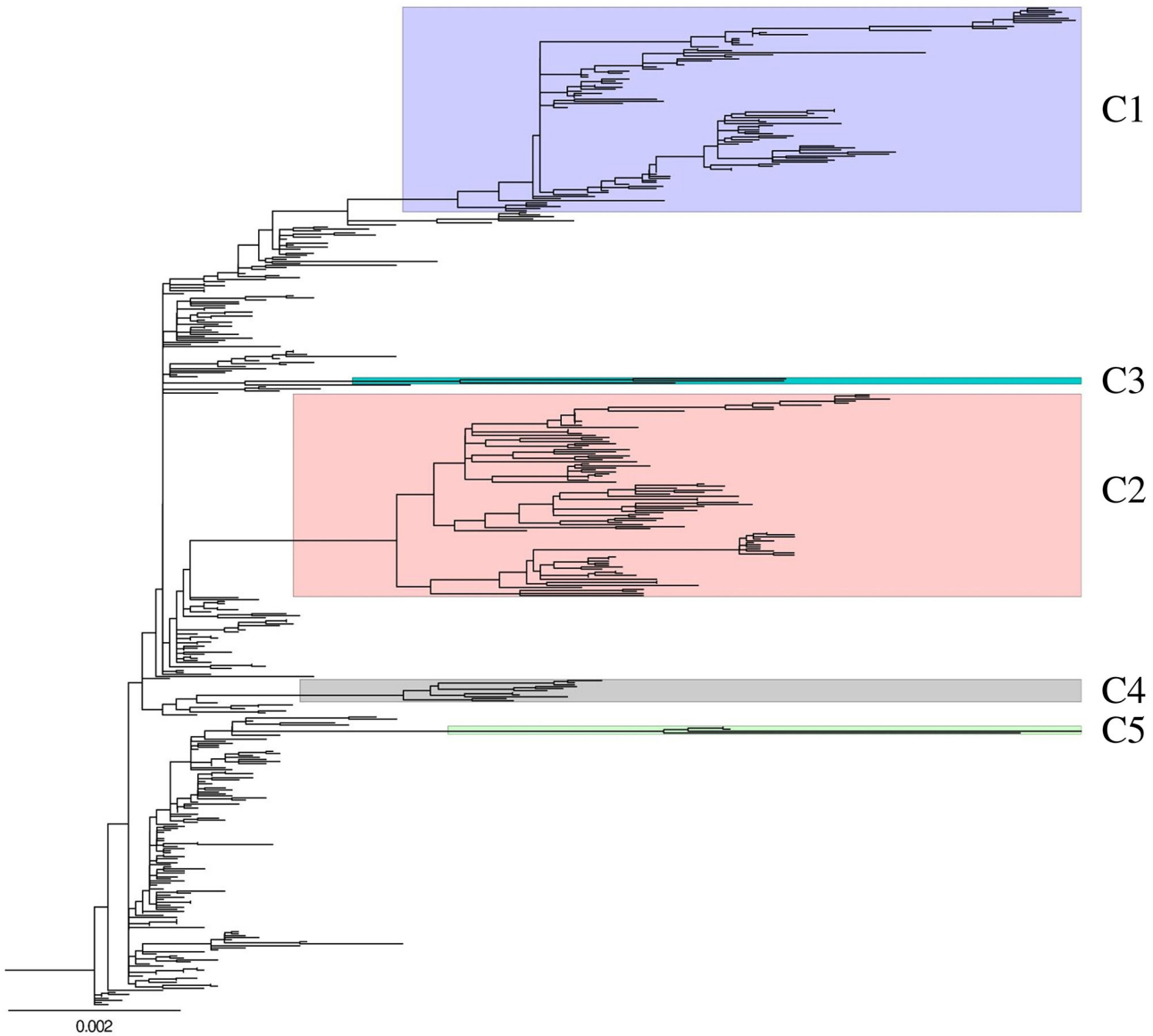


FIGURE 1 | Genetic characterization of H5N8 high pathogenicity avian influenza isolated in South Korea during 2014–2016. Maximum-likelihood phylogenetic tree of concatenated complete genome sequences. Whole genome sequences of 493 H5N8 high pathogenicity avian influenza viruses collected from 388 birds on domestic poultry farms and 105 wild birds between January 2014 and April 2016. Labels and highlights indicate genetic subgroups. C3 cluster had strains isolated from wild bird in only Japan. Non-labeled and highlighted viruses are belonged to subgroup C0. Phylogenetic analyses were performed with RAxML version 8.2.12 using a gamma distribution and a general time-reversible model.

RESULTS

ML Phylogenetic Analysis

A ML phylogenetic tree of concatenated whole genome sequences of 493 H5N8 HPAIVs isolated in South Korea and Japan during 2014–2016 was generated to determine the genetic relationship among isolates (**Figure 1**). The viruses isolated from wild bird and poultry in South Korea were classified into multiple genetic clusters, including primitive (subgroup C0) and further classified (subgroups C1, C2, C4, and C5) viruses. Viruses of subgroup C3, which were detected in North America, Taiwan, and Japan during 2014–2015, were not detected in South Korea.

Distribution in Geographic Region and Host

Of the 273 viruses belonging to subgroup C0 detected in South Korea from January to June 2014, 58 were detected in wild birds and 215 in domestic poultry, including in 162 ducks,

TABLE 1 | Pathogenicity and transmissibility of H1731, H2102, and H1924 in SPF chickens.

Virus	Group (No. of chickens)	BLD_{50} ($Log_{10}EID_{50}$/0.1 ml)	MDT (days)[a]	Mortality (%)[b]			
				10^3	10^4	10^5	10^6
H1731	Inoculated(5)	4.4	5.0	0/5(0)	1/5(20)	5/5(100)	5/5(100)
	Contact(3)	–	8.0	–	–	–	1/3(33)
H2102	Inoculated(5)	4.5	3.8	0/5(0)	0/5(0)	5/5(100)	5/5(100)
	Contact(3)	–	7.9	–	–	–	3/3(100)
H1924	Inoculated(5)	3.5	3.7	0/5(0)	5/5(100)	5/5(100)	5/5(100)
	Contact(3)	–	6.1	–	–	–	3/3(100)

[a]MDT was measured in groups inoculated with $10^6 EID_{50}$/0.1 ml of H1731, H2102, and H1924 viruses; [b]SPF chickens were intranasally inoculated with serial 10-fold dilutions, ranging from 10^3 to 10^6 EID_{50}/0.1 ml of H1731, H2102, and H1924 viruses: BLD, Bird infectious dose; MDT, mean death time.

42 chickens, and 11 minor poultry (**Supplementary Figure 1**). These 273 viruses included viruses from JL ($n = 82$), CC ($n = 72$), GG ($n = 41$), GS ($n = 17$), GW ($n = 2$), and JJ ($n = 1$) provinces (**Supplementary Figure 2**). Subgroup C1 ($n = 101$) was the predominant H5N8 HPAIV isolated from poultry from September 2014 to November 2015, especially in domestic ducks ($n = 74$), with 57, 25, 12, and 3 obtained from ducks in JL, CC, GG, and GS provinces, respectively (**Supplementary Figures 1**, **2**). Four viruses belonging to subgroup C5 were detected from May 2015 to April 2016, two in wild birds and two in domestic ducks, with all detected in GG province. Of the 87 subgroup C2 viruses detected in South Korea from December 2014 to April 2015, 18 were detected in wild birds and 69 in domestic poultry, including in 48 ducks, 19 chickens, and two minor poultry, with all detected in the western part of the country. Eight subgroup C4 viruses were detected from November 2014 to January 2015, three in wild birds and five in domestic ducks, including three in JL and two in GS province.

Pathogenicity and Transmissibility in Chickens

To investigate the differences in pathobiology among viral subgroups, chickens were intranasally inoculated with H1731, H2102, or H1924 virus, defined as the index viruses of the C1, C2, and C4 subgroups, respectively (**Table 1**). All chickens inoculated with 10^6 EID_{50}/0.1 ml of the H1731 virus died within 6 days with a mean death time (MDT) of 5 days. By contrast, all chickens inoculated with 10^6 EID_{50}/0.1 ml of the H2102 and H1924 viruses died in less than 4 days with MDTs of 3.8 and 3.7 days, respectively. The BLD_{50} of the H1731, H2102, and H1924 viruses were $10^{4.4}$, $10^{4.5}$, and $10^{3.5}$ EID_{50}/0.1 ml, respectively. Viral shedding of H1731 into the oropharynx and cloaca was detected from 2 to 6 dpi, whereas viral shedding of H2102 and H1924 were detected from 1 to 4 dpi (**Figures 2A,B**). The peak virus titers in oropharyngeal swabs were lower in H1731-inoculated chickens (3.4 $\log_{10}$ median tissue culture infectious dose [$TCID_{50}$/0.1 ml]) than in chickens inoculated with H2102 (5.25 $\log_{10}TCID_{50}$/0.1 ml) and H1924 (4.3 $\log_{10}TCID_{50}$/0.1 ml). None of the surviving SPF chickens shed virus or showed an HI response.

Analysis of viral transmission showed that one of the three chickens cohoused with the H1731-inoculated chickens died and shed virus (1.3–3.3 $\log_{10}TCID_{50}$/0.1 ml) into the oropharynx (**Figures 2C,D**). By contrast, all of the chickens cohoused with the H2102 and H1924-inoculated chickens died and shed virus into both the oropharynx (0.8–3.75 $\log_{10}TCID_{50}$/0.1 ml) and cloaca (1.0–2.5 $\log_{10}TCID_{50}$/0.1 ml). Chickens cohoused with the H1731- and H2102-inoculated chickens survived longer, with MDTs of 8 and 7.9 days, respectively, than chickens cohoused with H1924-inoculated chickens, with an MDT of 6.1 days.

All viral strains replicated in all collected tissues, including brain, heart, kidney, lung, cecal tonsil, pancreas, proventriculus, spleen, muscle, liver, thymus, and trachea (**Figure 3**). Viral titers in tissue samples were substantially lower in H1731-inoculated than in H2102- and H1924-inoculated chickens. In particular, the mean viral titers in the hearts of H1731-inoculated chickens were significantly lower than the viral titers in the hearts of H2102- and H1924-inoculated chickens ($p < 0.001$ each). In addition, viral titers in the kidney, lung, and cecal tonsil were significantly lower in H1731- than in H1924-inoculated chickens ($p < 0.05$ each), and viral titers in the thymus of H1731-inoculated chickens were significantly lower than titers in the thymus of chickens inoculated with H1924 ($p < 0.01$) and H2102 ($p < 0.05$) viruses.

Shannon Entropy Analysis and Selective Pressure

Calculations of Shannon entropy for all positions of the coding regions identified 53 dominant mutations in duck-derived H5N8 HPAIVs when compared with other host-derived H5N8 viruses (**Table 2**). H1731 (C1) virus had multiple mutations with significance, whereas Buan2 (C0), H2102 (C2), and H1924 (C4) viruses had no related mutation. Shannon entropy analysis of H1731 (subgroup C1) virus identified 18 amino acid substitutions (T184A, A442S, and D678G in PB2; V219I in PB1; V323I, V327G, and T618K in PA; S141P, S163N, D376N, Q455K, and A528V in HA; L23I, S69N, and T329I in NA; N80S and D171E in NS1; and M14K in NS2). Entropy analysis of the Buan2 (subgroup C0), H2102 (subgroup C2), and H1924 (subgroup C4) viruses, however, did not identify any amino acid substitutions. All of the mutations were founded in subgroup C1 and took

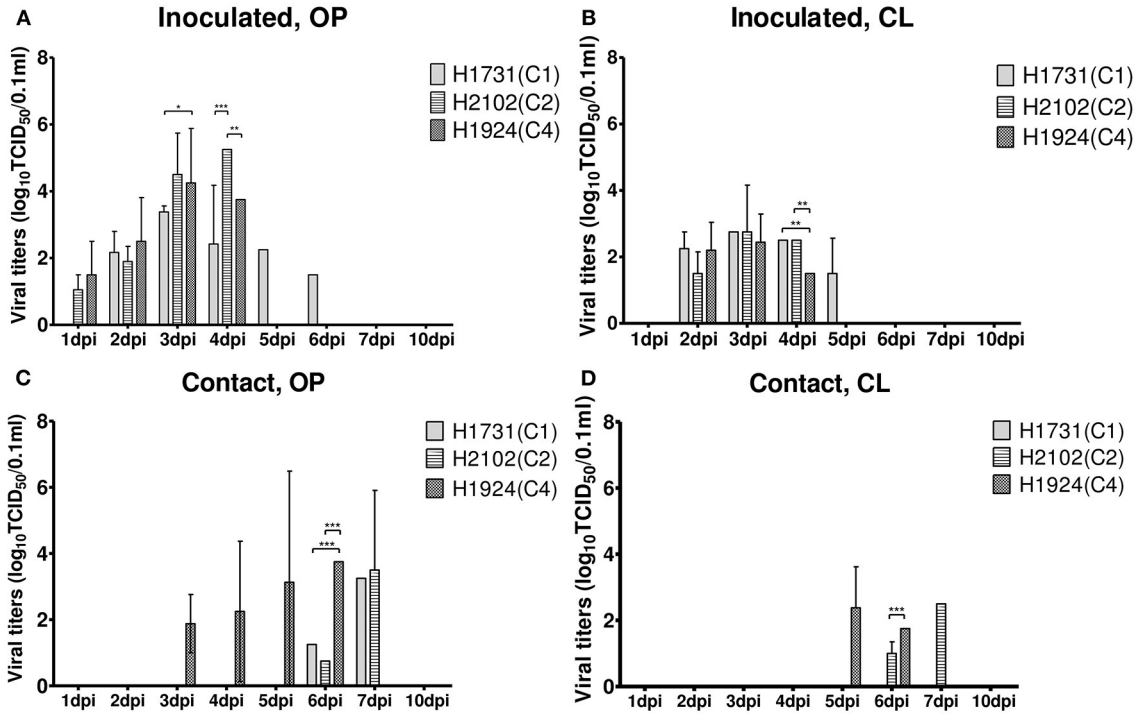


FIGURE 2 | Shedding of H1731, H1924, and H2102 viruses in SPF chickens. Virus was isolated from swab samples collected from **(A,B)** chickens intranasally inoculated with 10^6 EID_{50}/0.1 ml of virus and **(C,D)** from chickens cohoused with inoculated chickens. Viral titer was measured in DF-1 cells and shown as the mean ± standard deviation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. OP, oropharyngeal; CL, cloacal; SPF, specific pathogen-free; $TICD_{50}$, median tissue culture infectious dose.

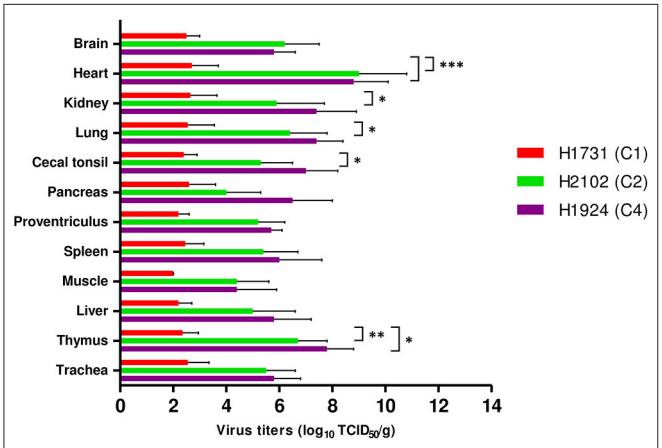


FIGURE 3 | Viral titers in tissues of chickens infected with H1731, H1924, and H2102 virus. Three SPF chickens were intranasally inoculated with 10^6 EID_{50}/0.1 mL of each virus and euthanized 3 days later. Tissue homogenates were inoculated into DF-1 cells. Viral titers were expressed as median tissue culture infectious dose ($TCID_{50}$). Data represent mean ± SD. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

up a high proportion (92.1–100%) except for A442S of PB2 (9.9%) compared with that of the other subgroups (0–17.8%) (**Supplementary Table 1**).

Analyses of the selection profiles of the H5N8 HPAIVs identified individual codons under positive selection pressure (**Supplementary Table 2**). Nine amino acids (aas) in the C0 subgroup, including aa 694 of PB2, aas 219 and 757 of PB1, aas 181 and 269 of HA [H5 numbering], aas 11 and 13 of M2, and aas 109 and 180 of NS1, were identified as being under positive selection. Eight aas in the C1 subgroup, including aa 184 and 559 of PB2, aas 219 and 757 of PB1, aa 113 of NA, aas 13 and 19 of M2, and aa 59 of NS1; five aas in the C2 subgroup, including aas 219 and 757 of PB1, aas 183 and 353 of NP, and aa 369 of NA; and two aas in the C4 subgroup, including aas 105 and 331 of NP, were also identified as being under positive selection pressure.

Bayesian Phylogenetic Analysis of the HA Gene

Based on the Bayesian phylogeny of HA gene, the time to most recent common ancestor (tMRCA) of all viruses identified in Korea was estimated to have arisen on July 2, 2013 (95% HPD: May 2 to August 11, 2013, with a posterior probability of 1), indicating that the H5N8 viruses of wild bird origin in South Korea most likely emerged during the breeding season of wild waterfowl in 2013 (**Figure 4**). The mean substitution rate of the HA gene was calculated as 7.12×10^{-3} substitutions/site/year (95% BCI: 5.53×10^{-3} to 8.80×10^{-3}), within the global range for AIV substitution rate (27).

TABLE 2 | Shannon entropy analysis between H5N8 highly pathogenic avian influenza viruses isolated from duck and other hosts.

Gene	nt position	nt change	nt proportion				*p*-value	Entropy difference between other hosts and duck	Amino acid change	Buan2 (C0)	H1731 (C1)	H2102 (C2)	H1924 (C4)
			Other hosts		Duck								
PB2	**550**	**A550G**	**A(154)**[a]	**G(37)**	**A(183)**	**G(110)**	**0**	**−0.17**	**T184A**	T	**A**	T	T
	1188	A1188T	A(189)	T(2)	A(275)	T(18)	0.01	−0.173	E396D	E	E	E	E
	1324	**G1324T**	**G(190)**	**T(1)**	**G(284)**	**T(9)**	**0.03**	**−0.105**	**A442S**	A	**S**	A	A
	1489	A1489C	A(169)	C(22)	A(219)	C(74)	0	−0.208	S497R	S	S	S	S
	2033	**A2033G**	**A(164)**	**G(27)**	**A(219)**	**G(74)**	**0**	**−0.158**	**D678G**	D	**G**	D	D
PB1	632	G632A	G(191)		G(287)	A(6)	0	−0.1	R211K	R	R	R	R
	655	**G655A**	**G(153)**	**A(38)**	**G(206)**	**A(87)**	**0**	**−0.109**	**V219I**	V	**I**	V	V
	1289	G1289A	G(191)		G(285)	A(8)	0.01	−0.125	R430K	R	R	R	R
	1439	A1439G	A(191)		A(286)	G(7)	0.03	−0.113	K480R	K	K	K	K
	1771	G1771A	G(191)		G(284)	A(9)	0	−0.137	V591I	V	V	V	V
	2207	A2207C	A(191)		A(287)	C(6)	0.05	−0.1	K736T	K	K	K	K
	2270	A2270G	A(190)	G(1)	A(281)	G(12)	0	−0.138	K757R	K	K	K	K
PA	176	A176T	A(191)		A(285)	T(8)	0.01	−0.125	K59I	K	K	K	K
	670	T670C	T(189)	C(2)	T(271)	C(22)	0	−0.231	S224P	S	S	S	S
	967	**G967A**	**G(150)**	**A(41)**	**G(198)**	**A(95)**	**0**	**−0.11**	**V323I**	V	**I**	V	V
	980	**T980G**	**T(190)**	**G(1)**	**T(284)**	**G(9)**	**0.03**	**−0.105**	**V327G**	V	**G**	V	V
	1010	C1010T	C(189)	T(2)	C(275)	T(18)	0.01	−0.173	A337V	A	A	A	A
	1076	A1076C	A(186)	C(5)	A(273)	C(20)	0.02	−0.128	N359T	N	N	N	N
	1159	G1159A	G(189)	A(2)	G(275)	A(18)	0.01	−0.173	V387I	V	V	V	V
	1853	**C1853A**	**C(165)**	**A(26)**	**C(219)**	**A(74)**	**0**	**−0.167**	**T618K**	T	**K**	T	T
	2047	C2047A	C(191)		C(288)	A(5)	0.05	−0.086	L683I	L	L	L	L
HA	**469**	**T469C**	**T(165)**	**C(26)**	**T(224)**	**C(69)**	**0**	**−0.148**	**S141P**	S	**P**	S	S
	515	C515T	C(189)	T(2)	C(280)	T(13)	0.01	−0.123	A156V	A	A	A	A
	536	**G536A**	**G(147)**	**A(44)**	**G(197)**	**A(96)**	**0**	**−0.093**	**S163N**	S	**N**	S	S
	610	A610G	A(189)	G(2)	A(278)	G(15)	0.01	−0.144	T188A	T	T	T	T
	633	C633A	C(191)		C(288)	A(4) T(1)	0.04	−0.095	D195E	D	D	D	D
	759	T759A	T(191)		T(284)	A(9)	0	−0.137	D236E	D	D	D	D
	865	C865A	C(191)		C(284)	A(9)	0	−0.137	H273N	H	H	H	H
	1171	**G1171A**	**G(151)**	**A(40)**	**G(197)**	**A(96)**	**0**	**−0.119**	**D376N**	D	**N**	D	D
	1408	**C1408A**	**C(164)**	**A(27)**	**C(219)**	**A(74)**	**0**	**−0.158**	**Q455K**	Q	**K**	Q	Q
	1456	T1456C	T(190)	C(1)	T(283)	C(10)	0.03	−0.116	Y471H	Y	Y	Y	Y
	1484	A1484G	A(189)	G(2)	A(278)	G(15)	0	−0.144	E480G	E	E	E	E
	1489	G1489A	G(179)	A(12)	G(250)	A(43)	0	−0.182	V482I	V	V	V	V
	1628	**C1628T**	**C(151)**	**T(40)**	**C(197)**	**T(96)**	**0**	**−0.119**	**A528V**	A	**V**	A	A
NP	1130	G1130A	G(191)		G(287)	A(6)	0.02	−0.1	S377N	S	S	S	S

(Continued)

TABLE 2 | Continued

Gene	nt position	nt change	nt proportion				*p*-value	Entropy difference between other hosts and duck	Amino acid change	Buan2 (C0)	H1731 (C1)	H2102 (C2)	H1924 (C4)
			Other hosts		Duck								
NA	**67**	**C67A**	**C(164)**	**A(27)**	**C(219)**	**A(74)**	**0**	**−0.158**	**L23I**	L	**I**	L	L
	206	**G206A**	**G(162)**	**A(29)**	**G(219)**	**A(74)**	**0**	**−0.139**	**S69N**	S	**N**	S	S
	220	G220A	G(190)	A(1)	G(285)	A(8)	0.05	−0.093	V74I	V	V	V	V
	239	G239A	G(189)	A(2)	G(275)	A(18)	0.01	−0.173	G80D	G	G	G	G
	986	**C986T**	**C(162)**	**T(29)**	**C(210)**	**T(83)**	**0**	**−0.17**	**T329I**	T	**I**	A	T
	1174	G1174A	G(191)		G(283)	A(10)	0	−0.149	V392M	V	V	V	V
NS1	167	C167A	C(189)	A(2)	C(274)	A(18) T(1)	0.01	−0.195	T56I	T	T	T	T
	175	C175T	C(182)	T(9)	C(260)	T(33)	0.02	−0.162	R59C	R	R	R	R
	226	G226A	G(189)	A(2)	G(276)	A(17)	0.01	−0.163	A76T	A	A	A	A
	239	**A239G**	**A(162)**	**G(29)**	**A(211)**	**G(82)**	**0**	**−0.167**	**N80S**	N	**S**	N	N
	409	A409G	A(191)		A(284)	G(9)	0.01	−0.137	I137V	I	I	I	I
	513	**T513A**	**T(166)**	**A(25)**	**T(221)**	**A(68) C(3) G(1)**	**0**	**−0.23**	**D171E**	D	**E**	D	D
	614	G614T	G(191)		G(286)	A(1) T(6)	0.03	−0.123	S205I	S	S	S	S
	625	G625A	G(189)	A(2)	G(275)	A(18)	0.01	−0.173	D209N	D	D	D	D
	637	T637C	T(188)	C(2) Y(1)	T(275)	C(18)	0.03	−0.14	S213P	S	S	S	S
NS2	**41**	**T41A**	**T(166)**	**A(25)**	**T(221)**	**A(68) C(3) G(1)**	**0**	**−0.23**	**M14K**	M	**K**	M	M
	142	G142T	G(191)		G(286)	A(1) T(6)	0.03	−0.123	A48S	A	A	A	A
	242	A242G	A(191)		A(284)	G(9)	0	−0.137	E81G	E	E	E	E

[a] *Nucleotide sequence at a significant nucleotide position (the number of H5N8 HPAI viruses having that mutation). Bold indicate the significant mutation that exclusively detected in H1731 virus.*

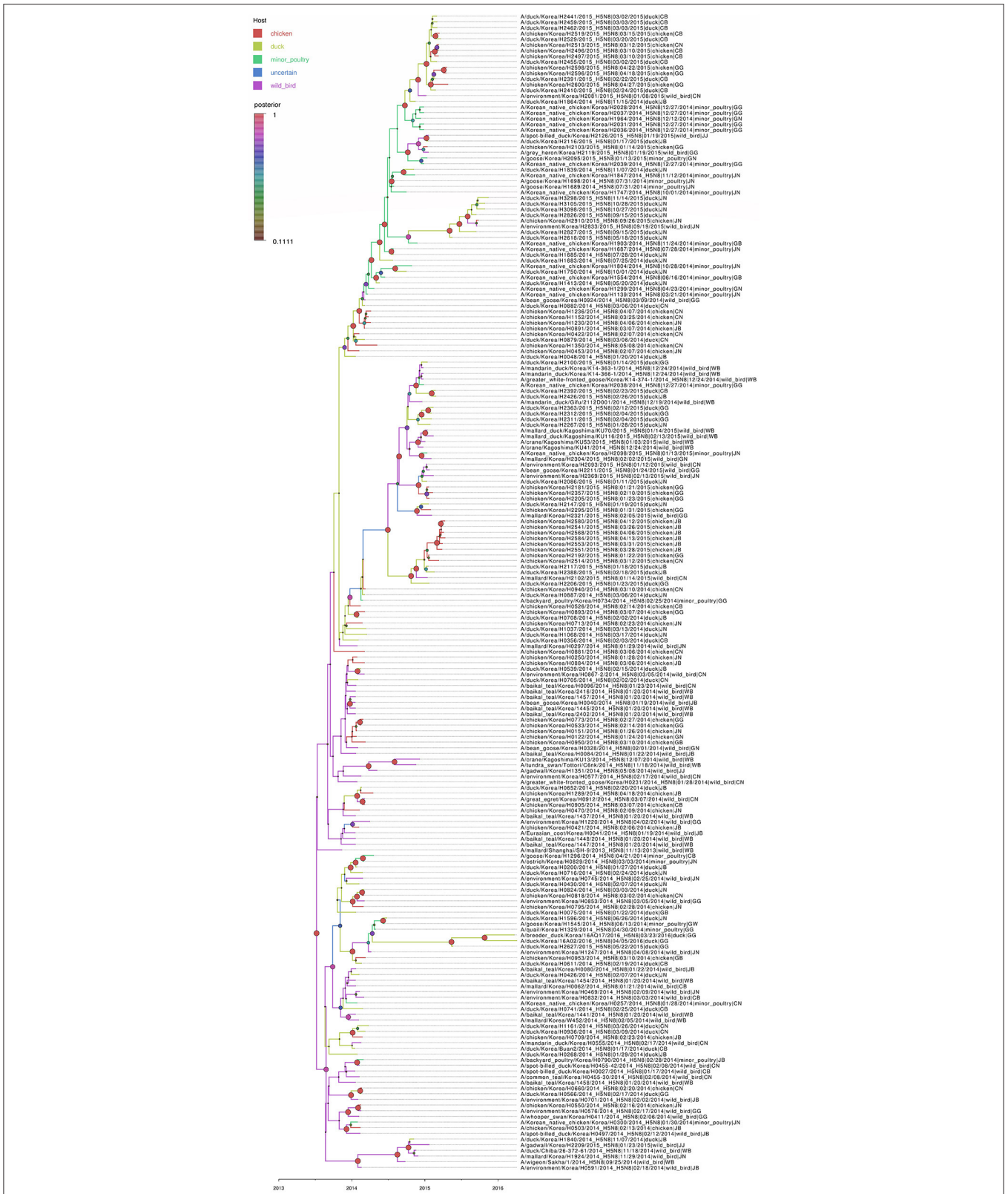


FIGURE 4 | Bayesian phylogenetic tree of the hemagglutinin gene of H5N8 high pathogenicity avian influenza viruses isolated in South Korea during 2014–2016. The host species were defined as discrete states. Posterior probabilities of nodes are indicated by the sizes and colors of the circles.

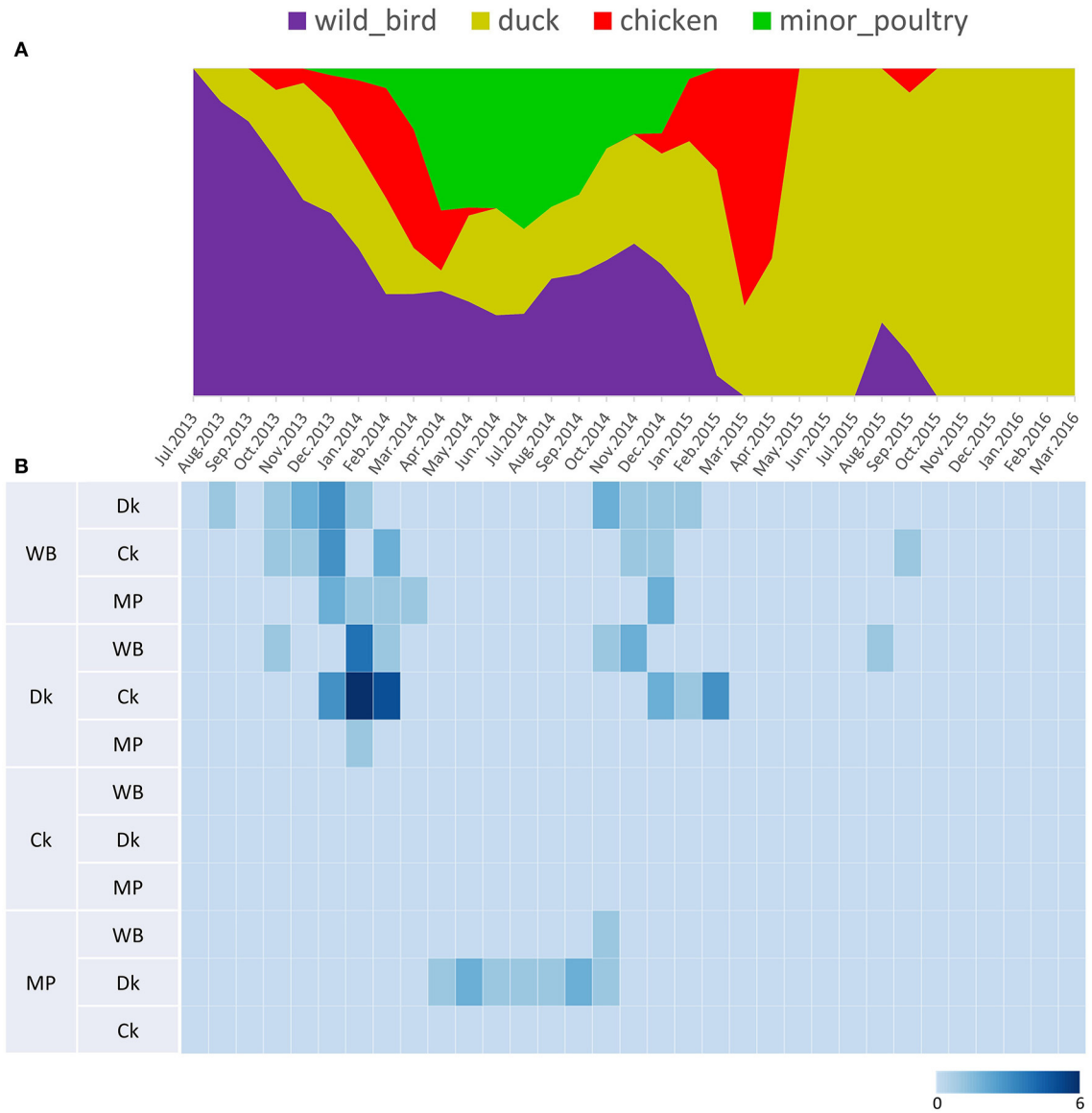


FIGURE 5 | Transition events and discrete state proportion over time of H5N8 high pathogenicity avian influenza viruses isolated in South Korea during 2014–2016. **(A)** Proportion of ancestral discrete states estimated on the phylogenetic trunk of H5N8 viruses over time. The host species were defined as discrete states. Shaded areas represent estimated ancestral discrete state proportions of each state. At each point in time, the width (y-axis) represents the mean proportion from 0 to 100% of each state. **(B)** Heat map showing the number of transition events between each discrete state (0–6) over time.

Host Dynamics

Source sink dynamics of the H5N8 HPAIVs identified in South Korea during 2014–2016 were analyzed by estimating the transition rate (TR), Markov jump count, and Markov rewards. Based on our DTA, well-supported transitions from wild birds to domestic ducks (TR: 1.525, BF: 84.5189, posterior probability: 0.9741), chickens (TR: 0.625, BF: 21.7051, posterior probability: 0.9061), and minor poultry (TR: 0.615, BF: 229.3770, posterior probability: 0.9903), indicate that multiple virus introductions from wild birds to poultry species occurred mainly from August 2013 to March 2014 and from October 2014 to January 2015 (**Figures 5**, **6**; **Table 3**). The well-supported transitions from domestic ducks to wild birds (TR: 1.065, BF: 79.3385, posterior probability: 0.9724), chickens (TR: 1.772, BF: 27329.6487, posterior probability: 0.9999), and minor poultry (TR: 0.251, BF: 3.5673, posterior probability: 0.6133) suggest the subsequent spread of viruses from domestic ducks to other species, including reverse spread to wild birds, during the winters of 2013–2014 and 2014–2015. Interestingly, minor poultry also contributed to virus spread, transmitting virus to domestic ducks

(TR: 1.473, BF: 2275.4089, posterior probability: 0.9990). The estimated total Markov reward time for each discrete state of poultry was estimated to be highest in domestic ducks (13.093, 95% HPD: 7.9203–20.0282), followed by minor poultry (5.499, 95% HPD: 4.0152–7.0553), and chickens (4.757, 95% HPD: 2.9123–7.0446) (**Figure 7A**). Collectively, these results indicate that domestic ducks and minor poultry played major roles in the maintenance and transmission of H5N8 HPAIV in South Korea after its multiple introductions by wild birds. GLM analysis revealed that sample size had a negligible effect on potential bias toward the origin or destination state in this analysis (source indicator: 0.156, source coefficient: 0.05453, sink indicator: 0.43, sink coefficient: 0.185).

Transmission History Between Regions

The DTA among provinces in South Korea involved seven discrete nominal categories: CB, CN, GG, GS, JB, JN, and WB. GLM analysis showed that sample size had a negligible effect on potential bias toward the origin or destination state in this analysis (source indicator: 0.13, source coefficient: 0.0601, sink indicator: 0.388, sink coefficient: 0.0996). The DTA indicated multiple introductions of H5N8 HPAIV from wild birds into domestic poultry in various regions in South Korea (**Figures 8–10**; **Table 4**). DTA also showed that virus was transmitted from WB to JB (TR: 1.836, BF: 115.3142, posterior probability: 0.9563) in August 2013, followed by introductions into JB and JN (TR: 1.628, BF: 549.5709, posterior probability: 0.9905), GS (TR: 0.724, BF: 25.2402, posterior probability: 0.8271), GG (TR: 1.559, BF: 229.3154, posterior probability: 0.9775), CN (TR: 1.089, BF: 44.8964, posterior probability: 0.8949), and CB (TR: 1.258, BF: 88.2501, posterior probability: 0.9436) during the winters of 2013–2014 and 2014–2015, indicating multiple introductions of virus from WB into poultry farms in various regions throughout the country. Viral transmission from JN to GG (TR: 0.359, BF: 5.4350, posterior probability: 0.5075), GS (TR: 0.497, BF: 13.4123, posterior probability: 0.7177), and WB (TR: 1.013, BF: 127.8281, posterior probability: 0.9604); from JB to GG (TR: 0.97, BF: 294.0055, posterior probability: 0.9824); from CN to GG (TR: 0.709, BF: 14.1282, posterior probability: 0.7281) and JN (TR: 1.219, BF: 396.1989, posterior probability: 0.9869); and from CB to JB (TR: 0.674, BF: 24.9877, posterior probability: 0.8257), GG (TR: 742, BF: 32.0996, posterior probability: 0.8589), and CN (TR: 0.943, BF: 160.1564, posterior probability: 0.9681) suggest subsequent spread of the viruses from JN, JB, CN, and CB to the other regions during the spring of 2014 and the winters of 2013–2014 and 2014–2015, including reverse spread from JN to wild birds during the summers of 2014 and 2015. The total Markov reward time for each discrete province was the highest for JN (9.616, 95% HPD: 7.0954–12.7019), followed by GG (6.638, 95% HPD: 4.9447–8.4321), JB (5.713, 95% HPD: 2.8411–13.2159), CB (5.263, 95% HPD: 2.9685–7.6259), CN (5.223, 95% HPD: 2.3518–11.09), and GS (2.048, 95% HPD: 1.1438–3.0173) (**Figure 7B**). The Markov reward time and viral transmission patterns indicate that poultry in JN, JB, CN, and CB may have acted as seeding populations of H5N8 HPAIV within South Korea.

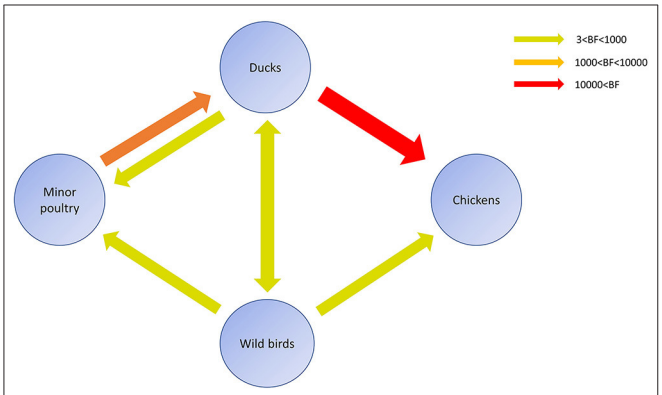


FIGURE 6 | Inferred transmission networks of H5N8 viruses between host species. Arrows show the directions of viral transmission. Line colors indicate the overall Bayes Factor test support for epidemiological linkage. Red, orange, and green lines indicate statistical support with BF >10,000 (very strong support), 1,000< BF <10,000 (strong support), and 3< BF <1,000, respectively.

DISCUSSION

AIVs evolve through the gradual accumulation of mutations and genome reassortment. H5N8 HPAIVs likely had opportunities to reassort with various AIV gene pools during the 28 months of

TABLE 3 | Viral transition of discrete traits (based on HA gene sequences) between different host species (ducks, wild birds, chickens, minor poultry) in South Korea.

Transition from	Transition to	Mean actual migration rate[a] (95% HPD)	Bayes factor	Posterior probability
Wild birds	Ducks	1.525 (0–3.0507)	84.5187	0.9741
Wild birds	Chickens	0.625 (0–1.5413)	21.7051	0.9061
Wild birds	Minor poultry	0.615 (0–1.2804)	229.3770	0.9903
Ducks	Wild birds	1.065 (0–2.3196)	79.3385	0.9724
Ducks	Chickens	1.772 (0.4169–3.421)	27329.6487	0.9999
Ducks	Minor poultry	0.251 (0–0.8906)	3.5673	0.6133
Minor poultry	Ducks	1.473 (0.3051–2.8596)	2275.4089	0.9990

[a]*Actual migration rates were calculated as the rate x indicator. HPD, highest probability density.*

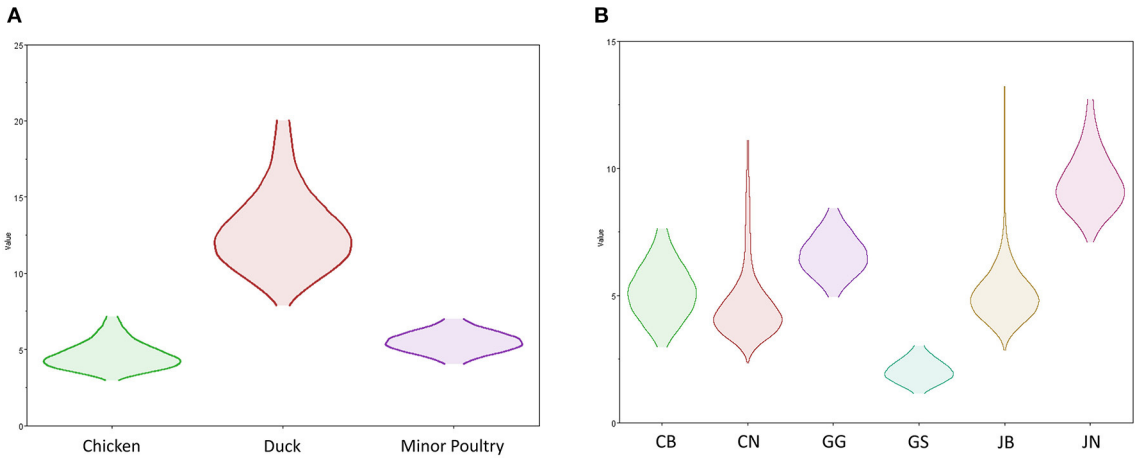


FIGURE 7 | Transmission dynamics between host species and geographical region of H5N8 high pathogenicity avian influenza viruses isolated in South Korea during 2014–2016 (Markov rewards). **(A)** The host species and **(B)** regions were defined as discrete states. Markov rewards were calculated for each posterior tree using a non-reversible continuous-time Markov chain model.

the H5N8 HPAI outbreak in 2014–2016 in South Korea, which could result in the generation of novel genotypes. However, there was no evidence of reassortment with other AIVs. ML phylogenetic analysis showed that the C0 subgroup had diverged into multiple subgroups within South Korea. The detection of four distinct subgroups, C1, C2, C4, and C5, diverging from C0 viruses in the phylogenetic tree suggests that each subgroup had evolved independently within South Korea during 2014–2016 without any reassortment. These findings indicate that the H5N8 HPAIV genotype introduced into South Korea was a fixed genome constellation with a high fitness produced in the Eurasian AIV gene pool. Interestingly, the H5N8 HPAIV spread into North America in late 2014 and reassorted with North American low-pathogenicity AIVs to produce the novel Eurasian-North American reassortant H5N1, H5N2, and H5N8 viruses (28).

The representative challenge viruses of the C1, C2, and C4 subgroups differed significantly in infectivity, transmissibility, and pathogenicity in SPF chickens. H1731 (C1) had a longer MDT, significantly lower viral titers in tissues, and lower transmissibility in chickens compared with H2102 (C2) and H1924 (C4). The virulence of H1731 in chickens was also lower than that of the buan2 (the index virus of C0) (29). In particular, the amounts of H1731 virus shed into the oropharynx and cloaca of inoculated SPF chickens ranged from $10^{1.5}$ to $10^{3.4}$ ($TCID_{50}$/0.1 ml) and from $10^{1.5}$ to $10^{2.8}$ ($TCID_{50}$/0.1 ml), respectively, whereas the amounts of Buan2 virus shed into these tissues ranged from $10^{4.2}$ to $10^{6.3}$ ($TCID_{50}$/0.1 ml) and from $10^{3.1}$ to $10^{7.5}$ ($TCID_{50}$/0.1 ml), respectively. Viral titers in tissue samples were also lower in chickens inoculated with H1731 than with buan2 virus (30). These findings suggest that viruses of the C1 subgroup have lower adaptation and transmissibility in chickens. Because viruses of the C1 subgroup had circulated mainly in domestic duck farms, they likely adapted to domestic ducks. AIVs acquire diverse mutations during adaptation to a particular host (31, 32). Shannon entropy analysis identified several dominant amino acid substitutions in H5N8 HPAIVs from domestic ducks compared with viruses from other species, with H1731 being the only challenge virus to have amino acid substitutions, further indicating that viruses of the C1 subgroup had become adapted to domestic ducks.

Long-term circulation of AIVs in a certain host species allows viruses to adapt to that species, affecting their virulence and host specificity via genetic evolution. Two amino acid residues (positions 219 and 757 in PB1) in C0, C1, and C2 were found to be under positive selection pressure, with Shannon entropy analysis showing significant mutations at these positions in H5N8 HPAIVs from ducks. Because viruses of the C0, C1, and C2 subgroups mainly circulated in domestic duck farms during the 2014–2015 outbreak, these results suggest that these viruses may have undergone positive selection in domestic ducks. Amino acid substitutions that occurred during the adaptation of H5 HPAIVs to a particular host were found to be related to viral pathogenicity in that host (33). H5N8 HPAIVs have been reported to be transmitted efficiently, with subclinical infection in domestic ducks contributing to the silent spread of virus to other hosts (34, 35). The linkage between long-term circulation of H5N8 HPAIV and asymptomatic infection in domestic ducks during the 2014–2016 outbreak has not been fully determined. Analyses of the putative role of selected amino acid substitutions in domestic ducks would require a determination of the molecular mechanism underlying viral adaptation to host and changes in pathobiology.

The estimated mean tMRCA of all H5N8 HPAIVs identified in Korea was July 2013, indicating that the H5N8 HPAIVs of wild bird origin detected in South Korea most likely emerged

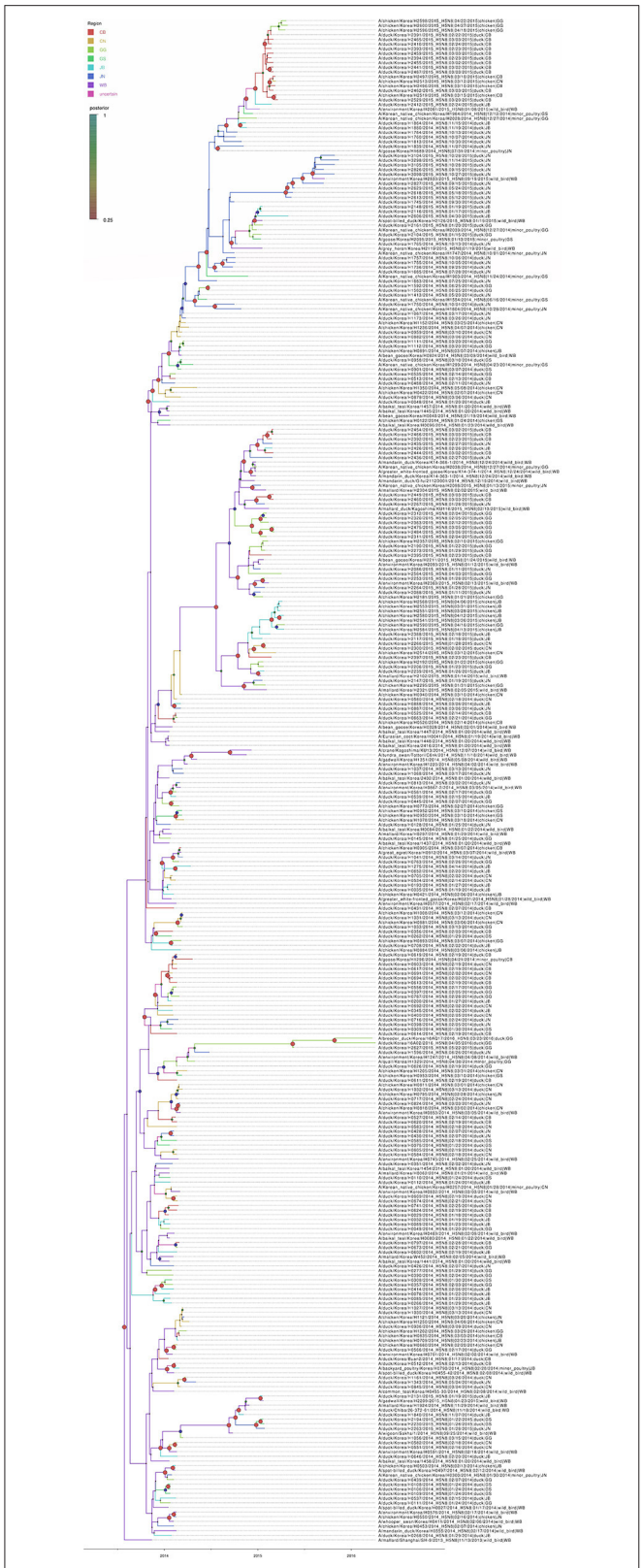


FIGURE 8 | Bayesian phylogenetic tree of the hemagglutinin gene of H5N8 high pathogenicity avian influenza viruses isolated in South Korea during 2014–2016. The geographic regions were defined in the analysis as discrete states. Posterior probabilities of nodes are indicated by the sizes and colors of the circles.

during the breeding season of wild waterfowl in 2013. These findings were consistent with other studies on the estimated tMRCA of H5N8 HPAIV (36, 37). Phylogeographic analysis suggests multiple introductions of the H5N8 HPAIVs from wild birds into JB province, followed by subsequent spread to other provinces in South Korea during the winters of 2013–2014 and 2014–2015. These findings highlight the major role of JN, JB, CN, and CB provinces in dissemination and persistence of the virus after its introductions from wild birds. The regional proportion of ancestral discrete states through time and Markov reward analyses also indicate that the H5N8 viruses were maintained predominantly in western South Korea, specifically in JN, JB, GG, CN, and CB provinces (**Supplementary Figure 3A**).

The western region of South Korea contains abundant habitats for wild waterfowl and a high density of domestic duck farms (8), creating an environment vulnerable to the transmission of HPAIVs (38, 39). In addition, surveillance studies have suggested that domestic ducks are a major contributor to the spread and maintenance of HPAIVs in South Korea (40). Consistent with previous findings, the present study showed that the transmission of H5N8 HPAIV from domestic ducks to other host species occurred frequently, further indicating that domestic ducks play major roles in the maintenance, amplification, and spread of HPAIVs (38–40). Although a previous study reported unidirectional transmission of H5N8 viruses from wild waterfowl to domestic ducks (40), the present study found exchange of the viruses between domestic ducks and wild birds (**Supplementary Figure 3B**). Geographical transmission dynamics showed dissemination of the virus into wild birds from JN, an area with large numbers of wild bird habitats adjacent to domestic duck farms. These findings suggest the need for enhanced levels of surveillance and biosecurity measures at domestic duck farms in this region to effectively monitor and prevent the introduction and spread of HPAIV.

Viral phylodynamic analysis also showed exchanges of HPAIVs between minor poultry populations and domestic ducks in South Korea, indicating that minor poultry may serve as reservoirs to maintain and disseminate HPAIV. Minor poultry are generally raised in small outdoor operations and backyards, and are usually marketed in live bird markets (LBMs), highlighting their importance as intermediary hosts in virus transmission under poor biosecurity conditions (41). LBMs include a wide variety of live poultry species, providing an ideal environment for the introduction, maintenance, and adaptation of viruses, as well as potential conditions for transmissions between duck farms and chicken farms (40).

Collectively, the present study presents unique molecular epidemiology and pathobiology of the H5N8 HPAIV outbreak in 2014–2016 in South Korea. Our data showed multiple introductions of the H5N8 HPAIV isolated in South Korea during 2014-2016 from wild waterfowl to poultry farms in multiple provinces. The virus, initially introduced into the western part of South Korea, which contains large populations of domestic ducks, was subsequently disseminated into other regions throughout the country. Furthermore, domestic ducks likely played a pivotal role in the persistent circulation of the virus, with minor poultry also serving

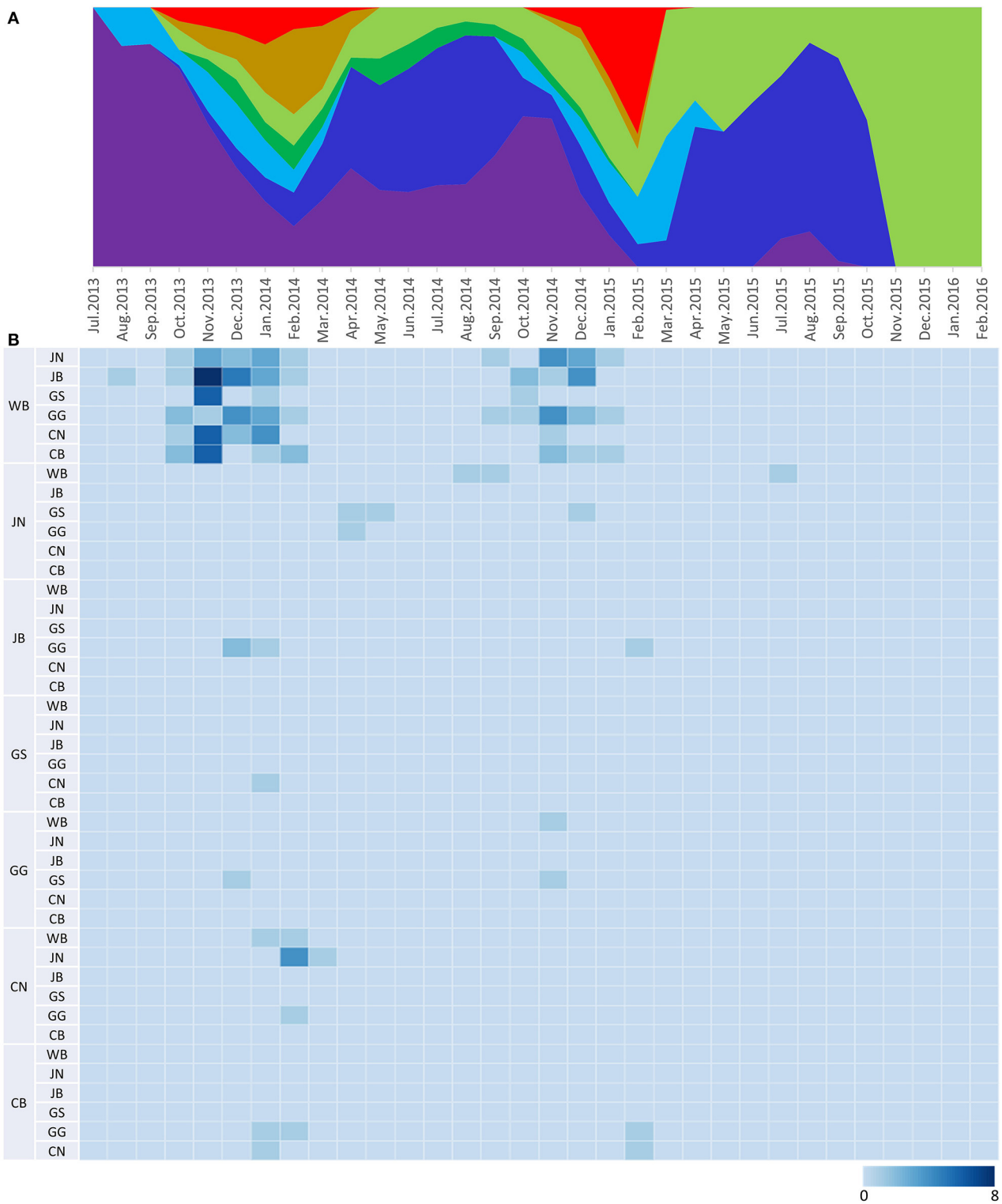


FIGURE 9 | Transition events and discrete state proportion over time of H5N8 high pathogenicity avian influenza viruses isolated in South Korea during 2014–2016. **(A)** Proportion of ancestral discrete states estimated on the phylogenetic trunk of H5N8 viruses over time. Geographic regions were defined as discrete states. Shaded areas represent estimated ancestral discrete state proportions of each state. At each point in time, the width (y-axis) represents the mean proportion from 0 to 100% of each state. **(B)** Heat map showing the number of transition events between each discrete state over time (0–8).

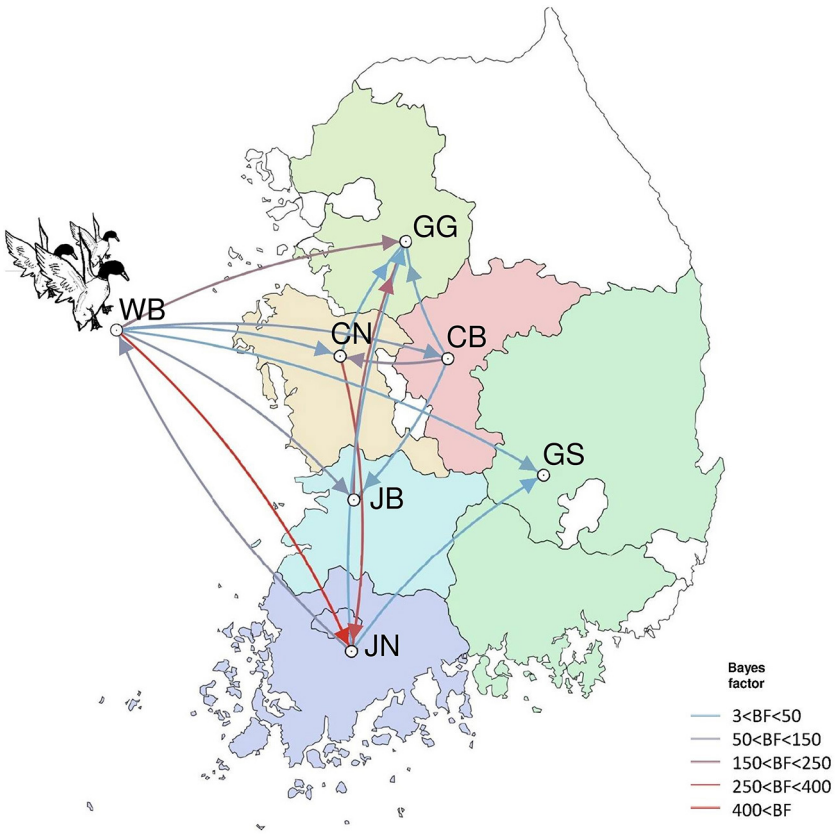


FIGURE 10 | Inferred transmission networks of H5N8 viruses between regions. Arrows show the directions of transmission. Line colors indicate the overall Bayes Factor test support for epidemiological linkage. WB, wild bird; JB, Jeollabuk-do; JN, Jeollanam-do; GG, Gyeonggi-do; GS, Gyeongsang-do; CN, Chungcheongnam-do; CB, Chungcheongbuk-do.

TABLE 4 | Viral transition of discrete traits (based on HA gene sequences) between different regions (WB, JN, JB, GG, GS, CB, and CN) in South Korea.

Transition from	Transtion to	Mean actual migration rate[a] (95% HPD)	Bayes factor	Posterior probability
WB	GS	0.724 (0–1.5599)	25.2402	0.8271
WB	CN	1.089 (0–2.1055)	44.8964	0.8949
WB	CB	1.258 (0–2.2904)	88.2501	0.9436
WB	JB	1.836 (0–3.1339)	115.3142	0.9563
WB	GG	1.559 (0–2.7841)	229.3154	0.9775
WB	JN	1.628 (0.4315–3.0045)	549.5709	0.9905
JN	GG	0.359 (0–1.3737)	5.4350	0.5075
JN	GS	0.497 (0–1.4033)	13.4123	0.7177
JN	WB	1.013 (0–2.1474)	127.8281	0.9604
JB	GG	0.97 (0–2.0953)	294.0055	0.9824
CN	GG	0.709 (0–1.9503)	14.1282	0.7281
CN	JN	1.219 (0–2.5162)	396.1989	0.9869
CB	JB	0.674 (0–1.7051)	24.9877	0.8257
CB	GG	0.742 (0–1.8076)	32.0996	0.8589
CB	CN	0.943 (0–2.1502)	160.1564	0.9681

[a]*Actual migration rates were calculated as the rate x indicator. HPD, highest probability density.*

as a source population. In addition, sequence analysis and *in vivo* experiments support that the possible adaptation of H5N8 HPAIV in domestic ducks likely reduced its virulence in chickens. Enhanced genomic surveillance and pathobiological characterization of the viruses are essential for better understanding of HPAI epidemiology and the design of prevention strategies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: GISAID EpiFlu database (http://www.gisaid.org) and accession no. EPI_ISL_157609-EPI_ISL_410213.

ETHICS STATEMENT

Experiments in animals were reviewed and approved by the Institutional Animal Care and Use Committee of the Animal and Plant Quarantine Agency (APQA) (approval no: 2018-398 and 2018-412).

AUTHOR CONTRIBUTIONS

Y-GB: writing. Y-JL: supervision and funding acquisition. E-KL: project administration. D-HL: writing—review and editing. Y-JS and Y-RP: NGS sequencing. DC: host transition analysis. J-HK: posterior tree analysis. G-BH: virus isolation. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Animal and Plant Quarantine Agency (APQA) and Ministry of Agriculture, Food and Rural Affairs of South Korea (MAFRA), via grant B-1543418-2021-23-01.

ACKNOWLEDGMENTS

The authors thank Da-Jung Kim for excellent technical assistance. We also thank the APQA, MAFRA, and Regional Office for Animal Disease Control for their efforts to implement avian influenza surveillance. We acknowledge our colleagues worldwide for their laboratory contributions, which are made available through GISAID and GenBank.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.906944/full#supplementary-material

Supplementary Figure 1 | Distribution of the H5N8 HPAIV isolated in South Korea during 2014-2016.

Supplementary Figure 2 | Geographical distribution of the H5N8 HPAIV isolated in South Korea during 2014-2016.

Supplementary Figure 3 | Transmission dynamics between discrete states. **(A)** Inferred mean transmission matrix of H5N8 viruses between the provinces. **(B)** Inferred mean transmission matrix of H5N8 viruses between the hosts.

Supplementary Table 1 | Amino acid differences between subgroups and proportion of H1731 mutation having a significance in Shannon entropy.

Supplementary Table 2 | Selective pressure analysis of H5N8 highly pathogenic avian influenza viruses isolated in South Korea during 2014-2016.

REFERENCES

1. Eagles D, Siregar ES, Dung DH, Weaver J, Wong F, Daniels P. H5N1 highly pathogenic avian influenza in Southeast Asia. *Rev Sci Tech.* (2009) 28:341–8. doi: 10.20506/rst.28.1.1864
2. Sonnberg S, Webby RJ, Webster RG. Natural history of highly pathogenic avian influenza H5N1. *Virus Res.* (2013) 178:63–77. doi: 10.1016/j.virusres.2013.05.009
3. Smith GJ, Donis RO, World Health Organization/World Organisation for Animal HF, Agriculture Organization HEWG. Nomenclature updates resulting from the evolution of avian influenza A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013-2014. *Influenza Other Respir Viruses.* (2015) 9:271–6. doi: 10.1111/irv.12324
4. Zhao K, Gu M, Zhong L, Duan Z, Zhang Y, Zhu Y, et al. Characterization of three H5N5 and one H5N8 highly pathogenic avian influenza viruses in China. *Vet Microbiol.* (2013) 163:351–7. doi: 10.1016/j.vetmic.2012.12.025
5. Kanehira K, Uchida Y, Takemae N, Hikono H, Tsunekuni R, Saito T. Characterization of an H5N8 influenza A virus isolated from chickens during an outbreak of severe avian influenza in Japan in April 2014. *Arch Virol.* (2015) 160:1629–43. doi: 10.1007/s00705-015-2428-9
6. Lee YJ, Kang HM, Lee EK, Song BM, Jeong J, Kwon YK, et al. Novel reassortant influenza A(H5N8) viruses, South Korea, 2014. *Emerg Infect Dis.* (2014) 20:1087–9. doi: 10.3201/eid2006.140233
7. Wu H, Peng X, Xu L, Jin C, Cheng L, Lu X, et al. Novel reassortant influenza A(H5N8) viruses in domestic ducks, eastern China. *Emerg Infect Dis.* (2014) 20:1315–8. doi: 10.3201/eid2008.140339
8. Hill SC, Lee YJ, Song BM, Kang HM, Lee EK. Hanna A, et al. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. *Infect Genet Evol.* (2015) 34:267–77. doi: 10.1016/j.meegid.2015.06.014
9. Ip HS, Torchetti MK, Crespo R, Kohrs P, DeBruyn P, Mansfield KG, et al. Novel Eurasian highly pathogenic avian influenza A H5 viruses in wild birds, Washington, USA, 2014. *Emerg Infect Dis.* (2015) 21:886–90. doi: 10.3201/eid2105.142020
10. Verhagen JH, Herfst S, Fouchier RA. Infectious disease. How a virus travels the world. *Science.* (2015) 347:616–7. doi: 10.1126/science.aaa6724
11. Kwon JH, Lee DH, Swayet al Hong WT, Jeong JH, Jeong S, et al. Highly Pathogenic Avian Influenza A(H5N8) viruses reintroduced into South Korea by migratory waterfowl, 2014-2015. *Emerg Infect Dis.* (2016) 22:507–10. doi: 10.3201/eid2203.151006
12. Global Consortium for HN, Related Influenza V. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science.* (2016) 354:213–7. doi: 10.1126/science.aaf8852
13. Mo IP, Bae YJ, Lee SB, Mo JS, Oh KH, Shin JH, et al. Review of Avian Influenza Outbreaks in South Korea from 1996 to 2014. *Avian Dis.* (2016) 60:172–7. doi: 10.1637/11095-041715-Review
14. Song BM, Lee EK, Lee YN, Heo GB, Lee HS, Lee YJ. Phylogeographical characterization of H5N8 viruses isolated from poultry and wild birds during 2014-2016 in South Korea. *J Vet Sci.* (2017) 18:89–94. doi: 10.4142/jvs.2017.18.1.89
15. Jeong JS, Kang HM, Lee EK, Song BM, Kwon YK, Kim HR, et al. Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its

relationship with migratory birds in South Korea during 2014. *Vet Microbiol.* (2014) 173:249–57. doi: 10.1016/j.vetmic.2014.08.002
16. Lee DH. Complete Genome Sequencing of Influenza A Viruses Using Next-Generation Sequencing. *Methods Mol Biol.* (2020) 2123:69–79. doi: 10.1007/978-1-0716-0346-8_6
17. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
18. OIE (World Organization for Animal Health). *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2021.* (2021). Available online at: https://www.oie.int/en/what-we-do/standards/codes-and-manuals/terrestrial-manual-online-access/ (accessed on March 18, 2022).
19. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* (2005) 22:1208–22. doi: 10.1093/molbev/msi105
20. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* (2013) 30:1196–205. doi: 10.1093/molbev/mst030
21. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* (2012) 8:e1002764. doi: 10.1371/journal.pgen.1002764
22. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT. Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* (2010) 26:680–2. doi: 10.1093/bioinformatics/btq003
23. Drummond AJ, Rambaut A. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evol Biol.* (2007) 7:214. doi: 10.1186/1471-2148-7-214
24. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* (2008) 25:1459–71. doi: 10.1093/molbev/msn090
25. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD. Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics.* (2011) 27:2910–2. doi: 10.1093/bioinformatics/btr481
26. Minin VN, Suchard MA. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci.* (2008) 363:3985–95. doi: 10.1098/rstb.2008.0176
27. Chen R, Holmes EC. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol.* (2006) 23:2336–41. doi: 10.1093/molbev/msl102
28. Lee DH, Bahl J, Torchetti MK, Killian ML, Ip HS, DeLiberto TJ, et al. Highly Pathogenic Avian Influenza Viruses and Generation of Novel Reassortants, United States, 2014-2015. *Emerg Infect Dis.* (2016) 22:1283–5. doi: 10.3201/eid2207.160048
29. Song BM, Kang HM, Lee EK, Jung J, Kang Y, Lee HS, et al. Pathogenicity of H5N8 virus in chickens from Korea in 2014. *J Vet Sci.* (2015) 16:237–40. doi: 10.4142/jvs.2015.16.2.237
30. Lee EK, Song BM, Kang HM, Woo SH, Heo GB, Jung SC, et al. Experimental infection of SPF and Korean native chickens with highly pathogenic avian influenza virus (H5N8). *Poult Sci.* (2016) 95:1015–9. doi: 10.3382/ps/pew028
31. Arafa A, Suarez D, Kholosy SG, Hassan MK, Nasef S, Selim A, et al. Evolution of highly pathogenic avian influenza H5N1 viruses in Egypt indicating progressive adaptation. *Arch Virol.* (2012) 157:1931–47. doi: 10.1007/s00705-012-1385-9
32. Hulse-Post DJ, Sturm-Ramirez KM, Humberd J, Seiler P, Govorkova EA, Krauss S, et al. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *Proc Natl Acad Sci.* (2005) 102:10682–7. doi: 10.1073/pnas.0504662102
33. Youk SS, Leyson CM, Seibert BA, Jadhao S, Perez DR, Suarez DL, et al. Mutations in PB1, NP, HA, and NA Contribute to Increased Virus Fitness of H5N2 Highly Pathogenic Avian Influenza Virus Clade 2344 in Chickens. *J Virol.* (2020) 95:e01675–20. doi: 10.1128/JVI.01675-20
34. Sun H, Pu J, Hu J, Liu L, Xu G, Gao GF, et al. Characterization of clade 2344 highly pathogenic H5 avian influenza viruses in ducks and chickens. *Vet Microbiol.* (2016) 182:116–22. doi: 10.1016/j.vetmic.2015.11.001
35. Pantin-Jackwood MJ, Costa-Hurtado M, Bertran K, DeJesus E, Smith D. Swayne DE. Infectivity, transmission and pathogenicity of H5 highly pathogenic avian influenza clade 2344 (H5N8 and H5N2) United States index viruses in Pekin ducks and Chinese geese. *Vet Res.* (2017) 48:33. doi: 10.1186/s13567-017-0435-4
36. Baek YG, Lee YN, Lee DH, Shin JI, Lee JH, Chung DH, et al. Multiple reassortants of H5N8 Clade 2.3.4.4b highly pathogenic avian influenza viruses detected in South Korea during the Winter of 2020-2021. *Viruses.* (2021) 13:490. doi: 10.3390/v13030490
37. Lee YN, Cheon SH, Kye SJ, Lee EK, Sagong M, Heo GB, et al. Novel reassortants of clade 2.3.4.4 H5N6 highly pathogenic avian influenza viruses possessing genetic heterogeneity in South Korea in late 2017. *J Vet Sci.* (2018) 19:850–4. doi: 10.4142/jvs.2018.19.6.850
38. Guinat C, Artois J, Bronner A, Guerin JL, Gilbert M, Paul MC. Duck production systems and highly pathogenic avian influenza H5N8 in France, 2016-2017. *Sci Rep.* (2019) 9:6177. doi: 10.1038/s41598-019-42607-x
39. Napp S, Majo N, Sanchez-Gonzalez R, Vergara-Alert J. Emergence and spread of highly pathogenic avian influenza A(H5N8) in Europe in 2016-2017. *Transbound Emerg Dis.* (2018) 65:1217–26. doi: 10.1111/tbed.12861
40. Kwon JH, Bahl J, Swayne DE, Lee YN, Lee YJ, Song CS, et al. Domestic ducks play a major role in the maintenance and spread of H5N8 highly pathogenic avian influenza viruses in South Korea. *Transbound Emerg Dis.* (2020) 67:844–51. doi: 10.1111/tbed.13406
41. Swayne DE. *Animal Influenza*. 2nd ed. Ames (Iowa): John Wiley & Sons (2016). p. 302–v27.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 12 July 2022
doi: 10.3389/fvets.2022.931272

Genome Sequence Variations of Infectious Bronchitis Virus Serotypes From Commercial Chickens in Mexico

Henry M. Kariithi[1,2], Jeremy D. Volkening[3], Christina M. Leyson[1], Claudio L. Afonso[3], Nancy Christy[4], Eduardo Lucio Decanini[4], Stéphane Lemiere[5] and David L. Suarez[1]**

[1] Exotic and Emerging Avian Viral Diseases Research Unit, Southeast Poultry Research Laboratory, U.S. National Poultry Research Center, USDA-ARS, Athens, GA, United States, [2] Biotechnology Research Institute, Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya, [3] BASE$_2$BIO, Oshkosh, WI, United States, [4] Boehringer Ingelheim Animal Health, Guadalajara, Mexico, [5] Boehringer Ingelheim Animal Health, Lyon, France

OPEN ACCESS

Edited by:
Nicola Pugliese,
University of Bari Aldo Moro, Italy

Reviewed by:
Joanna Sajdewicz-Krukowska,
National Veterinary Research Institute (NVRI), Poland
Cao Yong Chang,
Sun Yat-sen University, China

***Correspondence:**
Henry M. Kariithi
henry.kariithi@kalro.org
David L. Suarez
david.suarez@usda.gov

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 28 April 2022
Accepted: 20 June 2022
Published: 12 July 2022

Citation:
Kariithi HM, Volkening JD, Leyson CM, Afonso CL, Christy N, Decanini EL, Lemiere S and Suarez DL (2022) Genome Sequence Variations of Infectious Bronchitis Virus Serotypes From Commercial Chickens in Mexico. Front. Vet. Sci. 9:931272. doi: 10.3389/fvets.2022.931272

New variants of infectious bronchitis viruses (IBVs; *Coronaviridae*) continuously emerge despite routine vaccinations. Here, we report genome sequence variations of IBVs identified by random non-targeted next generation sequencing (NGS) of vaccine and field samples collected on FTA cards from commercial flocks in Mexico in 2019–2021. Paired-ended sequencing libraries prepared from rRNA-depleted RNAs were sequenced using Illumina MiSeq. IBV RNA was detected in 60.07% ($n = 167$) of the analyzed samples, from which 33 complete genome sequences were *de novo* assembled. The genomes are organized as 5'UTR-[Rep1a-Rep1b-S-3a-3b-E-M-4b-4c-5a-5b-N-6b]-3'UTR, except in eight sequences lacking non-structural protein genes (accessory genes) 4b, 4c, and 6b. Seventeen sequences have auxiliary S2' cleavage site located 153 residues downstream the canonically conserved primary furin-specific S1/S2 cleavage site. The sequences distinctly cluster into lineages GI-1 (Mass-type; $n = 8$), GI-3 (Holte/Iowa-97; $n = 2$), GI-9 (Arkansas-like; $n = 8$), GI-13 (793B; $n = 14$), and GI-17 (California variant; CAV; $n = 1$), with regional distribution in Mexico; this is the first report of the presence of 793B- and CAV-like strains in the country. Various point mutations, substitutions, insertions and deletions are present in the S1 hypervariable regions (HVRs I-III) across all 5 lineages, including in residues 38, 43, 56, 63, 66, and 69 that are critical in viral attachment to respiratory tract tissues. Nine intra-/inter-lineage recombination events are present in the S proteins of three Mass-type sequences, two each of Holte/Iowa-97 and Ark-like sequence, and one each of 793B-like and CAV-like sequences. This study demonstrates the feasibility of FTA cards as an attractive, adoptable low-cost sampling option for untargeted discovery of avian viral agents in field-collected clinical samples. Collectively, our data points to co-circulation of multiple distinct IBVs in Mexican commercial flocks, underscoring the need for active surveillance and a review of IBV vaccines currently used in Mexico and the larger Latin America region.

Keywords: lineage, NGS, mutation, recombination, hypervariable region, vaccine

INTRODUCTION

Infectious bronchitis virus (IBV; type species of the *Coronaviridae* family) causes the acute, highly contagious avian infectious bronchitis (IB) disease that primarily affects the upper respiratory tract in chickens of any age, but which can also cause urogenital or enteric disease, resulting in decreased production depending on virus strain and co-infecting pathogens, age, vaccination history, and immune competency of chickens (1, 2). The virus is shed by naturally infected and/or vaccinated birds and is transmitted via respiratory discharges (acute phase) and feces (disease recovery phase) to susceptible naïve birds (3). After initial respiratory tract infection, the virus can be disseminated to other tissues including trachea, lungs, kidney, oviduct, alimentary, and proventriculus (2).

The non-segmented, positive single-stranded RNA genome (∼ 27.6 kb in size) of IBV, which can also serve as a viral mRNA, comprises six genes flanked by 5′-/3′- untranslated regions (UTRs) (4, 5). Occupying about two thirds of the genomic 5′-end, gene 1 encodes the replicase/polymerase complex 1a and 1ab (Rep1a/1ab), which is produced by a programmed−1 ribosomal frameshifting mechanism that allows continuation of translation beyond the Rep1a stop codon. Rep1a/ab is proteolytically cleaved into 15 non-structural proteins by the virally-encoded accessory gene 3 (papain-like protease; PL ^{pro}) and gene 5 (3-C-like protease; 3CL ^{pro}) proteases (6, 7). Gene 2 encodes spike (S) glycoprotein, the largest structural and most divergent of all IBV proteins, which is proteolytically cleaved by cellular furin protease into subunits S1 and S2 (8). Genes 3 and 4 encode two accessory proteins and one membrane-binding structural protein each [3a/3b, and envelope (E), and 4b/4c, and membrane (M), respectively]. Gene 5 encodes accessory proteins 5a and 5b, while gene 6 encodes the structural nucleocapsid (*N*) protein and accessory protein 6. Accessory genes 3a, 3b, 4b, 4c, 5a, 5b, and 6b, which have a wide range of genomic configurations, are non-essential for virus replication (9). Accessory genes 4b, 4c, and 6b are rarely reported in the literature despite being present in many unpublished IBV genome sequences, but little is known about their roles in viral replication or pathogenesis (10, 11).

Whereas, spike subunit S2 is highly conserved amongst IBVs, the hypervariable regions (HVRs I-III) of the S1 harbor the majority of nucleotide (nt) heterogeneity between different strains (12, 13). Subunit S1 also contains the receptor-binding domain (RBD), which is essential for entering susceptible chicken cells and induction of host immune responses (14). New IBV variants, which can evade vaccine-induced immunity, continually emerge due to the mutations (caused by replication errors), and/or recombination events (caused by template switching) in the S1 sequences (15–17). Due to its close correlation with IBV serotypes, S1 sequence is used to classify IBV into 7 genogroups (GI-GVII) comprising at least 32 distinct lineages and several unassigned inter-lineage recombinants (1, 9, 18–20). "Genogroup" represents stable IBV categories and "lineage" is a descriptive dynamic serotype grouping (21).

The Mass-type serotype was first identified in the USA in the 1930s (22), but IBVs are now globally distributed in poultry, with regional-specific genotypes and serotypes that are often closely related to the live attenuated vaccine strains used in the regions, or are unique variants (15, 18, 23–25). Serotypes belonging to lineage GI-11 [South American 1 (SAI) serotypes] and lineage GI-16 (Q1-like serotypes) extensively circulate in South American chicken flocks (21). The uniquely South American SAI serotypes, which are associated with respiratory/enteric disease and reduced egg production, emerged in the 1950s in commercial poultry flocks in Brazil, and later spread to Argentina and Uruguay (18, 25). The Q1-like serotypes, which emerged in late 1970s in Asia, are more widespread than the SAI serotypes (18, 21). Both the SAI and Q1-like serotypes have low antigenic relatedness with the Mass-type vaccine strains (lineage GI-1) that are extensively used vaccination programs globally (16, 21). Other serotypes reported in South America include lineage GI-13 (793B or 4/91-like) in Brazil, Chile and Honduras (26–28). In Mexico, Mass-type (lineage GI-1), Holte/Iowa-97 (BL-56; lineage GI-3), Ark (lineage GI-9), and Q1-like serotypes have been reported in commercial flocks (29–31).

Live-attenuated (derived from virulent strains attenuated via serial passage in chicken eggs) or inactivated vaccines are routinely used in IB control (32–34). Routine use of serotype-specific live vaccines can result in evolution of novel variants (due to mutations, insertions, deletions, or recombination between co-infecting field and vaccine strains) that are serologically distinct from the vaccine strains, with potential negative impacts on vaccine efficacy (35–37). The apparent regionality of IBV diversity underscores the importance of viral characterization, which can be used to assess and properly deploy existing vaccines and potentially identify when new vaccines need to be developed.

In the current study, we sequenced 30 complete IBV genome sequences from clinical field samples collected from commercial flocks in Northern, Central and Southern Mexico in 2019–2021, along with 3 vaccine samples. We performed comparative analysis of sequence variation (vaccine vs. field sequences) and phylogenetic relationships with other IBVs, and assessed potential recombination events, point mutations, insertions, and deletions in the HVRs of the S glycoprotein. The data presented here expand current knowledge of the IBVs circulating in Mexico, which can inform vaccination strategies to control IB outbreaks in the country and in Latin America.

MATERIALS AND METHODS

Type, Origin, and Processing of Samples

The samples used in this study were randomly collected from "apparently healthy" (i.e., no observable clinical signs of disease at the time of sampling) commercial broiler (aged 21–42 days) and layer (aged 7 weeks) chicken flocks in central, northern, and southern regions of Mexico between April 2019 and December 2021. The samples were derived from respiratory (choana and lung), immunological (spleen and bursa) and digestive (cloaca) tissues from 100 birds per flock using sterile flocked swabs and pooled (25 samples per pool) in sterile 1.5 mL viral transport media. Samples were then spotted on 4-sample-area (125 μl per area) Whatman Flinders Technology Associates (FTA) cards® (Millipore-Sigma) within 24 h of collection and dried for at least 2 h at room temperature (RT; 15–25°C). After drying,

each sample-spotted FTA card was individually enclosed in double leak-proof zip-lock plastic bags with Whatman desiccant packets (GE healthcare). In addition to field samples, three vaccine samples (i.e., live attenuated Mass-type, 4/91 variant and Mass-type/Connecticut recombinant strains) from Boehringer Ingelheim Animal Health (BIAH), Mexico, were also FTA-spotted. The detailed information about what flocks were vaccinated and with what type of vaccine is not publicly available because of propriety and confidentiality between BIAH and their clients. All samples were shipped to the Southeast Poultry Research Laboratory (SEPRL), USDA-ARS, Athens, GA, and stored at −80°C in a BSL-3 laboratory until further processing.

RNA Extraction

From each sample-spotted FTA card, 24, 3-mm disks (i.e., 6 disks per spotted area) were punched out using sterile disposable biopsy punches (Robbins Instruments, USA) and incubated for 30 min at room temperature in 240 μL of nuclease-free TE buffer (10 mM Tris-HCl; 0.1 mM EDTA, pH 8.0) to elute nucleic acids. Total RNA was extracted from 100 μL of the TE eluate using MagMAX™-96 AI/ND Viral RNA Isolation Kit (Thermo Fisher Scientific, MA, USA) on an automated KingFisher Magnetic Particle Processor (Thermo Fisher, USA) following manufacturer's instructions. To selectively deplete abundant host-specific RNAs (18S, 28S and mitochondrial) and bacterial rRNAs (16S/23S), extracted RNAs were treated with an in-house RNaseH host rRNA depletion protocol we have recently described (38).

Library Preparation and NGS

Sequencing libraries were prepared using sequence-independent, single-primer amplification (SISPA) as previously described (39). Briefly, cDNAs were synthesized from 10 μL of RNA using random K-8N primer with SuperScript ™ IV First Strand synthesis Kit (Invitrogen, USA) and Klenow polymerase (NEB Inc., USA) kits, and then purified using Agencourt AMPure XP beads (Beckman Coulter Life Sciences, USA). Purified cDNAs were amplified by Phusion® High-Fidelity PCR Kit (NEB Inc., USA), and used to prepare sequencing libraries by the Nextera ™ DNA Flex kit (Illumina, USA), which were then quantified by Qubit™ dsDNA HS Assay Kit (Thermo-Fisher Scientific) and Agilent 4150 TapeStation HS D5000 System (Agilent Technologies, Inc.). Based on their concentrations and average fragment sizes, equimolar concentrations (4 nM, 8 μL of each library) of the libraries were pooled, then digested by incubation with 0.2 N NaOH (5 min at RT). Pooled libraries were further diluted to 10 pM final concentration, a control library added (5% PhiX library v 3) and paired-end (2 × 300 bp) sequencing performed using a 600-cycle MiSeq Reagent Kit v3 (Illumina, USA). Each NGS run consisted of 48 multiplexed samples.

Sequence Assembly

Customized workflows executed in Galaxy and Geneious Prime® 2021.2.2 platforms were used to assess the quality of the raw NGS reads (by FastQC) and trimming of adaptors (by Cutadapt) as previously described (40–43). Briefly, the host (chicken) and PhiX control reads were filtered out by mapping the trimmed reads against chicken (*Gallus gallus*) and PhiX174 reference genomes using BWA-MEM v 0.7.15.1 (44) with standard parameters. Trimmed/filtered overlapping forward and reverse read-pair sets were merged with PEAR v 0.9.6.1 (45); an in-house wrapper tool was then used to identify and remove chimeric Nextera reads. After digital normalization of the reads via k-mer abundances using khmer package (46), *de novo* sequence assembly was performed using MIRA v3.4.0 (47). BWA-MEM/samtools was used to re-call consensus sequences from the NGS reads aligned to the *de novo*-generated contigs (minimum coverage depth to call a base set at 3X).

Sequence Annotation and Phylogenetic Analysis

Open reading frames (ORFs; used here to refer to contiguous nt stretches from start to stop codons without interrupting in-frame stop codons) in the assembled consensus sequences were determined and annotated using ORF Finder (minimum ORF size set at 50 NT including start and stop codons) within the Geneious Prime® v 2022.1.1 (https://www.geneious.com). ORFs were confirmed by comparative analyses with annotated coronavirus (CoV) genes and genomes available at GenBank and/or PubMed. Classification of the assembled sequences was based on Valastro et al. (18). Putative protein domains were predicted using translated protein sequences with InterProScan v 2.0 (48) executed in Geneious Prime. For confirmation, putative annotations were aligned with similar annotations (coding regions and domains/motifs of translated amino acid sequences) of a Geneious Prime local database (containing reference sequences retrieved from GenBank and/or PubMed using BLASTp algorithm). ProPserver v1.0 (49) was used to predict putative cleavage site of the S glycoprotein at the border between subunits S1 and S2 (R-X-X-R↓S motif), and in S2′ site (R-X-R↓S motif) located upstream of subunit S2 ectodomain. Asparagine (N)-linked glycosylation sites in S1 protein sequences were predicted using NetNGlyc-1.0 (www.cbs.dtu.dk/services/NetNGlyc/) only sites with scores of at least 0.6 and supported by six of the nine predictive neural networks of the server were accepted.

The genome and specific gene sequences obtained from this study, together with sequences of representative serotypes within the IBV lineages [retrieved from the GenBank; lineages based on current IBV classification system (18)] were used for multiple sequence alignment using MAFFT v 7.490 (50). To minimize effects of poorly aligned regions, the multiply aligned sequences were trimmed using trimAl tool v 1.3 (51) with gappyout mode. Phylogenetic analysis was performed using maximum likelihood method executed in MEGA with 1,000 bootstrap replicates of the original data and the best model automatically identified by the software (52).

Analyses of Point Mutation, Insertions/Deletions, and Putative Recombination Events

Multiple alignments of amino acid sequences of subunit S1 HVRs I-III from this study, together with sequences of

representative serotypes within IBV lineages, were analyzed for presence of deletions/insertions (hereafter abbreviated as indels) and point mutations. For recombination events analysis, SplitsTree5 v 5.0.0_alpha (53) was used to determine the likelihood of recombination events in the complete S-gene sequences. Recombination events were further examined using seven heuristic recombination detection algorithms (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiSscan, and 3Seq) executed in the Recombination Detection Program 4 (RDP4) v4.101 software suite (54), at highest *p*-value of 0.05 with Bonferroni multiple correction and SEQ-GEN parametric data simulations. Confirmation of putative recombinant event was accepted only when the recombination breakpoints were detected by at least five of the seven algorithms, and with breakpoints of the transferred fragments (recombinant regions) supported by corrected *p*-values of $\leq 1 \times 10–6$.

RESULTS

Sequencing Libraries and Sequencing Data

In this study, we analyzed three vaccine and 275 field samples derived from immunological ($n = 126$), respiratory ($n = 141$), digestive ($n = 8$) tissues, and the vaccine samples ($n = 3$). Average fragment length distribution of the adaptor-ligated libraries (TapeStation estimates) ranged from 425 to 608 bp, but actual post-NGS average fragment length distribution post-FastQC (excluding adaptor sequences) were shorter, a discrepancy attributable to the fact that shorter fragments tend to cluster more efficiently than longer fragments (55). Total trimmed/filtered read counts ranged from 16,541 to 1.3 million reads. Proportions of chicken-specific reads ranged from 0.82 to 77.8%, with only six samples having over 50% of the reads mapping to the host genome.

Detection of Viral and Pathogenic Bacterial RNAs

IBV RNA was detected in 60.07% ($n = 167$) of the analyzed samples. Fifty-five of the samples (20 immunological tissue samples, 32 respiratory tissue samples, and the 3 vaccine samples) had enough IBV-specific reads to allow for assembly of complete or nearly-complete genome or *S*-gene consensus sequences (**Supplementary Table 1**). In addition to IBVs, 10 spleen/bursa and nine choanal/lung tissues contained RNA of avian viruses belonging to families *Astroviridae* (chicken astrovirus, serogroup 1b), *Birnaviridae* (infectious bursal disease virus, genogroup 2b), *Paramyxoviridae* (avian paramyxovirus type 1, subgenotype V.1), *Pneumoviridae* (avian metapneumovirus subtype A), *Reoviridae* (avian rotavirus serogroups A, D, and F), and *Picornaviridae* (avian nephritis virus, chicken gallivirus A, chicken megrivirus group C-3, and sicinivirus type A). Picornaviruses, and in particular SiV, were overrepresented (detected together with IBV in 73.68% of the 19 samples). In addition to viral agents, avian pathogenic bacterial species were detected in 29 out of the 55 samples, including *Bordetella*, *Enterococcus* spp, *Gallibacterium anatis*, *Salmonella enterica*, and *Streptococcus* spp.

Assembly of IBV Genome Sequences

Thirty-three complete genome sequences, six partial genome sequences, three complete *S*-gene sequences, and 13 partial *S*-gene sequences of IBVs were assembled (**Table 1**). Further analyses were restricted to the 33 complete genome sequences. One sequence was obtained from a spleen/bursa tissue sample swabbed from a 7-week old layer chicken, while 21 and nine sequences were from choanal/lung and spleen/bursa tissue samples, respectively, swabbed from broiler chickens aged between 21 and 42 days. All 33 genome sequences were supported by sufficient read depths (median read depths of 10–1,175 X) and genome coverage (99.98–100%). However, 19 of the genome sequences missed short stretches at the 5′- and/or the 3′-termini, which is not unusual for randomly primed viral genome sequencing (56).

Sequence Analyses

Genomic Organization and Features

All 33 complete sequences, with lengths ranging from 27,022 to 27,805 nt excluding poly(A) tails, contain the six IBV genes flanked by 5′- and 3′-UTRs (294–643 nt and 136–446 nt in length, respectively); 14 of the sequences, including the 3 vaccine sequences, have poly(A) tails of variable lengths (**Supplementary Table 2**). The genomes are organized as 5'UTR-[Rep1a-Rep1b-S-3a-3b-E-M-4b-4c-5a-5b-N-6b]-3'UTR, with 25 sequences having a cassette of seven "accessory" genes (3a, 3b, 4b, 4c, 5a, 5b, and 6b) interspersed variably downstream of gene 2 (S) genomic region. Genes 4b, 4c, and 6b are absent in eight sequences as follows: both genes 4b and 4c are absent in field sequence 2360/20 from Southern Mexico, 4b is absent in sequences 2359/20 and 2754/21 (from Southern Mexico) and 2723/21 (from Northern Mexico), and 6b is absent in the Mass-type vaccine strain 1616/19 and Mass-type/Conn recombinant vaccine strain 1623/19 sequences, and the field sequences 2523/21 and 2598/21 from Central and Northern Mexico, respectively.

Gene 1 (Rep1a/1ab Complex)

Lengths of gene 1 (containing 2 overlapping ORFs encoding Rep1a and Rep1ab) varies from 19,490 to 19,970 nt among the 33 sequences, but all sequences have a 4-nt overlap between Rep1a and 1b (**Table 2**). Whereas, the lengths of Rep1a vary (11,520–11,937 nt), 30 Rep1b sequences are 8,037 nt long; the field sequences 2721/21, 2353/20, and 2752/21 have comparatively shorter Rep1b lengths (7,935, 7,938, and 7,974 nt, respectively). Domain features and borders of accessory genes 2-16 produced from the cleavage of Rep1ab by the virally-encoded proteases PL ^{pro} (gene 3) and 3CL ^{pro} (gene 5) are shown in the **Supplementary Table 3**. As expected for CoVs (11, 57–60), cleavage sites and lengths of the accessory genes in Rep1a/1ab are conserved across all the sequences, including amino acid residues Q/S required by the 3CL ^{pro} for the cleavage of Rep1ab into Rep1a and 1ab, which releases products of genes10 (exonuclease; 145 amino acids), 11 (unknown function; 23 amino acids), and 12 (RdRp; 917 amino acids). Although the cleavage sites of PL ^{pro} are conserved in all sequenced, their length varies (1,529–1,619 amino acids) compared to the consistent lengths of the main CoV protease, 3CL ^{pro} (307 amino acid residues).

TABLE 1 | Sequence assembly coverage of 33 complete genome sequences assembled in this study.

Sequence	Sampling date	Sample origin	Flock (age)	Tissue	IBV-specific reads	Forward read quality[a]	Reverse read quality[a]	No. of IBV contigs	Coverage depth[a]	Reads used for consensus[b]	Consensus seq length (bases)[c]	Missing nt at 5'-end[d]	Missing nt at 3'-end[d]	percent coverage[e]
Live Mass-type vaccine strain/1616/19	23/Apr/19	BIAH Mexico	BIAH Mexico	Vaccine	8,839	2|2|2|2|38	2|2|2|25|36	4	0|15|**25**|39|407	5,135	27,469			100.%
4/91 vaccine variant/1619/19	23/Apr/19	BIAH Mexico	BIAH Mexico	Vaccine	2,490	2|2|21|25|36	2|2|2|2|38	28	0|6|**10**|16|112	2,053	27,656			100%
Mass-type-Conn recombinant vaccine strain/1623/19	23/Apr/19	BIAH Mexico	BIAH Mexico	Vaccine	11,636	2|2|2|25|35	2|2|2|2|38	2	0|21|**32**|51|572	6,899	27,474			100%
ck/MEX/2353/20	25/Nov/20	South Mexico	Broiler (42D)	Respiratory	6,072	2|33|37|38|38	2|35|37|38|38	9	0|22|**40**|71|390	6,032	27,410	184	138	99.98%
ck/MEX/2354/20	25/Nov/20	South Mexico	Broiler (42D)	Immunological	18,184	2|34|37|38|38	2|36|37|38|38	1	0|66|**131**|240|862	18,184	27,410			100%
ck/MEX/2359/20	25/Nov/20	South Mexico	Broiler (42D)	Respiratory	182,811	2|36|37|38|38	2|37|38|38|38	4	0|455|***1054***|2053|15505	174,143	27,634			100%
ck/MEX/2360/20	25/Nov/20	South Mexico	Broiler (42D)	Immunological	18,162	2|37|38|38|38	2|35|37|38|38	10	0|39|**116**|218|1929	17,859	27,167		83	99.99%
ck/MEX/2523/21	28/Jan/21	Central Mexico	Broiler (29D)	Respiratory	25,972	2|36|37|38|38	2|37|37|38|38	8	0|62|**148**|346|2345	25,154	27,449	21	3	99.99%
ck/MEX/2562/21	25-Feb-21	Central Mexico	Broiler (28D)	Respiratory	10,978	2|27|36|37|38	2|32|37|38|38	6	0|49|**71**|107|913	10,978	27,671			100%
ck/MEX/2563/21	25-Feb-21	Central Mexico	Broiler (28D)	Immunological	24,308	2|30|37|37|38	2|35|37|38|38	2	0|79|**152**|291|2404	24,308	27,805			100%
ck/MEX/2592/21	Apr/21	South Mexico	Broiler (21D)	Respiratory	20,491	2|21|28|37|38	2|26|34|37|38	2	0|32|**52**|100|409	17,806	27,700			100%
ck/MEX/2598/21	Apr/21	North Mexico	Broiler (27D)	Immunological	3,289	2|31|37|37|38	2|25|35|37|38	19	0|11|**24**|43|182	3,289	27,095		122	99.99%
ck/MEX/2602/21	Apr/21	South Mexico	Broiler (21D)	Immunological	5,422	2|28|36|37|38	2|23|33|37|38	10	0|23|**43**|68|293	5,422	27,623		11	100%
ck/MEX/2721/21	06/Jun/21	North Mexico	Broiler (21D)	Respiratory	14,548	2|33|37|37|37	2|27|36|37|37	7	0|63|**119**|210|523	14,465	27,280	54		99.99%
ck/MEX/2723/21	06/Jun/21	North Mexico	Broiler (28D)	Respiratory	65,561	2|33|36|37|37	2|36|37|37|37	3	0|246|**500**|953|3343	65,351	27,694	25		99.99%
ck/MEX/2725/21	06/Jun/21	North Mexico	Broiler (21D)	Respiratory	15,345	2|17|20|34|37	2|21|28|36|37	4	0|62|**101**|150|1016	14,551	27,594			100%
ck/MEX/2731/21	23/Jun/21	Central Mexico	Broiler (23D)	Respiratory	2,306	2|18|21|35|37	2|22|31|37|37	35	0|6|**11**|20|387	2,207	27,189	192		99.99%
ck/MEX/2742/21	11/Jul/21	South Mexico	Broiler (28D)	Immunological	24,335	2|18|31|36|37	2|24|35|37|37	2	0|75|**149**|279|3047	23,808	27,625	9		100%
ck/MEX/2743/21	11/Jul/21	South Mexico	Broiler (28D)	Respiratory	32,724	2|32|36|37|37	2|34|37|37|37	6	0|127|**261**|467|1855	32,606	27,509			100%
ck/MEX/2748/21	15/Jul/21	South Mexico	Broiler (28D)	Respiratory	6,983	2|18|25|36|37	2|23|33|37|37	8	0|33|**60**|96|325	6,891	27,619		12	100%

(Continued)

TABLE 1 | Continued

Sequence	Sampling date	Sample origin	Flock (age)	Tissue	IBV-specific reads	Forward read quality[a]	Reverse read quality[a]	No. of IBV contigs	Coverage depth[a]	Reads used for consensus[b]	Consensus seq length (bases)[c]	Missing nt at 5'-end[d]	Missing nt at 3'-end[d]	percent coverage[e]
ck/MEX/2752/21	20/Jul/21	South Mexico	Broiler (21D)	Respiratory	37,386	2|34|36|37|37	2|36|37|37|37	7	0|87|**310**|539|1686	37,229	27,022	180	59	99.99%
ck/MEX/2753/21	20/Jul/21	South Mexico	Broiler (28D)	Immunological	8,260	2|18|26|36|37	2|22|33|37|37	1	0|42|**64**|98|323	8,178	27,450	47		99.99%
ck/MEX/2754/21	20/Jul/21	South Mexico	Broiler (21D)	Respiratory	4,346	2|18|27|36|37	2|23|34|37|37	6	0|20|**35**|60|210	4,256	27,411			100%
ck/MEX/2818/21	9/10/21	North Mexico	Broiler (28D)	Respiratory	7,109	2|33|36|37|37	2|34|37|37|37	4	0|29|**51**|82|316	7,076	27,617			100%
ck/MEX/2819/21	9/10/21	North Mexico	Broiler (28D)	Immunological	8,690	2|33|36|37|37	2|33|36|37|37	3	0|36|**66**|122|449	8,635	27,638		17	99.99%
ck/MEX/2826/21	9/13/21	South Mexico	Broiler (28D)	Respiratory	18,838	2|33|36|37|37	2|35|37|37|37	1	0|73|**117**|212|1491	18,800	27,567	30	4	99.99%
ck/MEX/2833/21	9/27/21	North Mexico	Broiler (21D)	Respiratory	32,389	2|31|34|37|37	2|34|36|37|37	2	0|105|**203**|347|4082	30,900	27,694		7	100%
ck/MEX/2860/21	10/12/21	South Mexico	Broiler (28D)	Respiratory	60,382	2|23|35|37|37	2|31|36|37|37	3	0|278|**470**|747|2866	59,588	27,622	9		100%
ck/MEX/2930/21	23-Nov-21	Central Mexico	Broiler (21D)	Respiratory	22,389	2|33|37|37|37	2|36|37|37|37	14	0|54|**137**|260|883	21,704	27,719			100%
ck/MEX/2944/21	7-Dec-21	North Mexico	Broiler (28D)	Respiratory	57,769	2|31|35|37|37	2|35|37|37|37	1	0|237|**400**|654|3974	56,592	27,585	48		99.99%
ck/MEX/2956/21	14-Dec-21	South Mexico	Broiler (28D)	Respiratory	196,791	2|34|36|37|37	2|36|37|37|37	2	0|629|***1175***|2272|29429	193,735	27,697			100%
ck/MEX/2960/21	14-Dec-21	Central Mexico	Layer (7.6W)	Immunological	5,153	2|18|25|36|37	2|23|34|37|37	12	0|16|**31**|52|402	4,805	27,524		41	99.99%
ck/MEX/2961/21	15-Dec-21	Central Mexico	Broiler (21D)	Respiratory	19,584	2|19|32|37|37	2|27|36|37|37	4	0|66|**123**|262|910	18,347	27,577	4		100%

Shown are the quality of the reads, number of contigs, read coverage depth, lengths of the consensus genome sequences, and the percent genome coverage (fraction of expected full genome sequence vis-à-vis the consensus scaffold).

[a] *Number represent the minimum | lower quartile | median | upper quartile | maximum depth per position of the sequences.*

[b] *Numbers of paired-end reads that were used to re-call the consensus sequences.*

[c] *The genome sequence lengths excludes the poly A tails. Sequences that contained poly(A) tails are in bold.*

[d] *Number of nt missing at the 5'- and 3'-ends of the consensus sequences vis-à-vis the consensus scaffold.*

[e] *Percent genome coverage (fraction of expected full genome sequence vis-à-vis the consensus scaffold).*

TABLE 2 | Nucleotide overlaps between genes and ORFs in the 33 IBV sequences analyzed in this study.

Sequence	Gene 1	Genes 1 & 2	Genes 2 & 3	Gene 3		Genes 3 & 4	Gene 4		Genes 4 & 5	Gene 5	Genes 5 & 6	Gene 6
	Rep1a & 1b			3a & 3b	3b & E		M & 4b	4b & 4c		5a & 5b		N & 6b
Live Mass-type vaccine strain/1616/19	4	50	(VncRNA; 26)	1	20	29	0	80	17	4	58	N/A
4/91 vaccine variant/1619/19	4	50	(VncRNA; 26)	1	20	8	0	80	17	4	58	(VncRNA; 8)
Mass-type-Conn recombinant vaccine strain/1623/19	4	50	1	1	20	29	0	80	17	4	58	N/A
ck/MEX/2353/20	4	50	1	1	17	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2354/20	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2359/20	4	50	1	1	20	23	N/A	N/A	17	4	58	(VncRNA; 8)
ck/MEX/2360/20	4	50	1	1	20	23	N/A	N/A	N/A	4	58	(VncRNA; 8)
ck/MEX/2523/21	4	50	1	1	20	23	0	80	17	4	58	N/A
ck/MEX/2562/21	4	50	1	1	20	23	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2563/21	4	50	1	1	17	23	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2592/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2598/21	4	50	(VncRNA; 26)	1	20	29	0	80	17	4	58	N/A
ck/MEX/2602/21	4	50	1	1	20	29	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2721/21	4	50	1	1	20	20	(VncRNA; 24)	80	17	4	58	(VncRNA; 23)
ck/MEX/2723/21	4	50	1	1	20	20	N/A	N/A	17	4	58	(VncRNA; 23)
ck/MEX/2725/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 27)
ck/MEX/2731/21	4	50	1	1	20	23	(VncRNA; 13)	17	17	4	58	(VncRNA; 8)
ck/MEX/2742/21	4	50	1	1	20	23	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2743/21	4	50	1	1	38	29	0	95	17	4	58	(VncRNA; 23)
ck/MEX/2748/21	4	50	1	1	17	29	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2752/21	4	50	1	1	17	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2753/21	4	50	1	1	17	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2754/21	4	50	1	1	17	29	N/A	N/A	17	4	58	(VncRNA; 23)
ck/MEX/2818/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2819/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2826/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2833/21	4	50	1	1	20	20	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2860/21	4	50	1	1	20	29	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2930/21	4	50	1	1	20	(VncRNA; 16)	0	80	17	4	58	(VncRNA; 23)
ck/MEX/2944/21	4	50	1	1	20	4	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2956/21	4	50	1	1	20	23	0	80	17	4	58	(VncRNA; 8)
ck/MEX/2960/21	4	50	1	1	20	23	0	80	17	4	58	(VncRNA; 17)
ck/MEX/2961/21	4	50	1	1	20	29	0	80	17	4	58	(VncRNA; 8)

The numbers in brackets refer to nt in the VncRNAs.

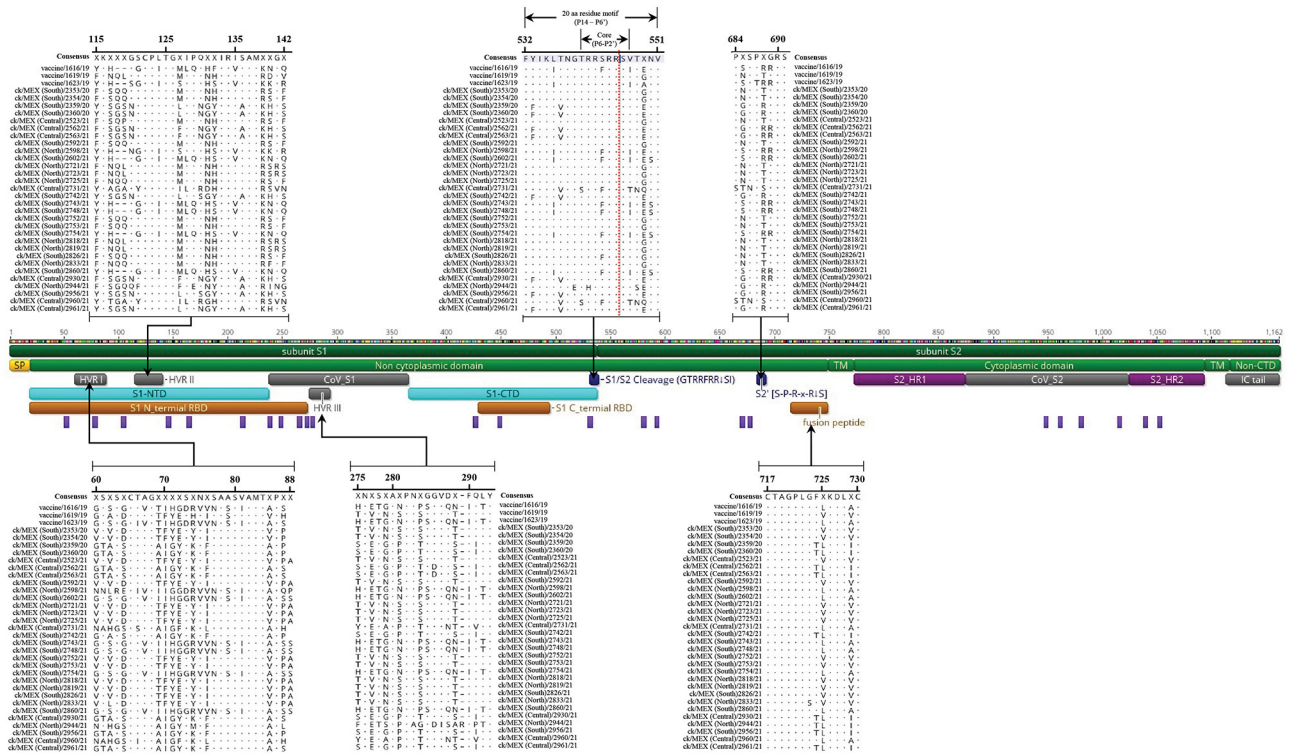


FIGURE 1 | Schematic and amino acid alignment of the overall features of the S glycoprotein. The scheme was constructed in Geneious Prime based on the S protein sequence of the field sequence 2602/21; all 33 S protein sequences in this study have similar general features (see **Supplementary Table 4**). Sequence IDs are shown on the alignments; the positions indicated at the top of the sequence alignments are in reference to ungapped amino acid residues; dots and dashes indicate identical and deleted amino acid residues, respectively. Features of subunit S1 include N-terminal signal peptide (SP), N-and C-terminal domains (S1-NTD/CTD), which harbor the receptor-binding domains (RBD), and hypervariable regions (HVR I-III). Shown is a 20- residue (running from positions P14–P6') motif furin S1/S2 cleavage site, consisting of a core region (8-residues; position P6–P2'), which harbors the canonical 4-residue motif (**R**x[**K/R**]-**R**↓**S**); the core region is flanked by two solvent accessible regions (8-residue; P7 to P14, and a 4-residue; P6 to P2'). Note the backward and forward numbering of P1-P14 and P1'-P6', respectively, starting at the conserved R immediately upstream of the cleavage S1/S2 site. The auxiliary S2′ (**P**S(G)S**PR**x**R**↓**S**) cleavage site positioned 153-residues downstream the primary S1/S2 cleavage site is also shown. Subunit S2 domains include fusion peptide (FP), heptad repeat regions (HR1 and HR2), transmembrane domain (TM), and intracellular (IC) tail. Purple vertical bars represent predicted N-linked glycosylation sites.

Gene 2 (S)

In all 33 sequences, gene 1 (Rep1a/1ab) and gene 2 (S) overlap by 50 nt (**Table 2**). Lengths of gene 2 (with a single ORF encoding S glycoprotein) vary, with the most sequences ($n = 13$) having 3,495 nt and others with either 3,510 nt ($n = 7$) or 3,498 nt ($n = 6$) (**Supplementary Table 4**). The shortest *S* genes are found in the Mass-type vaccine strain 1616/19 (3,462 nt) and the 4/91 vaccine variant 1619/19 (3,468 nt) sequences, and the field sequence 2598/21 (3,453 nt). Whereas, the lengths of subunit *S1* vary (1,602–1,632 nt), all subunit *S2* sequences have lengths of 1,878 nt, except the above-mentioned 3 sequences (1,851-nt long). **Figure 1** illustrates general S glycoprotein features. The lengths of HVR I, II and III (residues 60–88, 115–142, and 275–293, respectively) vary among the 33 sequences; HVR I of 27 or 28 residues (in 11 and 22 sequences, respectively), HVR II of 25 or 27 residues (in 8 and 25 sequences, respectively), and HVR III of 18 residues (in all sequences).

All 33 *S*-gene sequences contain the canonical 4 amino acid (aa) residue furin recognition consensus motif **R**-X-[**K**/**R**]-**R**↓**S** at the S1/S2 cleavage site (X is any residue, ↓ is cleavage position and underlined residues are conserved)—**Figure 1**. The canonical motif is within a 20 aa region (positions P14 to P6', with backward and forward numbering of P1-P14 and P1'-P6', respectively, starting at the conserved R immediate of the R↓S cleavage position), consisting of a core region (8 aa; P6-P2') flanked by two solvent accessible regions (8 aa P7-P14, and a 4 aa; P6-P2') (61). The S1/S2 cleavage motif contains critical physical properties required for furin cleavage and fusion efficiency (61), including absence of the acidic cysteine residue in the core region (P1-P6), presence of a positively-charged residue at position P4 (R is favored), hydrophilic residues in regions flanking the S1/S2 site (positions P7-P10 and P3′-P6′), small hydrophilic residue at position P1′ (S is preferred) and hydrophobic-aliphatic residue at position P2′ (V is favored). There are however, two exemptions in the 8 aa P6-P2′ region: firstly, nine sequences have the hydrophobic F-residue at P3 (instead of preferred hydrophilic residue), and secondly, sequences 2731/21 and 2960/21 have the hydrophilic T-residue at P2′ (instead of the preferred hydrophobic V or I residues). These exceptions are not unusual as the specific interactions of

the residues in P2′ and P3 within the furin cleavage pocket are unclear (61).

An auxiliary S2′ cleavage (motif **P**SxS**PR**x**R**↓**S**) is present in 17 sequences positioned 153 amino acid residues downstream of the primary S1/S2 cleavage site (**Supplementary Table 4** and **Figure 1**). Further, all sequences contain a membrane-fusion peptide (FP) in the conserved region flanked by 2 cysteine (C) residues located immediately downstream of the S2′ cleavage site, with the consensus motif **C**TAGPLGF/(T)XK**DL**X**C** (**Figure 1**); underlined residues C, D, L, and C are conserved across CoVs (62). Numbers of *N*-linked glycosylation sites (on conserved consensus NXS/T motif) varies among the 33 sequences (22–27 sites), with subunit S1 having more sites ($n = 12$–16) compared to subunit S2 ($n = 10$–12); these numbers are within the range of 19–39 sites reported in CoVs (9).

Gene 3 (3a/3b and E)

A one-nt overlap occurs between gene 2 (S) and gene 3 (accessory genes 3a-3b and E) in 30 of the 33 sequences. In the Mass-type vaccine 1616/19 and the 4/91 vaccine variant 1619/19 sequences, and the field sequence 2598/21, there is a 26-nt non-coding region (hereafter abbreviated as VncRNA to refer to genomic region between adjacent genes or ORFs without any ORF) between the two genes (**Table 2**). Within gene 3, there is a 1-nt overlap between 3a and 3b, while 3b and E overlap by 17, 20, and 38 nt in six, 26 and one sequences, respectively. In all sequences, 3a and 3b are of the same lengths (174 and 195 nt, respectively), but the length of the structural E varies (285–330 nt) among the sequences (**Supplementary Table 2**). Differences in lengths of E-gene is expected due to its extreme divergence in CoVs (9). The E protein sequences in all sequences contained two conserved cysteine residues at position 45 and 46, which serve as E protein palmitoylation sites (63).

Gene 4 (M, 4b/4c)

The overlap between gene 3 and gene 4 (M and accessory genes 4b/4c) varies among the sequences (**Table 2**). Twelve sequences have a 20 nt overlap and nine sequences each have 23-nt and 29-nt overlaps, while the field sequence 2944/21 and the 4/91 vaccine variant 1619/19 sequence have 4 and 8-nt overlaps between the two genes, respectively. There is a 16-nt VncRNA region between the two genes in sequence 2930/21. Gene 4 ORFs vary in length; 654–690 nt (M gene), 102–327 nt (4b), and 132–171 nt (4c; 25 4c are 171 nt in length). Amongst the 3 gene 4 ORFs, there is no overlap of M and 4b in 27 sequences. Sequences 2721/21 and 2731/21 have VncRNAs of 21 nt and 13 nt between M and 4b, respectively. Sequences 2359/20, 2360/20, 2723/2021, and 2754/21 lack 4b; sequence 2360/20 lacks both 4b and 4c (**Supplementary Table 2**). In 28 sequences, 4b and 4c overlap by 80 nt; the two 2 accessory genes overlap by 17 and 95 nt in sequences 2731/21 and 2743/21, respectively.

Gene 5 (5a and 5b)

In all the sequences, gene 4 and gene 5 (5a-5b) overlap by 17 nt; 5a and 5b are 198 and 249 nt in length, respectively, all overlapping by 4-nt (see **Table 2** and **Supplementary Table 2**). The presence and apparent conservation of 5a and 5b across all the 33 sequences in this study agree with reports that all avian CoVs contain gene 5, whose protein products are postulated to contribute to virus/host interactions (64).

Gene 6 (N and 6b)

Genes 5 and 6 (N and 6b) overlap by 58 nt in 29 of the 33 sequences (6b is absent in the Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences, and field sequences 2523/21 and 2598/21) (**Table 2**). Lengths of VncRNAs between N and 6b varies among the sequences; 8 nt ($n = 16$ sequences), 23 nt ($n = 11$ sequences), 17 nt (in sequence 2960/21), and 27 nt (in sequence 2725/21)—**Supplementary Table 2**. Whereas, all N gene sequences are of the same length (1,230 nt), the lengths of 6b varies widely from 129 to 321 nt amongst the sequences.

Sequence Comparison

Relationships Between Vaccine and Field Sequences

Supplementary Table 5 presents the results of comparative pairwise nt sequence identities of vaccine vs. field sequences identified in this study. Amongst the three vaccines, the highest identity (100%) is between gene 3 (3a and 3b) the Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences, and the lowest (77.31%) between subunit *S1*-gene of sequences of the Mass-type/Conn recombinant vaccine 1623/19 and the 4/91 vaccine variant 1619/19.

Comparing the vaccine vs. field sequences, the highest identities (99.43–100%) are between genes 3 and 4, and 5a and 5b of field sequence 2598/21 and their homologs in the vaccine sequences. The lowest identity is between 6b (60%) and the 3'-UTR (64.1%) of the field sequences 2723/21 and 2360/20, respectively, and their homologous genomic regions in the 4/91 vaccine variant 1619/19 sequence. For the *S1*-gene sequence, which is used for IBV classification (18), the highest identity (98.14%) is between field sequence 2860/21 and the Mass-type vaccine 1616/19 sequence, while lowest (75.78%) is between the field sequence 2944/21 and the 4/91 vaccine variant 1619/19 sequence. The two field sequences also showed similar identities in their complete *S*-gene sequences (highest and lowest identities of 98.15 and 80.15%, respectively) to the vaccine sequence.

Overall, the most conserved genomic regions amongst the 33 sequences are Rep1ab (88.5–93.24% nt identity) and 5b (93.98–100% nt identity), while the least conserved region are 6b (60–94.22% nt identity) and 4c (74.27–100% nt identity).

Relationships With Other Serotypes

Relationships between the 33 sequences in this study with other IBVs are summarized in **Table 3** (complete S-gene sequences) and **Supplementary Table 6** (complete genome sequence). Phylogenetic trees based on nt sequences of complete *S*-gene, *S1*-gene, and HVRs I-III classified the 33 sequences in this study within five different lineages (**Figure 2**).

Lineage GI-1 (Mass-Type). As shown in **Table 3**, six and two field and vaccine sequences are closely related to

TABLE 3 | BLASTn results of the 33 complete *S*-gene sequences in this study.

Sequence	Sampling date	Origin (region)	Flock age (days/weeks)	Tissue	Best BLASTn hit (isolate)	Lineage (serotype)	Seq. length	Hit start	Hit end	Query coverage	Bit-Score	Identity %
OM912698/live Mass-type vaccine strain/1616/19	23-Apr-19	n/a	Vaccine	BIAH Mexico	MK937828/ck/CN/I1124/16	GI-1 (Mass-type)	3,462	20,368	23,829	100%	6,558.98	99.51
OM912697/4/91 vaccine variant strain/1619/19	23-Apr-19	n/a	Vaccine	BIAH Mexico	MN548285/ck/UK/CR88/11	GI-13 (793B or 4/91)	3,468	20,371	23,838	100%	6,557.03	99.94
OM912696/Mass-type-Conn recombinant vaccine strain/1623/19	23-Apr-19	n/a	Vaccine	BIAH Mexico	MN696791/ck/TT/18RS1461-3/14	GI-1 (Mass-type)	3,489	1	3,489	100%	6,391.7	98.45
OM912680/ck/MEX/2353/20	25-Nov-20	South	Broiler (42D)	choanal/lung	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,720.48	100
OM912682/ck/MEX/2354/20	25-Nov-20	South	Broiler (42D)	spleen/bursa	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,720.48	100
OM912678/ck/MEX/2359/20	25-Nov-20	South	Broiler (42D)	choanal/lung	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,400.72	94.42
OM912677/ck/MEX/2360/20	25-Nov-20	South	Broiler (42D)	spleen/bursa	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,605.32	94.42
OM912695/ck/MEX/2523/21	28-Jan-21	Center	Broiler (29D)	choanal/lung	KP118891/ck/CH/LHLJ/111246/11	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,576.28	99.28
OM912694/ck/MEX/2562/21	25-Feb-21	Center	Broiler (28D)	choanal/lung	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,605.32	94.25
OM912693/ck/MEX/2563/21	25-Feb-21	Center	Broiler (28D)	spleen/bursa	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,576.48	94.19
OM912692/ck/MEX/2592/21	1-Apr-21	South	Broiler (21D)	choanal/lung	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,708.95	99.94
OM912685/ck/MEX/2598/21	1-Apr-21	North	Broiler (27D)	spleen/bursa	MK937828/ck/CN/I1124/16	GI-1 (Mass-type)	3,453	20,368	23,829	100%	6,335.94	98.5
OM912684/ck/MEX/2602/21	1-Apr-21	South	Broiler (21D)	spleen/bursa	KY626045/BR/Ma5/16	GI-1 (Mass-type)	3,489	20,314	23,802	100%	6,685.87	99.89
OM912691/ck/MEX/2721/21	6-Jun-21	North	Broiler (21D)	choanal/lung	KP118880/ck/CH/LHB/130927/13	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,628.19	99.54
OM912690/ck/MEX/2723/21	6-Jun-21	North	Broiler (28D)	choanal/lung	KP118880/ck/CH/LHB/130927/13	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,645.5	99.63
OM912689/ck/MEX/2725/21	6-Jun-21	North	Broiler (21D)	choanal/lung	KP118880/ck/CH/LHB/130927/13	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,383.14	99.63
OM912676/ck/MEX/2731/21	23-Jun-21	Center	Broiler (23D)	choanal/lung	GU393334/ck/US/Gray/60	GI-3 (Holte/Iowa-97)	3,498	20,382	23,874	99.69%	4,052.67	87.66
OM912688/ck/MEX/2742/21	11-Jul-21	South	Broiler (28D)	spleen/bursa	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,599.55	94.39
OM912683/ck/MEX/2743/21	11-Jul-21	South	Broiler (28D)	choanal/lung	KY626045/BR/Ma5/16	GI-1 (Mass-type)	3,489	20,314	23,802	100%	6,685.87	99.89
OM912687/ck/MEX/2748/21	15-Jul-21	South	Broiler (28D)	choanal/lung	KY626045/BR/Ma5/16	GI-1 (Mass-type)	3,489	20,314	23,802	100%	6,685.87	99.89
OM912679/ck/MEX/2752/21	20-Jul-21	South	Broiler (21D)	choanal/lung	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,708.95	99.94
OM912686/ck/MEX/2753/21	20-Jul-21	South	Broiler (28D)	spleen/bursa	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,708.95	99.94
OM912681/ck/MEX/2754/21	20-Jul-21	South	Broiler (21D)	choanal/lung	KY626045/BR/Ma5/16	GI-1 (Mass-type)	3,489	20,314	23,802	100%	6,685.87	99.89
OM912699/ck/MEX/2818/21	9/10/2021	North	Broiler (28D)	choanal/lung	KP118880/ck/CH/LHB/130927/13	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,377.6	99.6
OM912700/ck/MEX/2819/21	9/10/2021	North	Broiler (28D)	spleen/bursa	KP118880/ck/CH/LHB/130927/13	GI-13 (793B or 4/91)	3,495	20,314	23,808	100%	6,377.6	99.6
OM912701/ck/MEX/2826/21	9/13/2021	South	Broiler (28D)	choanal/lung	KP036503/ck/CH/LHB/121010/12	GI-13 (793B or 4/91)	3,495	20,314	23,808	99.99%	6,427	99.86
OM912702/ck/MEX/2833/21	9/27/2021	North	Broiler (21D)	choanal/lung	JN192154/ck/4/91(UK)	GI-13 (793B or 4/91)	3,495	1	3,492	99.91%	6,267	99.1
OM912703/ck/MEX/2860/21	10/12/2021	South	Broiler (28D)	choanal/lung	KY626045/BR/Ma5/16	GI-1 (Mass-type)	3,489	20,314	23,802	100%	50,190	99.46
OM912704/ck/MEX/2930/21	23-Nov-21	Center	Broiler (21D)	choanal/lung	DQ458217/AL/4614/98	GI-9 (Ark)	3,507	1	3,510	99%	5,441.35	94.65
OM912705/ck/MEX/2944/21	7-Dec-21	North	Broiler (28D)	choanal/lung	MK878536/ck/GA9977/19	GI-17 (CAV)	3,501	20,371	23,871	100%	6,377.6	99.54
OM912706/ck/MEX/2956/21	14-Dec-21	South	Broiler (28D)	choanal/lung	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,400.72	99.42
OM912707/ck/MEX/2960/21	14-Dec-21	Center	Layer (7.6 W)	spleen/bursa	GU393334/ck/US/Gray/60	GI-3 (Holte/Iowa-97)	3,498	20,382	23,874	100%	4,024.97	87.52
OM912708/ck/MEX/2961/21	15-Dec-21	Center	Broiler (21D)	choanal/lung	DQ458217/AL/4614/98	GI-9 (Ark)	3,510	1	3,510	100%	5,373.02	94.28

Lineage (serotype) classification is based on S1-gene sequence (18). Genome length excludes poly(A) tails. GenBank accession numbers are shown for the best BLASTn hit for each sequence.

FIGURE 2 | Maximum likelihood phylogenetic tree of nt sequences of subunit *S1*-gene, hypervariable regions I-III (HVRs I-III) of the S1-gene, and complete *S*-gene using GTR model in MEGA 6. The 33 sequences in this study (3 vaccine sequences highlighted in blue; 30 field sequences; color-coded based on sampling regions in Mexico) clustered with serotypes in 5 lineages of IBVs, i.e., lineages GI-1 (Mass-type serotypes; $n = 8$ sequences), GI-3 (Holte/Iowa/97 serotypes; $n = 2$ sequences), GI-9 (Ark-like serotypes; $n = 8$ sequences), GI-13 (793B also known as 4/91 serotypes; $n = 14$ sequences), and GI-17 (CAVs; $n = 1$ sequence). The analysis involved 94 sequences. All positions with <95% site coverage were eliminated. The final datasets had 1,576, 757, and 3,431 positions for the complete S1-gene, HVRs I-III, and complete S-gene, respectively.

three different Mass-type strains, and the Mass-type/Conn recombinant vaccine 1623/19 sequence is 98.45% similar to the pathogenic field strain MN696791/ck/T&T/18RS1461-3/14 isolated from a vaccinated broiler with respiratory disease (65). The Mass-type vaccine 1616/19 sequence and field sequence 2598/21 from Northern Mexico are 99.51 and 98.51% identical to strain MK937828/ck/CN/I1124/16, respectively. The remaining five field sequences from southern Mexico (2602/21, 2754/21, 2743/21, 2748/21, 2860/21), are 99.46–99.89% identical to a Brazilian government-licensed Ma5 vaccine strain [KY626045/BR/Ma5/16; (66)]. Phylogenetic clustering based on complete *S*-gene, *S1*-gene and HVRs I-III mirrored the BLASTn results, where the above-mentioned five sequences segregate in a subclade containing the Brazilian Ma5 and Canadian/European/Brazilian H120 vaccine strains (**Figure 2**). Based on the HVR I-III, the Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences are in a subclade containing the 2016 Chinese strain and a representative of Mass-type viruses (i.e., AY561711/Mas41), while the field sequence 2598/21 is in a distinct subclade containing the T&T and Connecticut vaccine strains. All of the 8 Mass-type sequences in this study contained an S2′ site downstream of the primary S1/S2 cleavage site (**Supplementary Table 4** and **Figure 1**).

Lineage GI-3 (Holte/Iowa-97). Based on complete *S*-gene sequences, viruses 2960/21 and 2731/21 from Central Mexico match closest (87.52 and 87.66% nt identities) to strain GU393334/ck/US/Gray/60 belonging to lineage GI-3, which comprises of respirotropic and nephropathogenic Holte, JMK, Gray and Iowa-97 viruses (18, 67). However, BLASTn searches using complete genome sequences returned closest hit (93.39 and 92.99%) to the pathogenic

field strain MH779860/ck/USA/Ark99/14 of lineage GI-9 (**Supplementary Table 6**). From **Figure 2**, complete *S*-gene does not phylogenetically place the two sequences with serotypes in either lineages GI-3 or GI-9; rather, the sequences are in a distinct subclade within a larger clade containing California variants (CAVs; lineage GI-17). However, based on the *S1*-gene and HVRs I-III, these sequences cluster with lineage GI-3 serotypes, but in a distinct subclade containing strain AF352831/BL-56/96, which has been previously described as uniquely Mexican (25).

Lineage GI-9 (Ark-Like). The North American lineage GI-9 serotypes were implicated in the rolling reactions in vaccinated flocks and the persistence of respiratory syndromes in flocks (25). Based on the complete *S*-gene sequences (**Table 4**), four sequences from Central Mexico (2930/21, 2961/21, 2562/21, and 2563/21), and four from Southern Mexico (2742/21, 2956/21, 2359/20, and 2360/20), are all closely-related to DQ458217/US/AL/4614/98 (94.19–94.65% nt identities), an Ark-DPI-derived vaccine virus originally isolated from a 40-day-old chicken with respiratory disease (68). Complete genome sequences returned similar hits, except that seven sequences are closest to an Ark pathogenic field strain (94.07–99.82% nt identity), while an attenuated ArkGA vaccine virus (MH779857/ck/USA/ArkGA-P20/15) is the closest match to sequence 2961/21 with 93.7% nt identity (**Supplementary Table 6**). Based on complete *S-*/*S1*-gene and HVRs I-III, all 8 sequences phylogenetically cluster in separate subclade distinct from other Ark-like strains, supported by bootstrap values of 100% (*S-*/*S-1* genes) and 83% (HVRs I-III). As in the case of Mass-type viruses, all 8 Ark-like sequences contained the S2′ cleavage (**Supplementary Table 4** and **Figure 1**).

Lineage GI-13 (793B, 4/91 or CR88). Genome sequences of 13 field samples and the 4/91 vaccine variant 1619/19 match closest to and cluster with lineage GI-13 serotypes (see **Table 4**, **Supplementary Table 6,** and **Figure 2**). Although field and vaccine 793B-like sequences have a near global presence, they have rarely been reported in Latin America. Both complete genome and S-gene of the 4/91 vaccine variant 1619/19 is closest MN548285/ck/UK/CR88/11 (99.94% and 99.98% nt identities, respectively). The 13 field sequences match to different strains depending on sampling regions (**Table 4**). Sequences 2353/20, 2354/20, 2592/21, 2752/21, 2753/21, and 2826/21 from Southern Mexico are 99.86–100% identical to KP036503/ck/CH/LHB/121010/12 based on complete *S*-gene sequences, while complete genome sequences are 99.87–99.97% identical to isolate MZ367369/ck/Belgium/4134 001/19. Six of the seven sequences from Northern Mexico (2721/21, 2723/21, 2725/21, 2818/21, and 2819/21) have nt identities of 99.54–99.63% to isolate KP118880/ck/CH/LHB/130927/13, while the field sequence 2833/21 is 99.1% identical to isolate JN192154/ck/4/91. Sequence 2523/21, from Central Mexico, is 99.28% similar to strain KP118891/ck/CH/LHLJ/111246/11 based on the complete S-gene, but the complete genome sequence is 92.96% identical MT665806/ck/Canada/17-038913/17. The BLASTn results are reflected in the phylogenetic clustering based on complete *S*-gene, *S1*-gene and the HVRs I-III, where the vaccine sequence group with CR88 strain is in a distinct subclade away from the field sequences (**Figure 2**).

Lineage GI-17 (California Variants; CAVs). Complete *S*-gene sequence of the field sequence 2944/21 from Northern Mexico is 99.54% identical to a Delmarva (DMV) virus isolated from broiler chickens [MK878536/ck/GA9977/19; (69)]—**Table 4**. The complete genome sequence is 96.04% identical to strain MN512437/ck/Can/18-048430/17. Phylogenetically, the *S-*/*S1*-gene trees group the Mexican field sequence with strains GA9977 and MT176422/ck/PA/P1810234-KD/18 in a subcluster distinct from other Canadian and USA viruses, but the HVRs I-III tree does not include the P1810234-KD strain (**Figure 2**). Again, like in the Mass-type and Ark-like viruses, the CAV-like sequence 2944/21 contained S2′ cleavage (**Supplementary Table 4** and **Figure 1**).

We also performed phylogenetic analyses based on complete genome and Rep1ab (**Figure 3**), structural genes E, M and N (**Supplementary Figure 1**), and 3ab, 5a/b, and 6b (**Supplementary Figure 2**). Notably, none of the 33 sequences cluster with lineage GI-11 (SAI serotypes), which have been described as serologically and phylogenetically unique to South America (29–31). However, the previously described Mexican SAI viruses group in distinct subclades within larger clades containing Mass-type (based on nt sequences of the complete genome, Rep1b and 3a/b) and 793B (based on nt sequences of the complete *S*-gene, structural genes E/M and 5a/b) strains.

Indels and Mutations in the HVRs of S1-Gene Sequence

Alignment of translated S1 HVR I (residues 37–88), HVR II (residues 115–146) and HVR III (residues 282–301) sequences revealed considerable variations among the 33 IBVs (**Figure 4**).

HVR I. Six residues in HVR I of the Mass-type prototype AY851295/M41 (i.e., N38, H43, S56, P63, I66, and T69) are critical for viral attachment to respiratory tract tissues (14). Within the Mass-type sequences, field sequence 2598/21 from Northern Mexico, and the Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences, have an asparagine in position 38 (N38) similar to the M41 sequence, while all five Mass-type sequences from Southern Mexico have an aspartic acid at position 38 (D38) similar to the Brazilian Ma5 vaccine, UK H120 and 2006 Mass41 sequences. The consensus histidine at position 43 (H43) is conserved across all 33 sequences, except in the 4/91 vaccine variant 1619/19 sequence (lineage GI-13), which has a substitution of histidine with tyrosine (H43Y) similar to the CR88 strain from UK. Serine and threonine residues at positions 56 (S56 and T56) are conserved amongst the 793B-like and Mass-type sequences, respectively, but varies among the Ark-like sequences. Further, the critical amino acid at position 63 (S63) is conserved in the sequences across the five IBV lineages used in the alignment, except in field sequences 2731/21 and 2960/21 (from Central Mexico) and sequence 2944/21 (from Northern Mexico), which have S63G substitution, and sequence 2598/21 (from Northern Mexico) with an S63R substitution. Position 66 is largely conserved across

TABLE 4 | Recombination events in the complete *S*-gene sequences in this study.

Potential recombinant seq		Breakpoints			"Major parent" sequence (nt)[a]		"Minor parent" sequence (nt)[b]		Confirmation of recombination event	
Sequence	**Lineage**	**Start -> end (nt length)**	**Start -> end (aa length)**	**Domain in aa seq**	**Sequence**	**Lineage**	**Sequence**	**Lineage**	**Detection algorithm[c]**	**Corrected av. *p*-value**
OM912707/ck/MEX (Central)/2960/21	GI-3 (Holte/Iowa-97)	2,965 -> 3,355 (391)	989 -> 1,118 (130)	HR2	FJ904713/ck/US/Mass41/06 (84%)	GI-1 (Mass-type)	OM912693/ck/MEX (Central)/2563/21 (99.5%)	GI-9 (Ark-like)	GENECONV (2), BootScan (2), MaxChi (2), Chimaera (2), SiSscan (2), 3Seq (2)	8.115 E-18
OM912676/ck/MEX (Central)/2731/21	GI-3 (Holte/Iowa-97)	2,965 -> 3,355 (391)	989 -> 1,118 (130)	HR2	FJ904713/ck/US/Mass41/06 (84%)	GI-1 (Mass-type)	OM912693/ck/MEX (Central)/2563/21 (99.1%)	GI-9 (Ark-like)	RDP (1), GENECONV (2), BootScan (2), MaxChi (2), Chimaera (2), SiSscan (2), 3Seq (2)	5.934 E-07
OM912696/Mass-type-Conn recombinant vaccine strain/1623/19	GI-1 (Mass-type)	1,868 -> 2,550 (683)	624 -> 850 (227)	FP – HR1	DQ834384/USA/M41/56 (98%)	GI-1 (Mass-type)	FJ904716/US/Conn46 vaccine/96 (99.1%)	GI-1 (Mass-type)	GENECONV (1), BootScan (1), MaxChi (1), Chimaera (1), SiSscan (1), 3Seq (1)	2.527 E-10
OM912693/ck/MEX (Central)/2563/21	GI-9 (Ark-like)	1,710 -> 2,395 (686)	571 -> 798 (228)	FP – HR1	Unknown* (GU393331/ck/USA/Cal56b/91)	GI-17 (CAV)	MN512438/ck/Can/18-049707/17 (97.2%)	GI-17 (CAV)	GENECONV (7), BootScan (7), MaxChi (7), Chimaera (7), SiSscan (7), 3Seq (4)	5.19 E-26
OM912703/ck/MEX (South)/2860/21	GI-1 (Mass-type)	1,749 -> 2,272 (524)	584 -> 757 (174)	FP	FJ888351/ck/NL/H120/60 (99.8%)	GI-1 (Mass-type)	MK937828/ck/CN/I1124/16 (100%)	GI-1 (Mass-type)	GENECONV (1), BootScan (1), MaxChi (1), Chimaera (1), SiSscan (1), 3Seq (1)	4.509 E-06
OM912704/ck/MEX (Central)/2930/21	GI-9 (Ark-like)	1,707 -> 2,392 (686)	570 -> 797 (228)	FP – HR1	Unknown* (GU393331/ck/USA/Cal56b/91)	GI-17 (CAV)	MN512438/ck/Can/18-049707/17 (93.7%)	GI-17 (CAV)	GENECONV (7), BootScan (7), MaxChi (7), Chimaera (7), SiSscan (7), 3Seq (4)	1284 E-15
OM912705/ck/MEX (North)/2944/21	GI-17 (CAV)	1,701 -> 2,386 (686)	568 -> 795 (228)	FP – HR1	Unknown* (GU393331/ck/USA/Cal56b/91)	GI-17 (CAV)	MN512438/ck/Can/18-049707/17 (95.5%)	GI-17 (CAV)	GENECONV (7), BootScan (7), MaxChi (7), Chimaera (7), SiSscan (7), 3Seq (4)	2.284 E-19
OM912697/4/91 vaccine variant strain/1619/19	GI-13 (793B or 4/91)	2,786 -> 3,156 (371)	930 -> 1,052 (123)	HR2	ck/MEX (North)/2833/21 (96.8%)	GI-13 (793B or 4/91)	Unknown* (FJ904716/US/Conn46 vaccine/96)	GI-1 (Mass-type)	RDP (1), GENECONV (1), BootScan (2), MaxChi (2), Chimaera (2), 3Seq (2)	7.266 E-13
OM912685/ck/MEX (North)/2598/21	GI-1 (Mass-type)	56 -> 681 (626)	20 -> 227 (208)	RBD (HVR-I – HVR-II)	MK937828/ck/CN/I1124/16 (99.8%)	GI-1 (Mass-type)	FJ904716/US/Conn46 vaccine/96 (99.8%)	GI-1 (Mass-type)	GENECONV (1), BootScan (1), MaxChi (1), Chimaera (1), SiSscan (1), 3Seq (1)	7.466 E-27

The analysis, which involved a total of 70 nucleotide sequences, was performed using RDP, GENECONV, MaxChi, BootScan, SiSscan, 3Seq, and Chimaera programs executed in the RDP v4.101 suite (54). The nt identities between the transferred fragment in the recombinant and the major/minor parent sequences are indicated. Phylogenetic relationships of the potential recombinants (based on alignment of the regions derived from the major parents), the major/minor parents and other viruses are presented in ***Figure 3****. GenBank accession numbers are shown for each sequence.*

[a] *"Major parent" indicates sequence in other viruses most closely related to the sequence surrounding the transferred fragment.*

[b] *"Minor parent" indicates the sequence closely related to the fragment in the recombinant.*

[c] *For each detection algorithm, the number of sequences in which the transferred fragment was detected is shown in brackets.*

**"Unknown" indicates that only one parent and a recombinant need be in the alignment for transferred fragment to be detectable (sequence in bracket was used to infer existence of the unknown parent).*

FP, membrane-fusion peptide; HR, heptad repeat region; HVR, hypervariable region; RBD, receptor-binding domains.

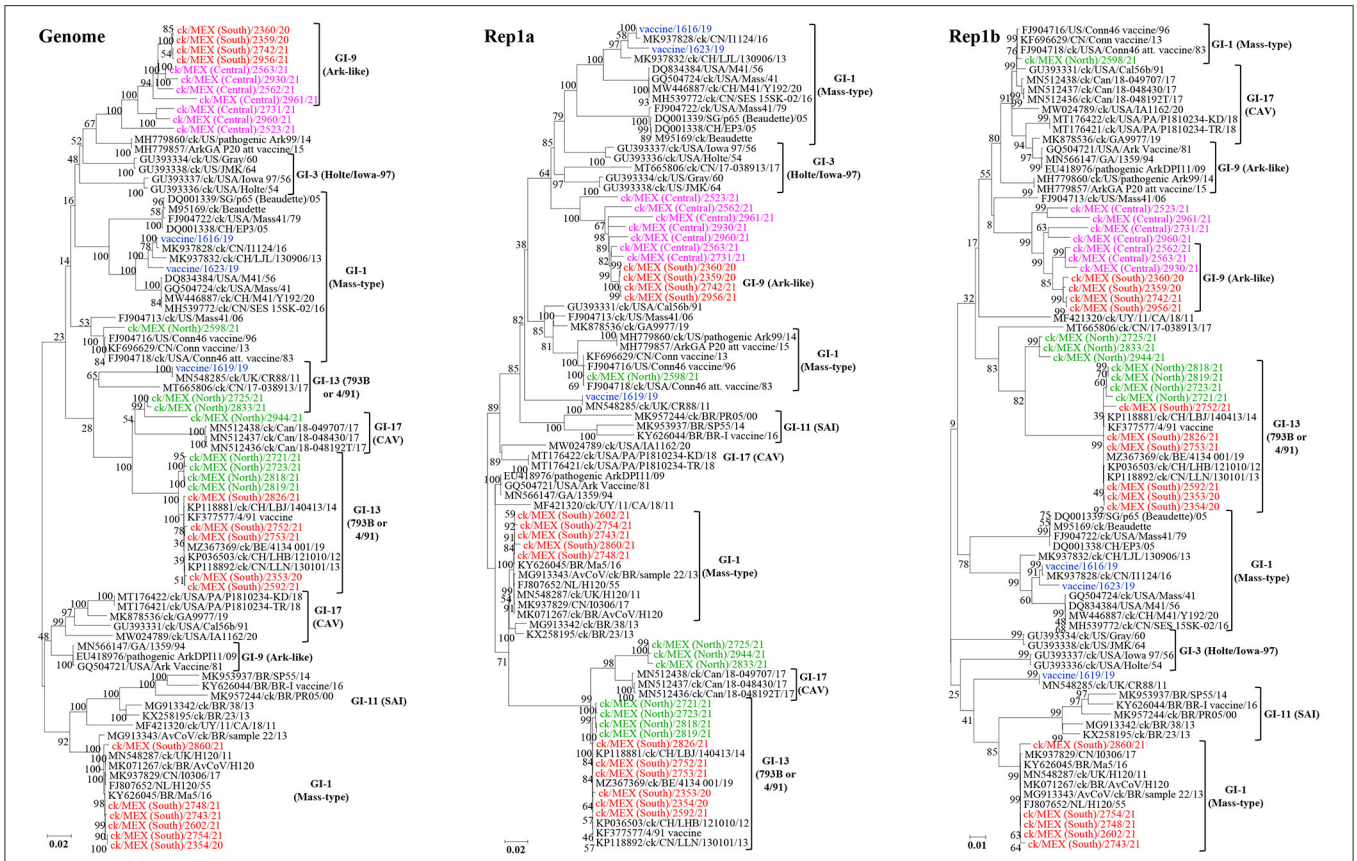


FIGURE 3 | Maximum likelihood phylogenetic tree nt sequences of complete genome, Rep1a, and Rep1b using GTR model in MEGA 6. The 3 vaccine sequences are highlighted in blue color; the 30 field sequences are color-coded based on sampling regions in Mexico. The analysis involved 83 sequences. All positions with <95% site coverage were eliminated. The final dataset had 26,895 (complete genome), 11,782 (Rep1a), and 8,037 (Rep1b) positions.

sequences in all 5 lineages, while position 69 appear to be lineage-specifically conserved, except in the Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences, which have T69 compared the six field sequences from Southern ($n = 5$) and Northern ($n = 1$) Mexico in the Mass-type group.

In addition to substitutions, HVR I has three instances of deletion. One is in field sequences 2731/21 and 2960/21 from Central Mexico (in lineage GI-3) with a 2-residue deletion (positions 61–62), which is also present in the Mexican BL-56 strain. Another instance is a 4-residue deletion (positions 57–60) in the field sequence 2944/21 from Northern Mexico (lineage GI-17), which is also present in the DMV strains MK878536/ck/GA9977/19 and MN512438/ck/Can/18-049707/17. The third is a 3-residue deletion (positions 60 to 62) in field Mass-type sequence 2598/21 from Northern Mexico, which is present in strain QKE31017/T&T/18RS1461-8/14) and three Conn vaccine strains from USA.

HVR II. Most amino acid variations in HVR II are between the positions 116–121 and 138–147 (**Figure 4**). For example, in lineage GI-13, all the six field sequences from Northern Mexico, and the 4/91 vaccine variant 1619/19 sequence have N117 (similar to CR88 strains from UK and a field variant from Ireland), while all six field sequences from Southern Mexico have S117 (similarly present in a pathogenic 4/91 strain from UK and two other Chinese strains). The Mass-type vaccine 1616/19 and Mass-type/Conn recombinant vaccine 1623/19 sequences, and all the field sequences in lineage GI-1 (one sequence from northern and five from Southern Mexico) have a 2-residue deletion between positions 117 and 120. Additionally, some of the field sequences from Northern ($n = 5$) and Central ($n = 2$) Mexico have a G145R substitution, also present in the Canadian and USA DMV-like strains. All eight field Ark-like sequences (four each from Central and Southern Mexico), and the CAV-like sequence 2944/21 from Northern Mexico have a single residue insertion between positions 142 and 143.

HVR III. One of the variations in the HVR III is in the Ark-like sequences, where all eight field sequences have E284V substitution compared to the USA pathogenic field strain AYA44731/US/Ark99. This substitution is also present in the CAV viruses, including in field sequence 2944/21 from Northern Mexico. Another variation is in the 4/91 vaccine variant 1619/19 sequence and the Holte/Iowa-97-like field sequence 2731/21 from Central Mexico, which have a single amino acid (alanine)

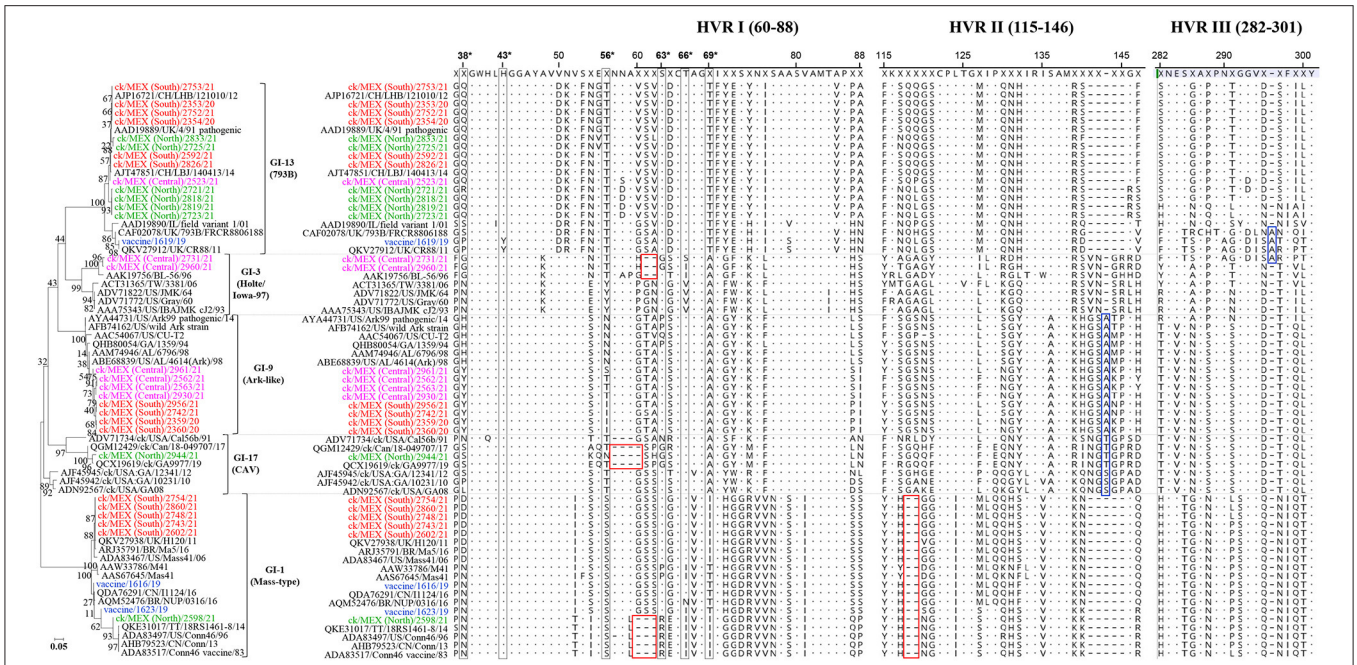


FIGURE 4 | Analyses of the *S1* glycoprotein HVRs I- III (amino acid residues 37-301). The phylogenetic tree was constructed using the JTT matrix-based model in MEGA and involved 67 sequences and a final dataset consisting of 253 positions. The 33 sequences in this study (3 vaccine sequences highlighted in blue; 30 field sequences; color-coded based on sampling regions in Mexico) assembled in this study are aligned with representatives of serotypes belonging to 5 IBV lineages. Dots and dashes in the alignments indicate identical and deleted amino acid residues, respectively. Amino acid residues critical for attachment of the Mass-type prototype (M41; GenBank accession No. AY851295) to chicken respiratory tract tissues are boxed in black [i.e., N38, H43, S56, P63, I66, and T69 (14)]; red and blue boxes indicate deleted and inserted amino acid residues, respectively.

insertion at position 296, which is also present in the CR88 strains AJ618986/UK/FR-88061-88 and MN548285/ck/UK/CR88/11.

Recombination Events

Neighbor-net analyses showed likelihood of intra-/inter-lineage recombination events between *S*-gene sequences. **Figure 5** represents a networked relationships among 70 *S*-gene sequences (including the 33 sequences from this study). Verification of recombination breakpoints using RDP4 resulted in identification of nine recombination events (**Table 4**), all of which reflected the results presented in **Figure 5**. All nine recombination events were statistically (corrected *p*-values of $\leq 1 \times 10^{-6}$) supported by at least six of the seven algorithms used for the analyses. Eight of the recombination events are located in the subunit S2 (in regions containing FP and/or HR1/2 domains), while the recombination event in the field sequence 2598/21 is in the N-terminal receptor domain, which contains HVRs I and II (**Table 4**).

Three recombination events are in Mass-type, i.e., in the Mass-type/Conn recombinant vaccine 1623/19 sequence, and the field sequences 2860/21 and 2598/21 from Southern and Northern Mexico, respectively. In all three events, both "major parent" sequence (i.e., sequence in other viruses most closely related to the sequence surrounding the recombinant/transferred regions) and "minor parent" sequence (i.e., sequence in other viruses most closely related to the recombinant regions) are Mass-type, with 99.8–100% nt identities between the recombinant region and the "parental" sequences (**Table 4**). Four other recombination events, all identified in Holte/Iowa-97-like sequences 2731/21 and 2960/21, and Ark-like sequences 2563/21 and 2930/21, all from Central Mexico, have their "parental" sequences in viruses from different lineages (**Table 4**). One of the recombination events (involving the 4/91 vaccine variant 1619/19 sequence) has field sequence 2833/21 as "major parent" and an unknown "minor" parent inferred to be a 1996 Mass-type vaccine strain (FJ904716/US/Conn46 vaccine/96). In this event, the nt identities between the recombinant region and the "parental" sequences is low (96–96.8%), which probably imply the recombinant regions could have accumulated further mutations, or that the recombination event happened long before the isolation and eventual divergence of the recombinant and parental viruses.

DISCUSSION

Infectious bronchitis is arguably one of the most important avian respiratory diseases in Mexico, and IBVs are commonly found in flocks, which, despite being vaccinated with the government-approved Mass-type vaccines, exhibit respiratory clinical signs consistent with IB. Ark-like, Mass-type, Holte/Iowa-97, and SAI strains have been previously identified in Mexico using molecular and serological assays (18, 29–31). Most of the Mexican variants have been reported from commercial flocks, but the likelihood of spillover to backyard poultry via introduction of surplus chickens from commercial enterprises, and via use

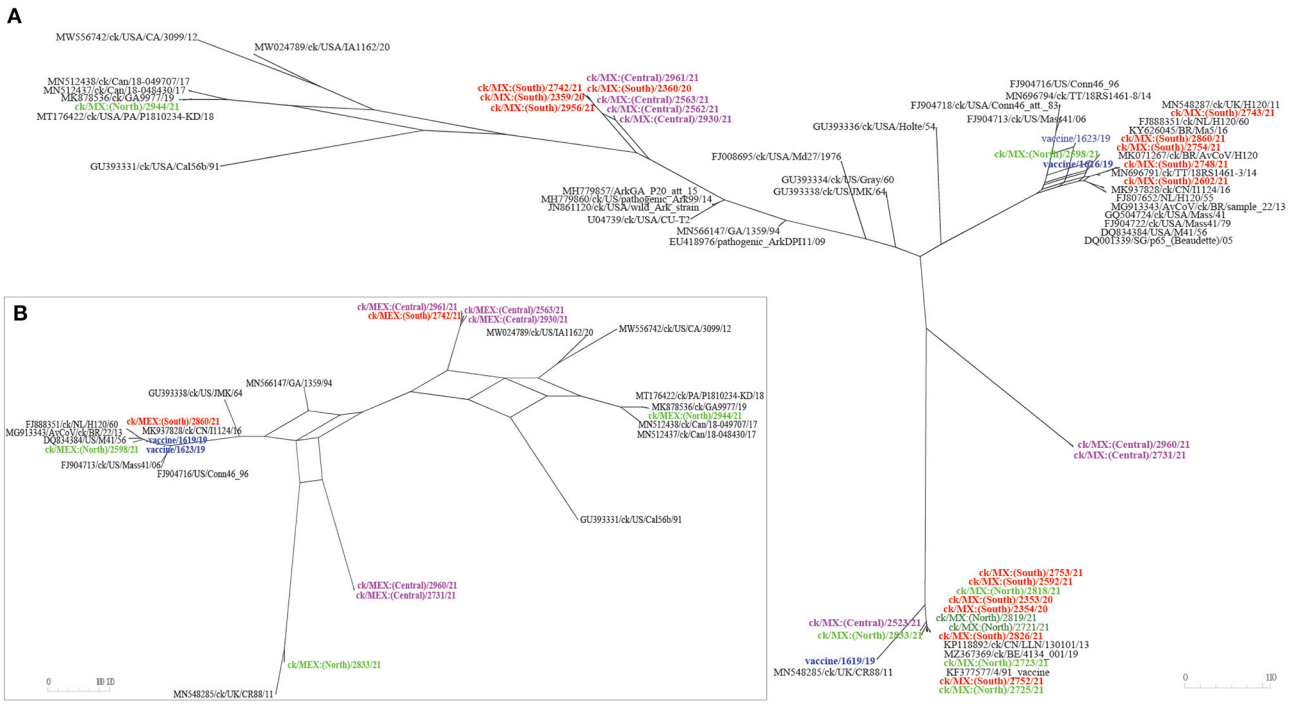


FIGURE 5 | Reticulate network using complete *S*-gene sequences constructed using SplitsTree5 v 5.0.0 (53). The network predicts putative evolutionary histories of the IBVs, where the internal nodes and the edges correspond to ancestral taxa and patterns of descents, respectively. Nodes with more than two parents represent reticulate events (e.g., recombination, horizontal gene transfer, or hybridization). "Split" is a partition of the taxa into two subsets with the edges separating the taxa subsets of the split from those on the other side of the split (the length of the edge in the network is proportional to the weight of the split it is associated with, which is analogous to branches in conventional phylogenetic trees). **(A)** The input contained 70 taxa and 101 trees constructed using Splits Network Algorithm with default options to obtain a splits network with 153 nodes and 165 edges. In the figure inset **(B)**, the input consisted of a subset of 28 taxa and 101 trees (from a splits network with 66 nodes and 72 edges). The recombination events of the sequences in panel B of this figure as determined using RDP4 are shown in **Table 4**.

of attenuated vaccines, cannot be ruled out. It is notable that all the IBVs reported in the current study were obtained from clinical samples from apparently healthy flocks. This observation may be due to mild clinical signs. Additionally, a lack of clinical signs of viral infection could be due to persistent (i.e., latent, chronic or slow) infection where the virus is not cleared from infected birds but remains in specific host cells. Although some studies have suggested occurrence of this type of infection for IBVs (70), the phenomenon remains to be convincingly demonstrated.

Studies have demonstrated the versatility of FTA cards as a low-cost option for collection, storage, transportation and preservation of genetic materials from field samples for the surveillance of various viral agents (71–74). Our study demonstrates that, despite the weakness of FTA cards in yielding lower quantities of mostly fragmented RNAs, resulting data is of sufficiently high quality to allow the assembly of full-length viral genomes, even when the clinical samples are prepared under field conditions. This opens up the advantages of FTA cards to NGS-based diagnostics of avian viruses via direct RNA sequencing from field-collected samples without the need of passage in eggs or transportation in liquid media. Avoiding egg-passage is advantageous for several reasons, including biosafety issues (there is no live virus manipulation, amplification, escape issues), and the absence of selective virus amplification (passage may select variants with a minority representation in the host, or introduce genetic changes).

High abundance of host/bacterial RNAs in samples prepared under field conditions, which, from our experiences and those of others, can constitute over 95% of NGS reads (38), results in low sensitivity in detection of viral RNAs. These challenges notwithstanding, our optimized protocols increased NGS sensitivity (high levels of virus-specific reads vs. low levels of host-/bacteria-specific reads), which produced complete genome and/or S-gene sequences (with optimal read coverage depth and coverage). Some of the assembled genome sequences were missing nt at the 5′- and/or 3′- termini, but all coding regions had sufficient depth coverage (≥10X). Furthermore, within single NGS runs (48 multiplexed samples per run), we obtained complete and/partial genome/gene sequences of various RNA viruses belonging to six taxonomic families, some of which were co-isolated with IBV.

Although the general genome organization (5′UTR-[Rep1a-Rep1b-S-3a-3b-E-M-4b-4c-5a-5b-N-6b]-3′UTR) slightly differs from many previously reported IBVs regarding presence of 4b, 4c, and 6b, it is consistent with some IBVs previously reported in various geographical regions across the globe (11, 75–77). The

three accessory genes are thought to be strain-dependent/species-specific, but their presence and/or genomic organization are not a prerequisite to production of viable virus progeny, and their role in viral pathogenesis remains undetermined (11, 78). The presence of these auxiliary and apparently non-essential accessory genes is attributable to the IBV's flexible genome, which tolerates not only for stable insertion of novel genes, but also for reorganization (78). The genomic organization of the essential genes is however canonically conserved, including the overlaps between the six genes, except in overlaps between genes 3 and 4 in the 4/91 vaccine variant 1619/19 sequence and the field sequence 2944/21 from Northern Mexico. Most of the variations in the genome organization are in overlaps among the accessory genes and in the presence of VncRNAs in gene 4 (between the structural M and 4b) and gene 6 (between the structural N and the 6b). The sources of gene overlaps remain unknown, but one school of thought is that they result from either utilization of previously unused ORFs to create novel genes, or by adjusting locations of start/stop codons into sequences of existing genes (79).

Both the length and key physical properties at specific positions within the 20-residue region harboring the S1/S2 cleavage motif in all the 33 sequences are evidence that the *S*-glycoprotein of the viruses undergo sufficient furin-mediated cleavage and viral-host membrane fusion (61). In addition to the S1/S2 cleavage, all eight Mass-type, eight Ark-like and the single CAV viruses assembled in this study contain the auxiliary S2'cleavage site downstream the primary S1/S2 cleavage site, which results in a 153-amino acid-long dual-cleavage peptide. Although it does not directly influence IBV pathogenesis, this peptide is postulated to influence the S glycoprotein conformation in some CoVs; deletion of the peptide was demonstrated to affect fusion and recovery of Vero cell-adapted Baudette mutants (8). Although the S2' site is not absolutely required, it is thought to enhance viral infection and replication in some CoVs (80). Nevertheless, it remains unclear why the S2' cleavage site is absent in the sequences belonging to lineages GI-3 ($n = 2$ sequences) and GI-13 ($n = 14$ sequences), regardless of the sample types.

Results of the phylogenetic classification agreed with the comparative pairwise nt sequence homologies between the vaccine vs. field sequences as well as the homologies shared with other IBVs in genomic databases. To the best of our knowledge, 793B- and CAV-like strains have not been previously reported in Mexico. However, there are unpublished reports f the use of 793B vaccines in Mexico. CAVs are indigenously North America currently. CAVs are indigenously North America currently consisting of 12 published viruses from Pennsylvania, California and Alabama, causing respiratory, renal and reproductive diseases (18). Except the 4/91 vaccine variant 1619/19 sequence and the field sequence 2523/21 from Central Mexico, 793B-like sequences are from Southern and Northern Mexico ($n = 6$ sequences each). The only CAV-like sequence is the field sequence 2944/21 from Northern Mexico, which clustered with, among others, isolate MN512437/ck/Can/18-048430/17. The Canadian 18-048430 virus is similar to a DMV strain originally isolated from a virus outbreak in a commercial broiler flock in the Delmarva peninsula in 2011 and the DMV-like strain MK878536/ck/GA9977/19 isolated from broiler chickens in Georgia, USA in 2019 (69, 81). None of the 33 sequences in this study clustered with indigenous SAI viruses based on the *S-/S1* and HVRs I-III trees. However, based on complete genome and other specific genomic regions, the SAI viruses clustered in distinct subclades within larger clades containing strains from other lineages, which, coupled with the clustering observed from S-/S1 and HVRs I-III trees, could be interpreted to imply decreased prevalence of the SAI viruses, resulting in the emergence of new, more fit field variants.

Since the coronaviruses can undergo recombination to result in novel variants, and since changes in the spike protein gene can result in shifts in antigenicity and/or tropism, we screened specifically for mutations and recombination in the S-gene (82). We noted various indels and point mutations in the subunit S1 HVRs I-III across the five IBV lineages. The most notable are in the HVR I (residues 38–69), which has been experimentally demonstrated in some IBVs to be critical for binding of S glycoprotein to host's respiratory tract tissues (14). Examples include 2-residue deletions between 60–63 and P63G substitutions in the sequences from Central (2731/21 and 2960/21) and Northern (2944/21) Mexico. Another example is a 3-residue deletion (between positions 69-63) coupled with a P63R substitution in sequence 2598/21 (Northern Mexico). Further, a H43Y (found in the sequence of the 4/91 vaccine variant 1619/19 in our study) was postulated to enhance the fitness of ArkDPI vaccine strain (via efficiency in binding to tracheal membranes) (83). Whether these changes in the S1 spike result in subsequent change in host cell binding, tissue tropism, or evasion of neutralizing antibodies remains to be determined.

The recombination events, presumably between vaccine and field viruses (intra- and inter-lineages), were found in the HVRs 1 and II of subunit S1 (in the field sequence 2598/21) and in the FP, HRs 1 and 2 domains of subunit S2 (in two vaccine and six field sequences). Recombination in the HVRs and in the above-mentioned domains, coupled with the various indels and mutations found in these regions, could modulate the S glycoprotein fusion properties, with the potential of variants to not only broaden their tissue-tropism (and possibly host-range), but also to efficiently adapt to naïve hosts as previously suggested (82). Additionally, these variants can potentially be a result of immune selection pressures (36). An example of this is the DE072 vaccine, which when first identified, showed more relatedness to a Dutch variant that to the US variants, but after undergoing a decade-long use, the strain evolved to a poor vaccine candidate (24).

CONCLUSIONS

This study has demonstrated FTA cards as adoptable, low-cost option for untargeted discovery and full-length sequencing of avian viruses from field-collected clinical samples from various tissues. Our data demonstrates that multiple distinct IBV serotypes/strains are co-circulating in Mexican commercial chickens, with the high likelihood of intra- and inter-lineage recombination, as well as indels and point mutations which are potentially driving the generation of new subpopulations of field variants capable spreading and adapting to chicken populations in the country. It should be further investigated whether strains

emerging from the commercial enterprises may have spilled over to backyard poultry, and whether they have further evolved into strains that are distantly related to the predominantly SAI strains. More importantly, our data reiterates the need for enhanced surveillance of IBVs in Mexico and the Latin America region, as well as a review of the vaccines currently used in the control of IBVs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, OM912676, OM912677, OM912678, OM912679, OM912680, OM912681, OM912682, OM912683, OM912684, OM912685, OM912686, OM912687, OM912688, OM912689, OM912690, OM912691, OM912692, OM912693, OM912694, OM912695, OM912696, OM912697, OM912698, OM912699, OM912700, OM912701, OM912702, OM912703, OM912704, OM912705, OM912706, OM912707, and OM912708.

AUTHOR CONTRIBUTIONS

CA, ED, and HK: conceptualization. CA, DS, and ED: funding acquisition. DS, SL, NC, and ED: project administration and coordination. HK: methodology and writing—original draft. HK, JV, and CL: data curation and formal analyses. CL, DS, and CA: writing—review and editing. All authors read and agreed to the published version of the manuscript.

FUNDING

This research was funded by Agricultural Research Service (ARS), USDA CRIS, and Oak Ridge Institute for Science and Education (ORISE) postdoctoral appointment. The funders had no role in data analyses or interpretation, or in the writing of the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge Dawn Williams-Coplin, Jesse Gallagher (SEPRL), and the field sampling teams (BIAH) for technical assistance in sample collection and preparation. The authors also acknowledge the contributions made by the BIAH field sample collection team (Alberto Carrillo, Alejandro Hori, Amauri Carrillo, David Garcia, and Marcos Jimenez).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.931272/full#supplementary-material

Supplementary Table 1 | Background information and sequencing data of the 55 samples in which IBV RNA was detected in this study. The quantification of the dsDNA (HiFi PCR products), sequencing libraries, proportions of the host (chicken) reads, and the IBV-specific reads obtained from the NGS data are shown. Complete genome sequences ($n = 33$), partial genome sequences ($n = 6$), complete *S*-gene sequences ($n = 5$), and partial *S*-gene sequences ($n = 11$) were assembled from NGS reads in the current study. Other identified microbial agents (viral and bacterial) that are of avian interest based on literature and expert knowledge are listed in the table.

Supplementary Table 2 | Genes, coding regions and nt lengths of the 33 IBV sequences identified in this study. The nt lengths (and gene intervals in brackets) of each of the genes and the 5'-/3'-UTRs are shown.

Supplementary Table 3 | The 15 accessory proteins (2–15) that are proteolytically processed from Rep1ab by the virally-encoded PL^{pro} and $3CL^{pro}$. For each of the 33 IBV sequences assembled in this study, the size (amino acid) of each cleavage product are indicated. The position (amino acid residue) of each product on the Rep1a (2–11) and Rep1b (12–16) are shown in brackets.

Supplementary Table 4 | Analysis of the *S*-gene nt sequences assembled in this study. The arrow (↓) indicates the S1/S2 proteolytic cleavage site. The conserved cysteine-flanked region in the S2 is indicated (amino acid residues in bold and underlined). The numbers in subscript indicate the position of the amino acid residues in the S protein sequence.

Supplementary Table 5 | Comparative pairwise homologies of IBV gene nt sequences between the field and vaccine sequences assembled in this study. Field sequences with the highest pairwise homologies to the vaccine sequences are in bold font, while sequences with the lowest are underlined. The lineage (serotype) classification is based on *S1*-gene sequence according to Valastro et al. (18).

Supplementary Table 6 | BLASTn results of the 33 complete genome nt sequences assembled in this study. The lineage (serotype) classification is based on S1-gene sequence according to Valastro et al. (18). GenBank accession numbers are shown for the best BLASTn hit for each sequence.

Supplementary Figure 1 | Maximum likelihood phylogenetic tree of nt sequences of the envelope (E), membrane (M) and nucleocapsid (N) genes using T92 model in MEGA 6. The 3 vaccine sequences assembled in this study are highlighted in blue color; the 30 field sequences are color-coded based on sampling regions in Mexico. The analysis involved 74 sequences. All positions with <95% site coverage were eliminated. The final datasets for the *E*, *M*, and *N* genes had 304, 659, and 1,215 positions, respectively.

Supplementary Figure 2 | Maximum likelihood phylogenetic tree of nt sequences of accessory genes in (A) gene 3 (3a/3b), gene 5 (5a/5b) and gene 6 (6b) using T92 model in MEGA 6. The 3 vaccine sequences assembled in this study are highlighted in blue color; the 30 field sequences are color-coded based on sampling regions in Mexico. The analysis involved 74 (3a/b and 5a/b) and 63 (6b) sequences. All positions with less than 95% site coverage were eliminated. The final datasets for the 3a/b, 5a/b, and 6 genes had 365, 443, and 130 positions, respectively.

REFERENCES

1. Jackwood MW, de Wit S. CHAPTER 4: Infectious Bronchitis. In: Swayne DE, Boulianne CM, Logue CM, McDougald LR, Nair V, Suarez DL, de Wit S, Grimes T, Johnson D, Kromm M, et al., editors. *Diseases of Poultry*. Hoboken, NJ: John Wiley & Sons, Ltd. (2020). p. 167–88.
2. Najimudeen SM, Hassan MSH, Cork SC, Abdul-Careem MF. Infectious bronchitis coronavirus infection in chickens: multiple system disease with immune suppression. *Pathogens*. (2020) 9:779. doi: 10.3390/pathogens9100779
3. Ignjatović J, Sapats S. Avian infectious bronchitis virus. *Rev Sci Tech Off Int Epiz*. (2000) 19:493–508. doi: 10.20506/rst.19.2.1228
4. Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev*. (2005) 69:635–64. doi: 10.1128/MMBR.69.4.635-664.2005
5. Ziebuhr J. The coronavirus replicase. In: Enjuanes LJ, editor. *Coronavirus Replication and Reverse Genetics. Current Topics in Microbiology and Immunology*. Heidelberg: Springer-Verlag GmbH (2005). p. 57–94.
6. Malone B, Urakova N, Snijder EJ, Campbell EA. Structures and functions of coronavirus replication–transcription complexes and their relevance

for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol.* (2022) 23:21–39. doi: 10.1038/s41580-021-00432-z
7. Brierley I, Boursnell M, Binns M, Bilimoria B, Blok V, Brown T, et al. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* (1987) 6:3779–85. doi: 10.1002/j.1460-2075.1987.tb02713.x
8. Yamada Y, Liu DX. Proteolytic activation of the spike protein at a novel RRRR/S motif is implicated in furin-dependent entry, syncytium formation, and infectivity of coronavirus infectious bronchitis virus in cultured cells. *J Virol.* (2009) 83:8744–58. doi: 10.1128/JVI.00613-09
9. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res.* (2006) 66:193–292. doi: 10.1016/S0065-3527(06)66005-3
10. Hodgson T, Britton P, Cavanagh D. Neither the RNA nor the proteins of open reading frames 3a and 3b of the coronavirus infectious bronchitis virus are essential for replication. *J Virol.* (2006) 80:296–305. doi: 10.1128/JVI.80.1.296-305.2006
11. Feng K, Chen T, Zhang X, Shao G, Cao Y, Chen D, et al. Molecular characteristic and pathogenicity analysis of a virulent recombinant avian infectious bronchitis virus isolated in China. *Poult Sci.* (2018) 97:3519–31. doi: 10.3382/ps/pey237
12. Cavanagh D, Davis PJ, Mockett AA. Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Res.* (1988) 11:141–50. doi: 10.1016/0168-1702(88)90039-1
13. Moore K, Jackwood M, Hilt D. Identification of amino acids involved in a serotype and neutralization specific epitope with in the s1 subunit of avian infectious bronchitis virus. *Arch Virol.* (1997) 142:2249–56. doi: 10.1007/s007050050239
14. Promkuntod N, Van Eijndhoven R, De Vrieze G, Gröne A, Verheije M. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology.* (2014) 448:26–32. doi: 10.1016/j.virol.2013.09.018
15. Moreno A, Franzo G, Massi P, Tosi G, Blanco A, Antilles N, et al. A novel variant of the infectious bronchitis virus resulting from recombination events in Italy and Spain. *Avian Pathol.* (2017) 46:28–35. doi: 10.1080/03079457.2016.1200011
16. Cook JK, Jackwood M, Jones R. The long view: 40 years of infectious bronchitis research. *Avian Pathol.* (2012) 41:239–50. doi: 10.1080/03079457.2012.680432
17. Zúñiga S, Cruz JL, Sola I, Mateos-Gómez PA, Palacio L, Enjuanes L. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol.* (2010) 84:2169–75. doi: 10.1128/JVI.02011-09
18. Valastro V, Holmes EC, Britton P, Fusaro A, Jackwood MW, Cattoli G, et al. S1 gene-based phylogeny of infectious bronchitis virus: an attempt to harmonize virus classification. *Infect Genet Evol.* (2016) 39:349–64. doi: 10.1016/j.meegid.2016.02.015
19. Kingham B, Keeler C Jr, Nix W, Ladman B, Gelb J Jr. Identification of avian infectious bronchitis virus by direct automated cycle sequencing of the *S-1* gene. *Avian Dis.* (2000) 44:325–35. doi: 10.2307/1592547
20. Ma T, Xu L, Ren M, Shen J, Han Z, Sun J, et al. Novel genotype of infectious bronchitis virus isolated in China. *Vet Microbiol.* (2019) 230:178–86. doi: 10.1016/j.vetmic.2019.01.020
21. Marandino A, Pérez R. Genetic and antigenic diversity of infectious bronchitis virus in South America. *Avian Dis.* (2021) 65:622–8. doi: 10.1637/aviandiseases-D-21-00103
22. Fabricant J. The early history of infectious bronchitis. *Avian Dis.* (1998) 42:648–50. doi: 10.2307/1592697
23. Cavanagh D. Coronavirus avian infectious bronchitis virus. *Vet Res.* (2007) 38:281–97. doi: 10.1051/vetres:2006055
24. Sjaak de Wit JJ, Cook JK, van der Heijden HM. Infectious bronchitis virus variants: a review of the history, current situation and control measures. *Avian Pathol.* (2011) 40:223–35. doi: 10.1080/03079457.2011.566260
25. Jackwood MW. Review of infectious bronchitis virus around the world. *Avian Dis.* (2012) 56:634–41. doi: 10.1637/10227-043012-Review.1
26. Villarreal L, Sandri TL, Souza S, Richtzenhain LJ, De Wit J, Brandao PE. Molecular epidemiology of avian infectious bronchitis in Brazil from 2007 to 2008 in breeders, broilers, and layers. *Avian Dis.* (2010) 54:894–8. doi: 10.1637/9218-121709-Reg.1
27. Guzmán M, Sáenz L, Hidalgo H. Molecular and antigenic characterization of GI-13 and GI-16 avian infectious bronchitis virus isolated in Chile from 2009 to 2017 regarding 4/91 vaccine introduction. *Animals.* (2019) 9:656. doi: 10.3390/ani9090656
28. Cook JK, Orbell SJ, Woods MA, Huggins MB. Breadth of protection of the respiratory tract provided by different live-attenuated infectious bronchitis vaccines against challenge with infectious bronchitis viruses of heterologous serotypes. *Avian Pathol.* (1999) 28:477–85. doi: 10.1080/03079459994506
29. Gelb J Jr, Ladman B, Tamayo M, Gonzalez M, Sivanandan V. Novel infectious bronchitis virus S1 genotypes in Mexico 1998–1999. *Avian Dis.* (2001) 45:1060–3. doi: 10.2307/1592889
30. Callison S, Jackwood M, Hilt D. Molecular characterization of infectious bronchitis virus isolates foreign to the United States and comparison with United States isolates. *Avian Dis.* (2001) 45:492–9. doi: 10.2307/1592994
31. Escorcia M, Jackwood M, Lucio B, Petrone V, Lopez C, Fehervari T, et al. Characterization of Mexican strains of avian infectious bronchitis isolated during 1997. *Avian Dis.* (2000) 44:944–7. doi: 10.2307/1593069
32. Cavanagh D. Severe acute respiratory syndrome vaccine development: experiences of vaccination against avian infectious bronchitis coronavirus. *Avian Pathol.* (2003) 32:567–82. doi: 10.1080/0307945031000162119 8
33. Finney PM, Box PG, Holmes HC. Studies with a bivalent infectious bronchitis killed virus vaccine. *Avian Pathol.* (1990) 19:435–50. doi: 10.1080/03079459008418698
34. Jackwood MW. Current and Future Recombinant Viral Vaccines for Poultry. In: Schultz RD, editor. *Advances in Veterinary Medicine*. London: Academic Press (1999). p. 517–22 doi: 10.1016/S0065-3519(99)80038-X
35. Cavanagh D, Davis P, Cook J. Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. *Avian Pathol.* (1992) 21:401–8. doi: 10.1080/03079459208418858
36. Jackwood MW, Hall D, Handel A. Molecular evolution and emergence of avian gammacoronaviruses. *Infect Genet Evol.* (2012) 12:1305–11. doi: 10.1016/j.meegid.2012.05.003
37. McKinley ET, Jackwood MW, Hilt DA, Kissinger JC, Robertson JS, Lemke C, et al. Attenuated live vaccine usage affects accurate measures of virus diversity and mutation rates in avian coronavirus infectious bronchitis virus. *Virus Res.* (2011) 158:225–34. doi: 10.1016/j.virusres.2011.04.006
38. Parris DJ, Kariithi HM, Suarez DL. Non-target RNA depletion strategy to improve sensitivity of next-generation sequencing for the detection of RNA viruses in poultry. *J. Vet. Diagn. Invest.* (in press). doi: 10.1177/10406387221102430
39. Chrzastek K, Lee D, Smith D, Sharma P, Suarez DL, Pantin-Jackwood M, et al. Use of Sequence-Independent, Single-Primer-Amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology.* (2017) 509:159–66. doi: 10.1016/j.virol.2017.06.019
40. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* (2018) 46:W537–44. doi: 10.1093/nar/gky379
41. 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. doi: 10.1186/s12985-017-0741-5
42. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* (2012) 28:1647–9. doi: 10.1093/bioinformatics/bts199
43. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* (2011) 17:10–2. doi: 10.14806/ej.17.1.200
44. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
45. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics.* (2013) 30:614–20. doi: 10.1093/bioinformatics/btt593

46. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, et al. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res.* (2015) 4:900. doi: 10.12688/f1000research.6924.1
47. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: *The German Conference on Bioinformatics, GCB'99*. Hanover: Citeseer (1999). p. 45–56.
48. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* (2014) 30:1236–40. doi: 10.1093/bioinformatics/btu031
49. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel.* (2004) 17:107–12. doi: 10.1093/protein/gzh013
50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* (2013) 30:772–80. doi: 10.1093/molbev/mst010
51. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* (2009) 25:1972–3. doi: 10.1093/bioinformatics/btp348
52. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis v6. *Mol Biol Evol.* (2013) 30:2725–9. doi: 10.1093/molbev/mst197
53. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* (2006) 23:254–67. doi: 10.1093/molbev/msj030
54. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* (2015) 1:vev003. doi: 10.1093/ve/vev003
55. Gohl DM, Magli A, Garbe J, Becker A, Johnson DM, Anderson S, et al. Measuring sequencer size bias using REcount: a novel method for highly accurate Illumina sequencing-based quantification. *Genome Biol.* (2019) 20:85. doi: 10.1186/s13059-019-1691-6
56. Alfson KJ, Beadles MW, Griffiths A. A new approach to determining whole viral genomic sequences including termini using a single deep sequencing run. *J Virol Methods.* (2014) 208:1–5. doi: 10.1016/j.jviromet.2014.07.023
57. Fang SG, Shen H, Wang J, Tay FP, Liu DX. Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. *Virology.* (2008) 379:175–80. doi: 10.1016/j.virol.2008.06.038
58. Liu X, Su J, Zhao J, Zhang G. Complete genome sequence analysis of a predominant infectious bronchitis virus (IBV) strain in China. *Virus Genes.* (2009) 38:56–65. doi: 10.1007/s11262-008-0282-5
59. Zhao F, Zou N, Wang F, Guo M, Liu P, Wen X, et al. Analysis of a QX-like avian infectious bronchitis virus genome identified recombination in the region containing the ORF 5a, ORF 5b, and nucleocapsid protein gene sequences. *Virus Genes.* (2013) 46:454–64. doi: 10.1007/s11262-013-0884-4
60. Hemida M, Barta J, Ojkic D, Yoo D. Complete genomic sequence of turkey coronavirus. *Virus Res.* (2008) 135:237–46. doi: 10.1016/j.virusres.2008.03.020
61. Tian S. A 20 residues motif delineates the furin cleavage site and its physical properties may influence viral fusion. *Biochem Insights.* (2009) 2:BCI-S2049. doi: 10.4137/BCI.S2049
62. Madu IG, Roth SL, Belouzard S, Whittaker GR. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J Virol.* (2009) 83:7411–21. doi: 10.1128/JVI.00079-09
63. Corse E, Machamer CE. The cytoplasmic tail of infectious bronchitis virus E protein directs Golgi targeting. *J Virol.* (2002) 76:1273–84. doi: 10.1128/JVI.76.3.1273-1284.2002
64. Casais R, Davies M, Cavanagh D, Britton P. Gene 5 of the avian coronavirus infectious bronchitis virus is not essential for replication. *J Virol.* (2005) 79:8065–78. doi: 10.1128/JVI.79.13.8065-8078.2005
65. Brown Jordan A, Fusaro A, Blake L, Milani A, Zamperin G, Brown G, et al. Characterization of novel, pathogenic field strains of infectious bronchitis virus (IBV) in poultry in Trinidad and Tobago. *Transbound Emerg Dis.* (2020) 67:2775–88. doi: 10.1111/tbed.13637
66. Brandão PE, Taniwaki SA, Berg M, Hora AS. Complete genome of avian coronavirus vaccine strains Ma5 and BR-I. *Genome Announc.* (2017) 5:e00201–17. doi: 10.1128/genomeA.00201-17
67. Thor SW, Hilt DA, Kissinger JC, Paterson AH, Jackwood MW. Recombination in avian gamma-coronavirus infectious bronchitis virus. *Viruses.* (2011) 3:1777–99. doi: 10.3390/v3091777
68. Toro H, Van Santen V, Li L, Lockaby S, Van Santen E, Hoerr F. Epidemiological and experimental evidence for immunodeficiency affecting avian infectious bronchitis. *Avian Pathol.* (2006) 35:455–64. doi: 10.1080/03079450601028811
69. Goraichuk IV, Kulkarni AB, Williams-Coplin D, Suarez DL, Afonso CL. First complete genome sequence of currently circulating infectious bronchitis virus strain DMV/1639 of the GI-17 lineage. *Microbiol Resour Announc.* (2019) 8:e00840–19. doi: 10.1128/MRA.00840-19
70. Naqi S, Gay K, Patalla P, Mondal S, Liu R. Establishment of persistent avian infectious bronchitis virus infection in antibody-free and antibody-positive chickens. *Avian Dis.* (2003) 47:594–601. doi: 10.1637/6087
71. Cortes AL, Montiel ER, Gimeno IM. Validation of Marek's disease diagnosis and monitoring of Marek's disease vaccines from samples collected in FTA® cards. *Avian Dis.* (2009) 53:510–6. doi: 10.1637/8871-041009-Reg.1
72. Kraus RHS, van Hooft P, Waldenström J, Latorre-Margalef N, Ydenberg RC, Prins HHT. Avian influenza surveillance with FTA cards: field methods, biosafety, and transportation issues solved. *J Vis Exp.* (2011) 54:e2832. doi: 10.3791/2832
73. 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) 35:93–8. doi: 10.1080/03079450600597410
74. Józwiak M, Wyrostek K, Domańska-Blicharz K, Olszewska-Tomczyk M, Smietanka K, Minta Z. Application of FTA® cards for detection and storage of avian influenza virus. *J Vet Res.* (2016) 60:1–6. doi: 10.1515/jvetres-2016-0001
75. Reddy VR, Theuns S, Roukaerts ID, Zeller M, Matthijnssens J, Nauwynck HJ. Genetic characterization of the Belgian nephropathogenic infectious bronchitis virus (NIBV) reference strain B1648. *Viruses.* (2015) 7:4488–506. doi: 10.3390/v7082827
76. Abozeid HH, Paldurai A, Khattar SK, Afifi MA, El-Kady MF, El-Deeb AH, et al. Complete genome sequences of two avian infectious bronchitis viruses isolated in Egypt: evidence for genetic drift and genetic recombination in the circulating viruses. *Infect Genet Evol.* (2017) 53:7–14. doi: 10.1016/j.meegid.2017.05.006
77. Hewson KA, Ignjatovic J, Browning GF, Devlin JM, Noormohammadi AH. Infectious bronchitis viruses with naturally occurring genomic rearrangement and gene deletion. *Arch Virol.* (2011) 156:245–52. doi: 10.1007/s00705-010-0850-6
78. de Haan CA, Volders H, Koetzner CA, Masters PS, Rottier PJ. Coronaviruses maintain viability despite dramatic rearrangements of the strictly conserved genome organization. *J Virol.* (2002) 76:12491–502. doi: 10.1128/JVI.76.24.12491-12502.2002
79. Simon-Loriere E, Holmes EC, Pagán I. The effect of gene overlapping on the rate of RNA virus evolution. *Mol Biol Evol.* (2013) 30:1916–28. doi: 10.1093/molbev/mst094
80. Cheng J, Zhao Y, Xu G, Zhang K, Jia W, Sun Y, et al. The S2 subunit of QX-type infectious bronchitis coronavirus spike protein is an essential determinant of neurotropism. *Viruses.* (2019) 11:972. doi: 10.3390/v11100972
81. Hassan MS, Ojkic D, Coffin CS, Cork SC, Van Der Meer F, Abdul-Careem MF. Delmarva (DMV/1639) infectious bronchitis virus (IBV) variants isolated in eastern Canada show evidence of recombination. *Viruses.* (2019) 11:1054. doi: 10.3390/v11111054
82. Fang SG, Shen S, Tay FP, Liu D. Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem Biophys Res Commun.* (2005) 336:417–23. doi: 10.1016/j.bbrc.2005.08.105
83. Leyson C, França M, Jackwood M, Jordan B. Polymorphisms in the S1 spike glycoprotein of Arkansas-type infectious bronchitis virus (IBV) show differential binding to host tissues and altered antigenicity. *Virology.* (2016) 498:218–25. doi: 10.1016/j.virol.2016.08.030

Conflict of Interest: JV and CA were employed by $BASE_2BIO$.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

TYPE Data Report
PUBLISHED 23 September 2022
DOI 10.3389/fvets.2022.1001204

OPEN ACCESS

EDITED BY
Iryna Goraichuk,
Institute of Experimental and Clinical Veterinary Medicine, Ukraine

REVIEWED BY
Susan Angela Nadin-Davis,
Canadian Food Inspection Agency, Canada
Gábor Kemenesi,
University of Pécs, Hungary

*CORRESPONDENCE
Dong-Hun Lee
donghunlee@konkuk.ac.kr

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 23 July 2022
ACCEPTED 07 September 2022
PUBLISHED 23 September 2022

CITATION
Chung DH, Helal Z, Desiato J, McGinnis H, Sims M, Hunt A, Kim J, Risatti GR and Lee D-H (2022) Genome sequencing and analysis of the raccoon variant rabies lyssaviruses directly from clinical samples, Connecticut, 2017–2019. *Front. Vet. Sci.* 9:1001204. doi: 10.3389/fvets.2022.1001204

Genome sequencing and analysis of the raccoon variant rabies lyssaviruses directly from clinical samples, Connecticut, 2017–2019

David H. Chung[1†], Zeinab Helal[1,2†], Julia Desiato[1], Holly McGinnis[1,2], Maureen Sims[1,2], Amelia Hunt[1,2], Junwon Kim[1], Guillermo R. Risatti[1,2] and Dong-Hun Lee[3*]

[1]Department of Pathobiology and Veterinary Science, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, United States, [2]Connecticut Veterinary Medical Diagnostic Laboratory, Department of Pathobiology and Veterinary Science, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, United States, [3]College of Veterinary Medicine, Konkuk University, Seoul, South Korea

KEYWORDS

rabies, multiplex RT-PCR, next-generation sequencing, genome assembly, phylogenetics

Introduction

Rabies lyssavirus is a zoonotic RNA virus typically transmitted through the bite of an infected animal *via* saliva. Infections attack the central nervous system and may result in symptoms such as respiratory arrest, paralysis, encephalitis, coma, or fatality if the patient is not treated with post-exposure prophylaxis promptly after being exposed (1). The virus is responsible for annual 59,000 human deaths globally, exhibiting great diversity in viral epidemiology and reservoir species distribution (2). Approximately 5,000 animal rabies cases are reported to the Centers for Disease Control and Prevention (CDC) each year, and most of the cases in the United States (U.S.) are from wild animals, which threaten transmission to domestic animals and even to humans (3, 4).

Lyssaviruses are detected in at least 30 identified reservoir species, consisting primarily of terrestrial carnivore species, including raccoons skunks and bats, and distributed worldwide in the Americas, Europe, Asia, Africa, and Australia (5, 6). In North America, the raccoon variant rabies lyssavirus (RRV) is of concern, given its relatively rapid spread in wildlife and its potential public health impact (7). In the U.S., the RRV's endemic geographic range has expanded in recent decades throughout the country's eastern seaboard (8). RRV was first reported in Connecticut in March 1991, when a rabid raccoon was found in Ridgefield, which borders New York State (9); since that, the RRV has spread throughout the state. In 2018 Connecticut reported 15.4% increase in the number of raccoon rabies cases detected compared with the number seen in 2017 (5).

Next-generation sequencing (NGS) approaches enable whole genome sequencing (WGS) of pathogens directly from clinical samples without isolation. Despite the potential and recent successes of whole genome metagenomics, target enrichment for specific viruses is often required to generate complete viral genome coverage while sequencing directly from clinical samples. The tiling amplicon panel schemes have been used to generate complete coverage of viral genomes (10, 11). For rabies lyssavirus, reverse transcription polymerase chain reaction (RT-PCR) based virus-specific genome enrichment methods have been used for complete genome sequencing (12, 13). Here, we optimized a multiplex tiling RT-PCR protocol with selective primer sets covering the entire coding region (CDS) of RRV in three reactions and successfully sequenced and analyzed 42 RRV-positive animal brain samples identified from Connecticut during 2017–2019.

Materials and methods

Rabies lyssavirus samples

In this study, we have used brain tissue samples from rabies cases submitted between 2017 and 2019 (Supplementary Table 1) to the Connecticut Veterinary Medical Diagnostics Laboratory (CVMDL). A total of 42 rabies-positive brain tissue samples of raccoon ($n = 28$), skunk ($n = 6$), woodchuck ($n = 2$), feline ($n = 2$), bobcat ($n = 1$), fox ($n = 1$), cow ($n = 1$), and deer ($n = 1$) confirmed rabies positive at CVMDL by direct fluorescent antibody test (14) and quantitative reverse transcription PCR (RT-qPCR) assay (15) were used for sequencing purposes.

Comparative *in-silico* selection of rabies primers

The selected primer sets targeting rabies lyssavirus were designed based on previous studies (13, 16). We retrieved 1,798 reference rabies viral genome sequences from North America encompassing the period 2009–2019 available on Virus Pathogen Database and Analysis Resource (www.viprbrc.org). The viral genome dataset was subsampled to 307 sequences using nucleotide sequence identity of 98.5% on the CD-HIT server (17). Sequences were aligned using Multiple Alignment with Fast Fourier Transformation (MAFFT) (18). RT-PCR primers for rabies lyssavirus reported previously were tested using *in-silico* PCR analysis with FastPCR program v6.6 (19) to evaluate them for mismatches and annealing temperature (Ta). Primers lacking nucleotide base-pair mismatch to reference rabies viral sequences were selected. The annealing temperatures of the primers predicted by the *in-silico* PCR setting ranged from 56 to 67°C.

In this study, a total of six primer pairs ranging 1,675–2,465 bp in amplicon sizes were used to cover the complete genome of RRVs. The forward-reverse primer pairs were then re-organized into 3 distinct sets for multiplex RT-PCR reactions, including Set 1; RVfor3-PR2a pair and LF8a-RVrev2 pair, Set 2; PF2a-RRVArev pair and LF3-RRVBrev pair, and Set 3; RRVBfor-LR3 pair and RRVCfor-LR8 pair (Supplementary Figure S1A; Supplementary Table S2). We assigned alternate target genome regions so that neighboring amplicons do not overlap within the same pool.

Viral RNA extraction and multiplex tiling RT-PCR

The viral RNA was extracted from samples using the TRIzol reagent (ThermoFisher Scientific, USA) according to manufacturer's instructions. One Taq one-step RT-PCR kit (New England BioLabs, USA, Cat# E5315S) was used to amplify genomic RNA. The RT-PCR reaction included 5 µl (500–1,000 ng) RNA template of RRV-positive samples, 10 µM of each primer (1 µl forward and 1 µl reverse primer), 12.5 µl reaction mix, 1 µl enzyme mix and 4.5 µl RNase free water to obtain a total volume of 25 µl. In this protocol, all RRV genome segments were amplified simultaneously. Amplicon panel length therefore ranges 1,676–2,485 bp together covering the complete RRV genome. Cycling conditions for the respective rabies RT-PCR were conducted as described: an initial primary reverse transcription step of 30 min at 48°C, then denaturation at 94°C for 1 min, followed by 35 cycles of 94°C for 15 s, 56°C for 30 s and 68°C for 2 min, and to conclude a final elongation step at 68°C for 5 min. PCR products were checked by agarose gel electrophoresis and purified using NucleoSpin gel and PCR clean-up kit (Macherey-Nagel, PA, USA) according to the manufacturer's instructions. After purification, DNA concentration was measured using the Qubit dsDNA HS Assay Kit and the Qubit 4 fluorometer (Thermoscientific, U.S.). DNA products (Multiplex PCR sets 1-3) were normalized and pooled together.

Genome sequencing

The Nextera DNA Flex Sample Preparation Kit (Illumina, U.S.) was used according to the manufacturer's instructions to generate multiplexed paired-end sequencing libraries of pooled PCR products of 42 samples. The dsDNA was fragmented and tagged with adapters by Nextera transposase. The fragmented PCR amplicons with added adaptor sequences enabled a 5-cycle PCR amplification to append additional unique dual index (i7 and i5) sequences at the end of each fragmented DNA for cluster formation. PCR fragments were purified on Sample Purification Beads included in the library prep kit. Fragments were analyzed

on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies, U.S.) before being loaded on the sequencing flow cell. Briefly, the libraries were adjusted to 1 nM concentration and equal volumes of 5 μl of each library were pooled. The pool was denatured with NaOH (0.2 N final concentration) and further diluted to 100 pM. Control library (5% PhiX library, Illumina, U.S.) was added to the pool. The library pool was loaded in the flow cell of the 151 cycle iSeq100 i1 Reagent Kit (Illumina, U.S.). The barcoded multiplexed library sequencing (2 × 151 bp) was performed on an Illumina iSeq100 platform (Illumina, U.S.).

Genome assembly

Reference guided genome assembly was performed using the Galaxy instance (20). We created an automated genome assembly workflow that integrated software tools for quality trimming, reference mapping, and consensus calling (Supplementary Figure S1B). Residual Nextera adapters and bases with low quality scores were removed from fastq files using Trimmomatic version 0.38.0 (21) with conservative parameters, which included removing bases from each read with a quality score <%Q30 and required a minimum read length of 100 bases each. Trimmed reads were mapped to a known reference rabies genome sequence using the Bowtie2 assembler version 2.3.4.3 (http://bowtie-bio.sourceforge.net/bowtie2). The consensus genome was called using iVar consensus version 1.2.3 (22). The complete coding genome sequences of the RRVs sequenced in this study have been deposited in the GenBank under Accession Numbers ON986428-30, ON986432-41, ON986443-55, ON986457, ON986467-69, ON986471-72, ON986474, ON986476-77, and ON986479-80.

Phylogenetic analysis

Using the Virus Pathogen Resource Database (VIPR), 916 RRVs full genome sequences were collected in North America between 2000 and 2020. A sequence from the recent RRVs collection from Connecticut was then run through the nucleotide BLAST on NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the 100 most genetically similar sequences of RRVs from surrounding states were combined with the 916 North American RRVs sequences and the newly sequenced samples from Connecticut to compare to each other. ElimDupes software (https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html) was used to eliminate any sequences that had 99.5% similar identity level to prune down to 203 sequences. Sequences were renamed using Microsoft Excel and BioEdit software (https://projects.ncsu.edu/cbirna/links.html) to format all names consistently. A maximum-likelihood (ML) phylogeny was created *via* Randomized Axelerated Maximum Likelihood (RAxML) with rapid bootstrap iterations of 1,000 (23).

Descriptive results

Dataset overview and description

The rabies-specific primers used in this study were selected from multiple candidates previously published. We compared the RT-PCR primers previously reported by Nadin-Davis et al. (13) and Campos et al. (16) against the RRV genome sequences collected from North America during the period 2009–2019 using *in-silico* PCR analysis and selected six pairs of forward-reverse primers from a previous paper published by Nadin-Davis et al. in 2017 (13). We successfully amplified 1,675–2,465 bp long amplicons that cover the complete coding sequences (CDS) of 42 RRVs using three distinct multiplex tiling RT-PCR reactions followed by sequencing of the amplicons using the Illumina iSeq100 NGS platform. The entire process from viral RNA isolation to genome assembly could be completed in 30 h leading to the generation of fastq files from 42 RNA samples, including 3 h of RT-PCR, 4 h of NGS library preparation, and 22 h of iSeq100 NGS run (Supplementary Table S3).

Sequencing read quality was determined by quality scores (%Q30) from the Illumina NGS run which indicates probability of incorrect base call leading to 99.9% inferred base call accuracy. A sum of 15,376,000 reads were generated with a total %Q30 rate of 89.95%, reads with identifiable barcode were 85.87%, both having typical rates for Illumina iSeq100 dataset (https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html), and each individual barcode was found to be distributed throughout the samples. The read distribution among the 42 samples exhibited a roughly even pattern, where samples with barcodes 7 and 18 had relatively lower identified read percentages (0.0945% and 0.1147%, respectively).

Genome assembly using the Galaxy workflow allowed the successful creation of 42 consensus sequences from the fastq file format. All samples had complete CDS. The number of mapped reads varied from 8,429 to 778,612 (Supplementary Tables S1, S3). Average sequence lengths were between 11,920 and 11,921 base pairs and complete CDS was obtained for each sequence. The mean depth coverage was varied across all sequences with a minimum value of 94.7 and maximum of 13,322.5.

The ML phylogeny indicated that RRVs in Connecticut fall into four genetically divergent subclades designated as CT Clades 1–4 (Figure 1). The viruses in CT Clade 1 detected in raccoons, skunks, bobcats, and deer clustered with viruses from cat, skunk, raccoon, and rodent from New York. The viruses in CT clades 2–4 were detected in raccoon, skunk, feline, fox, and woodchuck showed close relationship with RRV from New

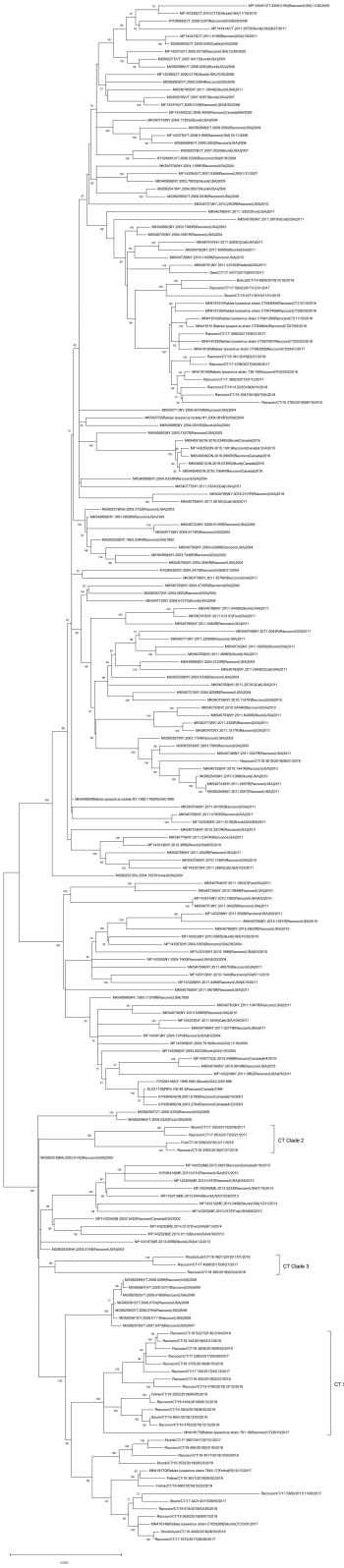


FIGURE 1
Maximum-likelihood analysis of 203 complete coding sequences of RRVs identified in North America including 42 *(Continued)*

FIGURE 1 (Continued)
viruses sequenced in this study. The scale bar shows the number of substitutions per site. The numerical values at each node represent 1,000 bootstrap replicate values expressed as a percentage.

England regions including Vermont, Massachusetts, Maine, New Brunswick (Canada), and Rhode Island. The Deer/CT/17-4477/2017 virus did not cluster together with viruses from Connecticut, but clustered with a 2011 rodent RRV from NY (MK540781|4316O|Rodent) as an outlier.

Discussion

Based on the 2019 rabies surveillance in the U.S., raccoons accounted for the highest percentage (32.9%) among all detected rabid wild animals (n = 94,770) followed by skunks (24). Consistent with previous findings (7, 24, 25), the high number of detections in raccoons and skunks in Connecticut during 2017–2019 suggests a key role in maintenance and spread of RRVs in the Northeastern U.S. These species are highly adaptive to urban environment raising the risk of rabies exposure to human and animal populations. Therefore, efficient genetic tools are in need for investigating RRV outbreaks if they should occur.

Based on our data and experience, rabies positive samples confirmed by direct fluorescent antibody test contain sufficient amount of viral RNA, mostly *Ct*-value < 30 in RT-qPCR, for multiplex tiling RT-PCR and genome sequencing. Our newly optimized multiplex tiling RT-PCR protocol was utilized with a pre-selected primer panel set that cover the complete CDS of the RRVs via three amplification reactions. While conventional NGS protocols involve high cost, labor, and complexity in the processes, our approach is both time-saving and cost-effective. The multiplex tiling RT-PCR combined with an affordable benchtop Illumina NGS system significantly enhanced experimental efficiency. The approach enables selective enrichment of viral genome from clinical samples with much less hands-on time and experimental steps.

The use of the described protocol allowed a fast and efficient genome sequencing of circulating RRV in the US northeast. The CDS-based ML phylogenetic analysis showed a high genetic diversity between RRVs detected in the State of Connecticut and those detected in neighboring states. Data generated here suggest possible viral transmissions of RRVs between Connecticut and in other States in Northeastern U.S. through movements of wild animals and independent evolution of each virus subclade in Connecticut.

In summary, we report passive surveillance WGS data for RRV detected in the state of Connecticut during 2017–2019 using our multiplex RT-PCR approach. The phylogenetic

analysis combined with the newly optimized methodology provide a useful methodological tool for future RRV surveillance that includes assessments of evolution and transmission of RRV in North America.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

Conceptualization, supervision, and funding acquisition: D-HL and GR. Methodology: DC and ZH. Sample preparation: HM, MS, AH, and ZH. Data analysis: DC, JD, and JK. Data curation: AH and ZH. Writing—original draft preparation: DC. Writing—review and editing: GR, JD, HM, MS, and D-HL. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the University of Connecticut's Office of the Vice President for Research (OVPR) through the Research Excellence Program.

Acknowledgments

We thank the unconditional support of staff and faculty from the Connecticut Veterinary Medical Diagnostic Laboratory (CVMDL), Department of Pathobiology and Veterinary Science, CAHNR, University of Connecticut.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.1001204/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Schematic representation of the complete protocol for multiplex tiling RT-PCR and reference guided genome assembly pipeline. **(A)** Representation of the 6 primer pairs ranging 1,675–2,465 bp in amplicon sizes used to cover the complete genome of RRVs. The forward-reverse primer pairs were designed into 3 distinct sets for multiplex RT-PCR reactions (Set 1: RVfor3-PR2a pair and LF8a-RVrev2 pair, Set 2: PF2a-RRVArev pair and LF3-RRVBrev pair, and Set 3: RRVBfor-LR3 pair and RRVCfor-LR8 pair); **(B)** description of automated workflow pipeline on the Galaxy instance used for reference guided genome assembly of raccoon variant rabies samples.

SUPPLEMENTARY TABLE S1
Summary of Raccoon variant rabies virus samples and genome assembly data.

SUPPLEMENTARY TABLE S2
Multiplex PCR primer pair sets for Raccoon variant Rabies RT-PCR.

SUPPLEMENTARY TABLE S3
Summary statistics for sequencing experiment of raccoon variant rabies lyssavirus samples using Illumina iSeq100 platform.

References

1. National Park Service. *One Health and Disease: Rabies.* Fort Collins, CO: National Park Service (2018).

2. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, et al. Estimating the global burden of endemic canine rabies. *PLoS Negl Trop Dis.* (2015) 9:e0003709. doi: 10.1371/journal.pntd.0003709

3. Centers for Disease Control and Prevention. *Animals and Rabies*. Atlanta, GA: Centers for Disease Control and Prevention (2020).

4. Centers for Disease Control and Prevention. *Human Rabies, Annual Surveillance Reports in the United States*. Atlanta, GA: Centers for Disease Control and Prevention (2020).

5. Ma X, Monroe BP, Cleaton JM, Orciari LA, Gigante CM, Kirby JD, et al. Public veterinary medicine: public health: rabies surveillance in the united states during 2018. *J Am Vet Med Assoc.* (2020) 256:195–208. doi: 10.2460/javma.256.2.195

6. Guerra MA, Curns AT, Rupprecht CE, Hanlon CA, Krebs JW, Childs JE. Skunk and raccoon rabies in the eastern United States: temporal and spatial analysis. *Emerg Infect Dis.* (2003) 9:1143–50. doi: 10.3201/eid0909.020608

7. Szanto AG, Nadin-Davis SA, Rosatte RC, White BN. Genetic tracking of the raccoon variant of rabies virus in eastern North America. *Epidemics.* (2011) 3:76–87. doi: 10.1016/j.epidem.2011.02.002

8. Trewby H, Nadin-Davis SA, Real LA, Biek R. Processes underlying rabies virus incursions across US-Canada border as revealed by whole-genome phylogeography. *Emerg Infect Dis.* (2017) 23:1454–61. doi: 10.3201/eid2309.170325

9. Centers for Disease Conyrol. Extension of the raccoon rabies epizootic–United States. *MMWR Morb Mortal Wkly Rep*. (1992) 41:661–4.

10. 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) 12:1261–76. doi: 10.1038/nprot.2017.066

11. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet.* (2019) 20:341–55. doi: 10.1038/s41576-019-0113-7

12. David D. Role of the RT-PCR method in ante-mortem & post-mortem rabies diagnosis. *Indian J Med Res.* (2012) 135:809–11. Available online at: https://journals.lww.com/ijmr/Fulltext/2012/35060/Role_of_the_RT_PCR_method_in_ante_mortem__.4.aspx

13. Nadin-Davis SA, Colville A, Trewby H, Biek R, Real L. Application of high-throughput sequencing to whole rabies viral genome characterisation and its use for phylogenetic re-evaluation of a raccoon strain incursion into the province of Ontario. *Virus Res.* (2017) 232:123–33. doi: 10.1016/j.virusres.2017.02.007

14. Meslin FX, Kaplan MM, Koprowski H, World Health Organization. *Laboratory Techniques in Rabies.* 4th ed. Geneva: World Health Organization (1996). p. 467.

15. Gigante CM, Dettinger L, Powell JW, Seiders M, Condori REC, Griesser R, et al. Multi-site evaluation of the LN34 pan-lyssavirus real-time RT-PCR assay for post-mortem rabies diagnostics. *PLoS ONE.* (2018) 13:e0197074. doi: 10.1371/journal.pone.0197074

16. Campos AC, Melo FL, Romano CM, Araujo DB, Cunha EM, Sacramento DR, et al. One-step protocol for amplification of near full-length cDNA of the rabies virus genome. *J Virol Methods.* (2011) 174:1–6. doi: 10.1016/j.jviromet.2011.03.030

17. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT accelerated for clustering the next-generation sequencing data. *Bioinformatics.* (2012) 28:3150–2. doi: 10.1093/bioinformatics/bts565

18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* (2013) 30:772–80. doi: 10.1093/molbev/mst010

19. Kalendar R, Khassenov B, Ramankulov Y, Samuilova O, Ivanov KI. FastPCR: an *in silico* tool for fast primer and probe design and advanced sequence analysis. *Genomics.* (2017) 109:312–9. doi: 10.1016/j.ygeno.2017.05.005

20. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* (2020) 48:W395–402. doi: 10.1093/nar/gkaa434

21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170

22. Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, et al. iVar, an interpretation-oriented tool to manage the update and revision of variant annotation and classification. *Genes.* (2021) 12:384. doi: 10.3390/genes12030384

23. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE).* New Orleans, LA (2010). p. 1–8.

24. Ma X, Monroe BP, Wallace RM, Orciari LA, Gigante CM, Kirby JD, et al. Rabies surveillance in the United States during 2019. *J Am Vet Med Assoc.* (2021) 258:1205–20. doi: 10.2460/javma.258.11.1205

25. Hanlon CA, Niezgoda M, Rupprecht CE. *Rabies in Terrestrial Animals.* London: Academic Press (2007).

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fvets.2022.1073919

OPEN ACCESS

EDITED BY
Ihab Habib,
United Arab Emirates University,
United Arab Emirates

REVIEWED BY
Sunil Kumar Mor,
University of Minnesota Twin Cities,
United States
Irit Davidson,
Kimron Veterinary Institute, Israel

*CORRESPONDENCE
Claudio L. Afonso
claudio.afonso@base2bio.com

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 19 October 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Butt SL, Kariithi HM, Volkening JD, Taylor TL, Leyson C, Pantin-Jackwood M, Suarez DL, Stanton JB and Afonso CL (2022) Comparable outcomes from long and short read random sequencing of total RNA for detection of pathogens in chicken respiratory samples. *Front. Vet. Sci.* 9:1073919. doi: 10.3389/fvets.2022.1073919

Comparable outcomes from long and short read random sequencing of total RNA for detection of pathogens in chicken respiratory samples

Salman L. Butt[1,2], Henry M. Kariithi[2,3], Jeremy D. Volkening[4], Tonya L. Taylor[2], Christina Leyson[2], Mary Pantin-Jackwood[2], David L. Suarez[2], James B. Stanton[5] and Claudio L. Afonso[4]*

[1]Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, United States, [2]Exotic and Emerging Avian Viral Diseases Research Unit, Southeast Poultry Research Laboratory, United States National Poultry Research Center, Agricultural Research Service, United States Department of Agriculture, Athens, GA, United States, [3]Biotechnology Research Institute, Kenyan Agricultural and Livestock Research Organization, Nairobi, Kenya, [4]BASE$_2$BIO, Oshkosh, WI, United States, [5]Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA, United States

Co-infections of avian species with different RNA viruses and pathogenic bacteria are often misdiagnosed or incompletely characterized using targeted diagnostic methods, which could affect the accurate management of clinical disease. A non-targeted sequencing approach with rapid and precise characterization of pathogens should help respiratory disease management by providing a comprehensive view of the causes of disease. Long-read portable sequencers have significant potential advantages over established short-read sequencers due to portability, speed, and lower cost. The applicability of short reads random sequencing for direct detection of pathogens in clinical poultry samples has been previously demonstrated. Here we demonstrate the feasibility of long read random sequencing approaches to identify disease agents in clinical samples. Experimental oropharyngeal swab samples (n = 12) from chickens infected with infectious bronchitis virus (IBV), avian influenza virus (AIV) and *Mycoplasma synoviae* (MS) and field-collected clinical oropharyngeal swab samples (n = 11) from Kenyan live bird markets previously testing positive for Newcastle disease virus (NDV) were randomly sequenced on the MinION platform and results validated by comparing to real time PCR and short read random sequencing in the Illumina MiSeq platform. In the swabs from experimental infections, each of three agents in every RT-qPCR-positive sample (Ct range 19–34) was detectable within 1 h on the MinION platform, except for AIV one agent in one sample (Ct = 36.21). Nine of 12 IBV-positive samples were assigned genotypes within 1 h, as were five of 11 AIV-positive samples. MinION relative abundances of the test agent (AIV, IBV and MS) were highly correlated with RT-qPCR Ct values (R range−0.82 to−0.98). In field-collected clinical swab samples, NDV (Ct range 12–37) was detected in all eleven samples within 1 h of MinION sequencing, with 10 of 11

samples accurately genotyped within 1 h. All NDV-positive field samples were found to be co-infected with one or more additional respiratory agents. These results demonstrate that MinION sequencing can provide rapid, and sensitive non-targeted detection and genetic characterization of co-existing respiratory pathogens in clinical samples with similar performance to the Illumina MiSeq.

KEYWORDS

MinION, MiSeq, Newcastle disease virus, avian influenza virus, infectious bronchitis virus, mycoplasma spp., clinical samples, respiratory disease

Introduction

Respiratory diseases are a continual significant threat to the global poultry industry (1). Newcastle disease virus (NDV), infectious bronchitis virus (IBV), avian influenza virus (AIV), *Ornithobacterium rhinotracheale* (ORT) (2), *Mycoplasma synoviae* (MS) and *M. gallisepticum* (MG) have been isolated from different avian species presenting similar clinical respiratory disease (3–7). Co-infections with these microbial pathogens produce respiratory disease complexes and complicate accurate disease diagnosis, when using target-specific approaches (5, 8, 9). For example, a commercial broiler flock, first diagnosed as infected with IBV, based on serological assays, was also co-infected with an atypical velogenic NDV, which was overlooked by relying on a single, target-specific detection approach (8). Experimental studies have shown that vaccine strains of IBV prolonged shedding of low pathogenic AIV (LPAIV) type H9N2 and increased the severity of clinical signs and postmortem lesions (10). Progressive pneumonia is a problem in commercial broiler flocks where ORT and H9N2 were primarily isolated; but it was difficult to establish the primary cause of the disease due to mixed infections (5). Co-infections complicate respiratory disease diagnostics and currently, diagnostic approaches to characterize the co-infecting viral and bacterial respiratory pathogens from chicken samples require both classical and molecular diagnostic tools.

One common approach to identify avian pathogens is isolation in embryonating eggs from specific-pathogen-free (SPF) chickens (11). However, coexistence of avian respiratory pathogens (i.e., NDV and AIV) in the same sample might present a diagnostic problem as it is possible to observe overwhelming growth of one agent over the other during isolation causing a biased characterization of clinical samples (12–14). Furthermore, some microbial pathogens associated with respiratory diseases such as Mycoplasmas are difficult and time consuming to culture under laboratory conditions.

A variety of polymerase chain reaction (PCR)-based rapid multiplexed diagnostic assays have been used for detection and molecular epidemiology of respiratory co-infecting pathogens (15–18). However, these assays were developed for specific pathogens, precluding the possibility of detecting unknown pathogens in the samples (19). Additionally, these conventional assays are sensitive to genetic variation, and mismatches on the pathogen's target sequence can lead to false negative results (20).

For decades, pathogen diagnostics and sequencing have been separate endeavors, with sequencing following diagnostics *via* PCR. Sanger sequencing has historically been the gold standard for sequence-based characterization of pathogens, but this approach is time consuming and expensive for complete identification of coinfecting agents in clinical samples (21, 22). More recently, the next generation sequencing (NGS) platforms have changed this paradigm by providing the possibility for simultaneous diagnostic testing and sequencing of novel and re-emerging pathogens directly from a clinical sample (23). We have recently optimized conditions for efficient detection or multiple respiratory pathogens in poultry by directly sequencing clinical samples with the Illumina short read sequences (24–26). The widespread application of these sequencing platforms for routine diagnostics is still limited due to the associated longer processing time, complex bioinformatics expertise of random sequencing data analysis, and higher cost, hence need for the improving recent alternative diagnostic methods to counter these challenges.

Targeted sequencing on the long-read sequencing platform (Oxford Nanopore Technologies) (27), MinION has recently been used to increase the utility of high-throughput sequencing as a tool for avian pathogen characterization (28, 29). The ability to perform near-real-time sequence analysis of long DNA molecules reduce the time from sample collection to outcome. MinION-based targeted sequencing has been used to genetically type respiratory pathogens such as NDV (30), IBV (31), AIV (32), and infectious laryngotracheitis (ILTV) (33). Recently, a random strand-switching approach was used to identify the novel avian paramyxovirus (APMVs) from cultured samples (34). However, a target-independent, multiplexed, single assay for these respiratory pathogens from uncultured swab samples has not been fully developed. Multiplexed, time- and cost-effective assays that require minimal equipment would be useful in rapidly diagnosing infections and co-infections. In the current study, a multiplexed, random sequencing approach based on

MinION nanopore sequencer was developed and compared to RT-qPCR and Illumina MiSeq for the detection of viral and bacterial co-infections in commercial poultry. Additionally, automated bioinformatics pipelines were developed for the rapid characterization of samples by non-experts.

Materials and methods

Samples

Clinical oropharyngeal swab samples (hereafter referred to as "clinical samples") were obtained from chickens from live bird markets in Kenya ($n = 11$) and submitted to the Southeast Poultry Research Laboratory (SEPRL), Athens, Georgia, USA as previously described (35). A second batch of archived chicken oropharyngeal swab samples ($n = 12$) were collected using standard procedures during an experimental coinfection study at SEPRL (hereafter referred to as "experimental samples") was used in the current study (Supplementary Table 1). Allantois fluid obtained from SPF eggs was used a negative control for both set of samples.

RT-qPCR assay on experimental swab samples

Total RNA was extracted from 200 µl of each of the virus isolation media used to collect the swab samples using the MagMax RNA extraction kit (Thermo Fisher Scientific, Waltham, MA, USA) as per manufacturer's instructions, and stored at −80°C until further use. The experimental samples were tested by three separate previously described RT-qPCR assays to detect IBV (*spike* gene) (36), AIV (*matrix* gene) (37), and *Mycoplasma* (16S-23S intergenic spacer region) (38) using the AgPath ID, One step RT-PCR kit (Thermo Fisher Scientific, Waltham, MA, USA) performed on the Applied Biosystems 7500 FAST.

MinION sample preparation

cDNA synthesis

For cDNA synthesis, prior to MinION library preparation, a reaction mixture of 11.5 µl of total RNA (∼50 ng RNA), 0.5 µl of 250 nM random hexamers (New England Biolabs, Ipswich, MA) and 1 µl of 10 mM dNTPs was incubated at 65 °C for 5 min, chilled on ice for 1 min followed by the addition of 7 µl of cDNA synthesis mix including SuperScript IV (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. The reaction mixture was incubated at 23°C for 10 min and, 55°C for 10 min for cDNA synthesis. The reaction was terminated at 80°C for 10 min, and then chilled on ice. To remove residual RNA, the cDNA solution was incubated with 1 µl of RNase H at 37°C for 20 min according to the manufacture's instruction.

Second strand DNA synthesis

The cDNA was immediately used for second strand synthesis. Briefly, 20 µl of cDNA solution was mixed with 10 µl of NEBNext (New England Biolabs, Ipswich, MA) second strand synthesis reaction buffer, 5 µl of NEBNext second strand enzyme mix and 45 µl of nuclease free water (NFW), incubated at 16°C for 1 h and cooled at 4°C. dsDNA was purified with AMPure XP beads (Beckman Coulter, Indianapolis, Indiana) at a bead: DNA volumetric ratio of 1.8:1 and eluted in 52 µl of NFW.

Adapter ligation

dsDNA was repaired and dA-tailed using 45 µl of dsDNA, 7 µl of Ultra II End-prep reaction buffer, 3 µl Ultra II End-prep enzyme mix (NEBNext Ultra End Repair/dA-Tailing Module, New England Biolabs) and 5 µl Nuclease-free water. The reaction mixture was incubated at 20°C for 5 min and 5 min at 65°C. AMPure bead purification was performed at 1:1 volumetric ratio of beads: DNA according to manufacturer's protocol. 15 µl of end-prepped DNA was mixed with 5 µl barcode adapter (1-96 barcoding kit, ONT) and 20 µl blunt/TA ligase master mix. The reaction mixture was incubated for 15 min at room temperature (RT). The adapter-ligated DNA was purified with AMPure bead at 0.4:1 volumetric ratio of bead: DNA and eluted in 26 µl NFW.

PCR-based barcoding was performed using 25 µl of adapter-ligated dsDNA, 2 µl of barcode and 50 µl of Long-Amp Taq 2X Master mix (New England Biolabs, Ipswich, MA). A reaction mixture of 100 µl was used for amplicon synthesis with the following conditions: denaturation at 95°C for 3 min; 17 cycles of denaturation at 95°C for 15 s, annealing at 62°C for 15 s and extension at 65°C for 90 s, and final extension of 65°C for 90 s and chilled at 4°C.

Library preparation and sequencing

The barcoded dsDNA was purified using AMPure beads (bead: DNA, 1:1 volumetric ratio), repaired by dA-tailing end prepped, purified (bead: DNA, 1.6:1), and adapter ligated by using 60 µl of 700 ng pooled (equal volume) barcoded sample, 10 µl of Adapter Mix (AMX 1D), 20 µl of NEBNext Quick Ligation Reaction Buffer (5X) and 10 µl of Quick T4 DNA Ligase (New England Biolabs, Ipswich, MA). The reaction mixture was mixed gently by flicking the tube, and incubated for 10 min at RT. After bead purifying the prepared DNA library was eluted in 15 µl of elution buffer.

A FLO-MIN106 R9.4 flow cell (27) was equilibrated to RT for 10 min and then primed with running buffer as per

manufacturer's instructions. The DNA libraries were prepared by combining 12 μL of the library pool with 2.5 μL NFW, 35 μL RBF (Running Buffer Fuel), 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 1D sequencing protocol was initiated using the MinKNOW software v.5.12.

MiSeq sample preparation

The same clinical and experimental samples were processed for MiSeq sequencing for a side-by-side comparison of sequencing data obtained from MinION and MiSeq. Briefly, DNA libraries were prepared from total RNA (used for MinION library synthesis) ($n = 25$) using the KAPA Stranded RNA-Seq Kit (Roche Sequencing Solutions, Inc., CA, USA) according to manufacturer's recommendations. Concentrations and distribution sizes (bp) of the cDNA in the KAPA libraries were assessed by Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA), and Agilent 2,100 Bioanalyzer (Agilent technologies Inc., Germany), respectively. Paired-end sequencing of the diluted pooled libraries (10 μl each; 4 nM final concentration) was performed on an Illumina MiSeq platform for 39 h using the 300 cycle MiSeq Reagent Kit v 2 (Illumina, USA) according to manufacturer's instructions (39).

MinION data analysis

Raw FAST5 data was basecalled, demultiplexed, and trimmed using Guppy 6.1.2, model "dna_r9.4.1_450bps_hac," barcode set "EXP-PBC096," with "–detect_mid_strand_barcodes," and "—trim_barcodes," without "–require_barcodes_both_ends." Reads were assigned taxonomic classifications using KrakenUniq (v0.5.8) (40) modified with local patches, against a hierarchical set of databases containing vector/contaminant sequences, host genome (*Gallus gallus* GRCg6a), human genome (GRCh38.p13), and the $BASE_2$BIO LLC (Oshkosh, WI, USA) untargeted database of microbial reference sequences. Classifications were further adjusted using a patched version of the "krakenuniq-filter" script, adjusting assignments up the taxonomic tree until the k-mer specificity was 0.05 for viral taxa and 0.25 for all other taxa. Each identified taxon was then further verified by BLASTn (41) search of a random subset of taxon-assigned reads against the full GenBank "nt" database and subsequent lowest common ancestor assignment by in-house tools. For taxa of interest (i.e., NDV, IBV, and AIV), genotypes were called using the standard $BASE_2$BIO genotyping module and curated agent databases. *De novo* assemblies of non-host reads were performed using MEGAHIT (v1.2.9) (42) with default settings, minimum contig length of 500 bp. All steps involving read mapping were performed with minimap2 (v2.24-r1122) (43). Tabulation, summarization, and visualization of correlation analysis results was performed in R v4.1.1 (https://www.R-project.org).

MiSeq data analysis

Analysis of Illumina MiSeq data was performed as described above for the MinION but with the following differences: Data was trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore) v06.7 to remove residual adapter sequences and low-quality 3' ends. Steps involving read mapping were performed using BWA MEM (v0.717-r1188) (44) with default settings.

Results

Rapid detection of pathogens from clinical samples

A total of twenty-three samples were used in this study, along with two negative controls. Samples 1 to 11 were clinical samples and 13 to 24 were experimental samples (Supplementary Table 1). In Table 1, comparative characterization of the clinical samples with MinION vs. MiSeq are presented as total reads, classified reads (belong to chicken and microbial genomes), the genotype calling based on the coverage breadth (3x) and the identification of major respiratory pathogens which included a selected list of known targeted and non-targeted agents as obtained from the known avian disease literature and identified using >1% of relative read abundance threshold. In addition to the full 8 h MinION run, a subset of reads corresponding to the first hour of the MinION sequencing run were analyzed separately to provide insight into the potential for rapid turnaround times. Prior to sequencing, the clinical samples 1 to 11 were NDV-positive using RT-qPCR targeting the Matrix gene, with Ct values ranging from 12.04 to 36.59. NDV was detected in all 11 samples in the first hour of MinION sequencing. Ten of the 11 samples had sufficient genome sequence coverage to accurately assign the genotype after 1 h–all 11 were correctly genotyped in the MiSeq run. In addition to genotype, determination of the fusion cleavage site amino acid motif and subsequent virulence classification was possible for some, or all samples depending on platform and run time (MiSeq: 11/11; MinION 8 h: 6/11; MinION 1 h: 3/11). The consensus sequences assembled from pathogen-specific reads showed presence of virulent NDV as predicted by the presence of amino acid motif (RRQKR↓F) at the cleavage site of the Fusion (F) protein (Supplementary Table 2). The co-infecting pathogens were present at the arbitrarily designed

TABLE 1 Pathogen identification from MinION and MiSeq sequencing from clinical samples collected from Kenya.

Sample	Total read count		Classified read count		NDV Ct	Called NDV genotype			Coverage breadth[a]			Major respiratory pathogens[b]	Relative abundance (%)[c]			Assigned read count		
	MiSeq	MinION	MiSeq	MinION		MiSeq	MinION (1 h)	MinION (8 h)	MiSeq	MinION (1 h)	MinION (8 h)		MiSeq	MinION (1 h)	MinION (8 h)	MiSeq	MinION (1 h)	MinION (8 h)
1	1,735,278	13,090	1,488,350	7,809	13.87	V	V	V	1	0.932	0.997	NDV	27.2	18.8	15.5	383,604	430	982
												Avi	0	1.6	1.6	263	37	100
												MP	1	1.6	1.4	13,723	37	89
												ORT	11.8	15.4	14.6	166,663	352	928
2	1,614,837	14,761	981,135	11,817	15.75	V	V	V	1	0.159	0.619	NDV	3.7	2.3	2.1	34,922	62	141
												Avi	0.2	1.6	1.5	1,497	44	98
												ORT	0.9	2.3	2.6	8,367	64	174
3	1,078,112	38,292	816,517	25,967	17.81	V	V	V	0.945	0	0.024	NDV	0.7	1	0.8	5,776	14	28
												Avi	0.2	9.6	10.6	1,891	137	366
												ORT	0.7	5.2	5.6	5,905	75	195
4	1,500,768	15,871	1,146,857	11,757	16.72	V	V	V	0.735	0.594	0.913	NDV	0.1	3.1	3.2	942	154	375
												Avi	0.9	1.7	1.7	10,426	84	197
												Gal	33.7	31.2	29.8	381,340	1,537	3,449
5	1,860,301	333,690	1,745,836	296,131	12.04	V	V	V	1	0.999	1	NDV	1.9	28.6	29.5	33,521	6,589	20,470
												Avi	1.1	14.1	13.6	18,494	3,243	9,475
6	1,767,122	4,297	1,517,558	3,459	15.43	V	V	V	0.824	0.078	0.316	NDV	0.1	2.4	2.4	1,581	34	78
												Avi	1.2	1.4	1.4	17,998	20	46
7	1,609,048	567,773	835,812	513,686	22.71	V	V	V	0.797	0.159	0.089	NDV	1.1	2.4	1.5	8,861	13	23
												Avi	0.1	9.2	8.2	1,014	50	128
8	1,796,809	299,197	1,484,728	278,030	16.68	V	V	V	1	0.866	0.987	NDV	60.4	34.2	33.5	804,278	361	925
												Avi	0.1	15.3	15.9	812	161	438
9	1,539,612	3,433	1,035,760	1,950	36.59	V	–	–	0.559	0	0	NDV	0.1	0.3	0.1	907	1	1
												Avi	0.8	6	6.8	5,033	19	50
												ORT	0.1	5	4.2	705	16	31
10	1,436,770	75,330	1,094,068	65,267	17.57	V	V	V	1	0.421	0.87	NDV	16.8	19.8	17.1	176,050	134	315
												Avi	0	1.9	1.8	89	13	34
												Gal	0	12.9	12.8	390	87	236
												MG	0.1	1.2	1.5	1,401	8	28
11	1,461,438	6,718	1,117,678	4,962	18.19	V	V	V	0.996	0.891	0.98	NDV	12.1	32.3	34.2	101,597	221	591
												Avi	0.5	1.3	1.2	4,502	9	20
												ORT	0.1	2	2.1	732	14	37
12[d]	N/A	953	N/A	5	N.T.	N/A	–	–	N/A	0	0	NDV	0	50	33.3	N/A	1	1

[a] As fraction of whole genome at > 3x depth of coverage. [b] Known respiratory pathogens detected on either platform at >1% abundance. [c] Relative abundance as a fraction of non-metazoan reads. [d] No-template control, sequenced only on MinION (no MiSeq data). NT, not tested; NDV, Newcastle disease virus also known as Avian orthoavulavirus; Avi, Avibacterium sp; Gal, Gallibacterium sp; ORT, Ornithobacterium rhinotracheale; MP, Mycoplasma pullorum.

10.3389/fvets.2022.1073919

1% threshold on at least one of the platforms–however, additional agents were detected in additional samples at levels below this cutoff. All 11 clinical samples contained other known avian respiratory pathogens, which were detected in addition to NDV at significant levels (>1% of total microbial reads), including *Mycoplasma* (*M. gallisepticum*, *M. synoviae,* and *M. pullorum*), *Avibacterium sp., Gallibacterium sp*., and *Ornithobacterium rhinotracheale*.

Rapid detection of pathogens from experimental samples

MinION sequencing of the twelve experimental samples and a negative template control is summarized in Table 2 as total reads, median read length, classified read counts, chicken host reads, fraction of host reads from total reads, read count for each pathogen from 8 to 1 h of sequencing, coverage breadth across the whole microbial genomes (AIV, IBV and MS). The fraction of host reads ranged from 0.80 to 0.96 of the total reads. The non-chicken reads (from 8 to 1 h sequencing run) were classified as microbial reads belonging to IBV, AIV and MS, alone or combined. These experimental samples were also tested using RT-qPCR for the presence of the suspected pathogens, and Ct values for the respective pathogen are included in Table 2. The Ct values for IBV fell in a narrow range from 21.94 to 25.87. IBV reads were detected in all twelve samples in the first hour of MinION sequencing. Coverage read depth was sufficient (3x) to accurately assign IBV genotypes in 9/12 samples after 1 h, and in 12/12 samples after 8 h (as well as in the MiSeq run). For AIV, reads were detected in 10/11 RT-qPCR-positive samples after 1 and 8 h. Full HA/NA subtypes were assigned for 5/11 samples after 1 h, 6/11 after 8 h, and 11/11 in the MiSeq run. For MS, reads were detected after 1 h in 10/10 RT-qPCR-positive samples. No target agents were detected in the mock control. Of all agent/sample combinations, in only one sample contained a single MS-specific read in the full 8 h sequencing run, but the sample was RT-qPCR-negative for the agent.

Reference-based and *de novo* genome determination

In addition to detection and genotyping, rapid NGS can provide detailed sequence information on the pathogens found in a sample. For each experimental sample, Table 2 lists the breadth of genome coverage (minimum 3x depth) calculated from read alignment. After 1 h of MinION sequencing, 3/12 samples had IBV coverage breadth > 50%, corresponding approximately with a Ct cutoff of 22. After 8 h, this fraction increased to 11/12. For MS, 9/10 samples had >50% coverage after 1 h, and 10/10 had >75% coverage after 8 h. For AIV, 3/11 samples had >50% coverage after a single hour, while 6/11 had similar coverage after 8 h.

Correlation of MinION and MiSeq sequencing with RT-qPCR on AIV, IBV, and MS

To model the relationship between RT-qPCR Ct values and NGS read abundance, relative read abundance (out of all microbial reads) for each of the four agents with RT-qPCR data available were plotted on a log2 scale against their known Ct values (Figures 1A–H). A linear least-squares model was fit to the data, and the lower 95% confidence interval of this model was used to estimate the lowest Ct value at which an agent read would be detected on average under several different sets of experimental assumptions. This model was also used to estimate the expected Ct value of an agent that is seen at an abundance of one read per thousand microbial reads. The estimated Ct thresholds at which a single read (MinION/MiSeq sequencing) per thousand microbial reads to be observed with 95% confidence at different run times and levels of host contamination, given current experimental conditions, and assuming 12 multiplexed samples was determined to be 27/27.5 for AIV, 26.5/26 for IBV and 36/36.5 for MS (Figures 1A–F, purple horizontal line). The three agents used in the experimental study all have strong correlation between Ct and log2 abundance (between−0.82 and−0.98). IAIV has the strongest correlation, and the largest range of Ct values (19–36) (Figures 1A,B). The other two agents in this study, IBV (Figures 1C,D) and MS (Figures 1E,F), had slightly weaker correlation, but also significant smaller Ct ranges (22–24 and 29–34, respectively).

For the clinical samples, correlation of NDV relative abundance and Ct is generally poor (Figures 1G,H). The overlaps of taxon IDs between MiSeq and MinION is demonstrated in Figure 2. Nearly parallel lines representing the number of taxon identification indicated a relationship (as expected) on the overlaps in detection as a function of taxon abundance and MinION run length. The comparison of per-taxon relative read abundances between Illumina MiSeq and MinION sequencing runs in combined clinical and experimental samples showed strong correlation ($R = 0.85$) as the identified taxa are highlighted with different color-coded symbols in Figure 3. Here, for simplicity the comparison was done with a selected number of viral and bacterial respiratory disease-causing agents. In addition, there was extremely high correlation ($R = 0.95$) between relative

TABLE 2 Microbial reads obtained at different hours (h) from random sequencing using MinION on experimental oropharyngeal swab samples.

Sample	Total reads	Median read length	Classified read count	Fraction host	Read count (8 h)			Read count (1 h)			Coverage breadth (8 h)[a]			Coverage breadth (1 h)			RT-qPCR (Ct)		
					IBV	MS	AIV	IBV	MS	AIV	IBV	MS	AIV	IBV	MS	AIV	IBV	MS	AIV
13	39,951	654	31,385	0.91	137	0	227	34	0	45	0.50	0.00	0.84	0.13	0.00	0.27	24.08	N.D.	22.08
14	45,258	705	39,587	0.92	292	294	2,006	41	49	359	0.84	0.97	0.96	0.12	0.76	0.95	23.15	33.58	18.71
15	22,151	459	14,844	0.63	143	46	17	27	6	2	0.57	0.78	0.12	0.09	0.00	0.00	24.17	33.14	25.63
16	67,481	509	38,080	0.96	159	202	321	31	33	48	0.65	0.94	0.90	0.11	0.62	0.30	24.45	33.91	21.87
17	36,239	488	19,999	0.78	684	3,037	4	114	560	1	0.98	0.99	0.00	0.39	0.98	0.00	22.38	29.34	27.75
18	42,522	576	34,922	0.87	1,360	1,893	3	263	355	2	1.00	0.98	0.00	0.66	0.97	0.00	21.94	30.97	28.79
19	24,190	294	9,007	0.80	262	662	0	56	104	0	0.70	0.98	0.00	0.21	0.92	0.00	22.42	29.87	36.21
20	26,991	470	15,809	0.90	135	1	487	30	0	73	0.62	0.00	0.95	0.04	0.00	0.58	24.02	N.D.	20.92
21	757,670	873	697,572	0.99	328	1,001	237	63	201	36	0.82	0.97	0.90	0.14	0.92	0.21	22.92	29.86	21.84
22	154,062	641	139,353	0.93	100	7,393	0	18	1,356	0	0.40	1.00	0.00	0.01	0.98	0.00	25.87	29.55	N.D.
23	124,015	609	81,368	0.88	954	1,322	727	159	234	123	1.00	0.98	0.95	0.55	0.97	0.75	22.4	31.54	21.48
24	86,752	977	79,936	0.94	736	2,061	4	152	386	1	1.00	0.99	0.00	0.62	0.96	0.00	22.52	30.12	29.17
NTC[b]	53,682	170	21	0.38	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	N.D.	N.D.	N.D.

[a]As fraction of whole genome at > 3x depth of coverage. [b]No-template control. ND, not detected; IBV, Infectious bronchitis virus; MS, Mycoplasma synoviae; AIV, Avian influenza virus.

10.3389/fvets.2022.1073919

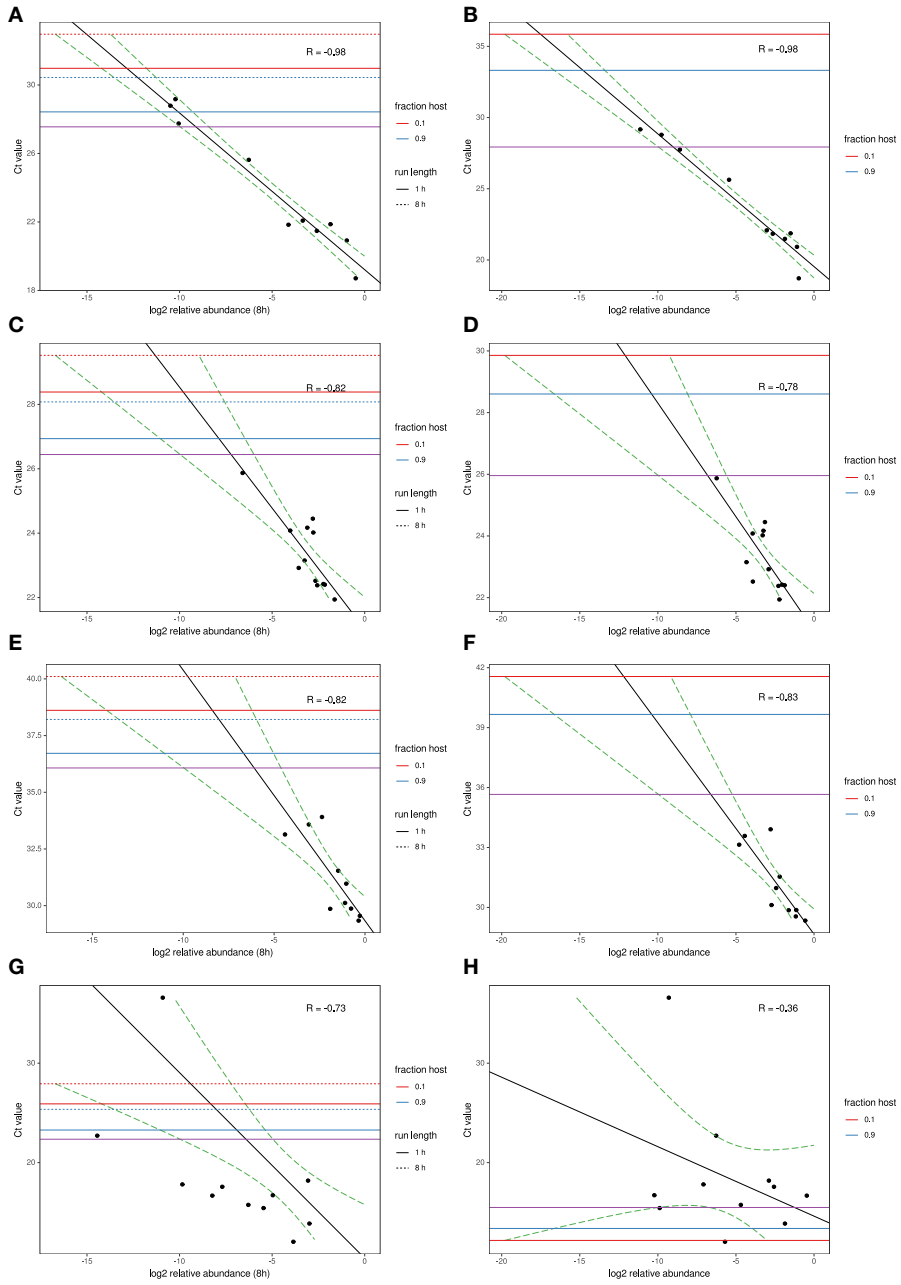


FIGURE 1
MinION and MiSeq read abundance vs RT-PCR Ct; for Avian Influenza virus **(A,B)**; for Infectious bronchitis virus **(C,D)**; for Mycoplasma synoviae **(E,F)** in experimental samples and for Newcastle disease virus **(G,H)** in clinical samples. Black line indicates best-fit linear regression model. Red/blue horizontal lines mark Ct thresholds at which a single read would be estimated to be observed with 95% confidence at different run times and levels of host contamination, given current experimental conditions, and assuming 12 multiplexed samples. Purple horizontal line marks Ct threshold corresponding to on average one agent read per thousand microbial reads at 95% confidence.

abundance estimates from MinION and MiSeq sequencing data from the experimental swab samples (Figure 4). The per-taxon relative read abundances comparison between Illumina MiSeq and MinION sequencing runs of clinical samples also showed a moderate correlation ($R = 0.79$) as shown in Figure 5.

Discussion

Sequence-based pathogen characterization approaches have evolved rapidly and are broadly accepted in the global research community. This study demonstrates the utility of random amplification for untargeted MinION nanopore sequencing

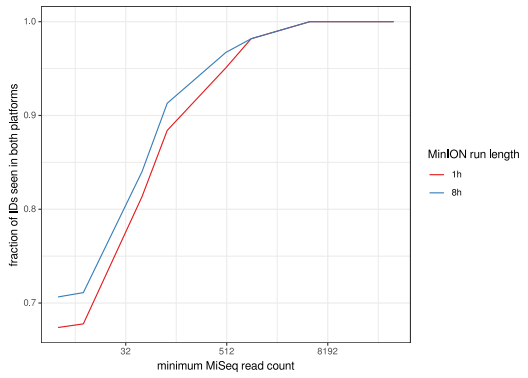


FIGURE 2
Overlaps of taxon IDs between MiSeq and MinION runs as a function of taxon abundance and MinION run length. A minimum k-mer count of 20 and a minimum b-score (subsampled BLAST agreement) of 0.7 was calculated from IDs.

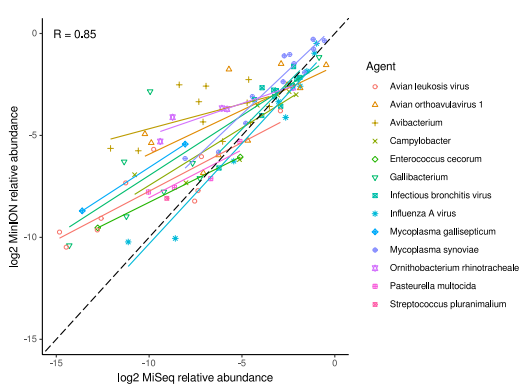


FIGURE 3
Correlation of per-taxon relative read abundances between Illumina MiSeq and MinION runs. Individual trend lines show least-squares regression models. Dashed black line indicates the identity (1:1) relationship. Data used was from experimental and clinical samples. Pearson's R value is calculated from the combined dataset.

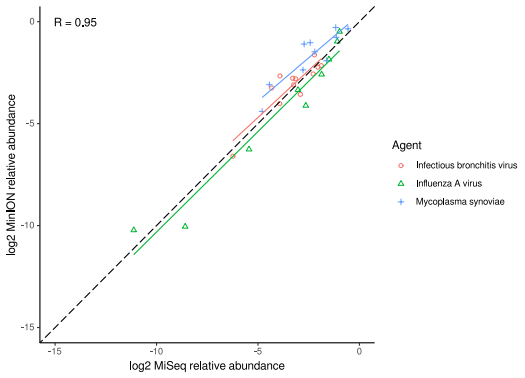


FIGURE 4
Correlation of per-taxon relative read abundances between Illumina MiSeq and MinION runs of experimental samples. Individual trend lines show least-squares regression models. Dashed black line indicates the identity (1:1) relationship. Pearson's R value is calculated from the combined dataset.

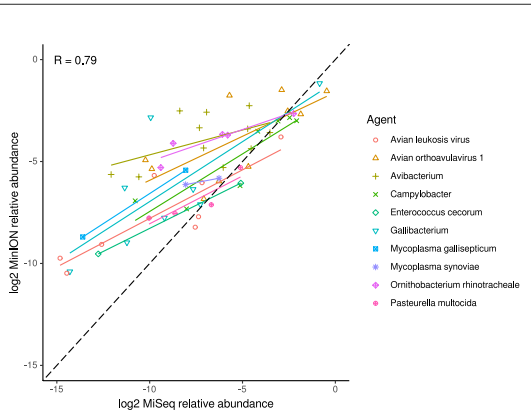


FIGURE 5
Correlation of per-taxon relative read abundances between Illumina MiSeq and MinION runs of clinical samples. Individual trend lines show least-squares regression models. Dashed black line indicates the identity (1:1) relationship. Pearson's R value is calculated from the combined dataset.

to achieve accurate identification and preliminary genetic characterization of viral and bacterial agents in co-infected clinical and experimental samples. We have shown that the MinION platform, as expected produces much smaller read output than MiSeq platform; however, in terms of positive-negative detection and sequence-based agent identification, it can approach the sensitivity of the MiSeq-based approach. This identification of the pathogen, when present at moderate abundance can be achieved with runtimes as short as 60 min, providing advantages in terms of cost, speed of data acquisition and processing time per sample in clinical settings. We have demonstrated strong quantitative correlation both between sequencing platforms and between sequence read abundance and RT-qPCR Ct values, indicating that detection sensitivity of nanopore sequencing is limited primarily by sequencing depth rather than any inherit weakness of the platform. A major advantage of this platform is the ability to adjust run times to suit requirements, even during a run. Therefore, it lends itself to cost optimization by balancing run times, multiplexing depth, and host depletion optimization to achieve target sensitivity levels.

Rapid and accurate detection and characterization of the microbial pathogens present in clinical samples has long been a major goal in diagnostic settings, and numerous advances have been made to improve tests. However, no single diagnostic

test is perfect and varying scenarios often require a variety of diagnostic tests. For rapid identification of avian respiratory pathogens, single or multiplexed PCR-based diagnostic assays have been developed and widely used. However, the target-specific nature of these assays makes them vulnerable to failure because of the ability of these pathogens to change, causing false negative results (20). Additionally, PCR-based rapid assays provide limited or no additional genetic information about the detected pathogens. Recently, a strand-switching based random MinION sequencing approach has been used on cultured viral pathogens (45). The work described here is a step further toward rapid characterization, as it demonstrates accurate identification and genetic typing of viral and bacterial pathogens from clinical samples. Also, it is demonstrated that a single targeted assay of the suspected viral pathogen (NDV) would have failed to identify co-infecting pathogens including *A. paragallinarum*, *M. pullorum*, and *G. anatis,* and which would have remained undetected in these clinical samples without additional testing. However, these bacterial pathogens were detected by untargeted MinION sequencing and in most samples confirmed by MiSeq sequencing.

Although multiplexed PCR based assays have been developed to detect multiple respiratory pathogens in a single assay, this incrementally improved approach still requires a prediction of what pathogens (or genetic variant in case of RNA viruses) are in a sample and will not identify unknown agents, which would result in the incomplete characterization of the clinical samples (8). The presence of more than one pathogen in clinical samples is known to occur and the diversity in the genetic material makes it necessary to perform additional assays to identify some of the pathogens despite the availability of multiplexed assays for respiratory viral pathogens. This MinION sequencing approach is target independent, which may reduce the chances of failure in detection of pathogens due to genetic change. Additionally, the ability to use total RNA (rRNA and mRNA) provides an opportunity to detect pathogens both from genomic viral RNA as well as the rRNAs of replicating bacterial pathogens and mRNA of replicating DNA viruses.

It has been reported that upgraded nanopore sequencing flow cells are capable of achieving as high as 95% raw accuracy (46). It is likely that as the sequencing technology and base calling algorithms will improve the single-read accuracy, further diminishing data analysis challenges specific to noisy long-read data as compared to short-read sequencing. Short-read-based metagenomics studies on platforms such as Illumina have experienced widespread use due to the high accuracy of sequencing. Although Sanger and Illumina (sequencing-by-synthesis) sequencing platforms are considered the gold standard in terms of accuracy (23), these approaches have limitations. Sanger sequencing is necessarily target specific and Illumina-based sequencing shares some similar limitations as MinION in terms of data management. It can become challenging to analyze hundreds of thousands to millions of individual reads due to the computational power and time required. In our current study, an automated pipeline running on cloud resources analyzed each sample in parallel in an average of 2.6 h (minimum 1.5, maximum 4.8) with no user intervention. This workflow overcomes the challenges associated with the lack of computational resources and speed of data analysis.

The primary current limitation of the non-targeted metagenomic sequencing assay is its lower sensitivity compared with targeted amplification. However, this study demonstrates that samples with Ct values into the 30s can be reliably detected from randomly amplified samples in as little as 1 h of sequencing time, often with depth sufficient to yield the genotypic classification. This observation is supported by linear modeling of NGS abundance and RT-qPCR thresholds, which backs the conclusion that at least some agents with Ct values $>$ 30 should be reliably identifiable under similar experimental conditions. The sensitivity of detection in these experimental samples was hampered both by a high degree of host contamination, which varied from 63 to 99% in the experimental MinION run, as well as multiplexed library sizes that varied by several orders of magnitude. Improvements in handling these challenges would be expected to significantly increase the sensitivity beyond that already observed and improve the reliability of the approach as a diagnostic tool.

One of the biggest challenges in sequencing-based diagnostics is the presence of nucleic acids in clinical samples from both host and pathogens, which may be sourced from a variety of genomic material. The total RNA sequencing approach adopted in this study allows the capture of broad population of RNAs from the clinical oral swab samples. Although oropharyngeal swab samples have comparatively lower host and commensal bacterial populations, there is still the background of many chicken-reads. Because clinical samples were collected at different time points and the quality of RNA may be low as well, these samples should be sequenced for around 8 h. It is notable that although no pre-enrichment approach was used to detect microbial RNA in the clinical samples, only 60 min of MinION sequence data was sufficient to detect all the test co-infecting viral and bacterial pathogens from the experimental samples. Specific reduction in the rRNA of chicken host will further increase the utility of this approach in recovery of viral genomes from metagenomic samples. At the moment, several companies offer kits for elimination of host and bacterial ribosomal RNAs, and the utilization of those are likely to improve the sensitivity of detection of pathogens (47). An RNaseH approach with probes targeted to rRNA from both chickens and bacteria has shown a significant increase in sensitivity and can be potentially used with this MinION approach as well (24). The utility of the assay for other species has not been tested but this protocol does not include any pre-enrichment of pathogen RNA which could also make it very useful for human pathogens as well.

Conclusions

The presence of important viral and bacterial pathogens in respiratory clinical samples of chickens was detected by direct extraction of total RNA followed by the use of Oxford Nanopore MinION or Illumina MiSeq sequencing technologies. NDV and various bacterial respiratory agents were detected in chickens from Kenyan live bird markets with both technologies. The MinION platform provided a rapid but still accurate characterization of the co-infecting viral and bacterial pathogens in experimental swab samples. Extensive testing on diverse clinical samples will further evaluate the viability of this protocol for diagnostic settings. In addition, because this MinION-based approach provides for rapid, multiplexed, and cost-effective detection of viral and bacterial pathogens in clinical samples with sufficient sensitivity for many applications, it represents a legitimate alternative for diagnostic laboratories that cannot afford more expensive equipment for next-generation diagnostics. Based on this work and related studies, the goal of a cost-effective, sensitive, and untargeted NGS-based diagnostic tool appears one step closer to reality.

Data availability statement

These dataset in this study are deposited in the SRA repository under BioProject PRJNA900571 and the accession numbers: SRR22262942- SRR22262990. These datasets can be accessed at link below: https://www.ncbi.nlm.nih.gov/sra/PRJNA900571.

Author contributions

SB processed the egg-grown and clinical samples, created the MinION libraries, analyzed the MinION data, and wrote the manuscript. HK contributed to the preparation and analysis of NGS data and manuscript preparation. JV developed the MinION data analysis workflow and assisted with manuscript preparation. TT and CL helped with the RT-qPCR for IBV, AIV, and MS. MP-J, DS, and JS assisted in data interpretation and manuscript preparation. CA was involved in the design of the study, data analysis, data interpretation, and writing of the manuscript. All authors were involved with editing the manuscript, read, and approved the final manuscript.

Funding

This project was supported by USDA CRIS 6040–32000-072.

Acknowledgments

Drs. Mark Jackwood and Naola Ferguson from Poultry Diagnostic and Research Center, College of Veterinary Medicine, University of Georgia, Athens, GA, United States helped provide the viral and bacterial reagents to optimize the protocol.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.1073919/full#supplementary-material

SUPPLEMENTARY TABLE 1
Details of clinical and experimental oropharyngeal swab samples.

SUPPLEMENTARY TABLE 2
Details of comparison of *de novo* contigs assigned per taxon from MinION and MiSeq sequencing platforms.

References

1. Saif F, Glisson M, David ES. *Diseases of Poultry 12th Edition*. Ames, IA: Blackwell Publishers (2008).

2. Guo Y, Krauss S, Senne D, Mo I, Lo K, Xiong X, et al. Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia. *Virology*. (2000) 267:279–88. doi: 10.1006/viro.1999.0115

3. Franca M, Howerth EW, Carter D, Byas A, Poulson R, Afonso CL, et al. Co-infection of mallards with low-virulence Newcastle disease virus and low-pathogenic avian influenza virus. *Avian Pathol.* (2014) 43:96–104. doi: 10.1080/03079457.2013.876530

4. Kammon A, Heidari A, Dayhum A, Eldaghayes I, Sharif M, Monne I, et al. Characterization of avian influenza and newcastle disease viruses from

poultry in libya. *Avian Dis.* (2015) 59:422–30. doi: 10.1637/11068-032215-Res Note.1

5. Pan Q, Liu A, Zhang F, Ling Y, Ou C, Hou N, et al. Co-infection of broilers with ornithobacterium rhinotracheale and H9N2 avian influenza virus. *BMC Vet Res.* (2012) 8:104. doi: 10.1186/1746-6148-8-104

6. Agnew-Crumpton R, Vaz PK, Devlin JM, O'Rourke D, Blacker-Smith HP, Konsak-Ilievski B, et al. Spread of the newly emerging infectious laryngotracheitis viruses in Australia. *Infect Genet Evol.* (2016) 43:67–73. doi: 10.1016/j.meegid.2016.05.023

7. Kouakou AV, Kouakou V, Kouakou C, Godji P, Kouassi AL, Krou HA, et al. Prevalence of Newcastle disease virus and infectious bronchitis virus in avian influenza negative birds from live bird markets and backyard and commercial farms in Ivory-Coast. *Res Vet Sci.* (2015) 102:83–8. doi: 10.1016/j.rvsc.2015.07.015

8. Umali DV, Ito H, Shirota K, Ito T, Katoh H. Atypical velogenic Newcastle disease in a commercial layer flock in Japan. li DV, Ito H, Shirota K*Poult Sci.* (2015) 94:890–7. doi: 10.3382/ps/pev011

9. Jones RC. Viral respiratory diseases (ILT, aMPV infections, IB): are they ever under control? *Br Poult Sci.* (2010) 51:1–11. doi: 10.1080/00071660903541378

10. Haghighat-Jahromi M, Asasi K, Nili H, Dadras H, Shooshtari A. Coinfection of avian influenza virus (H9N2 subtype) with infectious bronchitis live vaccine. *Arch Virol.* (2008) 153:651–5. doi: 10.1007/s00705-008-0033-x

11. Dufour-Zavala L. *A Laboratory Manual for the Isolation, Identification, and Characterization of Avian Pathogens.* Athens, GA: American Association of Avian Pathologists (2008).

12. Slemons R, Shieldcastle M, Heyman L, Bednarik K, Senne D. Type A influenza viruses in waterfowl in Ohio and implications for domestic turkeys. *Avian Dis.* (1991) 35:165–73. doi: 10.2307/1591309

13. Zowalaty MEE, Chander Y, Redig PT, El Latif HKA, Sayed MAE, Goyal SM. Selective isolation of Avian influenza virus (AIV) from cloacal samples containing AIV and Newcastle disease virus. *J Vet Diagn Invest.* (2011) 23:330–2. doi: 10.1177/104063871102300222

14. Costa-Hurtado M, Afonso CL, Miller PJ, Spackman E, Kapczynski DR, Swayne DE, et al. Virus interference between H7N2 low pathogenic avian influenza virus and lentogenic Newcastle disease virus in experimental co-infections in chickens and turkeys. *Vet Res.* (2014) 45:1. doi: 10.1186/1297-9716-45-1

15. Xie Z, Pang Y-s, Liu J, Deng X, Tang X, Sun J, et al. A multiplex RT-PCR for detection of type A influenza virus and differentiation of avian H5, H7, and H9 hemagglutinin subtypes. *Mol Cell Probes.* (2006) 20:245–9. doi: 10.1016/j.mcp.2006.01.003

16. Pang Y, Wang H, Girshick T, Xie Z, Khan MI. Development and application of a multiplex polymerase chain reaction for avian respiratory agents. *Avian Dis.* (2002) 46:691–9. doi: 10.1637/0005-2086(2002)046[0691:DAAOAM]2.0.CO;2

17. Xie Z, Luo S, Xie L, Liu J, Pang Y, Deng X, et al. Simultaneous typing of nine avian respiratory pathogens using a novel GeXP analyzer-based multiplex PCR assay. *J Virol Methods.* (2014) 207:188–95. doi: 10.1016/j.jviromet.2014.07.007

18. Tadese T, Potter AE, Fitzgerald S, Reed WM. Concurrent infection in chickens with fowlpox virus and infectious laryngotracheitis virus as detected by immunohistochemistry and a multiplex polymerase chain reaction technique. *Avian Dis.* (2007) 51:719–24. doi: 10.1637/0005-2086(2007)51[719:CIICWF]2.0.CO;2

19. Belák S, Thorén P, LeBlanc N, Viljoen G. Advances in viral disease diagnostic and molecular epidemiological technologies. *Expert Rev Mol Diagn.* (2009) 9:367–81. doi: 10.1586/erm.09.19

20. Cattoli G, De Battisti C, Marciano S, Ormelli S, Monne I, Terregino C, et al. 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.* (2009) 47:3791–2. doi: 10.1128/JCM.00895-09

21. Ansorge WJ. Next-generation DNA sequencing techniques. *N Biotechnol.* (2009) 25:195–203. doi: 10.1016/j.nbt.2008.12.009

22. 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. *J Pathol.* (2013) 229:535–45. doi: 10.1002/path.4145

23. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE.* (2009) 4:e7370. doi: 10.1371/journal.pone.0007370

24. Parris DJ, Kariithi H, Suarez DL. Non-target RNA depletion strategy to improve sensitivity of next-generation sequencing for the detection of RNA viruses in poultry. *J Vet Diagn Invest.* (2022) 34:638–45. doi: 10.1177/10406387221102430

25. Kariithi HM, Volkening JD, Leyson CM, Afonso CL, Christy N, Decanini EL, et al. Genome sequence variations of infectious bronchitis virus serotypes from commercial chickens in Mexico. *Front Vet Sci.* (2022) 9:931272. doi: 10.3389/fvets.2022.931272

26. Kariithi HM, Christy N, Decanini EL, Lemiere S, Volkening JD, Afonso CL, et al. Detection and genome sequence analysis of avian metapneumovirus subtype a viruses circulating in commercial chicken flocks in Mexico. *Vet Sci.* (2022) 9:579. doi: 10.3390/vetsci9100579

27. Rutvisuttinunt W, Chinnawirotpisan P, Simasathien S, Shrestha SK, Yoon I-K, Klungthong C, et al. Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq platform. *J Virol Methods.* (2013) 193:394–404. doi: 10.1016/j.jviromet.2013.07.001

28. Brown BL, Watson M, Minot SS, Rivera MC, Franklin RB. MinION^{TM} nanopore sequencing of environmental metagenomes: a synthetic approach. *Gigascience.* (2017) 6:1–10. doi: 10.1093/gigascience/gix007

29. 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. doi: 10.1186/s13742-015-0051-z

30. Butt SL, Taylor TL, Volkening JD, Dimitrov KM, Williams-Coplin D, Lahmers KK, et al. Rapid virulence prediction and identification of Newcastle disease virus genotypes using third-generation sequencing. *Virol J.* (2018) 15:179. doi: 10.1186/s12985-018-1077-5

31. Butt SL, Erwood EC, Zhang J, Sellers HS, Young K, Lahmers KK, et al. Real-time, MinION-based, amplicon sequencing for lineage typing of infectious bronchitis virus from upper respiratory samples. *J Vet Diagn Invest.* (2021) 33:179–90. doi: 10.1177/1040638720910107

32. Wang J, Moore NE, Deng Y-M, Eccles DA, Hall RJ. MinION nanopore sequencing of an influenza genome. *Front Microbiol.* (2015) 6:766. doi: 10.3389/fmicb.2015.00766

33. Spatz SJ, Garcia M, Riblet S, Ross TA, Volkening JD, Taylor TL, et al. MinION sequencing to genotype US strains of infectious laryngotracheitis virus. *Avian Pathol.* (2019) 48:255–69. doi: 10.1080/03079457.2019.1579298

34. Young KT, Stephens JQ, Poulson RL, Stallknecht DE, Dimitrov KM, Butt SL, et al. Putative Novel Avian Paramyxovirus (AMPV) and Reidentification of APMV-2 and APMV-6 to the Species Level Based on Wild Bird Surveillance (United States, 2016–2018). *Appl Environ Microbiol.* (2022) 88:e00466-22. doi: 10.1128/aem.00466-22

35. Kariithi HM, Welch CN, Ferreira HL, Pusch EA, Ateya LO, Binepal YS, et al. Genetic characterization and pathogenesis of the first H9N2 low pathogenic avian influenza viruses isolated from chickens in Kenyan live bird markets. *Infect Genet Evol.* (2020) 78:104074. doi: 10.1016/j.meegid.2019.104074

36. Callison SA, Hilt DA, Boynton TO, Sample BF, Robison R, Swayne DE, et al. Development and evaluation of a real-time Taqman RT-PCR assay for the detection of infectious bronchitis virus from infected chickens. *J Virol Methods.* (2006) 138:60–5. doi: 10.1016/j.jviromet.2006.07.018

37. Spackman E. Avian influenza virus detection and quantitation by real-time RT-PCR. *Methods Mol Biol.* (2014) 1161:105–18. doi: 10.1007/978-1-4939-0758-8_10

38. Raviv Z, Kleven SH. The development of diagnostic real-time TaqMan PCRs for the four pathogenic avian mycoplasmas. *Avian Dis.* (2009) 53:103–7. doi: 10.1637/8469-091508-Reg.1

39. Imai K, Tamura K, Tanigaki T, Takizawa M, Nakayama E, Taniguchi T, et al. Whole genome sequencing of influenza A and B viruses with the MinION sequencer in the clinical setting: a pilot study. *Front Microbiol.* (2018) 9:2748. doi: 10.3389/fmicb.2018.02748

40. Breitwieser FP, Baker DN, Salzberg SL. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* (2018) 19:198. doi: 10.1186/s13059-018-1568-0

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

42. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W, MEGAHIT. an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics.* (2015) 31:1674–6. doi: 10.1093/bioinformatics/btv033

43. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191

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

45. Young KT, Lahmers KK, Sellers HS, Stallknecht DE, Poulson RL, Saliki JT, et al. Randomly primed, strand-switching, MinION-based sequencing for the detection and characterization of cultured RNA viruses. *J Vet Diagn Invest.* (2021) 33:202–15. doi: 10.1177/1040638720981019

46. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* (2020) 21:1–16. doi: 10.1186/s13059-020-1935-5

47. Zeng Z, Huang B, Wang X, Fan J, Zhang B, Yang L, et al. A reverse transcriptase-mediated ribosomal RNA depletion (RTR2D) strategy for the cost-effective construction of RNA sequencing libraries. *J Adv Res.* (2020). doi: 10.1016/j.jare.2019.12.005

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fvets.2022.1080397

OPEN ACCESS

EDITED BY
Iryna Goraichuk,
United States Department of Agriculture (USDA), United States

REVIEWED BY
Lauro Velazquez-Salinas,
Agricultural Research Service (USDA), United States
Christopher Lewis Netherton,
The Pirbright Institute, United Kingdom

*CORRESPONDENCE
WeonHwa Jheong
purify@korea.kr

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 26 October 2022
ACCEPTED 19 December 2022
PUBLISHED 11 January 2023

CITATION
Kim G, Park J-E, Kim S-J, Kim Y, Kim W, Kim Y-K and Jheong W (2023) Complete genome analysis of the African swine fever virus isolated from a wild boar responsible for the first viral outbreak in Korea, 2019. *Front. Vet. Sci.* 9:1080397. doi: 10.3389/fvets.2022.1080397

Complete genome analysis of the African swine fever virus isolated from a wild boar responsible for the first viral outbreak in Korea, 2019

Garam Kim†, Jung-Eun Park†, So-Jeong Kim, Yeonji Kim, Wonjun Kim, Yong-Kwan Kim and WeonHwa Jheong*

Wildlife Disease Response Team, National Institute of Wildlife Disease Control and Prevention (NIWDC), Gwangju, Republic of Korea

African swine fever (ASF), a highly contagious and severe hemorrhagic viral disease in swine, is emerging as a major threat not only in Korea but also worldwide. The first confirmed case of ASF in Korea was reported in 2019. Despite the occurrence of ASF in Korea, only a few studies have genetically characterized the causative ASF virus (ASFV). In this study, we aimed to genetically characterize the ASFV responsible for the 2019 outbreak in Korea. The genome of the ASFV isolated during the first outbreak in Korea was analyzed. The Korea/YC1/2019 strain has 188,950 base pairs, with a GC content of 38.4%. The complete genome sequence was compared with other ASFV genomes annotated in the NCBI database. The Korea/YC1/2019 strain shared the highest similarity with Georgia 2007, Belgium 2018/1, and ASFV-wbBS01 strains. This study expands our knowledge of the genetic diversity of ASFV, providing valuable information for epidemiology, diagnostics, therapies, and vaccine development.

KEYWORDS

African swine fever, African swine fever virus, central variable region (CVR), complete genome, multigene families, phylogenetic analysis

1. Introduction

African swine fever (ASF) is caused by the African swine fever virus (ASFV) in pigs and wild boars; it is highly contagiousness and associated with high mortality (1). ASF is mainly transmitted through direct contact between infected and susceptible domestic pigs, and through indirect contact with contaminated pork, vehicles, and excrement and infected humans (2, 3). In 2007, ASFV was introduced to Georgia through the port of Poti as a potential contaminant in food used as swine feed (4). It has since spread rapidly to many countries including China, Vietnam, Cambodia, Hong Kong, North Korea, Laos, Myanmar, the Philippines, and South Korea (5, 6). The first case of ASF in Korea occurred in September 2019 at a pig farm in Paju, Gyeonggi-do (7). In the wild, ASF was first reported on Yeoncheon in October 2019 in wild boar (8). In 2019, after the outbreak, it

was limited to Yeoncheon, Cheolwon, and Hwacheon, but since 2020, ASF has spread to the east and south (9). During an ASF outbreak, analysis of the ASFV genome has proven to be the most useful tool for tracing the origin of ASFV (10). The full-length sequence of pig farm generated from farms in Paju (MT748042) and Yeoncheon (MW049116) in 2019 is registered with NCBI. The partial sequencing of the wild boar occurrence has been partially performed, but the genome information has not been revealed yet.

ASFV is a large double-stranded DNA virus belonging to the family *Asfaviridae* (11). The ASFV genome is 190–193 kb in size and encodes more than 150 open reading frames, with the central regions of the genes being highly conserved (12, 13). Although the biological functions of ASFV genes include nucleotide replication, messenger RNA processing, structural protein synthesis, and host defenses modulation, the functions of more than half of the genes in the ASFV genome are still unknown (12, 14). As the ASFV genome is large and complex, it is difficult to develop vaccines and drugs (15).

A comparative analysis of molecular properties of specific regions of the ASFV genome has proven useful in elucidating the origin and transmission pathways of ASFV during ASF outbreaks (16). Based on the p72 major capsid protein-encoding gene of ASFV *(B646L)*, the virus can be classified into at least 24 ASFV genotypes (16). The central variable region (CVR) within the *B602L* gene has been shown to be informative about the relationship between isolates at the genotype, national, and regional levels (17). A recent study revealed that a tandem repeat sequence (TRS) located in the intergenic region between *I73R* and *I329L* can be used to determine the origin of ASFV isolates (18). Additionally, the analysis of other sequences, such as *EP402R* and *E183L*, may help improve molecular epidemiological studies of ASFV (19, 20). To better understand and control the spread of ASF, it is necessary to analyze the genetic characteristics of ASFV strains. However, in South Korea, there have been no studies to characterize the complete genome of the ASFV strain responsible for the occurrence of ASF. In this study, we analyzed the complete genome sequence of the ASFV strain, which first occurred in Korea in 2019, and compared it with strains from neighboring countries.

2. Materials and methods

2.1. Analysis of distribution of ASF outbreaks

Data for this study were acquired from the Pig Progress website (http://www.pigprogress.net), and they are presented in Figure 1. The data pertaining to ASF outbreaks re-ported in China, Russia, and North Korea from 2018 to 2019 were retrieved from this database. A cartographical analysis of the geological location of ASF outbreaks was performed using the open-source Geographic Information System (GIS) software.

2.2. Detection of ASFV in wild boar samples

DNA was extracted from the blood of a wild boar using the Maxwell RSC Viral Total Nucleic Purification Kit (Promega, Madison, WI, USA) following the manufacturer's instructions. The presence of ASFV DNA was detected via polymerase chain reaction (PCR) using the ASFV diagnostic primers *PPA1* (5′-AGTTATGGGAAACCCGACCC-3′), *PPA2* (5′-CCCTGAATCGGAGCATCCT-3′) (21), P72D (5′-GTACTGTAACGCAGCACAG-3′), and *P72U* (5′-GGCACAAGTTCGGACATGT-3′) (16), which partially amplified *B646L* (*p72*).

2.3. Genetic characterization of ASFV and phylogenetic analysis of *B646L* (*p72*)

To demonstrate the phylogenetic organization of the ASFV, we selected 16 related viruses from the initial phylogeny and built a maximum-likelihood phylogenetic tree of the whole genome sequences with RAxML version 8.0.0 (REF) using default parameters and a general time-reversible model with gamma-distributed rate variation among sites. The entire genome sequences were aligned using the MAFFT in Geneious Prime software. To assess relatedness support, maximum likelihood (ML) phylogenetic tree of the p72 gene was constructed under the MEGA. ML bootstrapping was performed with 1,000 replicates to assess the robustness of tree topologies.

The nucleotide sequence of *B646L* (*p72*) of the ASFV Korea/YC1/2019 strain was aligned with that of other ASFV strains representing the *B646L* (*p72*) genotype using the ClustalW algorithm in MEGA X. The evolutionary history was inferred using the Maximum Composite Likelihood model method. Phylogenetic analysis was performed using the neighbor-joining method with 1,000 bootstrap replications.

2.4. Complete genome sequencing of ASFV

The total DNA was extracted directly from 200 μl of whole blood from a wild boar using the Maxwell Viral Total Nucleic Purification Kit (Promega) following the manufacturer's recommendations. To detect the presence of ASFV, PCR amplification of the samples was performed as described in the World Organization for Animal Health (OIE) manual using primers for *PPA1/PPA2* and *P72*. gDNA was sheared

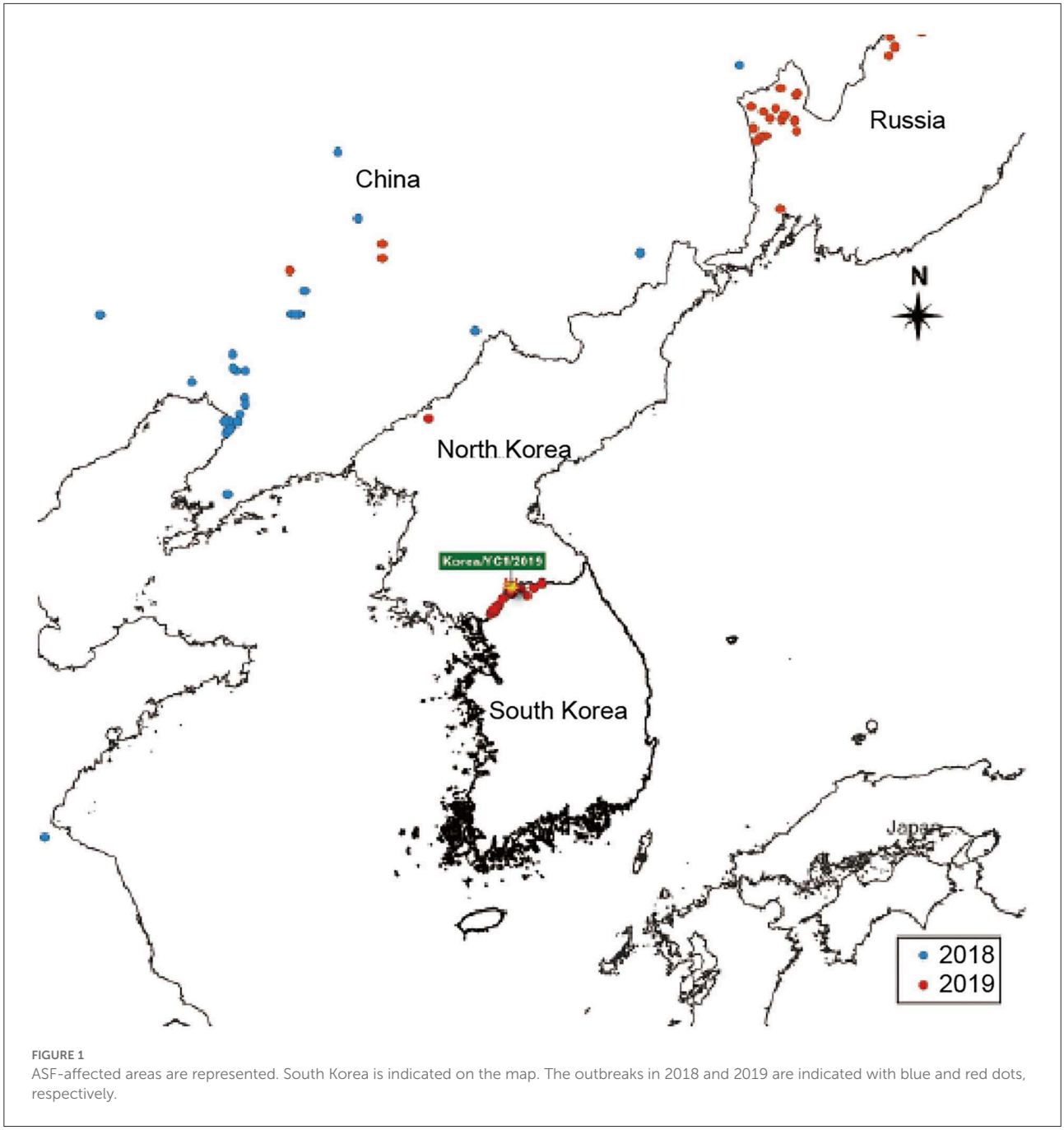


FIGURE 1
ASF-affected areas are represented. South Korea is indicated on the map. The outbreaks in 2018 and 2019 are indicated with blue and red dots, respectively.

and made library preparation using Enzymatic Preparation Kit (Celemics, Seoul, Republic of Korea). Prepared gDNA library and capture probes were hybridized to capture target regions through the use of Celemics target enrichment kit (Celemics, Seoul, Republic of Korea). Capture probes were designed and chemically synthesized to hybridize target region. Captured regions were then further amplified by post-PCR to enrich the amount of sample. The target-captured library were then sequenced on an Illumina NextSeq550 instrument (Illumina, San Diego, CA, USA) using the read layout 2 × 150 bp. The adaptor sequences and low quality bases were first trimmed using Fastx Toolkit (fastx_toolkit 0.0.14). The exact sequence that trimmed with the AdapterRemoval (version 2.2.2). The reads were mapped on reference ASFV genomes (accession number: FR682468) using Burrows-Wheeler Aligner software version 0.7.10. SNP, InDel, and SV variations were detected using GATK (Genome Analysis TK 4.0.4.0). The quality of read alignment was assessed using the SAMtools software (samtools

TABLE 1 Summary of the Korea/YC1/2019 strain genomic sequencing data.

Strain	Sample type	Raw reads	Filtered read	ASFV read (pre-filtered)	Mean coverage (pre-filtered)	ASFV read (post-filtered)	Mean coverage (post-filtered)
Korea/YC1/2019	Blood	13,496,754	13,451,232	13,028,940	7,511	10,048,618	5,917

1.1) and the Python software package (numpy 1.11.0). The ASFV complete genome Korea/YC1/2019 was annotated with the genome annotation transfer utility (GATU) software using the genome of ASFV Georgia2007 as the reference.

3. Results

3.1. First outbreak of ASF in wild boars in the Korean Peninsula

To analyze the epidemic situation in Korea and neighboring countries, data on ASF outbreaks notified by the OIE from 2018 to 2019 were retrieved. The ASF outbreaks in Korea and neighboring countries from 2018 to 2019 are shown in Figure 1. The first ASF outbreak in Asia occurred in China in 2018 and subsequently spread to Mongolia, Vietnam, Cambodia, North Korea, Laos, Philippines, and Myanmar (5, 6). As shown in Figure 1, ASF was detected in a pig farm in North Korea (in Jagang-do) on May 25, 2019. Five months later, ASF occurred in South Korea. ASF spread to Yeoncheon, South Korea in September 2019. Since then, the distribution of ASF has spread rapidly. Continued outbreaks of ASF in South Korea have raised awareness regarding the negative effects of ASF on the pork industry.

3.2. Characteristics of the complete genome sequence of ASFV Korea/YC1/2019

The genome of Korea/YC1/2019 strain was successfully obtained from the blood of the first ASFV-positive wild boars in Korea, Yeoncheon, 2019 (10). In order to characterize the ASFV Korea/YC1/2019 strain, Illumina reads were aligned against the Georgia 2007/1 reference sequence (Table 1). The complete genome generated via genome assembly was 188,950 bp with a GC content of 38.4%, and 183 open reading frames were annotated using CGView (Circular Genome Viewer) (Figure 2). Coverage for the forward and reverse strands is shown in the outer and middle circles, respectively. The ORFs of the Korea/YC1/2019 strain contains those that encode 4 proteins involved in host cell interactions, 16 structural proteins, and 25 proteins involved in nucleotide metabolism, transcription, replication, and repair. Furthermore, 47 MGF members were identified within the genome of Korea/YC1/2019 strain including MGF100 (3 members), MGF110 (12 members), MGF300 (3 members), MGF360 (19 members), and MGF505 (10 members) (Figure 2).

To determine genetic relationships at the whole-genome level, we performed multiple sequence alignments of the whole-genome sequences. The 29 strains are listed by accession strain name, accession number, country, and year of isolation in Table 2. The newly determined Korea/YC1/2019 genome with 188,950 bp is shorter than Georgia 2007 (190,584 bp) and Belgium 2018/1 (190,599 bp) genomes (Table 2). The complete genome of Korea/YC1/2019 exhibited 99.9% nucleotide identity with Georgia 2007/1, China/CAS19-01/2019, HLJ/2018, POL/2015/Podlaskie, and China/2018/AnhuiXCGQ genomes (Table 2). The nucleotide identity with other ASFV p72 genotype II derived from wild boars, such as Belgium 2018/1 and ASFV-wbBS01, was 100% (Table 2). In contrast, the Estonia 2004 strain (genotype II) exhibited 97.8% nucleotide identity with the Korea/YC1/2019 strain (Table 2). Unlike ASFV genotype II viruses, including the Korea/YC1/2019 strain, ASFV Estonia 2004 exhibited deletion and specific rearrangement at the 5′ end, which resulted in reduced virulence (26).

In previous studies, the Korea/YC1/2019 strain was identified as a member of the genotype II group on the basis of *B646L* encoding the capsid protein p72 (10). The partial nucleotide sequence of *B646L* of the Korea/YC1/2019 strain was aligned with the sequence of the indicated ASFV strains using the ClustalW algorithm in MEGA X. The phylogenetic analysis based on the *B646L* gene sequences revealed that the 29 strains were grouped into eight genotypes (Figure 3). A comparison of ASFV sequences revealed 100% similarity between the Korea/YC1/2019 strain and the Georgia 2007 and China/CAS19-01/2019 strains.

Several studies have demonstrated that tandem repeat sequences (TRS) located in the CVR of *B602L* are suitable as genetic markers for distinguishing the genotype of ASFVs (16). The CVR tetrameric repeats of ASFV in the Korea/YC1/2019 strain included CADT, NVDT, CASM, CAST, and CSTS, which correspond to the B, N, D, and A codes, respectively. The results showed a single TRS profile of 10 amino acid tetramers "BNDBNDBNAA" in Korea/YC1/2019, with 100% sequence identity to the Georgia 2007 strain (Table 3).

The ASFVs detected in Korea clustered together in the whole genome ML phylogeny, suggesting high genetic relatedness of

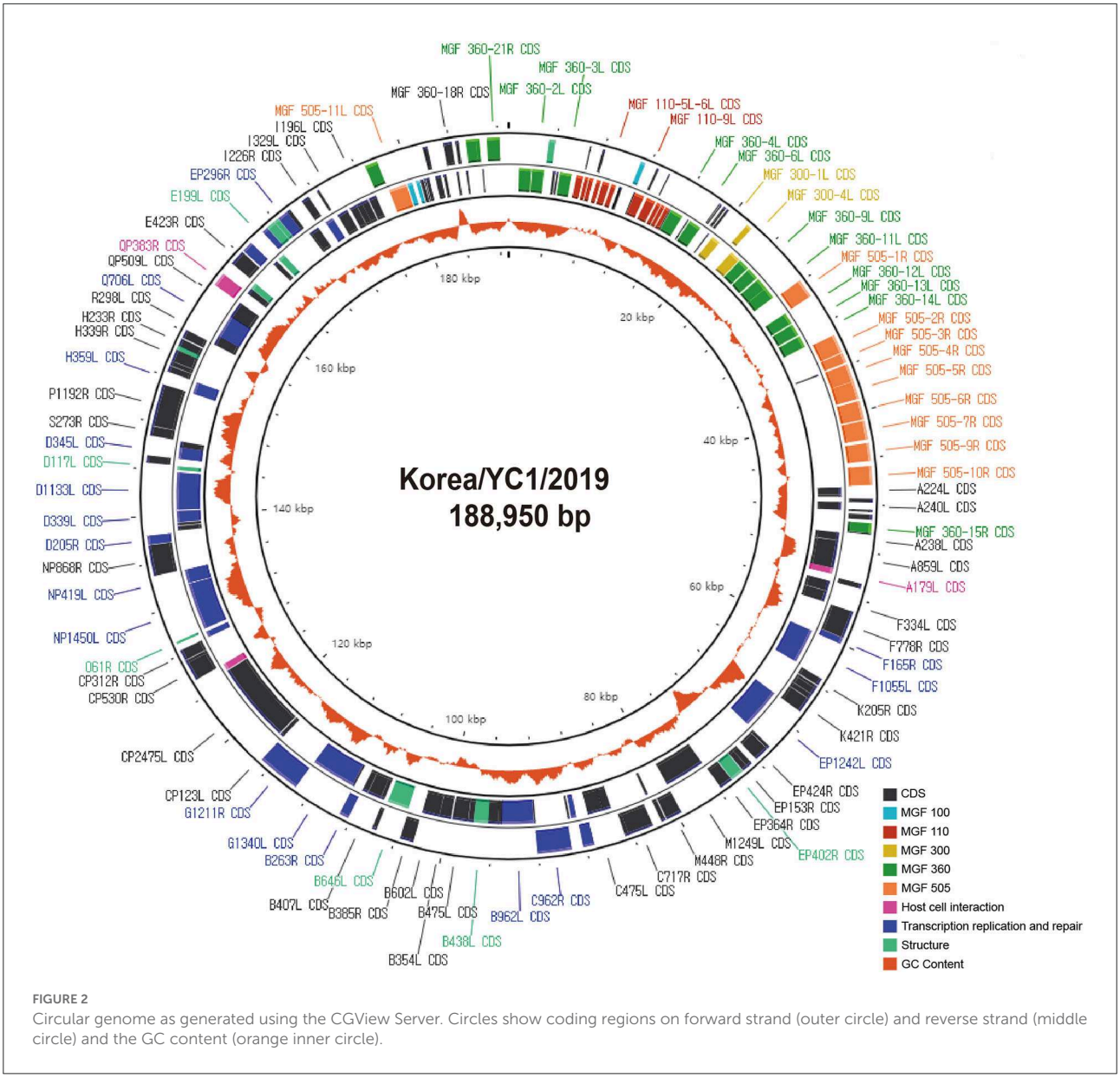


FIGURE 2
Circular genome as generated using the CGView Server. Circles show coding regions on forward strand (outer circle) and reverse strand (middle circle) and the GC content (orange inner circle).

the viruses. The ASFVs detected in Korea and China shared a common ancestry and formed a well-supported monophyletic cluster with high bootstrap support (>70%), showing that the ASFV isolates detected in Korea are most likely descendants of the viruses that circulated in China (Figure 4).

Moreover, the Korea/YC1/2019 strain had numerous insertions corresponding to the 10-nucleotide sequence TATATAGGAA, a TRS between *I73R* and *I329L* (Figure 5). A previous study has reported that the insertion in this region has no relationship with attenuation or virulence of the ASFV (18). This insertion was present in the genome of Belgium 2018/1 (accession no. LR356725), but absent in the genome of POL/2015/Podlaskie (accession no. MH681419) and Georgia 2007/1. The generated complete genome sequences were submitted to GenBank and assigned an accession number (ON075797 for Korea/YC1/2019).

4. Discussion

ASFV partial sequencing data are commonly used to determine ASFV genotypes and distinguish related ASFV strains. However, studies on complete genome sequencing are required to obtain adequate information about ASFV genetic variation and molecular evolution events. In this study, we characterized the genome of the ASFV strain that caused the first outbreak of ASF in Korea in 2019. A comparison of the genome sequence of the ASFV Korea/YC1/2019 strain

TABLE 2 Comparison of genome features of different ASFV strains and Korea/YC1/2019.

Strain	Accession number	Country	Year	Length (bp)	Genotype	Identity to Korea/YC1/2019	References
Georgia 2007/1	FR682468.2	Georgia	2007	190584	II	99.9	(22)
Arm/07/CBM/c2	LR812933	Armenia	2007	190145	II	99.9	(23)
Tanzania/Rukwa/2017/1	LR813622	Tanzania	2007	183186	II	99.9	(24)
Odintsovo_02/14	KP843857	Russia	2014	189333	II	99.9	(25)
Estonia 2014	LS478113	Estonia	2014	182446	II	97.8	(26)
POL/2015/Podlaskie	MH681419	Poland	2015	189394	II	99.9	(27)
Belgium 2018/1	LR536725	Belgium	2018	189404	II	99.9	(28)
ASFV-wbBS01	MK645909	China	2018	189394	II	100	Unpublished
China/2018/AnhuiXCGQ	MK128995	China	2018	189393	II	99.9	(29)
ASFV_HU_2018	MN715134	China	2018	190601	II	99.9	(30)
Pig/HLJ/2018	MK333180	China	2018	189404	II	99.9	(19)
CAS19-01/2019	MN172368	China	2019	189405	II	99.9	(31)
Wuhan2019-1	MN393476	China	2019	190576	II	99.9	Unpublished
Ulyanovsk 19/WB-5699	MW306192	Russia	2019	189263	II	99.9	(32)
Portugal/L60	KM262844	Portugal	1960	182362	I	92.7	(33)
BA71	KP055815	Spain	1971	180365	I	90.9	(34)
Spain/E75	FN557520	Spain	1975	181187	I	92.6	(35)
Mkuzi1979	AY261362	South Africa	1979	192714	I	95.7	Unpublished
Benin 97/1	NC_044956	Benin	1997	182284	I	92.7	(11)
47/Ss/2008	KX354450	Italy	2008	184638	I	92.4	Unpublished
Warmbaths	AY261365	South Africa	1987	190773	III	94.5	Unpublished
Namibia/Warthog	AY261366	Namibia	1980	186528	IV	93.1	(36)
Ken06.Bus	KM111295	Kenya	2006	184368	IX	84.7	(37)
Uganda/R35	MH025920	Uganda	2015	188629	IX	86.8	Unpublished
Tengani62	AY261364	Malawi	1962	185689	V	91.9	Unpublished
MalawiLil-20/1	AY261361	Malawi	1983	187162	VIII	87.7	(38)
Kenya1950	AY261360	Kenya	1950	193886	X	86.1	(37)
BUR/18/Rutana	MW856067	Burundi	2018	176564	X	83.5	(39)
ASFV Ken.rie1	LR899131	Kenya	2019	190592	X	86.8	Unpublished

with the sequences of other p72 genotype II strains showed that it shared similarities with Georgia and China strains. As a result of their recent introduction to Georgia and the subsequent rapid dissemination north of the Russian Federation and east to China and Southeast Asia, the genotype II p72 group ASFV represents by far the most geographically widespread of the 24 viral genotypes (16, 22). The virus responsible for the 2019 ASF outbreak in Korea clustered into p72 genotype II and showed high nucleotide identity with ASFV strains causing outbreaks in neighboring countries, suggesting that the same ASFV strains are causing outbreaks across borders. Moreover, recent molecular epidemiological analyses have indicated that only genotype II appears to be widespread in Korea. Molecular characterization of the ASFV Korea/YC1/2019 strain showed that it is highly homologous and almost identical to the samples that were obtained from Georgia (2007), Russia (2012), Estonia (2014), and China (2018) (40). The onset of epidemics in different countries with similar characteristics, despite very different dates of initiation, suggests that outbreaks originated from a single source and then

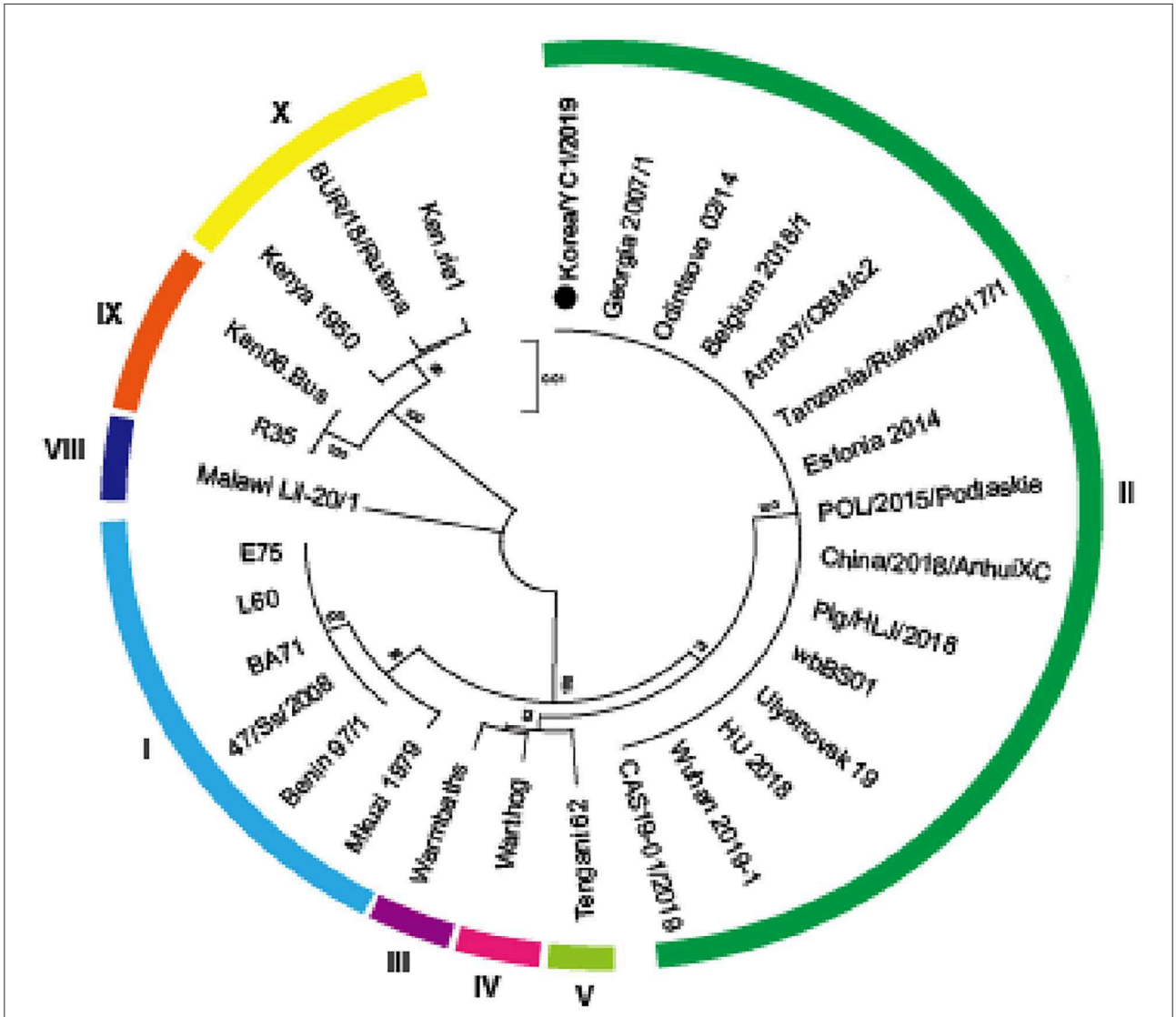


FIGURE 3
Phylogenetic relationship of ASFV strains based on *p72* (*B646L*). A phylogenetic tree based on the full-length *p72* sequence alignment of the Korea/YC1/2019 strain and 29 public ASFV strains. ASFV Korea/YC1/2019 isolated in this study is marked with a black dot. ASFV strains are named by their isolate names. Genotypes I, II, III, IV, V, VIII, IX, and X are labeled in sky blue, green, purple, pink, yellow-green, blue, yellow, and orange, respectively. Phylogeny was inferred following 1,000 bootstrap replications and node values show percentage bootstrap support. The scale bar indicates nucleotide substitutions per site.

propagated across Asia. The spread of ASF across Asia has been largely unidirectional.

The gene encoding p72, *B646L*, is relatively conserved. Therefore, p72 clustering is the preferred method for identifying the origin of ASFV because it can help trace the source of the virus at the molecular level, offering insights into possible transmission routes. Other genotypes can also be used, such as *E183L* (p54), *CP204L* (p30), *B602L* (CVR), and TRS (16–18, 41–43). The previous studies provide useful information for molecular characterization of ASFV strains. According to previous studies on partial gene sequencing, the B602L, CVR, and TRS genotypes were found in these ASFV strains circulating in Europe and recently in China (42). A comparison of the IGR sequences of the Korea/YC1/2019 strain and the Korean-type Paju strains Korea/19S3965/wb/2019 (MT300324) and Korea/19S5464/wb/2019 (MT300325) revealed a deletion or insertion in each Paju strain, suggesting that it may have a different origin.

The dsDNA genome (170–194 kb) of ASFV, with repeats and scattered invert-ed-terminal-repeats (ITR), hamper whole-genome sequencing (14). The total length of the ASFV Georgia 2007/1 genome is 190,594 bp, 5′-ITR region is 956 bp, 3′-ITR region is 231 bp, and total length of each ITR region is 1,378 bp, with ORF DP60R and ASFV G ACD 01990 annotated

TABLE 3 Tetrameric amino acid sequence in the central variable region (CVR) of *B602L* in ASFV strains.

Strain	CVR (aa sequence)	No. of repeats	IGR type
Korea/YC1/2019	BNDBNDBNAA	10	II
Georgia 2007/1	BNDBNDBNAA	10	I
China/CAS19-01/2019	BNDBNDBNAA	10	II
Belgium 2018/1	BNDBNDBNAA	10	II
ASFV-wbBS01	BNDBNDBNAA	10	I
Arm/07/CBM/c2	BNDBNDBNAA	10	I
Wuhan2019-1	BNDBNDBNAA	10	II
Tanzania/Rukwa/2017/1	BNDBNDBNAA	10	I
China/2018/AnhuiXCGQ	BNDBNDBNAA	10	II
POL/2015/Podlaskie	BNDBNDBNAA	10	I
Odintsovo_02/14	BNDBNDBNAA	10	II
Estonia 2014	BNDBNDBNAA	10	II
ASFV_HU_2018	BNDBNDBNAA	10	I
Ulyanovsk 19/WB-5699	BNDBNDBNAA	10	III
Pig/HLJ/2018	BNDBNDBNAA	10	III
Benin 97/1	ABNAAAAFBNAAAAAFBNAAAAAFBNAAAAFBNAFA	36	I
Mkuzi1979	BVWAFNBNAAAF	12	I
Porutugal/L60	ABNAAAAFBNABNABNABNVNTDBNAFA	25	I
47/Ss/2008	ABNAAADBNAFA	12	I
Spain/E75	ABNAAAAAFBNABNABNAB	26	I
BA71	ABNAAAAFBNABNABNABNABNVNTDBNAFA	28	I
Warmbaths	BVWVWVVNAABAG	13	I
Namibia/Warthog	BNAB	4	I
Tengani62	ABNBBMA	7	III
MalawiLil-20/1	AVSVSOVNAVNOVVNVOVNAVNOVVNOVOOV	31	I

Letters in CVR sequence represent the TRS previously characterized in ASFV isolates: A = CAST, CVST, CTST, CASI; B = CADT, CADI, CTDT, CAGT, CVDT; F = CANT, CAAT; N = NVDT, NVGT, NVDI; T = DVDT, NVNT; S = SAST; O = NANI, NADI, NASI; V = NAGT, NAST, NAVT, NADT, NANT, NANV; D = CASM; G = CTNT; M = NEDT; W = SADT, SVDT.

ITR regions (44). Inverted terminal repeats were missing from both ends of the Korea/YC1/2019 genome sequence, presumably because of the complexity of the ASFV material or region in the limited number of sequenced DNA samples and/or the difficulty of sequencing/assembly. In addition, the Illumina sequencing platform only generates short reads, which makes it difficult to effectively sequence and assemble repeat regions, as short repeat repeats may collapse (45).

ASF was first discovered in Korea in 2019, and its outbreaks have been occurring since then. In this study, we analyzed the genome of ASFV in wild boars in Korea in 2019 for the first time. The results of whole-genome sequencing of ASFV in this study can provide information on genetic variation and help track sources and entry points. In Korea, 50 wild boars were infected with ASFV in 2019, 814 in 2020, and 964 in 2021. From 2019 to 2021, the number of benign cases in hunting animals seems to have steadily increased to 5, 40, and 116 cases in 2019, 2020, and 2021, respectively. Accordingly, with the increasing incidence of ASF in wild boars, more studies are being conducted to trace the genetic mutations in ASFV. We plan to analyze the whole ASFV genomes by year and region to continuously track the transmission route and genetic variation.

5. Conclusions

Here, we report the genome characterization of the ASFV Korea/YC1/2019 strain. This whole-genome characterization of ASFV may contribute to tracing the evolution of ASFV during its spread. The genetic analysis showed that the

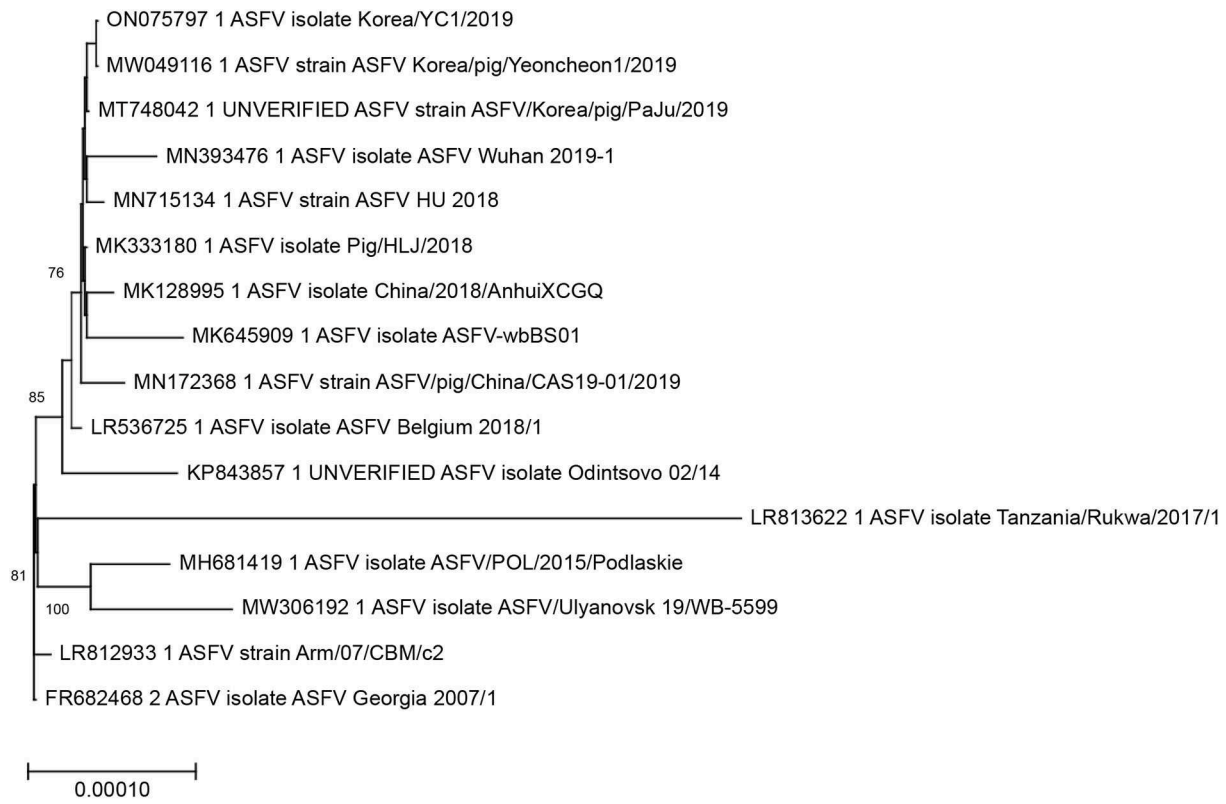


FIGURE 4
Maximum-likelihood analysis of 16 complete coding sequences of ASFV including the Korea/YC1/2019 strain sequenced in this study. The phylogeny was rooted at the Georgia/2007 virus. The scale bars show the number of substitutions per site. The numerical values represent 1,000 bootstrap replicate values >70 expressed as a percentage.

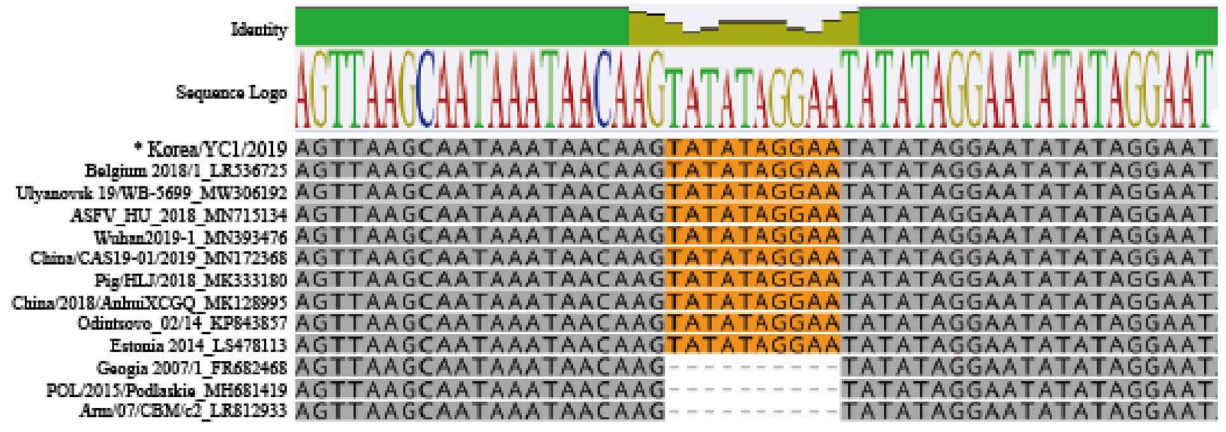


FIGURE 5
Sequence alignment of intergenic region between *I73R* and *I329L* in ASFV genotype II and Korea/YC1/2019. Orange shading on the sequences indicates the 10-nucleotide tandem repeat sequence (TRS) insertion.

Korea/YC1/2019 strain is closely clustered with genotype II ASFVs, providing insights into the 2019 ASF outbreak. In addition, analysis of these results will provide valuable information for the improvement of ASF diagnostic methods and vaccine development, as well as epidemiological evidence that can be used to trace the virus in a specific region. Further

studies of the emerging ASFVs are needed to provide more insights into genetic characterization and variations.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, ON075797.

Author contributions

Conceptualization: WJ and GK. Formal analysis: YK. Investigation: J-EP, WK, and Y-KK. Data curation: S-JK. Writing—original draft preparation, writing—review and editing, and project administration: GK. Supervision: WJ. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by a grant from the National Institute of Wildlife Disease Control and Prevention (NIWDC), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIWDC-2022-SP-02).

Acknowledgments

We would like to thank the wildlife response team for laboratory analysis of the field sample. We are grateful to Donghun Lee for insightful discussions and useful suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Arias M, Sánchez-Vizcaíno JM. African swine fever. *Trends Emerg Viral Infect Swine.* (2002) 2002:119–24.

2. Costard S, Wieland B, de Glanville W, Jori F, Rowlands R, Vosloo W, et al. African swine fever: How can global spread be prevented? *Philos Trans R Soc Lond B Biol Sci.* (2009) 364:2683–96. doi: 10.1098/rstb.2009.0098

3. Olesen AS. *Investigation of transmission dynamics and virulence of new African swine fever virus strains* (Ph.D. thesis). Technical University of Denmark, Denmark (2019).

4. Rowlands RJ, Michaud V, Heath L, Hutchings G, Oura C, Vosloo W, et al. African swine fever virus isolate, Georgia, 2007. *Emerg Infect Dis.* (2008) 14:1870–4. doi: 10.3201/eid1412.080591

5. Lu G, Pan J, Zhang G. African swine fever virus in Asia: its rapid spread and potential threat to unaffected countries. *J Infect.* (2020) 80:350–71. doi: 10.1016/j.jinf.2019.11.011

6. Kedkovid R, Sirisereewan C, Thanawongnuwech R. Major swine viral diseases: An Asian perspective after the African swine fever introduction. *Porcine Health Manag.* (2020) 6:20. doi: 10.1186/s40813-020-00159-x

7. Kim HJ, Cho KH, Lee SK, Kim DY, Nah JJ, Kim HJ, et al. Outbreak of African swine fever in South Korea, 2019. *Transbound Emerg Dis.* (2020) 67:473–475. doi: 10.1111/tbed.13483

8. Kim SH, Kim J, Son K, Choi Y, Jeong HS, Kim YK, et al. Wild boar harboring African swine fever virus in the demilitarized zone in South Korea, 2019. *Emerg Microbes Infect.* (2020) 9:628–30. doi: 10.1080/22221751.2020.1738904

9. Jo YS, Gortázar C. African swine fever in wild boar: assessing interventions in South Korea. *Transbound Emerg Dis.* (2021) 68:2878–89. doi: 10.1111/tbed.14106

10. Gallardo MC, Reoyo AD, Fernández-Pinero J, Iglesias I, Muñoz MJ, Arias ML. African swine fever: a global view of the current challenge. *Porcine Health Management.* (2015) 1:21. doi: 10.1186/s40813-015-0013-y

11. Chapman DAG, Tcherepanov V, Upton C, Dixon LK. Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *J Gen Virol.* (2008) 89:397–408. doi: 10.1099/vir.0.83343-0

12. Alejo A, Matamoros T, Guerra M, Andrés G. A proteomic atlas of the African swine fever virus particle. *J Virol.* (2018) 92:e01293–18. doi: 10.1128/JVI.01293-18

13. Yáñez RJ, Rodríguez JM, Nogal ML, Yuste L, Enríquez C, Rodriguez JF, et al. Analysis of the complete nucleotide sequence of African swine fever virus. *Virology.* (1995) 208:249–78. doi: 10.1006/viro.1995.1149

14. Dixon LK, Chapman DA, Netherton CL, Upton C. African swine fever virus replication and genomics. *Virus Res.* (2013) 173:3–14. doi: 10.1016/j.virusres.2012.10.020

15. Sang H, Miller G, Lokhandwala S, Sangewar N, Waghela SD, Bishop RP, et al. Progress toward development of effective and safe African swine fever virus vaccines. *Front Vet Sci.* (2020) 7:84. doi: 10.3389/fvets.2020.00084

16. Bastos AD, Penrith ML, Crucière C, Edrich JL, Hutchings G, Roger F, et al. Genotyping field strains of African swine fever virus by partial p72 gene characterisation. *Arch Virol.* (2003) 148:693–706. doi: 10.1007/s00705-002-0946-8

17. Gallardo C, Mwaengo DM, Macharia JM, Arias M, Taracha EA, Soler A, et al. Enhanced discrimination of African swine fever virus isolates through nucleotide sequencing of the p54, p72, and pB602L (CVR) Genes. *Virus Genes.* (2009) 38:85–95. doi: 10.1007/s11262-008-0293-2

18. Sanna G, Dei Giudici S, Bacciu D, Angioi PP, Giammarioli M, De Mia GM, et al. Improved strategy for molecular characterization of African swine fever viruses From Sardinia, based on analysis of p30, CD2V and I73R/I329L variable regions. *Transbound Emerg Dis.* (2017) 64:1280–6. doi: 10.1111/tbed.12504

19. Shi K, Liu H, Yin Y, Si H, Long F, Feng S. Molecular characterization of African swine fever virus from 2019–2020 outbreaks in Guangxi Province, southern China. *Front Vet Sci.* (2022) 9:912224. doi: 10.3389/fvets.2022.912224

20. Rodríguez JM, Yáñez RJ, Almazán F, Viñuela E, Rodriguez JF. African swine fever virus encodes a CD2 homolog responsible for the adhesion of erythrocytes to infected cells. *J Virol.* (1993) 67:5312–20. doi: 10.1128/JVI.67.9.5312-5320.1993

21. Agüero M, Fernández J, Romero L, Sánchez Mascaraque C, Arias M, Sánchez-Vizcaíno JM. Highly sensitive PCR assay for routine diagnosis of African swine fever virus in clinical samples. *J Clin Microbiol.* (2003) 41:4431–4. doi: 10.1128/JCM.41.9.4431-4434.2003

22. Chapman DA, Darby AC, Da Silva M, Upton C, Radford AD, Dixon LK. Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. *Emerg Infect Dis.* (2011) 17:599–605. doi: 10.3201/eid1704.101283

23. Pérez-Núñez D, Castillo-Rosa E, Vigara-Astillero G, García-Belmonte R, Gallardo C, Revilla Y. Identification and isolation of two different subpopulations within African swine fever virus Arm/07 stock. *Vaccines.* (2020) 8:625. doi: 10.3390/vaccines8040625

24. Njau EP, Domelevo Entfellner JB, Machuka EM, Bochere EN, Cleaveland S, Shirima GM, et al. The first genotype II African swine fever virus isolated in Africa provides insight into the current Eurasian pandemic. *Sci Rep.* (2021) 11:13081. doi: 10.1038/s41598-021-92593-2

25. Pershin A, Shevchenko I, Igolkin A, Zhukov I, Mazloum A, Aronova E, et al. A long-term study of the biological properties of ASF virus isolates originating from various regions of the Russian Federation in 2013–2018. *Vet Sci.* (2019) 6:99. doi: 10.3390/vetsci6040099

26. Zani L, Forth JH, Forth L, Nurmoja I, Leidenberger S, Henke J, et al. Deletion at the 5′-end of Estonian ASFV strains associated with an attenuated phenotype. *Sci Rep.* (2018) 8:6510. doi: 10.1038/s41598-018-24740-1

27. Mazur-Panasiuk N, Wozniakowski G, Niemczuk K. The first complete genomic sequences of African swine fever virus isolated in Poland. *Sci Rep.* (2019) 9:4556. doi: 10.1038/s41598-018-36823-0

28. Pikalo J, Schoder ME, Sehl J, Breithaupt A, Tignon M, Cay AB, et al. The African swine fever virus isolate Belgium 2018/1 shows high virulence in European wild boar. *Transbound Emerg Dis.* (2020) 67:1654–9. doi: 10.1111/tbed.13503

29. Bao J, Wang Q, Lin P, Liu C, Li L, Wu X, et al. Genome comparison of African swine fever virus China/2018/Anhui XCGQ strain and related European p72 genotype II strains. *Transbound Emerg Dis.* (2019) 66:1167–76. doi: 10.1111/tbed.13124

30. Olasz F, Mészáros I, Marton S, Kaján GL, Tamás V, Locsmándi G, et al. A simple method for sample preparation to facilitate efficient whole-genome sequencing of African swine fever virus. *Viruses.* (2019) 11:1129. doi: 10.3390/v11121129

31. Jia L, Jiang M, Wu K, Hu J, Wang Y, Quan W, et al. Nanopore sequencing of African swine fever virus. *Sci China Life Sci.* (2020) 63:160–4. doi: 10.1007/s11427-019-9828-1

32. Mazloum A, van Schalkwyk A, Shotin A, Igolkin A, Shevchenko I, Gruzdev KN, et al. Comparative analysis of full genome sequences of African swine fever virus isolates taken from wild boars in Russia in 2019. *Pathogens.* (2021) 10:521. doi: 10.3390/pathogens10050521

33. Portugal R, Coelho J, Höper D, Little NS, Smithson C, Upton C, et al. Related strains of African swine fever virus with different virulence: genome comparison and analysis. *J Gen Virol.* (2015) 96:408–19. doi: 10.1099/vir.0.070508-0

34. Rodríguez JM, Moreno LT, Alejo A, Lacasta A, Rodríguez F, Salas ML. Genome sequence of African swine fever virus BA71, the virulent parental strain of the nonpathogenic and tissue-culture adapted BA71V. *PLoS ONE.* (2015) 10:e0142889. doi: 10.1371/journal.pone.0142889

35. De Villiers EP, Gallardo C, Arias M, da Silva M, Upton C, Martin R, et al. Phylogenomic analysis of 11 complete African swine fever virus genome sequences. *Virology.* (2010) 400:128–36. doi: 10.1016/j.virol.2010.01.019

36. Zsak L, Borca MV, Risatti GR, Zsak A, French RA, Lu Z, et al. Preclinical diagnosis of African swine fever in contact-exposed swine by a real-time PCR assay. *J Clin Microbiol.* (2005) 43:112–9. doi: 10.1128/JCM.43.1.112-119.2005

37. Njau EP, Machuka EM, Cleaveland S, Shirima GM, Kusiluka LJ, Okoth EA, et al. African swine fever virus (ASFV): biology, genomics and genotypes circulating in sub-Saharan Africa. *Viruses.* (2021) 13:2285. doi: 10.3390/v13112285

38. Rodriguez F, Alcaraz C, Eiras A, Yáñez RJ, Rodriguez JM, Alonso C, et al. Characterization and molecular basis of heterogeneity of the African swine fever virus envelope protein p54. *J Virol.* (1994) 68:7244–52. doi: 10.1128/JVI.68.11.7244-7252.1994

39. Ge S, Li J, Fan X, Liu F, Li L, Wang Q, et al. Molecular characterization of African swine fever virus, China, 2018. *Emerg Infect Dis.* (2018) 24:2131–3. doi: 10.3201/eid2411.181274

40. Nix RJ, Gallardo C, Hutchings G, Blanco E, Dixon LK. Molecular epidemiology of African swine fever virus studied by analysis of four variable genome regions. *Arch Virol.* (2006) 151:2475–94. doi: 10.1007/s00705-006-0794-z

41. Forth JH, Forth LF, King J, Groza O, Hübner A, Olesen AS, et al. A deep-sequencing workflow for the fast and efficient generation of high-quality African swine fever virus whole-genome sequences. *Viruses.* (2019) 11:846. doi: 10.3390/v11090846

42. Meireles M, Costa JV. Nucleotide sequence of the telomeric region of the African swine fever virus genome. *Virology.* (1994) 203:193–6. doi: 10.1006/viro.1994.1474

43. Dixon LK, Twigg SR, Baylis SA, Vydelingum S, Bristow C, Hammond JM, et al. Nucleotide sequence of a 55-Kbp region from the right end of the genome of a pathogenic African swine fever virus isolate (Malawi LIL20/1). *J Gen Virol.* (1994) 75:1655–84. doi: 10.1099/0022-1317-75-7-1655

44. Hakizimana JN, Ntirandekura JB, Yona C, Nyabongo L, Kamwendo G, Chulu JLC, et al. Complete genome analysis of African swine fever virus responsible for outbreaks in domestic pigs in 2018 in Burundi and 2019 in Malawi. *Trop Anim Health Prod.* (2021) 53:438. doi: 10.1007/s11250-021-02877-y

45. Bishop RP, Fleischauer C, de Villiers EP, Okoth EA, Arias M, Gallardo C, et al. Comparative analysis of the complete genome sequences of Kenyan African swine fever virus isolates within p72 genotypes IX and X. *Virus Virus Genes.* (2015) 50:303–9. doi: 10.1007/s11262-014-1156-7

frontiers | Frontiers in Veterinary Science

TYPE Brief Research Report
PUBLISHED 12 January 2023
DOI 10.3389/fvets.2022.1073648

OPEN ACCESS

EDITED BY
Moh A. Alkhamis,
Kuwait University, Kuwait

REVIEWED BY
Mohamed Nayel,
University of Sadat City, Egypt
Laxmi Narayan Sarangi,
National Dairy Development Board, India

*CORRESPONDENCE
Jiong Huang
jh124@163.com

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 18 October 2022
ACCEPTED 13 December 2022
PUBLISHED 12 January 2023

CITATION
Wei Y-R, Ma W-G, Wang P, Wang W, Su X-H, Yang X-Y, Mi X-Y, Wu J-Y and Huang J (2023) Retrospective genomic analysis of the first Lumpy skin disease virus outbreak in China (2019). *Front. Vet. Sci.* 9:1073648. doi: 10.3389/fvets.2022.1073648

Retrospective genomic analysis of the first Lumpy skin disease virus outbreak in China (2019)

Yu-Rong Wei[1], Wen-Ge Ma[1], Ping Wang[1], Wen Wang[2], Xiao-Hui Su[2], Xue-Yun Yang[1], Xiao-Yun Mi[1], Jian-Yong Wu[1] and Jiong Huang[1]*

[1]Xinjiang Key Laboratory of Animal Infectious Diseases, Institute of Veterinary Medicine, Xinjiang Academy of Animal Science, Urumqi, China, [2]Center for Animal Disease Prevention and Control of Xinjiang, Urumqi, China

Lumpy skin disease caused by Lumpy skin disease virus (LSDV) is a severe systemic disease affecting cattle and other ruminants. Lumpy skin disease was first reported in northwest China in August 2019 and has severely threatened the cattle breeding industry in China. However, there have been limited genomic studies of LSDV from the first outbreak and its subsequent epidemics. This study aims to characterize the comparative genomic evolution of the LSDV strain from the first outbreak in China. The etiological agent was isolated in a Madin-Darby bovine kidney cell culture and subsequently identified by PCR and Sanger sequencing of six selected genes. The genome sequence was determined using Illumina sequencing and analyzed through genome alignment and phylogenetic tree. The results showed that all six genes were successfully amplified and genetically clustered into LSDV. The virus presented the highest homology to strain China/GD01/2020, which shared 100% identities among 150 open reading frames (ORFs), and 97.1–99.7% identities among additional 6 ORFs. Bayesian inference tree analysis revealed that the virus shared a common ancestor with LSDV strains from China and Vietnam. The study provides an additional genomic data for LSDV tracking and control in China and neighboring countries.

KEYWORDS

Lumpy skin disease virus, Illumina sequencing, phylogenetic analysis, genome sequence, genome alignment

1. Introduction

Lumpy skin disease (LSD) is a significant transboundary viral disease that affects cattle water buffalos and giraffe, and other ruminants (1–4). The disease is caused by the Lumpy skin disease virus (LSDV), a member of the family Poxviridae, genus Capripoxvirus (5, 6). The virus appears to be mechanically transmitted by blood-sucking arthropods such as flies, mosquitoes and ticks, and to a lesser extent by direct contact between cattle (7, 8). In addition, LSDV-contaminated milk, blood, nasal secretions, and saliva are alternative sources of infection through feeding or drinking routes (9). The affected animals mainly manifest fevers and nodular lesions, and they produce

dramatically less milk and undergo weight loss (10). LSDV can cause a high incidence of 5–45% when introduced into a herd, and the case fatality rate ranges from 0.5 to 7.0% (11, 12). As a result, LSD poses a significant economic threat to the cattle-breeding industry.

LSD was initially described in Zambia in 1929 and identified as a communicable disease in the 1940's (6). In the 1950's, LSDV spread rapidly through Central and Eastern Africa. It then spread out of Africa into the Middle East in the 1980's (13). Since 2015, LSD outbreaks have occurred in countries neighboring China, such as Kazakhstan, and Russia (14, 15). In August 2019, the first outbreak of LSD was reported in the Xinjiang Uygur Autonomous Region, Northwest China, in which borders Kazakhstan (16). The disease was then reported in eight provincial administrative regions (Anhui, Fujian, Hong Kong, Guangdong, Jiangxi, Sichuan, Taiwan, and Zhejiang), resulting in 10 LSDV outbreaks in China (16–18). From January 2013 to July 2021, there were 28,442 LSDV outbreaks worldwide, resulting in 326,300 cases and 15,500 deaths (19). The global dissemination of LSDV has resulted in serious risk of this contagious disease affecting the large cattle population (more than 95.6 million cattle in stock) in China.

The genome of LSDV is a linear double-stranded 145–152 kb DNA molecule that contains 150–156 predicted open reading frames (ORFs). The first complete genome sequence was determined in 2001 from primary lamb testicle cells of the Neethling type strain 2,490, which contained 156 annotated genes (20). Of these encoded genes, the G protein-coupled chemokine receptor (GPRC) and the RNA polymerase 30 kDa subunit (RP030) were recognized as markers for differentiating the poxviruses at the family and genus levels (21, 22). In addition, the LSDV contains 90 core genes conserved in all chordopoxviruses, and they have been used for phylogenetic analyses (17). Since 2001, approximately 40 LSDV genome sequences of different origins have been sequenced and deposited in GenBank, but comparative genomic data and evolutionary studies are still limited.

The genomes of Chinese strains from Guangdong and Hong Kong in South China have been sequenced (17, 23), and previous studies of these circulating Chinese LSDV strains showed close genetic relationship with the LSDV/Russia/Saratov/2017 (accession no. MH646674.1) strain or Neethling vaccine strain based on single or multiple genes (11). Thus, the genetic relationships between the LSDV isolate present in the first outbreak in China and those of subsequent epidemics in China and other countries still needs to be clarified. In this study, we employed next-generation sequencing to obtain the complete genome sequence of the LSDV/China/XJ01/2019 strain isolated from the only cow that died during the first LSDV outbreak in China. We also performed a detailed genomic comparison of LSDV/China/XJ01/2019 and related genomic sequences reported before and after this outbreak. This study provides insight into the spread of LSDV during the epidemic in China.

2. Materials and methods

2.1. Specimens

In August 2019, the first ever LSD outbreak occurred in Ili Kazakh Autonomous Prefecture, Xinjiang Uygur Autonomous Region, Northwest China, which shares an approximately 50-km border with Kazakhstan (Supplementary Figure S1). During the outbreak, a Holstein cow was found to have died from LSD. After an examination and dissection of the dead dairy cow, a skin nodule sample was collected. It was then transported and shipped to the laboratory under cold conditions and immediately stored at −80°C for further testing. The owner of the animal was informed about the purpose and process of this study. The farmer agreed to allow the skin nodule samples to be collected from his died cow.

2.2. DNA extraction and amplification

The virus was cultured in Madin-Darby bovine kidney cell, and after three generations of culturing, the viral genome was extracted by using a QIAamp DNA Mini Kit (QIAGEN). The RPO30 and GPRC genes were amplified using the primers listed in Supplementary Table S1. Four additional primers (containing deletion or insertion sequences compared with Goatpox and Sheeppox viruses) were designed to amplify Ankyrin repeat protein (LSDV152), Interleukin-1 receptor-like protein (LSDV013), Putative alpha amanitin-sensitive protein (LSDV009) and Putative late transcription factor (LSDV076) genes that can be used to differentiate the genus Capripoxvirus referred to as the strain, LSDV/Russia/Saratov/2017 (accession no. MH646674.1) (Supplementary Table S1). These six genes were purified using Quick Gel columns (QIAGEN), and then ligated into pMD19-T (Tiangen Biotech) and transformed into Escherichia coli DH5α competent cells (Tiangen Biotech). Quintuplicate positive clones were extracted for Sanger sequencing (Sangon Biotech). Nucleotide sequences of the above six genes were downloaded from the National Center for Biotechnology Information, USA (Supplementary Table S2). Phylogenetic relatedness analyses were carried out using the MEGA 11 (https://megasoftware.net/) using the Maximum Likelihood method and the best fitting DNA model with 1,000 bootstrap replicates (24).

2.3. Genome sequencing and analysis

The LSDV genomic DNA was used for library construction and next-generation sequencing (Novagene). The raw reads were processed using a standard in-house pipeline (Novagene) to remove adapters, host sequences, chimeras, short reads,

and low-quality reads. The clean data were further mapped to LSDV/Russia/Saratov/2017 (accession no. MH646674.1) by using Geneious prime software (25). The resulting contigs were assembled into a whole-genome sequence, which was then mapped onto a reference genome, resulting in a draft genome sequence. The clean reads were mapped to the gaps between the draft genome sequence and LSDV/Russia/Saratov/2017 (accession no. MH646674.1) using the medium sensitivity/Fast mode and iterated up to five times using Geneious prime 2020.0.3 software (25). The mapped sequences were used to generate consensus sequences to obtain the primary genomic sequence, which was then manually checked. Genome annotation were performed using GATU software (26) with the 20L81_Bang-Thanh VNM 20 (accession no. MZ577076.1) and Kubash KAZ 16 (accession no. MN642592.1) genomes as reference. The annotations were manually verified and curated using the Ugene software package (27).

The genome sequence of the strain LSDV/China/XJ01/2019 was aligned to a set of reference LSDV sequences retrieved from GenBank using ClustalX 2.1 (http://www.clustal.org/clustal2/). A phylogenetic tree was generated using the alignment and Bayesian approaches in MrBayes v. 3.2.7 (28) to evaluate the relationships between LSDV/China/XJ01/2019 and reference genome sequences in GenBank (Supplementary Table S3). Phylogenetic reconstruction was performed using the GTR evolutionary model including a Γ distribution and two runs of four chains each. The chain convergence was evaluated after 200,000 generations. Results were considered stable if an EES value was >200.

3. Results

3.1. Virus identification

The GPCR and PRO30 genes were successfully amplified from the cell culture, primers for four other genes (Ankyrin repeat protein, Interleukin-1 receptor-like protein, Putative alpha amanitin-sensitive protein, and Putative late transcription factor) were designed to confirm the presence of LSDV, and finally, all six genes were tested as positive. After Sanger sequencing and phylogenetic analysis, all six genes were clustered with LSDV; consequently, the virus was identified as LSDV (Figure 1) and named LSDV/China/XJ01/2019.

3.2. Genome assembly

The complete genome sequence of the LSDV/China/XJ01/2019 strain was analyzed using an Illumina NovaSeq sequencer (Illumina, USA) generating 150-bp single reads. A total of 7,496,974 (150 × 150 bp) PE150 clean reads were obtained from the Illumina NovaSeq sequencer. Consensus sequences were generated by *de novo* assembling and mapping to LSDV/Russia/Saratov/2017 (accession no. MH646674.1). The clean reads were subsequently mapped to the consensus sequences and the average coverage was determined to be 728.6× (Supplementary Figure S1). The viral genome sequence was then submitted and deposited in GenBank under accession no. OM105589.

3.3. Genomic comparisons between LSDV/China/XJ01/2019 and other strains in and around China

Pairwise genome sequence comparison revealed that LSDV/China/XJ01/2019 strain shared the highest similarity (99.9960%) with 20L43_Ly-Quoc/VNM/20, followed by 20L43_Ly-Quoc/VNM/20 (99.9954%), 20L42_Quyet-Thang/VNM/20 (99.9947%), 20L81_Bang-Thanh/VNM/20 (99.9947%) and China/GD01/2020 (99.9893%). The mVista program (http://genome.lbl.gov/vista/mvista/submit.shtml) was used to analyze the genome-wide differences among the strains in and around China. Seven regions of the viral genome were extremely variable, and it contained deletions and mutation mainly in genes LSDV008 (Putative soluble interferon-gamma receptor gene), LSDV011 (G protein-coupled chemokine receptor-like protein gene), LSDV126 (putative EEV glycoprotein gene), LSDV145 (Ankyrin repeat protein gene) and LSDV146 (Phospholipase D-like protein gene). Specifically, complete genome sequences of Chinese strains LSDV/China/XJ01/2019, China/GD01/2020, and LSDV/HongKong/2020 are greatly similar to the genomes of the Vietnamese strains 20L42_Quyet-Thang/VNM/20, 20L43_Ly-Quoc/VNM/20, 20L70_Dinh-To/VNM/20 and 20L81_Bang-Thanh/VNM/20, with only three significantly different regions observed (Figure 2).

3.4. Open reading frame comparisons of the LSDV/China/XJ01/2019 with other strains in and around China

The genome sequence was annotated referred to the strain 20L81_Bang-Thanh VNM 20 (accession no. MZ577076.1), a total of 156 open reading frames (ORFs) were identified in the strain, LSDV/China/XJ01/2019. Comparing the ORFs of LSDV/China/XJ01/2019 with the most closely related strain China/GD01/2020 revealed 150 ORFs sharing 100% sequence identities and 6 ORFs sharing 97.1–99.7% sequence identities (Table 1).

A total of 17 variable loci in the genomic sequence, resulting in variations in 7 ORFs and 2 non-coding regions, were found between LSDV/China/XJ01/2019 and the most closely related

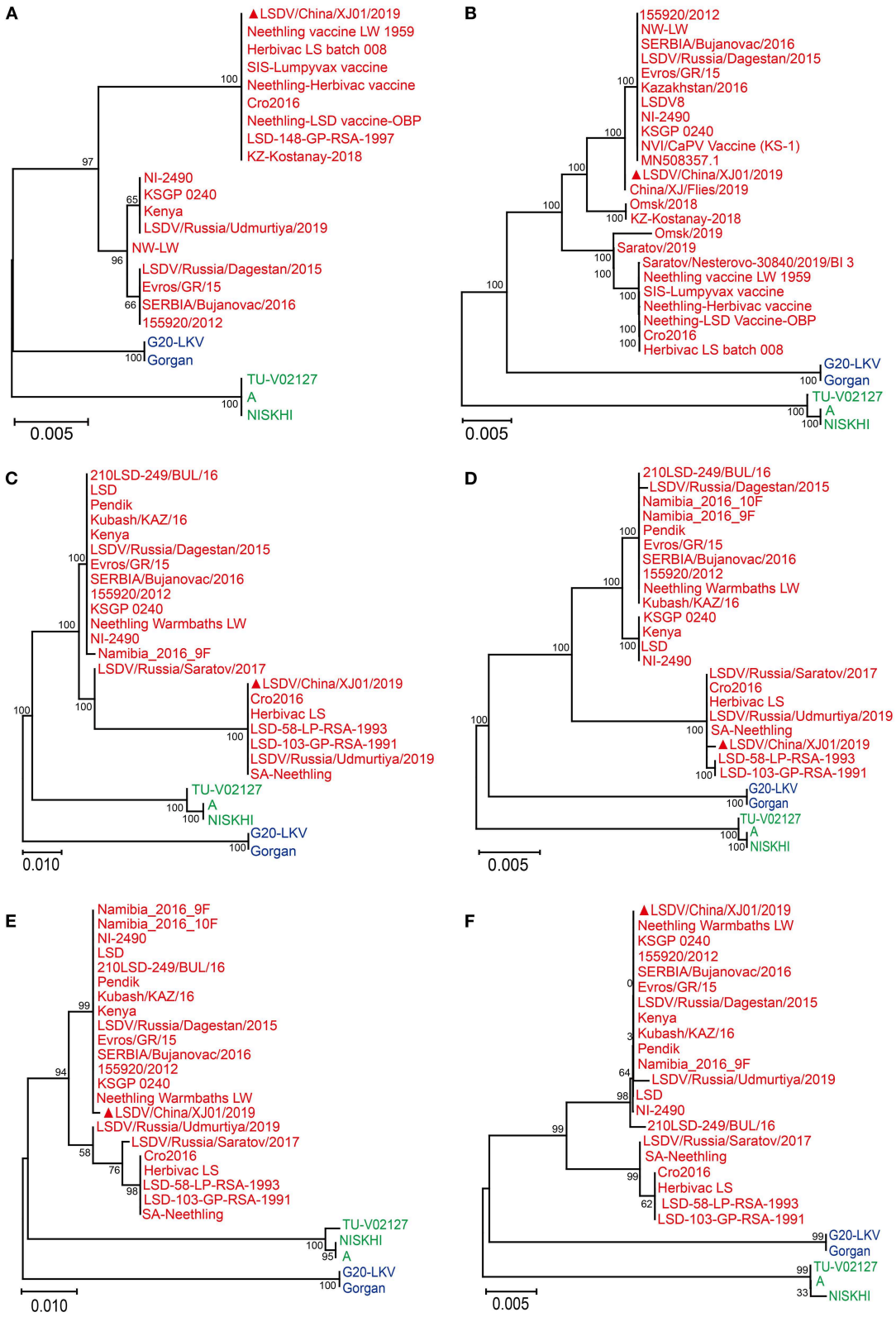


FIGURE 1
Phylogenetic diagram of RPO30 **(A)**, GPCR **(B)**, Putative alpha amanitin-sensitive protein **(C)**, Ankyrin repeat protein gene **(D)**, Interleukin-1 receptor-like protein gene **(E)** and Putative late transcription factor **(F)** genes of Lumpy skin disease virus. The phylogenetic tree was constructed using MEGA 11 with the Maximum Likelihood method. The Lumpy skin disease virus, Sheeppox virus, Goatpox virus are labeled with red, blue and green colors, respectively. The sequences obtained from this study are indicated by solid red triangles.

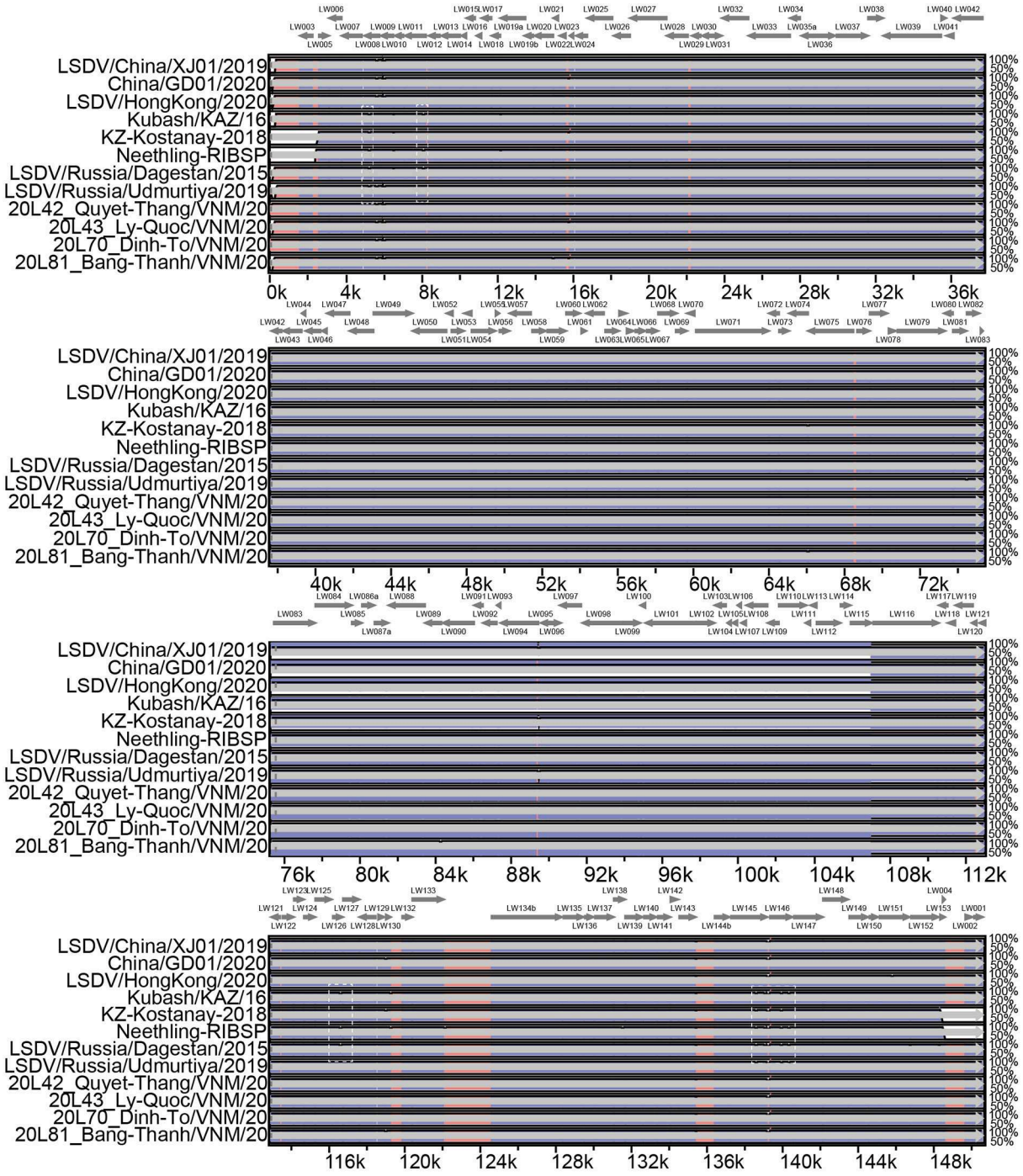


FIGURE 2
Multiple genomic sequence alignment of the LSDV strains in and around China. Graph showing sequence similarities between LSDV/China/XJ01/2019 and the strains 20L42_Quyet-Thang/VNM/20, 20L43_Ly-Quoc/VNM/20, 20L70_Dinh-To/VNM/20, 20L81_Bang-Thanh/VNM/20, China/GD01/2020, Kubash/KAZ/16, LSDV/China/XJ01/2019, LSDV/HongKong/2020, LSDV/Russia/Dagestan/2015, LSDV/Russia/Saratov/2017 and LSDV/ Russia/Udmurtiya/2019, which are plotted in a sliding 100-bp window. The seven differentiated regions between the genomes of the Chinese strains and reference strain LSDV/Russia/Saratov/2017 are indicated by white dotted boxes.

TABLE 1 ORFs that are not identical between LSDV/China/XJ01/2019 and other Chinese, Kazakhstan, Russian and Vietnamese Lumpy skin disease virus strains.

Gene	Product	China	Vietnam	Kazakhstan	Russia
LSDV001	Hypothetical protein	99.4	100	100	100
LSDV005	Interleukin-10-like protein	100	100	100	98.8–100
LSDV006	Interleukin-1 receptor-like protein	100	100	99.6–100	97.8–00
LSDV008	Putative soluble interferon gamma receptor	100	100	95.6–96.4	95.6–100
LSDV009	Putative alpha amanitin-sensitive protein	100	100	97.0	97.0–100
LSDV010	LAP|PHD-finger protein	100	100	98.8	98.8–100
LSDV011	G protein-coupled chemokine receptor-like protein	100	100	98.7–99.7	98.7–100
LSDV012	Ankyrin repeat protein	100	100	99.5–100	99.5–100
LSDV013	Interleukin-1 receptor-like protein	100	100	98.8–100	98.8–100
LSDV017	anti-apoptotic membrane protein	100	100	97.2–98.3	97.2–100
LSDV018	dUTPase	100	100	99.3	99.3–100
LSDV019	Kelch-like protein	100	100	99.3–100	99.3–100
LSDV020	Ribonucleotide reductase small subunit	100	100	99.7–100	99.7–10
LSDV021	Hypothetical protein	100	100	96.5–100	96.5–100
LSDV022	Hypothetical protein	100	100	97.4–100	97.4–100
LSDV024	S-S bond formation pathway protein	100	100	99.5–100	99.5–100
LSDV025	Ser-Thr kinase	99.8–100	100	100	99.8–100
LSDV026	Hypothetical protein	100	100	75.2–100	100
LSDV027	EEV maturation protein	100	100	99.5–100	100
LSDV028	Palmytilated EEV membrane glycoprotein	100	100	99.7–100	100
LSDV032	Poly(A) polymerase large subunit	100	100	99.6–100	99.6–100
LSDV033	Hypothetical protein	100	100	99.6–99.7	99.5–100
LSDV034	Double-strand RNA-binding protein	100	100	99.4–99.6	99.4–100
LSDV035	RNA polymerase subunit	100	100	99.4–100	100
LSDV036	Hypothetical protein	100	100	92.8–100	92.5–100
LSDV037	Hypothetical protein	100	100	92.8–99.8	99.8–100
LSDV038	Putative membrane protein	100	100	99.8–100	100
LSDV039	DNA polymerase	100	100	99.9–100	99.8–100
LSDV040	Sulfhydryl oxidase	100	100	99.9–100	100
LSDV041	Putative virion core protein	100	100	99.2–100	99.2–100
LSDV042	Hypothetical protein	100	100	99.2–99.6	99.4–100
LSDV043	Putative DNA-binding virion core protein	100	100	99.4–100	99.7–100
LSDV044	Hypothetical protein	100	100	99.7–100	100
LSDV045	Putative DNA-binding phosphoprotein	100	100	99.3–100	99.3–100
LSDV046	Putative IMV membrane protein	100	100	99.3–100	100
LSDV047	Hypothetical protein	100	100	99.5–100	99.7–100
LSDV048	Putative virion core protein	100	100	99.7–100	100
LSDV049	RNA helicase NPH-II	100	100	99.7–100	99.7–100
LSDV050	Putative metalloprotease	100	100	99.7–100	99.8–100

(Continued)

TABLE 1 (Continued)

Gene	Product	China	Vietnam	Kazakhstan	Russia
LSDV051	Hypothetical protein	100	100	99.1–100	99.1–100
LSDV052	Putative transcriptional elongation factor	100	100	99.1–100	100
LSDV054	Hypothetical protein	100	100	99.8–100	100
LSDV057	Putative virion core protein	100	100	99.7–100	99.7–100
LSDV058	Putative late transcription factor	100	100	99.7–100	100
LSDV059	Poxvirus myristoylprotein	99.7	99.7	99.7–100	99.7
LSDV060	Putative myristylated IMV envelope protein	100	100	99.7–100	100
LSDV061	Hypothetical protein	100	100	98.9–100	97.8–100
LSDV062	Hypothetical protein	100	100	98.9–100	100
LSDV064	Putative membrane protein	100	100	98.5–100	98.5–100
LSDV065	Hypothetical protein	100	100	98.5–100	98.6–100
LSDV066	Thymidine kinase	100	100	98.6–100	100
LSDV067	Putative host range protein	87.8–100	100	99.0–100	87.8–100
LSDV068	Poly(A) polymerase small subunit	100	100	99.0–100	99.7–100
LSDV069	RNA polymerase subunit	100	100	99.7–100	100
LSDV071	RNA polymerase subunit	100	100	99.9–100	99.9–100
LSDV072	Putative protein-tyrosine phosphatase	100	100	99.9–100	100
LSDV073	Putative viral membrane protein	100	99.5–100	100	100
LSDV074	Putative IMV envelope protein	100	100	99.7–100	99.7–100
LSDV075	RNA polymerase-associated protein	100	100	99.7	99.6–100
LSDV076	Late transcription factor VLTF-4	86.5–100	100	97.8–100	86.5–100
LSDV079	mRNA capping enzyme large subunit	99.9–100	100	99.9–100	99.9–100
LSDV080	Hypothetical protein	100	100	100	99.4–100
LSDV081	Putative virion protein	100	100	100	98.0–100
LSDV083	Putative NTPase	100	100	99.6–100	100
LSDV089	mRNA capping enzyme small subunit	100	100	99.7–100	99.7–100
LSDV090	Putative rifampicin resistance protein	99.8–100	100	99.7–100	99.8–100
LSDV094	Putative virion core protein	100	100	99.8–100	99.8–100
LSDV095	Virion core protein	100	100	99.8–100	100
LSDV096	RNA polymerase subunit	100	100	99.4–100	100
LSDV097	Hypothetical protein	99.7–100	100	99.4–100	99.7–100
LSDV098	Putative early transcription factor large subunit	100	100	99.7–100	100
LSDV100	Putative IMV membrane protein	100	100	99.7–100	100
LSDV102	Hypothetical protein	100	100	99.9–100	99.7–100
LSDV103	Putative virion core protein	98.4–100	100	99.5–100	98.4–99.5
LSDV104	Putative IMV membrane protein	100	100	99.5–10	100
LSDV107	Hypothetical protein	100	100	98.9–100	98.9–100
LSDV108	Putative myristylated membrane protein	100	100	98.9–100	100
LSDV109	Putative phosphorylated IMV membrane protein	100	100	99.5–100	99.5–100
LSDV110	Putative DNA helicase transcriptional elongation factor	100	100	99.4–99.6	99.6–100

(Continued)

TABLE 1 (Continued)

Gene	Product	China	Vietnam	Kazakhstan	Russia
LSDV111	Hypothetical protein	100	100	99.6–100	100
LSDV113	IMV membrane protein	100	100	98.3–100	100
LSDV113	Putative DNA polymerase processivity factor	100	100	99.5–100	99.8–100
LSDV114	Hypothetical protein	94.4–100	100	93.9–99.8	93.9–94.4
LSDV115	Putative intermediate transcription factor subunit	100	100	93.9–99.5	99.5–100
LSDV116	RNA polymerase subunit	100	100	99.5–100	100
LSDV122	EEV glycoprotein	100	100	99.0–100	99.0–100
LSDV123	IEV and EEV membrane glycoprotein	100	100	99.0–100	100
LSDV126	EEV glycoprotein	73.8–100	100	94.5–100	73.8–100
LSDV127	Hypothetical protein	100	100	94.5–100	99.6–100
LSDV128	CD47-like protein	99.7–100	100	98.0–99.7	99.7–100
LSDV129	Hypothetical protein	100	100	98.0–100	98.4–100
LSDV130	Hypothetical protein	100	100	96.3–100	96.3–100
LSDV131	Superoxide dismutase-like protein	93.6–100	100	96.3–100	93.6–100
LSDV132	Hypothetical protein	100	100	97.7–100	97.7–100
LSDV133	DNA ligase-like protein	100	100	97.7–99.8	99.8–100
LSDV134	Variola virus B22R-like protein	60.6–100	100	99.6–99.8	60.6–99.7
LSDV135	Putative IFN-alpha|beta binding protein	100	100	98.9	98.9–100
LSDV136	Hypothetical protein	100	100	98.9–99.3	99.3–100
LSDV137	Hypothetical protein	100	100	99.3–99.4	99.4–100
LSDV138	Ig domain OX-2-like protein	100	100	98.9–99.4	98.9–100
LSDV139	Putative ser-thr protein kinase	100	100	95.7–99.0	99.0–100
LSDV140	N1R-p28-like protein	100	100	98.3–99.0	98.3–100
LSDV141	EEV host range protein	100	100	98.3–99.6	99.1–100
LSDV142	Putative secreted virulence factor	100	100	98.5–99.1	98.5–100
LSDV143	Tyrosine protein kinase-like protein	97.1–97.1	97.1	96.5–98.5	96.5–97.1
LSDV144	Kelch-like protein	98.9–99.3	100	96.5–100	98.9–100
LSDV145	Ankyrin repeat protein	100	100	99.1–100	99.8–100
LSDV146	Phospholipase D-like protein	100	100	99.3–99.7	99.3–100
LSDV147	Ankyrin repeat protein	100	100	99.3–1,100	100
LSDV148	Ankyrin-like protein	100	100	100	99.6–100
LSDV149	Serpin-like protein	100	100	99.7–100	99.4–100
LSDV150	Hypothetical protein	100	100	99.7–100	100
LSDV151	Kelch-like protein	99.3–100	100	99.3–100	97.8–99.6
LSDV152	Ankyrin-like protein	100	100	99.2–99.6	98.4–100
LSDV153	Hypothetical protein	100	100	99.2–100	100
LSDV154	Putative ER-localized apoptosis regulator	99.6	100	100	100
LSDV155	Hypothetical protein	97.7	100	100	100

strain, China/GD01/2020. The changed ORFs represent seven proteins, hypothetical protein (LSDV001), putative myristylated protein (LSDV059), superoxide dismutase-like protein (LSDV131), tyrosine protein kinase-like protein (LSDV143), kelch-like protein (LSDV144), putative ER-localized apoptosis regulator (LSDV154), and hypothetical protein (LSDV155). In total, seven ORFs were missing in the LSDV/China/XJ01/2019 genome compared to the China/GD01/2020 genome. Of these seven missing ORFs, all (LSDV004, LSDV023, LSDV044, LSDV055, LSDV057, LSDV106 and LSDV107) had both start and stop codons in the China/GD01/2020 strain, which may be annotation issues.

The LSDV/China/XJ01/2019 strain shared 153 identical ORFs with the Vietnamese strains (20L42_Quyet-Thang/VNM/20, 20L43_Ly-Quoc/VNM/20, 20L70_Dinh-To/VNM/20 and 20L81_Bang-Thanh/VNM/20) and 3 ORFs had 97.1–99.7% sequence identities. The changed ORFs included three proteins, putative myristylated protein (LSDV059), putative viral membrane protein (LSDV073) and tyrosine protein kinase-like protein (LSDV143). The LSDV/China/XJ01/2019 strain showed no additional ORFs compared to the Vietnamese strains.

A comparison of LSDV/China/XJ01/2019 with the Russian strains (LSDV/Russia/Dagestan/2015, LSDV/Russia/Saratov/2017 and LSDV/Russia/Udmurtiya/2019), identified 75 identical ORFs and 81 ORFs sharing 60.6–99.9% sequence identities. Two ORFs (LSDV114 and LSDV131) in LSDV/Russia/Dagestan/2015 and one ORF (LSDV081) in LSDV/Russia/Udmurtiya/2019 were missing in our genomic sequence.

A comparison of LSDV/China/XJ01/2019 with the Kazakhstani strains (Kubash/KAZ/16, Neethling-RIBSP, and KZ-Kostanay-2018), revealed 53 identical ORFs and 103 ORFs sharing 75.2–99.9% sequence identities. In total, 1 ORF (LSDV023) in Kubash/KAZ/16, 10 ORFs (LSDV001, LSDV002, LSDV003, LSDV004, LSDV023, LSDV027, LSDV135, LSDV148, LSDV155, and LSDV156) in Neethling-RIBSP and 10 ORFs (LSDV001, LSDV002, LSDV003, LSDV004, LSDV114, LSDV131, LSDV153, LSDV154, LSDV155, and LSDV156) in KZ-Kostanay-2018 were missing in the genomic sequence due to lacking of the complete $5^{'}$- and $3^{'}$-terminal sequences of Neethling-RIBSP and KZ-Kostanay-2018.

3.5. Phylogenetic analyses

LSDV/China/XJ01/2019 together with China/GD01/2020, LSDV/HongKong/2020, 20L42_Quyet-Thang/VNM/20, 20L43_Ly-Quoc/VNM/20, 20L70_Dinh-To/VNM/20, and 20L81_Bang-Thanh/ VNM/20 belonged to the same clade, showing their elevated similarity and comprised a monophyletic group with short tree branches. The phylogenetic tree confirmed that Chinese and Vietnamese strains belong to the same evolutionary lineage compared to other LSDV strains (Figure 3).

4. Discussion

LSD is a severe systemic disease that was first reported in China 3 years ago, and it has severely impacted the cattle-breeding industry and brought significant economic losses to the affected areas (11). Before and after it, LSD outbreaks had occurred in several Eurasian countries, including Vietnam, Kazakhstan, Russia, India, and South Korea, and affected cattle, Korean water deer, and giraffe (4, 29–36), indicating a wave of highly contagious epidemic. Here we present the genomic characterization of the virulent strain of LSDV, together with its homologous viruses that were subsequently discovered in China and Vietnam. This study provides additional genomic data for LSDV evolution and is crucial for virus tracking and vaccine development.

Genomic comparisons of the LSDV/China/XJ01/2019 strain with those in and around China showed that it showed the highest level of genomic similarity with the Chinese and Vietnamese strains. Phylogenetic analysis and genomic similarity comparisons using whole genome sequences indicated that the strain LSDV/China/XJ01/2019 was most closely related to the strains China/GD01/2020, LSDV/HongKong/2020, 20L42_Quyet-Thang/VNM/20, 20L43_Ly-Quoc/VNM/20, 20L70_Dinh-To/VNM/20 and 20L81_Bang-Thanh/VNM/20, inferred that these LSDV strains might originated from a common ancestor. Genome sequencing of more LSDV strains circulating in East and Southeast Asia may help pinpoint their origins.

A total of 17 variable loci in the genomic sequence caused variations in 7 ORFs and 2 non-coding regions between LSDV/China/XJ01/2019 and the most closely related strain, China/GD01/2020. The changed ORFs encoded seven proteins, which are hypothetical protein (LSDV001), putative myristylated protein (LSDV059), superoxide dismutase-like protein (LSDV131), tyrosine protein kinase-like protein (LSDV143), kelch-like protein (LSDV144), putative ER-localized apoptosis regulator (LSDV154) and hypothetical protein (LSDV155). According to the whole-genome comparison of LSDV, the putative myristylated protein (LSDV059) and tyrosine protein kinase-like protein (LSDV143) were highly variable, indicating that the diversity of these two proteins was probably resulted from adaptive evolutionary pressure. In future, it would be instructive to reveal the association between gene mutations and viral pathogenicity or transmissibility, which would be helpful in understanding LSDV adaptive evolution.

Data on the movements of cattle into the epidemic zone during the first LSD outbreak in China and on

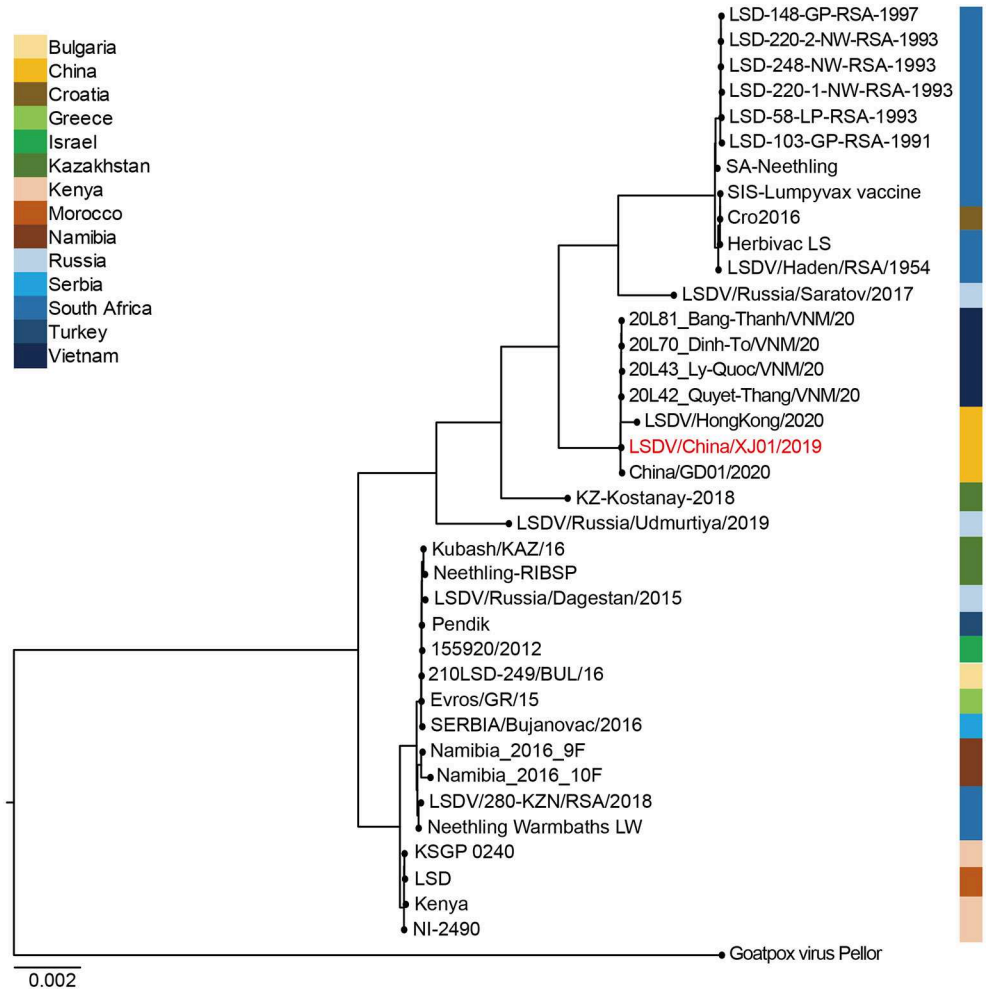


FIGURE 3
Bayesian maximum clade credibility tree of whole-genome sequences of Lumpy skin disease virus. The whole-genome MCC tree was constructed using the BEAST software package (v. 2.4.3) and then visualized by FigTree (v1.4.3). The LSDV genome sequenced in this study are labeled in red.

LSDV-infected cattle in region bordering China and Kazakhstan are still lacking, indicating the emergence of LSDV in Xinjiang, northwest China may be unexpectedly complex and difficult to trace, as it is still unknown whether the outbreak was imported or localized, highlighting the need for further research. Moreover, the phylogeny of the LSDV strains in Xinjiang, northwest China, and Vietnam, which are thousands of kilometers apart, highlights how little we know about the spread of this lineage and its introduction to these regions.

In summary, we analyzed the LSDV genome from a dairy cow that died during the first LSD outbreak in China and found that the strain LSDV/China/XJ01/2019 strain was genetically close to additional LSDV strains in China and Vietnam. Molecular epidemiological investigations of LSDV in susceptible animals and vectors at national, regional, and global levels are desired to understand the evolution and transmission routes of the latest global LSD epidemic.

Data availability statement

The data presented in the study is deposited in the Genbank, accession number OM105589.

Ethics statement

Ethical review and approval was not required for the animal study because specimens were collected from a dead cow,

approved by the Animal Husbandry and Veterinary Bureau of Xinjiang Uygur Autonomous Region. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

JH conceptualized the study. Y-RW, W-GM, PW, WW, X-HS, X-YY, X-YM, and J-YW carried out the laboratory assays. Y-RW analyzed the data and drafted the paper. All authors approved the final version of the manuscript to be published.

Funding

This work was supported by the National Science and Technology Major Project (Grant No. 2021YFD1800505) and the Science and Technology Major Project of the Xinjiang Uygur Autonomous Region of China (2020A01001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.1073648/full#supplementary-material

References

1. Das M, Chowdhury M, Akter S, Mondal A, Uddin M, Rahman M, et al. An updated review on lumpy skin disease: a perspective of Southeast Asian countries. *J Adv Biotechnol Exp Therap.* (2021) 4:322. doi: 10.5455/jabet.2021.d133

2. Molini U, Boshoff E, Niel AP, Phillips J, Khaiseb S, Settypalli TBK, et al. Detection of lumpy skin disease virus in an asymptomatic eland (Taurotragus oryx) in Namibia. *J Wildl Dis.* (2021) 57:708–11. doi: 10.7589/JWD-D-20-00181

3. Ko YS, Oh Y, Lee TG, Bae DY, Tark D, Cho HS. Serological and molecular prevalence of lumpy skin disease virus in Korean water deer, native and dairy cattle in Korea. *Korean J Vet Serv.* (2022) 45:133–7. doi: 10.7853/kjvs.2022.45.2.133

4. Tung Duy D, Long Hoang T, Hiep Dinh N, Thuy Thi H, Giang Hoang N, Kien Viet Dung T, et al. Characterization of Lumpy skin disease virus isolated from a C in Vietnam. *Transbound Emerg Dis.* (2022) 69:E3268–72. doi: 10.1111/tbed.14583

5. Vandenbussche F, Mathijs E, Haegeman A, Al-Majali A, Van Borm S, De Clercq K. Complete genome sequence of capripoxvirus strain KSGP 0240 from a commercial live attenuated vaccine. *Genome Announc.* (2016) 4:e01114–01116. doi: 10.1128/genomeA.01114-16

6. Sprygin A, Artyuchova E, Babin Y, Prutnikov P, Kostrova E, Byadovskaya O, et al. Epidemiological characterization of lumpy skin disease outbreaks in Russia in 2016. *Transbound Emerg Dis.* (2018) 65:1514–21. doi: 10.1111/tbed.12889

7. Tuppurainen ESM., Venter EH, Coetzer JW, Bell-Sakyi L. Lumpy skin disease: Attempted propagation in tick cell lines and presence of viral DNA in field ticks collected from naturally-infected cattle. *Ticks Tick-Borne Dis.* (2015) 6:134–40. doi: 10.1016/j.ttbdis.2014.11.002

8. Sanz-Bernardo B, Haga IR, Wijesiriwardana N, Basu S, Larner W, Diaz AV, et al. Quantifying and modeling the acquisition and retention of lumpy skin disease virus by hematophagus insects reveals clinically but not subclinically affected cattle are promoters of viral transmission and key targets for control of disease outbreaks. *J Virol.* (2021) 95:e02239–e02220. doi: 10.1128/JVI.02239-20

9. Sprygin A, Pestova Y, Wallace DB, Tuppurainen E, Kononov AV. Transmission of lumpy skin disease virus: a short review. *Virus Res.* (2019) 269:197637. doi: 10.1016/j.virusres.2019.05.015

10. Gupta T, Patial V, Bali D, Angaria S, Sharma M, Chahota R. A review: lumpy skin disease and its emergence in India. *Vet Res Commun.* (2020) 44:111–8. doi: 10.1007/s11259-020-09780-1

11. Liu P, Li J, Chen R, Cheng Z, Shi Y, Baosuyalatu, et al. The first outbreak investigation of lumpy skin disease in China. *China Anim Health Inspection.* (2020) 37:1–5. doi: 10.3969/j.issn.1005-944X.2020.01.001 (in Chinese)

12. Wang Y, Zhao L, Yang J, Shi M, Nie F, Liu S., et al. Analysis of vaccine-like lumpy skin disease virus from flies near the western border of China. *Transbound Emerg Dis.* (2022) 69:1813–23. doi: 10.1111/tbed.14159

13. Morgenstern M, Klement E. The effect of vaccination with live attenuated neethling lumpy skin disease vaccine on milk production and mortality-an analysis of 77 dairy farms in Israel. *Vaccines.* (2020) 8:324. doi: 10.3390/vaccines8020324

14. Salnikov N, Usadov T, Kolcov A, Zhivoderov S, Morgunov Y, Gerasimov V, et al. Identification and characterization of lumpy skin disease virus isolated from cattle in the Republic of North Ossetia-Alania in 2015. *Transbound Emerg Dis.* (2018) 65:916–20. doi: 10.1111/tbed.12818

15. Mathijs E, Vandenbussche F, Saduakassova M, Kabduldanov T, Haegeman A, Aerts L, et al. Complete coding sequence of a lumpy skin disease virus strain isolated during the 2016 outbreak in Kazakhstan. *Microbiol Resour Announc.* (2020) 9:e01399–19. doi: 10.1128/MRA.01399-19

16. Lu G, Xie J, Luo J, Shao R, Jia K, Li S. Lumpy skin disease outbreaks in China, since August 3, 2019. *Transbound Emerg Dis.* (2021) 68:216–9. doi: 10.1111/tbed.13898

17. Flannery J, Shih B, Haga IR, Ashby M, Corla A, King S, et al. A novel strain of lumpy skin disease virus causes clinical disease in cattle in Hong Kong. *Transbound Emerg Dis.* (2022) 69:e336–e343. doi: 10.1111/tbed.14304

18. Wu J, Chen Y, Che C, Li B. Epidemiological investigation and analysis of an epidemic of Lumpy skin disease in Sichuan, China. *China Anim Health.* (2021) 21:20–1. doi: 10.3969/j.issn.1008-4754.2021.01.014 (in Chinese)

19. Yang Z, Hou W, Mo Q, Yin J, Yuan D, Zhou R, et al. International and domestic prevalence of Lumpy skin disease. *Shandong J Anim Sci Vet Med.* (2021) 42:66–71. doi: 10.3969/j.issn.1007-1733.2021.10.020 (in Chinese)

20. Tulman ER, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. Genome of lumpy skin disease virus. *J Virol.* (2001) 75:7122–30. doi: 10.1128/JVI.75.15.7122-7130.2001

21. El-Tholoth M, El-Kenawy AA. G-Protein-Coupled Chemokine Receptor Gene in Lumpy Skin Disease Virus Isolates from Cattle and Water Buffalo

(Bubalus bubalis) in Egypt. *Transbound Emerg Dis.* (2016) 63:e288–95. doi: 10.1111/tbed.12344

22. Chibssa TR, Sombo M, Lichoti JK, Adam TIB, Liu Y, Elraouf YA, et al. Molecular analysis of east african lumpy skin disease viruses reveals a mixed isolate with features of both vaccine and field isolates. *Microorganisms.* (2021) 9:1142. doi: 10.3390/microorganisms9061142

23. Ma J, Yuan Y, Shao J, Sun M, He W, Chen J, et al. Genomic characterization of lumpy skin disease virus isolated in southeast China. *Transbound Emerg Dis.* (2022) 69:2788–99. doi: 10.1111/tbed.14432

24. Koichiro T, Glen S, Sudhir K. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* (2021) 38:3022–7. doi: 10.1093/molbev/msab120

25. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* (2012) 28:1647–9. doi: 10.1093/bioinformatics/bts199

26. Tcherepanov V, Ehlers A, Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics.* (2006) 7:150. doi: 10.1186/1471-2164-7-150

27. Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* (2012) 28:1166–7. doi: 10.1093/bioinformatics/bts091

28. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* (2012) 61:539–42. doi: 10.1093/sysbio/sys029

29. Acharya KP, Subedi D. First outbreak of lumpy skin disease in Nepal. *Transbound Emerg Dis.* (2020) 67:2280–1. doi: 10.1111/tbed.13815

30. Kumar N, Chander Y, Kumar R, Khandelwal N, Riyesh T, Chaudhary K, et al. Isolation and characterization of lumpy skin disease virus from cattle in India. *PLoS ONE.* (2021) 16:e0241022. doi: 10.1371/journal.pone.0241022

31. Orynbayev MB, Nissanova RK, Khairullin BM, Issimov A, Zakarya KD, Sultankulova KT, et al. Lumpy skin disease in Kazakhstan. *Trop Anim Health Prod.* (2021) 53:166. doi: 10.1007/s11250-021-02613-6

32. Tran HTT, Truong AD, Dang AK, Ly DV, Nguyen CT, Chu NT, et al. Lumpy skin disease outbreaks in vietnam, 2020. *Transbound Emerg Dis.* (2021) 68:977–80. doi: 10.1111/tbed.14022

33. Byadovskaya O, Prutnikov P, Shalina K, Babiuk S, Perevozchikova N, Korennoy F, et al. The changing epidemiology of lumpy skin disease in Russia since the first introduction from 2015 to 2020. *Transbound Emerg Dis.* (2022) 69:E2551–62. doi: 10.1111/tbed.14599

34. Chouhan CS, Parvin S, Ali MY, Sadekuzzaman M, Chowdhury MGA, Ehsan MA, et al. *Epidemiology and economic impact of lumpy skin disease of cattle in Mymensingh and Gaibandha districts of Bangladesh. Transbound Emerg Dis.* (2022). doi: 10.1111/tbed.14697

35. Odonchimeg M, Erdenechimeg D, Tuvshinbayar A, Tsogtgerel M, Bazarragchaa E, Ulaankhuu A, et al. Molecular identification and risk factor analysis of the first Lumpy skin disease outbreak in cattle in Mongolia. *J Vet Med Sci.* (2022) 84:1244–52. doi: 10.1292/jvms.22-0250

36. Sariya L, Paungpin W, Chaiwattanarungruengpaisan S, Thongdee M, Nakthong C, Jitwongwai A, et al. Molecular detection and characterization of lumpy skin disease viruses from outbreaks in Thailand in 2021. *Transbound Emerg Dis.* (2022) 69:E2145–52. doi: 10.1111/tbed.14552

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 19 January 2023
DOI 10.3389/fvets.2023.1026296

OPEN ACCESS

EDITED BY
Kimberly VanderWaal,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Chantal Snoeck,
Luxembourg Institute of Health, Luxembourg
Ursula Höfle,
University of Castilla-La Mancha, Spain

*CORRESPONDENCE
Denys Muzyka
✉ dmuzyka77@gmail.com

SPECIALTY SECTION
This article was submitted to
Veterinary Epidemiology and Economics,
a section of the journal
Frontiers in Veterinary Science

RECEIVED 23 August 2022
ACCEPTED 02 January 2023
PUBLISHED 19 January 2023

CITATION
Goraichuk IV, Gerilovych A, Bolotin V, Solodiankin O, Dimitrov KM, Rula O, Muzyka N, Mezinov O, Stegniy B, Kolesnyk O, Pantin-Jackwood MJ, Miller PJ, Afonso CL and Muzyka D (2023) Genetic diversity of Newcastle disease viruses circulating in wild and synanthropic birds in Ukraine between 2006 and 2015. *Front. Vet. Sci.* 10:1026296. doi: 10.3389/fvets.2023.1026296

Genetic diversity of Newcastle disease viruses circulating in wild and synanthropic birds in Ukraine between 2006 and 2015

Iryna V. Goraichuk[1,2], Anton Gerilovych[1], Vitaliy Bolotin[1], Olexii Solodiankin[1], Kiril M. Dimitrov[2], Oleksandr Rula[1], Nataliia Muzyka[1], Oleksandr Mezinov[3], Borys Stegniy[1], Olena Kolesnyk[1], Mary J. Pantin-Jackwood[2], Patti J. Miller[2], Claudio L. Afonso[2] and Denys Muzyka[1,4]*

[1]National Scientific Centre, Institute of Experimental and Clinical Veterinary Medicine, Kharkiv, Ukraine, [2]Exotic and Emerging Avian Viral Diseases Research Unit, Southeast Poultry Research Laboratory, US National Poultry Research Center, Agricultural Research Service, USDA, Athens, GA, United States, [3]The F.E. Falz-Fein Biosphere Reserve "Askania Nova", National Academy of Agrarian Sciences of Ukraine, Askania-Nova, Kherson Oblast, Ukraine, [4]Department of Zoology, H.S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine

Newcastle disease virus (NDV) infects a wide range of bird species worldwide and is of importance to the poultry industry. Although certain virus genotypes are clearly associated with wild bird species, the role of those species in the movement of viruses and the migratory routes they follow is still unclear. In this study, we performed a phylogenetic analysis of nineteen NDV sequences that were identified among 21,924 samples collected from wild and synanthropic birds from different regions of Ukraine from 2006 to 2015 and compared them with isolates from other continents. In synanthropic birds, NDV strains of genotype II, VI, VII, and XXI of class II were detected. The fusion gene sequences of these strains were similar to strains detected in birds from different geographical regions of Europe and Asia. However, it is noteworthy to mention the isolation of vaccine viruses from synanthropic birds, suggesting the possibility of their role in viral transmission from vaccinated poultry to wild birds, which may lead to the further spreading of vaccine viruses into other regions during wild bird migration. Moreover, here we present the first publicly available complete NDV F gene from a crow (genus *Corvus*). Additionally, our phylogenetic results indicated a possible connection of Ukrainian NDV isolates with genotype XXI strains circulating in Kazakhstan. Among strains from wild birds, NDVs of genotype 1 of class I and genotype I of class II were detected. The phylogenetic analysis highlighted the possible exchange of these NDV strains between wild waterfowl from the Azov-Black Sea region of Ukraine and waterfowl from different continents, including Europe, Asia, and Africa.

KEYWORDS

NDV, *Avian orthoavulavirus 1*, surveillance, sequencing, bird migration, synanthropic, pigeon, Ukraine

1. Introduction

Newcastle disease virus (NDV), or *Avian orthoavulavirus 1*, is a member of the recently separated subfamily *Avulavirinae* in the *Paramyxoviridae* family (1). NDV is capable of infecting a wide range of bird species and is of great economic importance to both large poultry enterprises and private farms. Since the discovery of the first NDV in the 1920s, an array of data has accumulated illustrating the genetic and pathogenic diversity of the viral strains (2). According to the Terrestrial Manual (World Organization for Animal Health, formerly the Office International des Epizooties), (WOAH, founded as OIE) Terrestrial Manual, NDV strains are divided into three main pathogenicity groups: velo-, meso-, and lentogenic (3, 4). Velogenic strains cause hemorrhages, reduced egg production, intestinal lesions, and neurological symptoms with high mortality in infected chickens. Mesogenic strains cause disease with respiratory or neurological symptoms but little mortality. Lentogenic strains can cause subclinical respiratory or intestinal infections and are considered low-virulent. *Avian orthoavulavirus 1* strains are divided into two classes based on the nucleotide sequence of the fusion (F) protein gene. Class I consists mainly of lentogenic strains isolated from wild birds worldwide and includes only one genotype (5). Class II contains 21 genotypes with velo-, meso- and lentogenic strains that have been detected in a wide variety of host species around the world (6).

The fusion protein is an important determinant of NDV pathogenicity (7, 8). It is synthesized as an inactive precursor (F0), which is proteolytically cleaved by host proteases into two polypeptides (F1 and F2) for the virus particles to be infectious. The efficiency of proteolytic cleavage is dependent on the host cell and the virus strain (9–11). F proteins of velogenic and mesogenic NDV virus strains are characterized by the presence of multiple basic amino acids at the F0 cleavage sequence which are recognized by ubiquitous host cell proteases (12–14). In contrast, the F0 protein of lentogenic strains has a monobasic cleavage site, which is cleavable by a restricted number of certain host proteases. These differences in the F gene sequences which correlate with different virulence phenotypes are prime targets for the development of molecular biological approaches to identify and characterize NDV isolates. However, the viral population may acquire mutations and adaptive changes in response to different pressures by the host's immune system (15, 16). Together with large genetic diversity it can affect real-time PCR specificity for NDV detection, particularly virulence determination (17, 18) and vaccine efficacy (19, 20) since both are typically based on the genetic sequence of the F protein, most often affected by vaccine pressure.

Although a wealth of information about this virus has accumulated to date, the role of wild birds in the global circulation and epidemiology of certain NDV genotypes remains unclear. The expansion of the surveillance program for *Avian orthoavulavirus 1* will improve our understanding of the global epidemiological picture, as well as help effectively prevent Newcastle Disease (ND). A large-scale study of the potential host range of NDV among wild birds in Africa showed the year-round presence of viruses, as well as their phylogenetic relationship with strains of domestic birds in the study area (21). Currently, in the territory of Eurasia, few studies of the potential host range of *Avian orthoavulavirus 1* have been carried out, and the seasonality of infection has not been established. Even though there are many reports of NDV isolated from wild birds in this region, more than 30 countries remain unstudied or with limited data (22–29). Therefore, studies performed at the major stopping point locations for migratory birds can be of great contribution toward gaining more knowledge on the potential host range of NDV therein.

Ukraine occupies a unique geographical location in central and Eastern Europe, where the West Asia-East Africa flyways of wild migratory birds cross the Black Sea-Mediterranean, and East Atlantic flyways (30, 31). The natural conditions such as climate and the abundance of wetlands with an area of more than 590,000 ha in Ukraine contribute to the year-round presence of a large number of wild birds. This is especially evident during the period of seasonal migrations and wintering, when numerous wild bird species pass from North Asia and Europe to the Mediterranean, Africa, and Southwest Asia, and also cross from the Baltic and Caspian Seas to the Black and Mediterranean Seas, and from western Siberia and Kazakhstan to Western Europe and North Africa (32). The Azov-Black Sea region is one of the densest territories of Eastern Europe from an ornithological point of view. This region is historically an area of nesting, flight, migratory stops, and wintering for many bird species. Therefore, a large number of waterfowl and waterbirds from the Central part of Eurasia winter in the Azov-Black Sea region of Ukraine or stop there during migrations. As a result, wild bird populations from Ukraine and West Europe can interact with birds from Asia and Africa (33).

Domestic poultry also plays a role in the circulation of *Avian orthoavulavirus 1*. According to data from the United Nations Food and Agriculture Organization (FAO), the total poultry presence in the world (chickens, ducks, turkeys, geese, and guinea fowl) was ~35 billion birds in 2020 (34). Chickens accounted for 94% of the world's poultry population with about 46% of these located in the territory of Asia. To help keep ND under control in developed countries, industrial poultry farms are recommended to vaccinate their birds (mainly chickens and turkeys) and isolate them from the external environment to avoid contact with wild and synanthropic bird species (35, 36). But cases of possible transmission of the NDV vaccine strain from poultry to wild birds have been described (23, 37, 38). There are also reports of velogenic strains in wild birds (39). Currently, it is not well studied if lentogenic strains of NDV from wild birds may cause respiratory infection in poorly vaccinated poultry birds following exposure. However, a few cases of the change in the virus virulence were well-documented. Retrospective studies showed that cumulative mutations at the fusion protein cleavage site acquired through natural transmission at the poultry farms lead to change in the virulence of originally lentogenic virus during NDV outbreak of genotype I in Australia (40–42). Another study demonstrated an initially non-pathogenic strain from a wild duck acquiring pathogenicity through passaging in chickens (43).

Currently, the issue of the introduction and circulation of lentogenic NDV vaccine strains into the wild bird population is under consideration and discussion, thus it is important to assess the presence of vaccine spillovers to wild and synanthropic birds as well (23). Therefore, this article presents surveillance data of *Avian orthoavulavirus 1* circulation in a natural reservoir from eight different regions in the South and East of Ukraine, including migratory bird's stopping points located in the Azov-Black Sea region. Part of these results was previously reported (5, 23, 27, 29, 44).

2. Materials and methods

2.1. Sample collection

Samples from wild and synanthropic birds were collected in Ukraine between 2006 and 2015 during active and passive surveillance (29, 45). Wild birds were defined as those occurring in a natural habitat other than poultry and synanthropic birds were defined as undomesticated birds that live in close association with people. A total of 21,924 samples were collected, among them, 21,854 samples were collected from 103 species of wild birds during active surveillance and 70 samples from synanthropic birds during passive surveillance. The sampling sites were located in the Azov-Black Sea region [Kherson, Mykolaiv, Odesa, Zaporizhzhya, and the Autonomous Republic (AR) of Crimea regions] and the East of Ukraine (Donetsk, Dnipro, and Kharkiv regions) (Figure 1). The samples' background data, including bird species and location, were recorded (Supplementary Tables S1, S2).

Sampling from wild birds was carried out in cooperation with ornithologists, who helped identify the bird species. Cloacal swabs were collected from apparently healthy live-trapped and hunted wild birds. Fresh feces were collected from places of mass bird accumulation. Feces were collected only if the origin and type of bird had been established. Immediately after sampling, samples of the biological material from wild birds were placed in tubes with a transport medium (Hank's balanced salt solution containing 0.5% lactalbumin, 10% glycerol, 200 U penicillin, 0.200 mg streptomycin, 100 U polymyxin, 0.250 mg gentamicin, and 50 U nystatin per ml) then stored and transported on ice to the laboratory and further stored in liquid nitrogen.

Samples of spleen, brain, liver, and intestines were collected from dead synanthropic birds (pigeons and gray crows) in Ukraine between 2006 and 2015. All dead synanthropic birds were found in locations of their natural habitat within cities in Kharkiv, Dnipro, Donetsk, Odesa, and AR Crimea regions. Samples were chilled at 4°C, transported to the lab, and stored at −80°C. Before analysis, samples were thawed and suspended in transport media (10% w/v).

2.2. Virus isolation and identification

Each fecal/cloacal swab medium or tissue suspension supernatant was inoculated (0.2 ml) into five 9- to 11-day-old specific-pathogen-free (SPF) embryonated chicken eggs (ECEs) using standard methods as described previously (46, 47). Allantoic fluids from all inoculated ECEs were harvested and tested for hemagglutination activity with chicken red blood cells using a hemagglutination assay (HA) (48). All HA-positive samples were analyzed in the hemagglutination inhibition (HI) assay using reference antisera to influenza A virus subtypes H1–H16 and avian paramyxoviruses (APMV-1–APMV-9) according to previous recommendations (47, 49, 50). The identification of avian influenza viruses and APMV-4, APMV-6, APMV-7, and APMV-13 have been previously described (29, 32, 51, 52).

Pathogenicity evaluation was performed on eleven selected NDV isolates (KF851268–KF851270, KJ914671, KJ914672, KU133362–KU133365, KY042127, KY042128, MZ101338) using the intracerebral pathogenicity index (ICPI) assay on 1-day-old SPF chickens following established procedures at the Southeast Poultry Research Laboratory (SEPRL), U.S. Department of Agriculture (USDA), Athens, GA, USA (3).

2.3. RNA extraction and sequencing

All HI-identified NDV isolates were subject to sequencing to determine their genotype and virulence. Viral RNA was extracted from infected allantoic fluids using TRIzol LS (Invitrogen, USA) following the manufacturer's instructions. The nucleotide sequences of the complete coding region of the F protein gene were determined by utilizing the RT-PCR/sequencing approach (53). The amplification reaction was performed using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Life Technologies, USA) with overlapped primer pairs for sequencing the complete F gene of different NDV genotypes from class I and II as described previously by Miller et al. (24). All RT-PCR products were subjected to electrophoresis in a 1% agarose gel (0.5X TBE). The appropriately sized DNA bands were excised from the gel and purified using the QuickClean II Gel Extraction Kit (GenScript, USA) and subjected to DNA sequencing. Nucleotide sequencing was performed for 19 samples on an ABI Sanger sequencer (Applied Biosystems, USA) with fluorescent dideoxy-nucleotide terminators at SEPRL, USDA, Athens, GA, USA. Sequence editing and assembly were performed using the SeqMan software of the LaserGene package (DNASTAR, USA).

2.4. Phylogenetic analysis

Genotype and sub-genotype identification were based on the phylogenetic topology and evolutionary distances between different taxonomic groups using a pilot dataset as described by Dimitrov et al. (5). Multiple sequence alignments of the NDV complete F gene sequences were produced using the MAFFT version 7 software (54). The pilot tree of class I and II NDV isolates ($n = 98$) was constructed using the Maximum-likelihood method based on the General Time-Reversible (GTR) model with a discrete gamma distribution (+G) and allowing for invariant sites (+I) with statistical analysis based on 1,000 bootstrap replicates, as implemented in MEGA7 (55). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. For all analyses, the codon positions included were 1st, 2nd, 3rd, and non-coding, and all positions containing gaps and missing data were eliminated. A total of 1,661 positions were included in the pilot analysis of the complete F gene dataset.

For each genotype detected (1 of class I and I, II, VI, VII, and XXI of class II), more detailed phylogenetic trees were constructed using the most closely related sequences detected in BLAST. The Roman numerals presented in the name of each sequence in the phylogenetic tree represent the respective sub-genotype, followed by the GenBank accession number, host name, country of isolation, strain designation, and year of isolation (if available).

The complete F gene data set used for the phylogenetic analysis was also used to estimate the average evolutionary distances comparing Ukrainian NDV isolates to other relative strains. Pair-wise analysis was conducted using the maximum composite likelihood model using MEGA7 software (56). The rate variation among sites was modeled with a gamma distribution (shape parameter = 4).

10.3389/fvets.2023.1026296

FIGURE 1
Sample collection number, bird type, and location in Ukraine. Sampled regions are indicated in green color. The number of collected samples are indicated in parenthesis under the name of each region. Genotypes detected in this study are shown in red (genotype 1, class I), pink (genotype I, class II), teal (genotype II, class II), blue (genotype VI, class II), purple (genotype VII, class II), and brown (genotype XXI, class II).

3. Results

3.1. Geographic distribution of the viruses sequenced

The serological examination of biological material collected from 21,854 wild and 70 synanthropic birds belonging to 105 species and 11 different orders was conducted during 2006–2015. The largest number of samples was collected from birds of the order *Anseriformes* (15,013 samples), followed by *Charadriiformes* (4,737 samples), and *Passeriformes* (1,562 samples). The main sampling sites (~99% of the samples) for wild birds were located in the Azov Black Sea region. This region is the meeting point of the transcontinental migration routes of different wild birds from Siberia, Africa, Europe, and Asia. The rest of the biological samples were collected in the eastern region of Ukraine (Figure 1). NDV was identified in 31 (0.14%) samples from asymptomatic wild birds and 24 (34.29%) dead synanthropic birds by the HI test, part of which was previously reported (22, 29). All three overlapping regions of the complete F gene were successfully amplified by RT-PCR for 18 isolates, previously classified as NDV based on the serological HI test (Table 1). Two complete F genes of different NDV genotypes were amplified and sequenced for one isolate, which brought the total number of obtained sequences to 19. Some of the viruses were sequenced and previously published (5, 23, 27, 29, 44). Sequenced NDV isolates were obtained from dead synanthropic birds and asymptomatic wild waterfowl in different regions of Ukraine (Table 1). Twelve NDV isolates were attained from synanthropic birds (pigeons and crows). Isolates from synanthropic birds were obtained from four different regions: three originated from the Kharkiv region and one each from the Dnipro region, the Donetsk region, and the Autonomous Republic of Crimea (Figure 1). The six isolates from wild birds all came from the Kherson region in the south of Ukraine.

3.2. Pathotype characterization

NDV pathogenicity markers were examined for all 19 NDV sequences. The deduced amino acid sequences of the fusion protein cleavage site revealed that sequences of seven Ukrainian isolates from this study presented four basic amino acids at the C-terminus of the F2 protein from residues 112–116 and phenylalanine at residue 117 (Table 1, highlighted in bold), and based on the WOAH definition of vNDV molecular pathotyping (3) they are considered as virulent strains. The cleavage site motifs of these Ukrainian isolates were ^{112}KRQKR↓^{117} (n = 5) and ^{112}RRQKR↓F^{117} (n = 2). All seven viruses were isolated from synanthropic pigeons in the Dnipro, Donetsk, Kharkiv, and AR Crimea regions of Ukraine. The rest of the cleavage site motifs were represented by sequences ^{112}GRQGR↓L^{117} (n = 6), ^{112}GKQGR↓L^{117} (n = 5), and ^{112}ERQGR↓L^{117} (n = 1) that are typical for viruses of low virulence.

TABLE 1 Background information data for NDV isolates recovered in Ukraine between 2006 and 2015.

Species	Class/sub-genotype	Region	Isolate	Year of isolation	ICPI[a]	Cleavage site motif	GenBank accession number
Pigeon	II/II	Kharkiv	Kharkiv/1	2007	1.09	GRQGR↓L	KU133364
Pigeon	II/II	Kharkiv	Kharkiv/2	2007	0.29	GRQGR↓L	KU133365
Pigeon	II/II	Dnipro	Dnipropetrovsk/07	2007	1	GRQGR↓L	KU133363
Pigeon	II/XXI.1.1	Donetsk	Doneck/3/968	2007	1.48	**KRQKR↓F**	KY042128
	II/II					GRQGR↓L	KU133362
Crow	II/II	Kharkiv	Izum/8-15[c]	2007	N/A	GRQGR↓L	MZ101343
Pigeon	II/II	Donetsk	Doneck/10/26-6[c]	2008	N/A	GRQGR↓L	MZ101344
Teal	II/I.2	Donetsk	Krasnooskilsky/5-11	2009	0.08	GKQGR↓L	KF851269
Mallard	I/1.2	AR Crimea	Krasnoperekopsk/18-23-10	2010	0.16	ERQER↓L	KF851268
Ruddy shelduck	II/I.2	Kherson	Askania-Nova/3-20-11[c]	2010	0.05	GKQGR↓L	MZ101338
Pigeon	II/VII.1.1	AR Crimea	Simferopol/2-26-11	2011	N/A	**RRQKR↓F**	KU710277
Ruddy shelduck	II/I.2	Kherson	Askania-Nova/37-15-02	2011	0.76	GKQGR↓L	KF851270
Pigeon	II/XXI.1.1	Dnipro	Dnipropetrovsk/1-18-11[c]	2011	1.15	**KRQKR↓F**	KJ914671
Pigeon	II/XXI.1.1	AR Crimea	Ukromne/3-26-11[c]	2011	1.36	**KRQKR↓F**	KJ914672
Pigeon	II/XXI.1.1	Kharkiv	Kharkiv/23-01/967	2013	1.7	**KRQKR↓F**	KY042127
White-fronted Goose	II/I.2	Kherson	Askania-Nova/72-28-03[c]	2013	N/A[b]	GKQGR↓L	MZ101339
Mediterranean Gull	II/I.2	Kherson	Smalanyi/5-11-07[c]	2013	N/A	GKQGR↓L	MZ101340
Pigeon	II/XXI.1.1	Kharkiv	Kuksov/13-05[c]	2014	N/A	**KRQKR↓F**	MZ101341
Pigeon	II/VI.2.1.1.2.2	Kharkiv	Petruk/15-01[c]	2015	N/A	**RRQKR↓F**	MZ101342

[a]ICPI, intracerebral pathogenicity index.
[b]N/A, not available.
[c]Sequenced in this study.
The virulent cleavage sites are highlighted in bold.

Pathogenicity evaluation was performed for eleven Ukrainian viruses using ICPI. The ICPI values ranged from 0.05 to 1.7 (Table 1). Four viruses had an ICPI value below 0.7, which characterizes them as lentogenic (avirulent) which is in agreement with the predictions based on deduced amino acid cleavage site sequences. ICPI values equal to or above 0.7 indicate a virulent strain. Virulent strains with ICPI values below 1.5 are mesogenic (moderate virulence) and those with values above 1.5 are velogenic NDV strains (3, 55, 57). Thus, six viruses were classified as mesogenic and one (pigeon/Ukraine/Kharkiv/23-01-/967/2013) as velogenic (3, 55).

3.3. Genotypic characterization

A dataset of 79 complete F gene coding sequences of class I and II NDV isolates retrieved from GenBank was added to the 19 sequences from this study and used to construct a pilot tree for the preliminary identification of NDV genotypes as previously described (5). Based on the pilot tree, one NDV isolate belonged to class I and 18 to class II (Figure 2). Among class II isolates, five were of genotype I, six sequences were of genotype II, one was of genotype VI, one was of genotype VII, and five were of genotype XXI (Figure 2).

3.4. Genotype 1 isolates

The single NDV isolate of class I was collected from a mallard in AR Crimea in 2010 (KF851268) and belongs to sub-genotype 1.2 of class I (Figure 3), previously known as sub-genotype 1c (5, 58). The highest homology of this isolate is with the two strains—pochard in Finland from 2006 (EU493454) and duck in China from 2007 (JF893453), sharing 98.81% nucleotide identity.

3.5. Genotype I isolates

All five isolates of class II genotype I were classified as sub-genotype I.2 (Figure 4). These strains were collected from wild

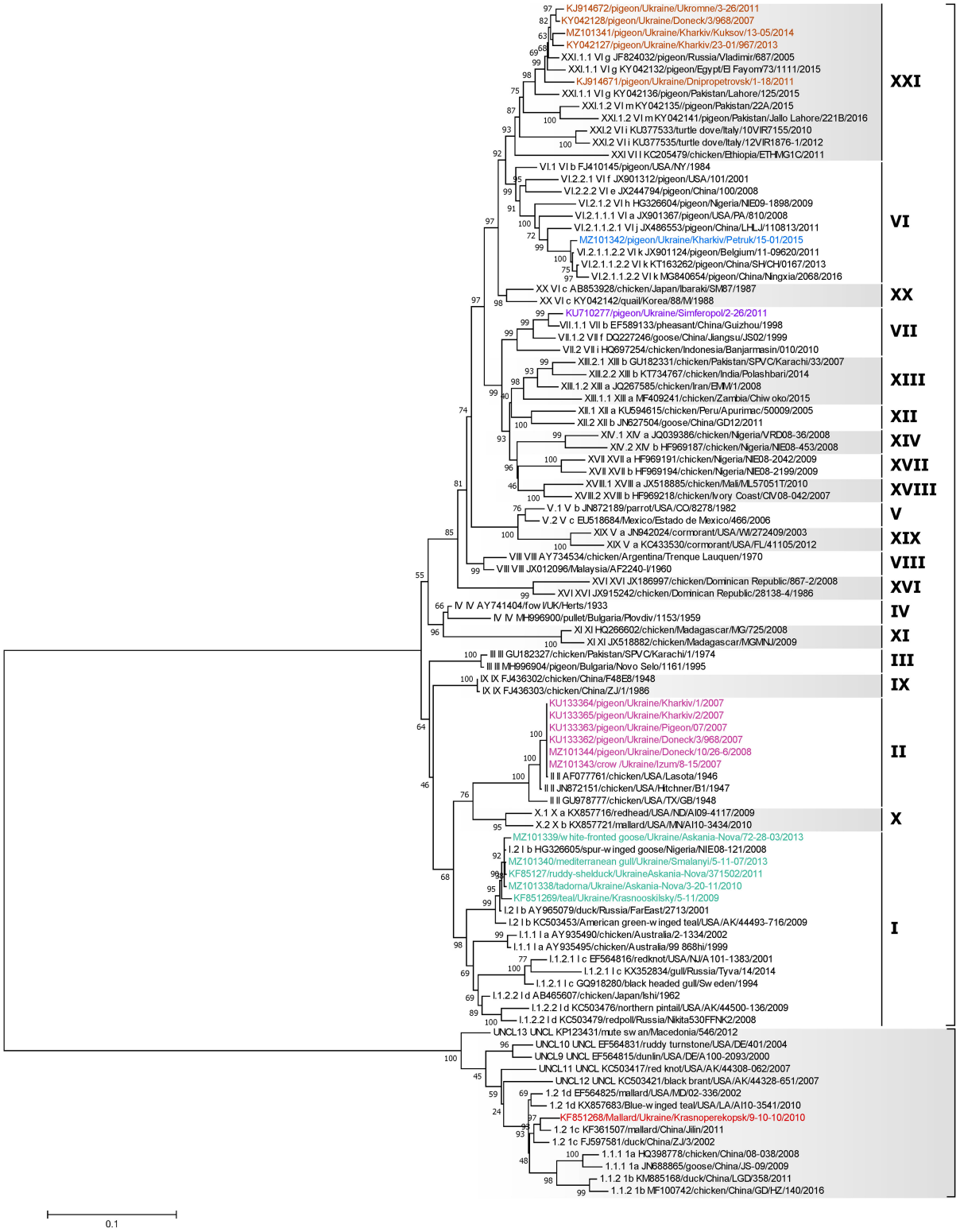


FIGURE 2
Phylogenetic analysis of NDV class I and II isolates based on the complete fusion gene sequences constructed with the maximum likelihood method, based on the general time-reversible model in MEGA v. 7.0.26. The percentage of trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The analysis involved 98 nucleotide sequences, 79 represent all class I and II sub-genotypes described by Dimitrov et al. (5) and 19 were collected from wild and synanthropic birds in Ukraine. All positions containing gaps and missing data were eliminated. There were a total of 1,661 positions in the final data set. The isolates used in this study are shown in colors. The Roman numerals presented in the taxa names in the phylogenetic tree represent the respective sub-genotype for each isolate, followed by the GenBank identification number, host name, country of isolation, strain designation, and year of isolation (if available).

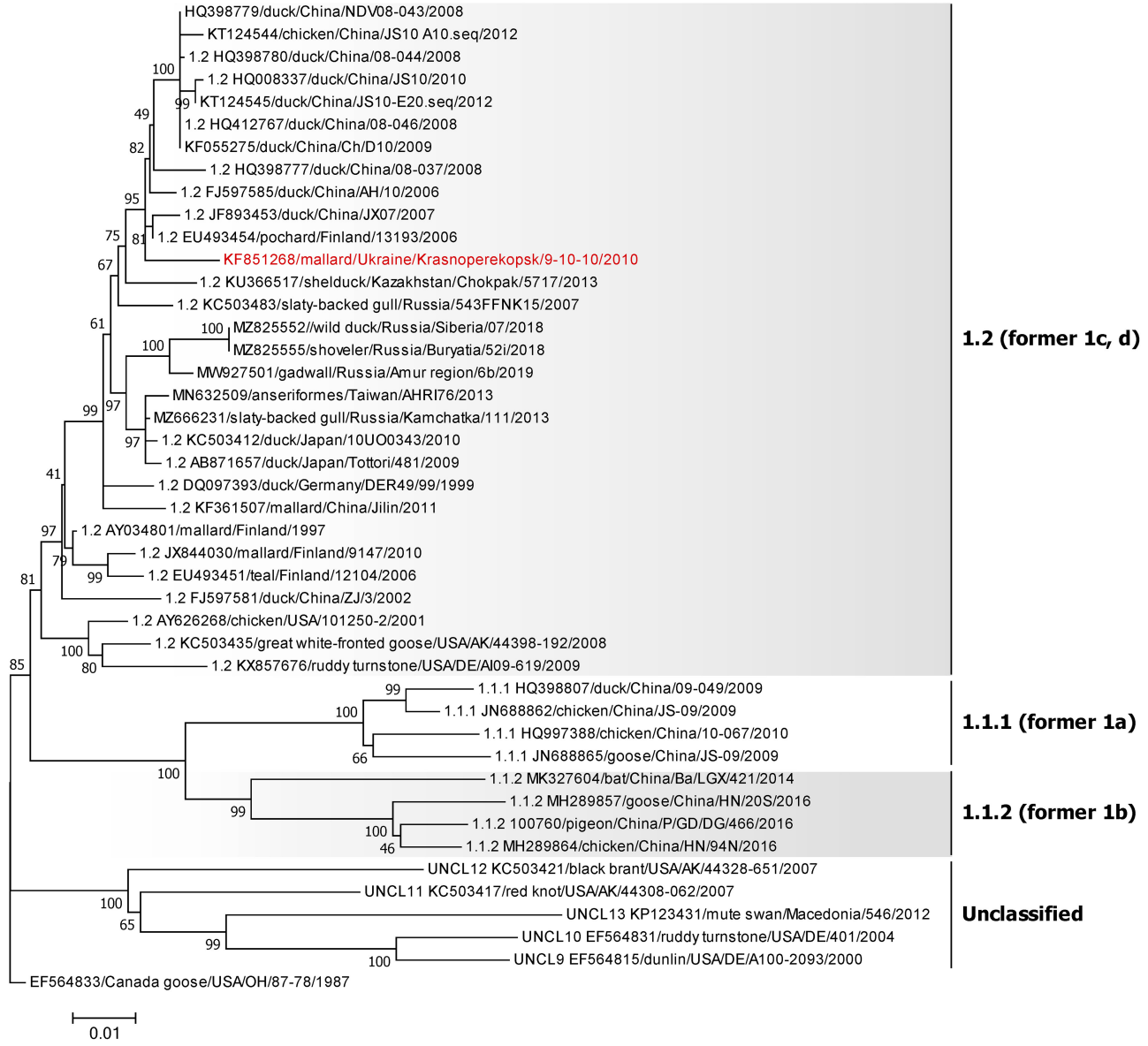


FIGURE 3
Phylogenetic analysis of NDV class I isolates based on the complete fusion gene sequences constructed with the maximum likelihood method with 1,000 bootstrap replicates. The analysis involved 44 nucleotide sequences. The tree was rooted to the oldest class I NDV isolate EF564833/Canada goose/USA/OH/78/1987. All positions containing gaps and missing data were eliminated. There were a total of 1,662 positions in the final data set. The isolates used in this study are shown in red. The Roman numerals on the right of the phylogenetic tree represent the respective sub-genotype for each isolate in accordance with the current classification (5) and the former nomenclature (58) is listed in parenthesis for reference.

birds in the Kherson and Donetsk regions and shared between 98.50 and 99.82% nucleotide identity. Among them, the isolate collected from a white-fronted goose (MZ101339) had the highest nucleotide identity of 99.4% with the isolate collected from a Mediterranean gull (MZ101340) in 2013, which had an even higher nucleotide identity of 99.82% with another Ukrainian strain collected from ruddy shelduck (genus *Tadorna*) in 2011 (KF851270) (29). All these viruses were collected in the Kherson region. The fourth strain (MZ101338) from the Kherson region, isolated from ruddy shelduck in 2010, was most similar to an isolate originating from central Eurasia (Novosibirsk region) in 2010 (KX352836) (25) and shared 99.88% nucleotide identity. The only virus from the Donetsk region was collected from teal in 2009 (KF851269) and had the highest nucleotide homology with the strain isolated from mallard in Luxembourg in 2008 (HE972213) (59).

3.6. Genotype II isolates

All six NDV genotype II isolates (KU133362–KU133365, MZ101343, and MZ101344) were 100% identical (Figure 4). These isolates were collected from dead pigeons in Dnipro, Donetsk, and Kharkiv regions and a dead crow (genus *Corvus)* in the

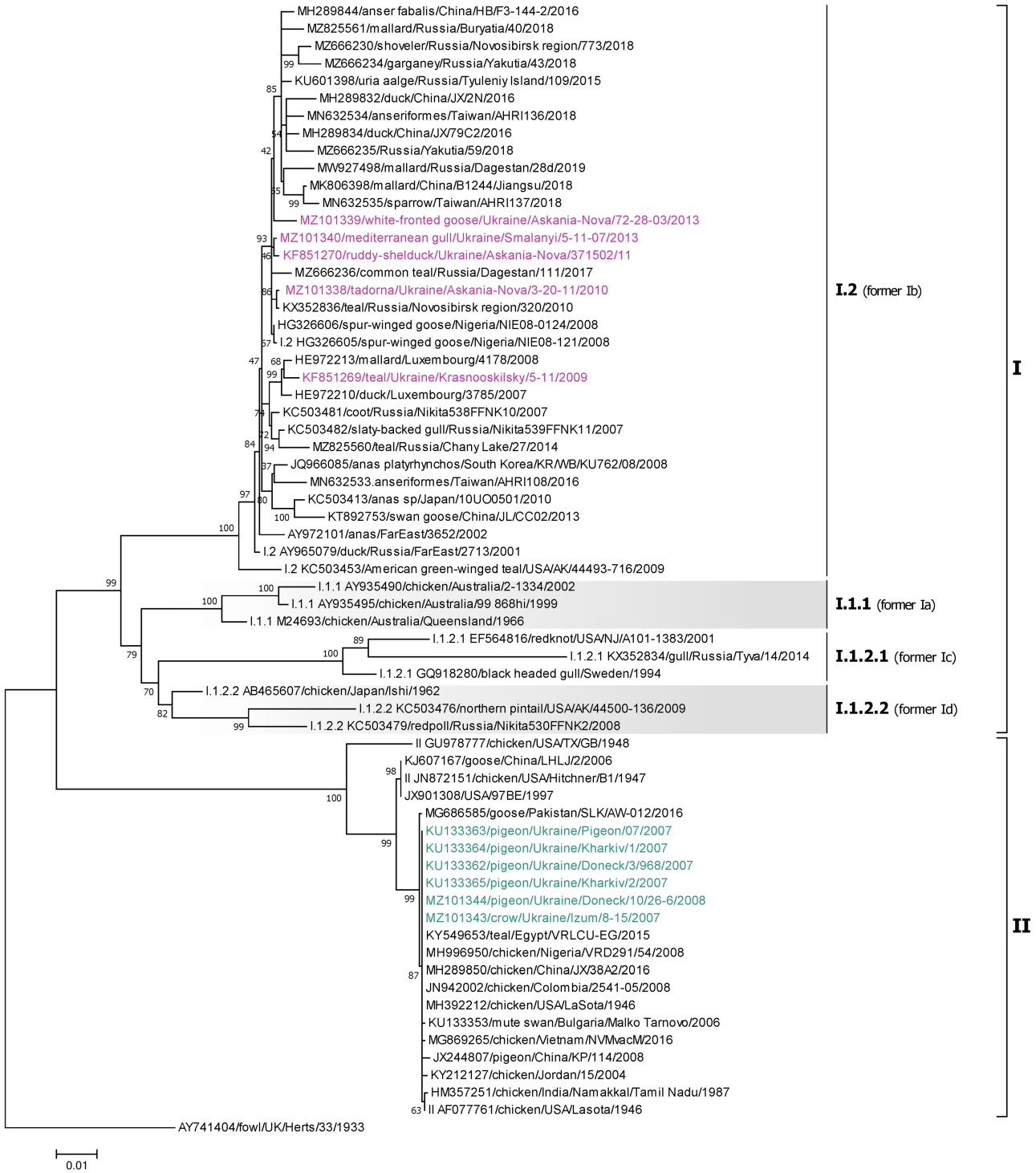


FIGURE 4
Phylogenetic analysis of NDV genotype I and II class II isolates based on the complete fusion gene sequences constructed with the maximum likelihood method with 1,000 bootstrap replicates. The analysis involved 65 nucleotide sequences (a sequence from genotype IV AY741404/fowl/UK/Herts/33/1933 was included as an outgroup). All positions containing gaps and missing data were eliminated. There were a total of 1,661 positions in the final data set. The isolates used in this study are shown in pink (genotype I) and teal (genotype II). The Roman numerals on the right of the phylogenetic tree represent the respective sub-genotype for each isolate in accordance with the current classification (5) and the former nomenclature (58) is listed in parenthesis for reference.

Kharkiv region in 2007 (23). These viruses were identical to the LaSota vaccine strain (MH392212) (60) and other vaccine strains isolated from different species of birds across different continents (37).

3.7. Genotype VI isolates

The single NDV genotype VI isolate was derived from a pigeon located in the Kharkiv region of

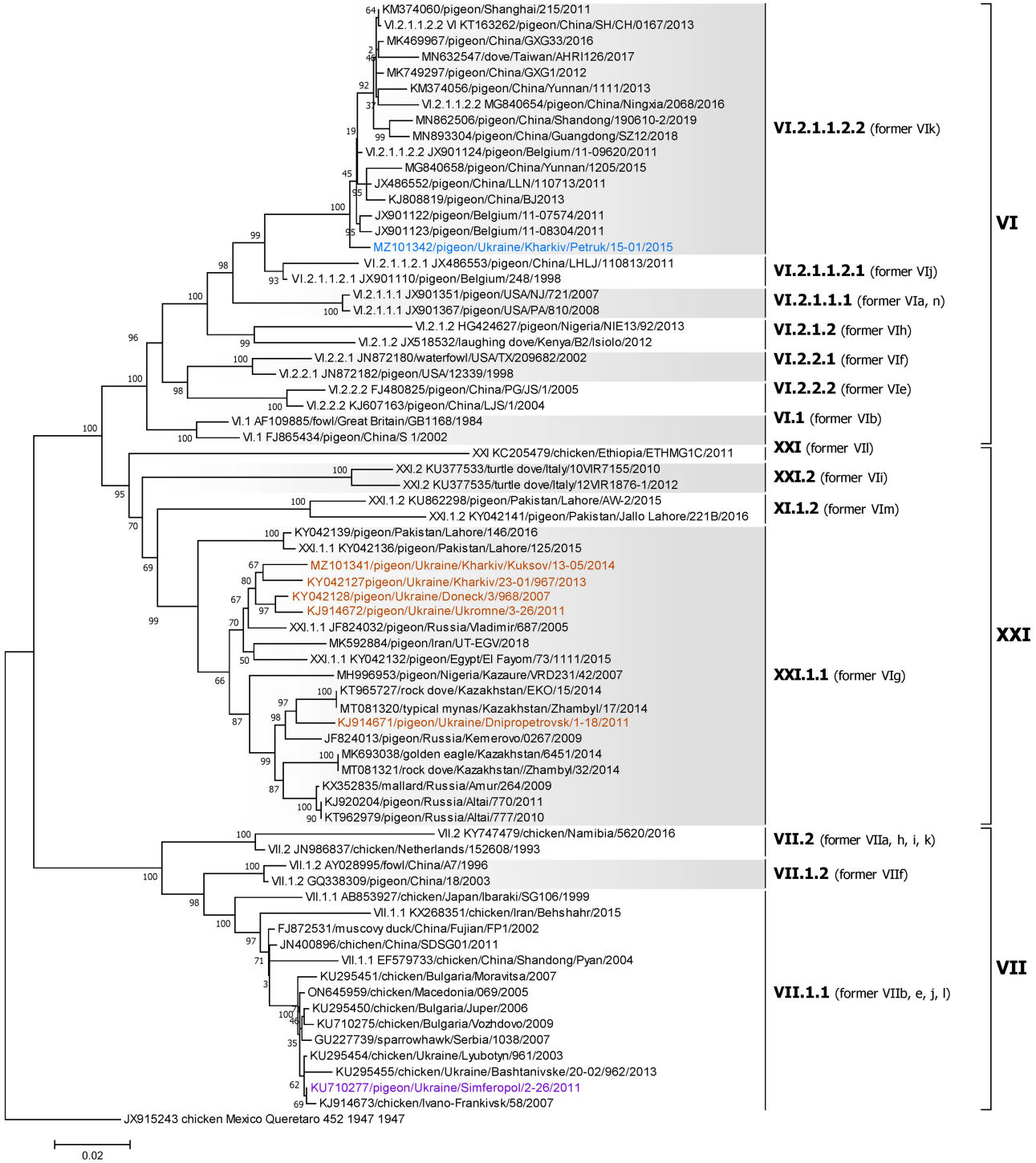


FIGURE 5
Phylogenetic analysis of NDV genotype VI, VII, and XXI class II isolates based on the complete fusion gene sequences constructed with the maximum likelihood method with 1,000 bootstrap replicates. The analysis involved 71 nucleotide sequences (a sequence from genotype XVI JX915243/chicken/Mexico/Queretaro/452/1947 was included as an outgroup). All positions containing gaps and missing data were eliminated. There were a total of 1,660 positions in the final data set. The isolates used in this study are shown in blue (genotype VI), purple (genotype VII), and brown (genotype XXI). The Roman numerals on the right of the phylogenetic tree represent the respective sub-genotype for each isolate in accordance with the current classification (5) and the former nomenclature (58) is listed in parenthesis for reference.

Ukraine (MZ101342) and belongs to sub-genotype VI.2.1.1.2.2, formerly classified as sub-genotype VIk (Figure 5) (5, 58). The highest homology of this isolate is to a strain isolated from a pigeon in Belgium in 2011 (JX901124) (61), sharing 99.16% nucleotide identity.

3.8. Genotype VII isolates

The single NDV genotype VII isolate was derived from a pigeon located in Simferopol, AR Crimea (KU710277) and belongs to sub-genotype VII.1.1, formerly classified as sub-genotype VIId (Figure 5) (5, 58). The highest homology of this isolate is to a strain isolated from a chicken in the Kharkiv region of Ukraine in 2003 (KU295454) (44), sharing 99.88% nucleotide identity.

3.9. Genotype XXI isolates

All five genotype XXI isolates were detected in Ukrainian pigeons and belong to sub-genotype XXI.1.1, previously known as VIg (Figure 5) (5, 58). These strains were collected in Donetsk, Dnipro, Kharkiv, and AR Crimea regions and shared between 95.91 and 98.98% nucleotide identity. Of these, three (KJ914672, KY042127, and MZ101341), that were collected in Kharkiv and AR Crimea, had the highest homology to another Ukrainian strain (KY042128) isolated from a pigeon in the Donetsk region in 2007 (27), sharing 98.98, 98.13, and 98.07% nucleotide identity, respectively (27, 28). Another Ukrainian sub-genotype XXI.1.1 sequence from the Dnipro region (KJ914671) shared 98.01% homologies with two strains isolated from pigeons in Kazakhstan in 2014 (KT965727 and MT081320) (62).

4. Discussion

Our previous studies, as well as this one, provide essential information on the epidemiology of the NDV isolates from synanthropic and wild birds in Ukraine. In this study, we used Sanger sequencing to obtain complete F gene sequences of NDV in Ukraine from 2006 to 2015. Our study reports repeated NDV detection of sub-genotype I.2 in wild birds collected at the stopping points of migratory birds in Ukraine, as well as the first occurrence of sub-genotype VI.2.1.1.2 and the continuous presence of sub-genotypes II and XXI.1.1 in Ukrainian synanthropic birds. Additionally, we present the first publicly available complete NDV F gene from a crow (genus *Corvus*).

The Azov-Black Sea region of Ukraine is part of three transcontinental wild bird migration routes: the West Asia-East Africa, East Atlantic, and Black Sea-Mediterranean flyways (30, 31, 45). This region is comprised of areas for transit, stops during migration, and nesting for many bird species, which makes it one of the highly important regions in Eurasia for monitoring and studying the global circulation of NDV and predicting the emergence of new strains possibly transmitted by wild birds.

The single class I NDV isolate was detected from wild waterfowl at a major stopping point location in AR Crimea and belonged to the 1.2 sub-genotype, formerly known as 1c. This virus clustered together with strains also isolated from wild waterfowl of the order *Anseriformes* (family *Anatidae*) in China and Finland. The high identity between these isolates from Europe and Asia supports our hypothesis of intercontinental viral transmission by migratory birds.

All detected class II NDV isolates from wild waterfowl and shorebirds collected at the major stopping point locations of migratory routes in Ukraine, belonged to genotype I, which is consistent with our previous study (29, 63) and confirms the continuous threat of virus introduction from wild birds. All viruses were obtained from members of the orders *Anseriformes* (family *Anatidae*) and Charadriiformes (family *Laridae*). Of the five genotype I strains analyzed, all were classified as sub-genotype I.2. These isolates grouped together with isolates previously detected in different hosts at various locations, including (25, 29). Luxembourg, the North Caucasian, Siberian, and Far East Federal Districts of Russia, China, Taiwan, and Nigeria which confirms the intercontinental spread of the virus by migratory birds. Similar findings were reported in our previous studies where epidemiological connections of Avian Paramyxoviruses between Europe and Africa were shown (29). The high identity between isolates from wild birds in Europe and Asia, and their close phylogenetic relationship with strains from Africa, support our hypothesis of virus exchange along the Black Sea-Mediterranean and Asian-East African migratory flyways and highlight the possibility of intercontinental viral transmission.

Special attention should be given to the data obtained from the phylogenetic analysis of NDVs isolated from synanthropic birds in Ukraine. Six of these isolates, which were isolated from pigeons and a crow in three Eastern and Southern regions of Ukraine in 2007 and 2008, were classified as genotype II. To the best of our knowledge, there was no complete NDV F gene of any sub-genotype from crows in public databases prior to this study. However, partial F gene sequences of viruses of genotypes II isolated from crows in India in 2002 (AY339400) and Pakistan in 2017 (MN728799–MN728801) were available in the GenBank database (64). Interestingly, Ukrainian viruses were identical to the vaccine strain LaSota, which is widely used in Ukraine and around the world as a live vaccine against NDV (37). Industrial and backyard poultry farming is very developed in these Eastern regions of Ukraine. Furthermore, one of the last NDV outbreaks in Ukraine was previously recorded in one of those regions (Kharkiv region) in 2006 (3, 29). Because of that outbreak, a number of anti-epizootic measures were implemented, involving vaccination against NDV in industrial and backyard farms (including those where poultry had direct contact with synanthropic birds). We supposed the presence of a vaccine virus found in the synanthropic birds could be a result of contact with vaccinated poultry. In recent decades, the number of genotypes has increased. It is likely that the spread of a vaccine strain could contribute to an increase in the genetic diversity of NDV (32). Therefore, special attention must be paid to the distribution of vaccine strains in wild birds in order to understand the consequences of global vaccination.

NDV isolates of XXI genotype were identified in five dead pigeons with clinical signs of disease, which not only confirmed that sub-genotype XXI.1.1 is seemingly maintained in pigeons in the East of Ukraine (Donetsk and Kharkiv regions) from 2007 to 2014 (27) but also was detected for the first time in the South of Ukraine (Dnipro and Simferopol regions). Even though the isolates collected in Donetsk, Kharkiv, and AR Crimea regions grouped together, the isolate collected in the Dnipro region in 2011 was highly similar to isolates from pigeons and wild birds from genus *Acridotheres* (KT965727 and MT081320) from Kazakhstan in 2014 (62). This further highlights the possibility of continuous intercontinental viral spread. This may also indicate an additional link in the distribution of genotype XXI strains or its association with a host preference for pigeons. Interestingly, two different NDV sub-genotypes (II and

XXI.1.1) were detected in a pigeon from Donetsk collected in 2007. This was possible due to the utilization of two different sets of primers to amplify the complete F gene of different NDV genotypes. However, this is rare, because in some cases we were not able to amplify a complete F gene even though APMV-1 was confirmed in these isolates by serological methods. We are speculating that it was due to acquired mutations in the NDV F gene which resulted in the primer's mismatch. Also, we can't exclude the possibility of the cross-reactivity of different serotypes by serology, which we previously observed and reported in an isolate collected from a white-fronted goose in the Kherson region in 2011 (51, 52). This isolate weakly cross-reacted with APMV-1 and APMV-7 antisera in serology, but we were unable to amplify NDV's F gene by utilizing the set of primers for routine NDV detection. However, random whole-genome next-generation sequencing allowed us to assemble the complete genome and discover a new serotype of APMV (named APMV-13), which explained the previous inability to amplify the complete F gene using primers specific for NDV. This highlights the need for implementing random next-generation sequencing as a routine diagnostic tool in order to better perceive the complete epidemiological situation.

To the best of our knowledge, only viruses of sub-genotypes II, VII.1.1, and XXI.1.1 have been previously reported to circulate in pigeons in Ukraine (23, 27, 44, 65). In this study, we report the first identification of genotype VI NDV in Ukraine. This virulent NDV isolate, classified as sub-genotype VI.2.1.1.2, was obtained from a deceased pigeon in the Kharkiv region in 2015.

To obtain a complete picture of the distribution of NDV along the migratory flyways and to determine all circulating genotypes among wild waterfowl, it is necessary to continue the annual monitoring of NDV in the Azov-Black Sea region, as one of the major stopping points for migratory birds. Additional data will help to assess the degree of involvement of wild birds in the spread of virulent strains, that can be especially dangerous for poultry production. Further monitoring of NDV in synanthropic bird species will provide useful data for the study of vaccine strains widespread globally.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MZ101338, MZ101339, MZ101340, MZ101341, MZ101342, MZ101343, MZ101344, KF851268, KF851270, KJ914671, KJ914672, KU710277, KF851269, KU133362, KU133363, KU133364, KU133365, KY042127, and KY042128.

Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of the National Scientific Center Institute of Experimental and Clinical Veterinary Medicine.

Author contributions

Conceptualization: IG, AG, DM, and CA. Funding acquisition: CA and BS. Project administration and coordination: CA, AG, and DM. Methodology: IG. Sample preparation: IG, VB, OS, OR, NM, OM, and OK. Data curation and formal analyses: IG, KD, PM, and CA. Writing—original draft: IG and DM. Writing—review and editing: PM, CA, and MP-J. All authors have read and agreed to the published version of the manuscript.

Funding

The work with Ukrainian isolates was supported by the U.S. Defense Threat Reduction Agency and by the USDA, ARS CRIS Project 6040-32000-064. Part of the research was funded by USDA Project P444, through the Ukrainian Science and Technology Center. Part of the research was done in the frame of the Joint Ukrainian-Austrian R&D Project (contract #M7-2021, 11.11.2021 and #M37-2022, 24.05.2022).

Acknowledgments

We would like to acknowledge Timothy Olivier and Dawn Williams-Coplin for their technical assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1026296/full#supplementary-material

SUPPLEMENTARY TABLE S1
Number of samples of biological material collected from wild birds of different species in different regions of Ukraine from 2006 to 2015.

SUPPLEMENTARY TABLE S2
Number of samples of biological material collected from synanthropic birds of different species in different regions of Ukraine from 2006 to 2015.

References

1. Rima B, Balkema-Buschmann A, Dundon WG, Duprex P, Easton A, Fouchier R, et al. ICTV virus taxonomy profile: paramyxoviridae. *J Gen Virol.* (2019) 100:1593–4. doi: 10.1099/jgv.0.001328

2. Kaleta EF, Baldauf C. *Newcastle Disease in Free-Living and Pet Birds*. Newcastle: Springer (1988), 197–246.

3. OIE. *Newcastle Disease (Infection with Newcastle Disease Virus). Manual of Diagnostic Tests and Vaccinesfor Terrestrial Animals*. Geneva: World Organization for Animal Health (2021).

4. OIE. *Infection with Newcastle Disease Virus. Terrestrial Animal Health Code*. Paris: OIE (2022).

5. Dimitrov KM, Abolnik C, Afonso CL, Albina E, Bahl J, Berg M, et al. Updated unified phylogenetic classification system and revised nomenclature for Newcastle disease virus. *Infect Genet Evol.* (2019) 74:103917. doi: 10.1016/j.meegid.2019.103917

6. 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. doi: 10.1016/j.meegid.2016.01.008

7. Oberdorfer A, Werner O. Newcastle disease virus: detection and characterization by PCR of recent German isolates differing in pathogenicity. *Avian Pathol.* (1998) 27:237–43. doi: 10.1080/03079459808419330

8. Russell PH, Samson ACR, Alexander DJ. Newcastle disease virus variations. *Appl Virol Res.* (1990) 2:177–95.

9. Nagai Y, Klenk HD, Rott R. Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of Newcastle disease virus. *Virology.* (1976) 72:494–508. doi: 10.1016/0042-6822(76)90178-1

10. Nagai Y, Shimokata K, Yoshida T, Hamaguchi M, Iinuma M, Maeno K, et al. The spread of a pathogenic and an apathogenic strain of Newcastle disease virus in the chick embryo as depending on the protease sensitivity of the virus glycoproteins. *J Gen Virol.* (1979) 45:263–72. doi: 10.1099/0022-1317-45-2-263

11. Wang Y, Bi Y, Yu W, Wei N, Wang W, Wei Q, et al. Two mutations in the HR2 region of Newcastle disease virus fusion protein with a cleavage motif "RRQRRL" are critical for fusogenic activity. *Virol J.* (2017) 14:185. doi: 10.1186/s12985-017-0851-0

12. Toyoda T, Sakaguchi T, Imai K, Inocencio NM, Gotoh B, Hamaguchi M, et al. Structural comparison of the cleavage-activation site of the fusion glycoprotein between virulent and avirulent strains of Newcastle disease virus. *Virology.* (1987) 158:242–7. doi: 10.1016/0042-6822(87)90261-3

13. Pritzer E, Kuroda K, Garten W, Nagai Y, Klenk HD. A host range mutant of Newcastle disease virus with an altered cleavage site for proteolytic activation of the F protein. *Virus Res.* (1990) 15:237–42. doi: 10.1016/0168-1702(90)90031-6

14. Panda A, Huang Z, Elankumaran S, Rockemann DD, Samal SK. Role of fusion protein cleavage site in the virulence of Newcastle disease virus. *Microb Pathog.* (2004) 36:1–10. doi: 10.1016/j.micpath.2003.07.003

15. Miller PJ, Kim LM, Ip HS, Afonso CL. Evolutionary dynamics of Newcastle disease virus. *Virology.* (2009) 391:64–72. doi: 10.1016/j.virol.2009.05.033

16. Kim SH, Wanasen N, Paldurai A, Xiao S, Collins PL, Samal SK. Newcastle disease virus fusion protein is the major contributor to protective immunity of genotype-matched vaccine. *PLoS ONE.* (2013) 8:e74022. doi: 10.1371/journal.pone.0074022

17. 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. doi: 10.1177/104063870601800601

18. 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) 48:2440–8. doi: 10.1128/JCM.02213-09

19. Miller PJ, King DJ, Afonso CL, Suarez DL. Antigenic differences among Newcastle disease virus strains of different genotypes used in vaccine formulation affect viral shedding after a virulent challenge. *Vaccine.* (2007) 25:7238–46. doi: 10.1016/j.vaccine.2007.07.017

20. Cardenas-Garcia S, Diel DG, Susta L, Lucio-Decanini E, Yu Q, Brown CC, et al. Development of an improved vaccine evaluation protocol to compare the efficacy of Newcastle disease vaccines. *Biologicals.* (2015) 43:136–45. doi: 10.1016/j.biologicals.2014.11.003

21. Cappelle J, Caron A, Servan De Almeida R, Gil P, Pedrono M, Mundava J, et al. Empirical analysis suggests continuous and homogeneous circulation of Newcastle disease virus in a wide range of wild bird species in Africa. *Epidemiol Infect.* (2015) 143:1292–303. doi: 10.1017/S095026881400185X

22. Muzyka D, Pantin-Jackwood M, Stegniy B, Afonso C, editors. *Avian Paramyxovirus Serotypes Circulating in Wild Bird Populations of the Azov-Black Sea Region of Ukraine in 2006–2011*. Meeting Abstract (2013).

23. Ayala AJ, Dimitrov KM, Becker CR, Goraichuk IV, Arns CW, Bolotin VI, et al. Presence of vaccine-derived newcastle disease viruses in wild birds. *PLoS ONE.* (2016) 11:e0162484. doi: 10.1371/journal.pone.0162484

24. 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. doi: 10.3389/fpubh.2015.00235

25. Yurchenko KS, Zhou P, Kovner AV, Zavjalov EL, Shestopalova LV, Shestopalov AM. Oncolytic effect of wild-type Newcastle disease virus isolates in cancer cell lines *in vitro* and *in vivo* on xenograft model. *PLoS ONE.* (2018) 13:e0195425. doi: 10.1371/journal.pone.0195425

26. Glushchenko AV, Alikina TY, Yurchenko KS, Shekunov EV, Gulyaeva MA, Matsuno K, et al. Nearly complete genome sequence of a Newcastle disease virus strain isolated from a wild garganey. *Microbiol Resour Announc.* (2019) 8:19. doi: 10.1128/MRA.01072-19

27. 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. doi: 10.1186/s12917-017-1211-4

28. Pchelkina IP, Manin TB, Kolosov SN, Starov SK, Andriyasov AV, Chvala IA, et al. Characteristics of pigeon paramyxovirus serotype-1 isolates (PPMV-1) from the Russian Federation from 2001 to 2009. *Avian Dis.* (2013) 57:2–7. doi: 10.1637/10246-051112-Reg.1

29. Muzyka D, Pantin-Jackwood M, Stegniy B, Rula O, Bolotin V, Stegniy A, et al. Wild bird surveillance for avian paramyxoviruses in the Azov-Black Sea regions of Ukraine (2006–2011) reveals epidemiological connections with Europe and Africa. *Appl Environ Microbiol.* (2014) 80:5427–38. doi: 10.1128/AEM.00733-14

30. Diadicheva E, Matsievskaya N. Migration routes of waders using stopover sites in the Azov-Black Sea region, Ukraine. *Vogelwarte.* (2000) 40:161–78. Available online at: https://www.zobodat.at/pdf/Vogelwarte_40_1999_0161-0178.pdf

31. Kulak MV, Ilinykh FA, Zaykovskaya AV, Epanchinzeva AV, Evstaphiev IL, Tovtunec NN, et al. Surveillance and identification of influenza A viruses in wild aquatic birds in the Crimea, Ukraine (2006–2008). *Avian Dis.* (2010) 54:1086–90. doi: 10.1637/9272-020510-ResNote.1

32. Muzyka D, Pantin-Jackwood MJ, Spackman E, Smith DM, Rula O, Muzyka N, et al. Isolation and genetic characterization of avian influenza viruses isolated from wild birds in the Azov-Black Sea region of Ukraine (2001–2012). *Avian Dis.* (2016) 60:365–77. doi: 10.1637/11114-050115-Reg

33. Boere GC, Galbraith CA, Stroud DA. *Waterbirds Around the World: A Global Overview of the Conservation, Management and Research of the World's Waterbird Flyways*. London: Stationery Office (2006).

34. FAO. *Gateway to Poultry Production and Products*. New York, NY: FAO (2020). Available online at: https://www.fao.org/poultry-production-products/production/poultry-species/en/ (accessed August 3, 2022).

35. Miller PJ, Koch G. Newcastle disease. In: Swayne DE, Glisson JR, McDougald LR, Nolan LK, Suarez DL, Nair VL, editor. *Diseases of Poultry*. 13 ed. Ames, IA: Wiley-Blackwell in Partnership with the American Association of Avian Pathologists (2013). p. 89–138.

36. Dimitrov KM, Afonso CL, Yu Q, Miller PJ. Newcastle disease vaccines: a solved problem or a continuous challenge? *Vet Microbiol.* (2017) 206:126–36. doi: 10.1016/j.vetmic.2016.12.019

37. Welch CN, Shittu I, Abolnik C, Solomon P, Dimitrov KM, Taylor TL, et al. Genomic comparison of Newcastle disease viruses isolated in Nigeria between 2002 and 2015 reveals circulation of highly diverse genotypes and spillover into wild birds. *Arch Virol.* (2019) 164:2031–47. doi: 10.1007/s00705-019-04288-9

38. Cardenas Garcia S, Navarro Lopez R, Morales R, Olvera MA, Marquez MA, Merino R, et al. Molecular epidemiology of Newcastle disease in Mexico and the potential spillover of viruses from poultry into wild bird species. *Appl Environ Microbiol.* (2013) 79:4985–92. doi: 10.1128/AEM.00993-13

39. Vidanović D, Sekler M, Asanin R, Milić N, Nisavić J, Petrović T, et al. Characterization of velogenic Newcastle disease viruses isolated from dead wild birds in Serbia during 2007. *J Wildl Dis.* (2011) 47:433–41. doi: 10.7589/0090-3558-47.2.433

40. Afonso CL. Virulence during Newcastle disease viruses cross species adaptation. *Viruses.* (2021) 13:115. doi: 10.3390/v13010110

41. Polard D. Newcastle disease outbreak in Australia. *World Poult.* (1999) 15:36–7.

42. Kirkland PD. Virulent Newcastle disease virus in Australia: in through the 'back door'. *Aust Vet J.* (2000) 78:331–3. doi: 10.1111/j.1751-0813.2000.tb11786.x

43. Shengqing Y, Kishida N, Ito H, Kida H, Otsuki K, Kawaoka Y, et al. Generation of velogenic Newcastle disease viruses from a nonpathogenic waterfowl isolate by passaging in chickens. *Virology.* (2002) 301:206–11. doi: 10.1006/viro.2002.1539

44. Dimitrov KM, Bolotin V, Muzyka D, Goraichuk IV, Solodiankin O, Gerilovych A, et al. Repeated isolation of virulent Newcastle disease viruses of sub-genotype VIId from backyard chickens in Bulgaria and Ukraine between 2002 and 2013. *Arch Virol.* (2016) 161:3345–53. doi: 10.1007/s00705-016-3033-2

45. Muzyka D, Pantin-Jackwood M, Spackman E, Stegniy B, Rula O, Shutchenko P. Avian influenza virus wild bird surveillance in the Azov and Black Sea regions of Ukraine (2010–2011). *Avian Dis.* (2012) 56:1010–6. doi: 10.1637/10157-040912-ResNote.1

46. Alexander DJ. *Newcastle Disease Virus and Other Avian Paramyxoviruses. A Laboratory Manual for the Isolation and Identification of Avian Pathogens*. 4 ed. Kennett Square, PA: The American Association of Avian Pathologists (1998). p. 156-63.

47. OIE. *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*. 8th ed. Paris: OIE (2018).

48. American Association of Avian Pathologists. *A Laboratory Manual for the Isolation, Identification and Characterization of Avian Pathogens*. 5th ed. Athens, GA: American Association of Avian Pathologists (2008).

49. Capua I, Alexander DJ. *Avian Influenza and Newcastle Disease: A Field and Laboratory Manual*. 1st ed. Milan: Springer (2009). p. 186. doi: 10.1007/978-88-470-0826-7_1

50. Spackman. *Animal Influenza Virus*. New York, NY: Humana Press (2014). doi: 10.1007/978-1-4939-0758-8

51. Goraichuk I, Sharma P, Stegniy B, Muzyka D, Pantin-Jackwood MJ, Gerilovych A, et al. Complete genome sequence of an avian paramyxovirus representative of putative new serotype 13. *Genome Announc*. (2016) 4:16. doi: 10.1128/genomeA.00729-16

52. Goraichuk I, Poonam S, Dimitrov K, Stegniy B, Muzyka D, Pantin-Jackwood M, et al. Phylogenetic analysis of the complete genome of the APMV-13 isolate from Ukraine. In: *Proceedings of the 17th International Congress on Infectious Disease 7th International Congress on Infectious Diseases: Journal of Infectious Disease*. International Journal of Infectious Diseases (2016). p. 459. doi: 10.1016/j.ijid.2016.02.972

53. Diel DG, Miller PJ, Wolf PC, Mickley RM, Musante AR, Emanueli DC, et al. Characterization of Newcastle disease viruses isolated from cormorant and gull species in the United States in 2010. *Avian Dis*. (2012) 56:128–33. doi: 10.1637/9886-081111-Reg.1

54. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. (2013) 30:772–80. doi: 10.1093/molbev/mst010

55. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 70 for bigger datasets. *Mol Biol Evol*. (2016) 33:1870–4. doi: 10.1093/molbev/msw054

56. 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) 101:11030–5. doi: 10.1073/pnas.0404206101

57. Alexander DJ. Newcastle disease and other avian paramyxoviruses. *Rev Sci Tech*. (2000) 19:443–62. doi: 10.20506/rst.19.2.1231

58. 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) 12:1770–9. doi: 10.1016/j.meegid.2012.07.012

59. Snoeck CJ, Marinelli M, Charpentier E, Sausy A, Conzemius T, Losch S, et al. Characterization of newcastle disease viruses in wild and domestic birds in Luxembourg from 2006 to 2008. *Appl Environ Microbiol*. (2013) 79:639–45. doi: 10.1128/AEM.02437-12

60. Butt SL, Taylor TL, Volkening JD, Dimitrov KM, Williams-Coplin D, Lahmers KK, et al. Rapid virulence prediction and identification of Newcastle disease virus genotypes using third-generation sequencing. *Virol J*. (2018) 15:179. doi: 10.1186/s12985-018-1077-5

61. Van Borm S, Rosseel T, Steensels M, van den Berg T, Lambrecht B. What's in a strain? Viral metagenomics identifies genetic variation and contaminating circoviruses in laboratory isolates of pigeon paramyxovirus type 1. *Virus Res*. (2013) 171:186–93. doi: 10.1016/j.virusres.2012.11.017

62. Orynbayev MB, Fereidouni S, Sansyzbai AR, Seidakhmetova BA, Strochkov VM, Nametov AM, et al. Genetic diversity of avian avulavirus 1 (Newcastle disease virus genotypes VIg and VIIb) circulating in wild birds in Kazakhstan. *Arch Virol*. (2018) 163:1949–54. doi: 10.1007/s00705-018-3815-9

63. Gerilovych A, Stegniy B, Potkonjak A, Bolotin V, Solodyankin O. Genotyping of Newcastle disease virus strains, allocated in Ukraine in 1967–2007 (Genotypes 1, 2 and 4). *Contemp Agric*. (2009) 58:58–67. Available online at: http://polj.uns.ac.rs/wp-content/uploads/arhiva-savremena-poljoprivreda/2009Savremenapoljoprivreda34.pdf

64. Tirumurugaan KG, Vinupriya MK, Vijayarani K, Kumanan K. Analysis of the fusion protein cleavage site of Newcastle disease virus isolates from India reveals preliminary evidence for the existence of II, VI, and VII genotypes. *Indian J Virol*. (2011) 22:131–7. doi: 10.1007/s13337-011-0044-1

65. Gerilovych A, Potkonjak A. Molecular evolution of newcastle disease virus in Ukraine. *Contemp Agric*. (2009) 58:46–55. Available online at: http://polj.uns.ac.rs/wp-content/uploads/arhiva-savremena-poljoprivreda/2009Savremenapoljoprivreda12.pdf

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fvets.2023.1112850

OPEN ACCESS

EDITED BY
Christina Leyson,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Ali Mazloum,
Federal Center for Animal Health (FGBI ARRIAH), Russia
Antoinette Van Schalkwyk,
Agricultural Research Council of South Africa (ARC-SA), South Africa

*CORRESPONDENCE
Carmina Gallardo
✉ gallardo@inia.csic.es

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 30 November 2022
ACCEPTED 06 January 2023
PUBLISHED 25 January 2023

CITATION
Gallardo C, Casado N, Soler A, Djadjovski I, Krivko L, Madueño E, Nieto R, Perez C, Simon A, Ivanova E, Donescu D, Milicevik V, Chondrokouki E, Nurmoja I, Frant M, Feliziani F, Václavek P, Pileviciene S and Marisa A (2023) A multi gene-approach genotyping method identifies 24 genetic clusters within the genotype II-European African swine fever viruses circulating from 2007 to 2022. *Front. Vet. Sci.* 10:1112850. doi: 10.3389/fvets.2023.1112850

A multi gene-approach genotyping method identifies 24 genetic clusters within the genotype II-European African swine fever viruses circulating from 2007 to 2022

Carmina Gallardo[1]*, Nadia Casado[1], Alejandro Soler[1], Igor Djadjovski[2], Laura Krivko[3], Encarnación Madueño[1], Raquel Nieto[1], Covadonga Perez[1], Alicia Simon[1], Emiliya Ivanova[4], Daniel Donescu[5], Vesna Milicevik[6], Eleni Chondrokouki[7], Imbi Nurmoja[8], Maciej Frant[9], Francesco Feliziani[10], Petr Václavek[11], Simona Pileviciene[12] and Arias Marisa[1]

[1]European Union Reference Laboratory for ASF (EURL-ASF): Centro De investigación en Sanidad Animal (CISA-INIA, CSIC), Madrid, Spain, [2]Faculty of Veterinary Medicine, University Ss. Cyril and Methodius in Skopje, Skopje, North Macedonia, [3]Latvia NRL: Laboratory of Microbiology and Pathology, Institute of Food Safety, Animal Health and Enviroment, BIOR, Riga, Latvia, [4]Bulgaria NRL: National Diagnostic and Research Veterinary Medical Institute (NDVRI), Sofia, Bulgaria, [5]Romania NRL: Institute for Diagnostic and Animal Health, Bucharest, Romania, [6]Republic of Serbia NRL: Institute of Veterinary Medicine of Serbia, Belgrade, Serbia, [7]Greece NRL: Greek Ministry of Rural Development and Food FMD, Virological, Rickettsial & Exotic Diseases, Athens, Greece, [8]Estonian NRL: National Centre for Laboratory Research and Risk Assessment (LABRIS), Tartu, Estonia, [9]Poland NRL: National Veterinary Research Institute, Puławy, Poland, [10]Italy NRL: Istituto Zooprofilattico Sperimentale (IZS) dell'Umbria e delle Marche, Perugia, Italy, [11]Czech Republic NRL: State Veterinary Institute Jihlava, Jihlava, Czechia, [12]Lithuania NRL: National Food and Veterinary Risk Assessment Institute (NFVRAI), Vilnius, Lithuania

Introduction: African swine fever (ASF) is a contagious viral disease of pigs and wild boar that poses a major threat to the global swine industry. The genotype II African swine fever virus (ASFV) entered the European Union (EU) in 2014 and since then fourteen countries have been affected, Italy and North Macedonia being the last in 2022. While whole genome sequencing remains the gold standard for the identification of new genetic markers, sequencing of multiple loci with significant variations could be used as a rapid and cost-effective alternative to track outbreaks and study disease evolution in endemic areas.

Materials and methods: To further our understanding of the epidemiology and spread of ASFV in Europe, 382 isolates collected during 2007 to 2022 were sequenced. The study was initially performed by sequencing the central variable region (CVR), the intergenic region (IGR) between the *I73R* and *I329L* genes and the *O174L* and *K145R* genes. For further discrimination, two new PCRs were designed to amplify the IGR between the *9R* and *10R* genes of the multigene family 505 (MGF505) and the IGR between the *I329L* and *I215L* genes. The sequences obtained were compared with genotype II isolates from Europe and Asia.

Results: The combination of the results obtained by sequencing these variable regions allowed to differentiate the European II-ASFV genotypes into 24 different groups. In addition, the SNP identified in the IGR *I329L-I215L* region, not previously described, grouped the viruses from North Macedonia that caused the 2022 outbreaks with viruses from Romania, Bulgaria, Serbia and Greece, differentiating from other genotype II isolates present in Europe and Asia. Furthermore, tandem repeat sequence (TRS) within the *9R-10R* genes of the multigene family 505 (MGF505) revealed eight different variants circulating.

Discussion: These findings describe a new multi-gene approach sequencing method that can be used in routine genotyping to determine the origin of new introductions in ASF-free areas and track infection dynamics in endemic areas.

KEYWORDS

ASFV, genotyping, TRS, SNP, genetic groups

1. Introduction

African swine fever (ASF) is considered one of the most devastating disease of pigs and wild boar. The ASF virus (ASFV) is a large, enveloped virus, member of the family Asfarviridae (1). The genome of ASFV is a linear double-stranded DNA (dsDNA) molecule with a length of 171–193 kb with terminal inverted repeats and hairpin loops (2). The size differences between the different strains are due to insertions or deletions at the terminal regions of the genome where the multigene families (MGF) are located. Variations in the conserved central region related to single nucleotide polymorphism (SNP) or the presence of tandem repeat sequences (TRS) have also been described (2). ASFV isolates are classified into 24 genotypes by comparative analysis of the C-terminal end of the *B646L* gene, which encodes the p72 protein (3, 4). All 24 genotypes are present in Africa, where ASF was first described a century ago (5). Outside of Africa, genotype I was related to historical ASFVs circulating in Europe and America until the mid-1990s. This genotype has remained endemic only in Sardinia (Italy) since 1978, although its presence has been recently associated to domestic pig's outbreaks in China in 2021 (6).

In 2007, the presence of ASFV genotype II was confirmed in the Caucasus region of Georgia (7). From there, ASFV gradually spread to neighboring countries (i.e., Armenia, Azerbaijan, Russia, Ukraine, Moldova, and Belarus) affecting domestic pigs and wild boar. In the European Union (EU) the presence of ASFV genotype II was first reported in 2014 in Lithuania and Poland (8). Since then, genotype II of ASFV has been notified in Belgium, Bulgaria, the Czech Republic, Estonia, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Poland, Romania, and Slovakia, causing serious concerns. So far, only two European countries have managed to eradicate the disease: Belgium (event resolved in March 2020) and the Czech Republic (event resolved in April 2018). Furthermore, no ASF outbreaks in domestic pigs nor cases in wild boar have been reported in Greece since February 2020. The disease has also been reported in Serbia and North Macedonia, so there is a constant risk of re-introduction for European countries that are sharing borders (9). In August 2018, ASFV genotype II was detected in China (People's Republic of), marking the first occurrence of ASF in Asia (10). As of the end of October 2022, ASF has been reported in 32 provinces in China and 16 Asian countries; the last Thailand in January 2022. In September 2019, the ASFV appears in Oceania, in Timor-Leste, followed by Papua New Guinea (March 2020). In July 2021, another transcontinental leap in ASF occurs with the reappearance of genotype II in the Americas after an absence of almost 40 years, with outbreaks detected in the Dominican Republic and Haiti (11).

The key for understanding the diversity of the ASFV, including its evolution, is to analyze its genetic variations by sequencing specific genetic markers. The carboxy terminal end of the p72 gene (*B646L*) is sequenced to place ASFVs within one of 24 genotypes (3, 4, 12). This method allows relatively quick and easy typing of ASFV strains and remains the first method to identify the origin of an outbreak in case of introduction into new territories. However, the *B646L* gene-based genotyping method does not always provide adequate typing resolution or the ability to discriminate between closely related viruses. For intra-genotypic differentiation, the central variable region (CVR) of the *B602L* gene, is one of the most widely used markers. It is characterized by the presence of tandemly repeated sequences (TRS) which allows up to 31 subgroups of ASFV to be distinguished (13). However, despite the large number of ASFV genotype II outbreaks in the EU, only three CVR variants have been identified in Estonia, with variant 1 (Georgia 2007-type) being predominant throughout the EU (14). The analysis of additional genetic markers such as TRS present in the *O174L* gene (15, 16), or in the intergenic regions (IGR) between *I73R-I329L* (8), has made possible to differentiate between closely related genotype II viruses (17–19). Similarly, sequencing of single nucleotide polymorphisms (SNPs) within the *K145R* gene has also been described as a useful molecular tool to track the spread of ASFV in Poland (16).

It was hypothesized that these regions could be used as a rapid and cost-effective method to investigate the epidemiology, evolution, and molecular relationship of a large number of EU strains. The aim of this study was therefore to open the spectrum of characterized viruses and perform a molecular genotyping of 382 ASFV isolates from Europe, collected during 2007 to 2022. The study was performed with the initial analysis of the TRS located in the CVR, in the IGR between *I73R–I329L* and in the *O174L* gene, and by the sequencing of the SNP identified in the *K145R* gene. For additional discrimination, two new PCRs were designed to amplify new markers characterized by the presence of TRS, the IGR between the *9R−10R* genes of the multigene family 505 (MGF505) (20, 21), and the IGR between the *I329L* and *I215L* genes (22).

This is the first detailed report on the molecular characterization of the ASFV strains circulating in all affected EU countries since 2014 to 2022. Moreover, the multi-gene approach strategy followed in this study distinguish 24 genetic groups and revealed previously undescribed variants circulating in the EU that allowed us, for the first time, to trace variants with genomic epidemiology to regional clusters.

2. Materials and methods

2.1. Samples selection

A total of 345 clinical samples collected in the EU from ASFV-positive wild boar ($n = 244$) and domestic pigs ($n = 101$) between 2014 and 2022 sent by the EU National Reference Laboratories

(NRLs) to the EU Reference laboratory (EURL) for ASF (CISA, INIA-CSIC, Madrid Spain) for ASF laboratory confirmation were included in this study. Specimens were selected to represent all geographic areas from affected EU countries, including index cases, where ASFV has been present up to now. Thirty seven ASFVs from the neighboring non-EU Serbia, North Macedonia, Moldova, Ukraine, Belarus, Russia Federation, Armenia and Georgia were also included resulting in a final panel of 382 ASFVs characterized. The detailed characteristics of ASFVs, as well as obtained results for individual viruses, are presented in Supplementary Table S1.

2.2. PCR amplification for routine ASF diagnostic

To confirm the presence of the ASFV genome in samples received, the DNA was extracted from clinical samples (sera, whole blood, tissues), using the High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH, Roche Applied Science, Mannheim, Germany). Briefly, 10% (w/v) clarified homogenized tissue suspensions or blood were prepared in phosphate-buffered saline (PBS). For ASF routine diagnosis, the Universal Probe Library (UPL) real-time PCR (23, 24) was performed using undiluted extracted DNA from each sample. Extracted DNA was stored at −20°C until further analysis.

2.3. PCR amplification for sequencing

For genetic characterization, PCR was performed on nucleic acid extracted from ASFV positive samples using published primers and protocols to initially amplify four independent regions on ASFV genome; (i) the CVR of the *B602L* gene using the CVR1 and CVR2 primer pair (26), (ii) TRS located between the *I73R* and *I329L* (IGR) genes using primers ECO1A and ECO1B (8), (iii) TRS located in the *O174L* gene using the *O174L*F and *O174L*F (15, 16), and (iv) partial *K145R* gene using primers *K145R*-F and *K145R*-R (16). Amplification conditions were previously described (8, 15, 17, 26).

We designed two additional sets of primers to amplify; (i) a 551 base pair (bp) amplicon that includes the TRS located in the IGR between the MGF505 *9R* and *10R* genes (Figure 3A) using primers MGF505U and MGF505L (20), and (ii) a 604 bp fragment containing the 55 bp at the 3′ end of the *I329L*, 288 bp of the IGR between the *I329L* and *I215L* genes, and 261 bp at the 5′ end of the *I215L* gene using the primers named ECO2A and ECO2B. The primer binding sites were based on the Georgia ASFV genome (Accession No. FR682468.2). Conditions for the PCR assays were as follows; 10–50 ng of sample DNA, 1x PCR buffer II (50 mM KCl, 10 mM Tris-HCl), 2.5 mM $MgCl_2$, 0.2 mM concentrations of the four deoxynucleoside triphosphates (Roche Molecular Biochemicals), 0.4 μM concentrations of the primers and 0.025 U/μl of Taq Gold polymerase (Applied Biosystems). The amplification programs were identical to that used for the CVR amplification (25) but with annealing temperatures of 56°C for the IGR MGF505 *9R-10R* and 55°C for the IGR *I329L-I215L* amplification. Primers specific to selected regions were designed by Primer3 (http://primer3.ut.ee). Detailed characteristics of the primers used and regions amplified are summarized in Table 1.

2.4. Sequence analysis

Amplicons of predicted size were excised and purified by Quiaex gel extraction (QUIAGEN) and the nucleotide sequence of purified products determined using the same primers as used for amplification on an automated 3730 DNA analyzer (Applied Biosystems).

Sequence quality assessment was done using Chromas (www.technelysium.com.au). The individual forward and reverse sequences were assembled using the CLUSTALW algorithm implemented in MEGA v11 software (26). The generated nucleotide sequences were compared to publicly available sequences using the Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignment was done using the CLUSTALW algorithm implemented in MEGA v11 software. For the TRS analyses nucleotide sequence or deduced amino acid sequences were manually aligned with gaps being inserted to optimize the alignment. Comparisons were made using previous TRS as described (8, 13–16). The sequences were compared with ASFV genotype II homologous sequences from Europe and Asia available in the GenBank, giving final data sets of 540 sequences for CVR, 910 for IGR, 636 for *O174L*, 540 for *K145R*, and 442 for the MGF and ECO2 regions.

Sequences generated within this study were deposited in GenBank (Supplementary Table S1).

3. Results

3.1. Analysis of the CVR within the *B602L* gene (CVR region)

For CVR amplification of the *B602L* gene, 338 samples from ASF cases in wild boar and pig outbreaks collected between 2014 and 2022 in the EU were sequenced. CVR sequences were initially compared with data obtained in previous studies at the EURL of 32 ASFVs (8, 13). Three hundred and seventy ($n = 370$) ASFVs sequenced in this study showed 100% sequence identity to the Georgia 2007/1 strain, thus presenting the CVR region genotype II-variant I (CVR1) with 10 amino acid TRS (BNDBNDBNAA). Sequence alignment revealed the presence in two ASFVs from Poland and Lithuania of two different non-synonymous SNPs resulting in amino acid changes. The identified SNPs differed from the CVR1/SNP1 variant previously described in Estonia (14). Sample Pol17/WB/CASE316 from Poland contained a nonsynonymous (A/T) SNP (CVR1/SNP2) at nucleotide (nt) position 46 of the amplified CVR that is at position 514 of the complete *B602L* protein, resulting in an exchange of methionine (M) for leucine (L) at amino acid position 16 of the CVR (172 of the *B602L* protein). In ASFV from Lithuania (Lt17/WB/Kupiskis/18) a different SNP (CVR1/SNP3) was identified, where thymine (T) was replaced with cytosine (C) at nt position 541 of the complete gene. This transition also resulted in an amino acid change at CVR position 25 that is at position 181 in the complete *B602L* protein, where cysteine (C) was replaced by arginine (R; Figure 1). These variants

10.3389/fvets.2023.1112850

TABLE 1 Characteristics of the primers used in the study.

ID	Name	Sequence 5′ → 3′	Position referring to Georgia 2007/1 (FR682468.2)	Amplicon length (nt)	Reference
CVR	CVR1	ACTTTGAAACAGGAAACWAATGATG	102,943–102,968	491	(25)
	CVR2	ATATTTTGTAATATGTGGGCTGCTG	102,520–102,524		
IGR	Eco1A	CTATTTATCCCCCRCTTTGG	173,272–173292	356	(8)
	Eco1B	TCGTCATCCTGAGACAGCAG	173,607–173,627		
O174L	*O174L*-F	TGGCTCAGACGATATTTCAACTC	128,160–128,182	63	(15)
	O174L-R	GCCTCCACCACTTGAACCAT	128,832–128,813		
K145R	*K145R*-F	TTTCAGGCTGAAAACTTTTTAT	65,030–65,051	22	(16)
	K145R-R	AAAGTTTTCAATGGTTGTTAGC	65,312–65,291		
MGF	505U	AGAAACCGCAGATGAATGTA	45,069–45,089	51	This study
	505L	TACAGCCCTAGTTGTTGAAG	45,567–45,587		
ECO2	Eco2A	TCCTACCTGTTAAGCCACTTCC	174,452–174,472	604	This study
	Eco2B	GCAAATGTGGATGCAGCTAA	175,035–175,055		

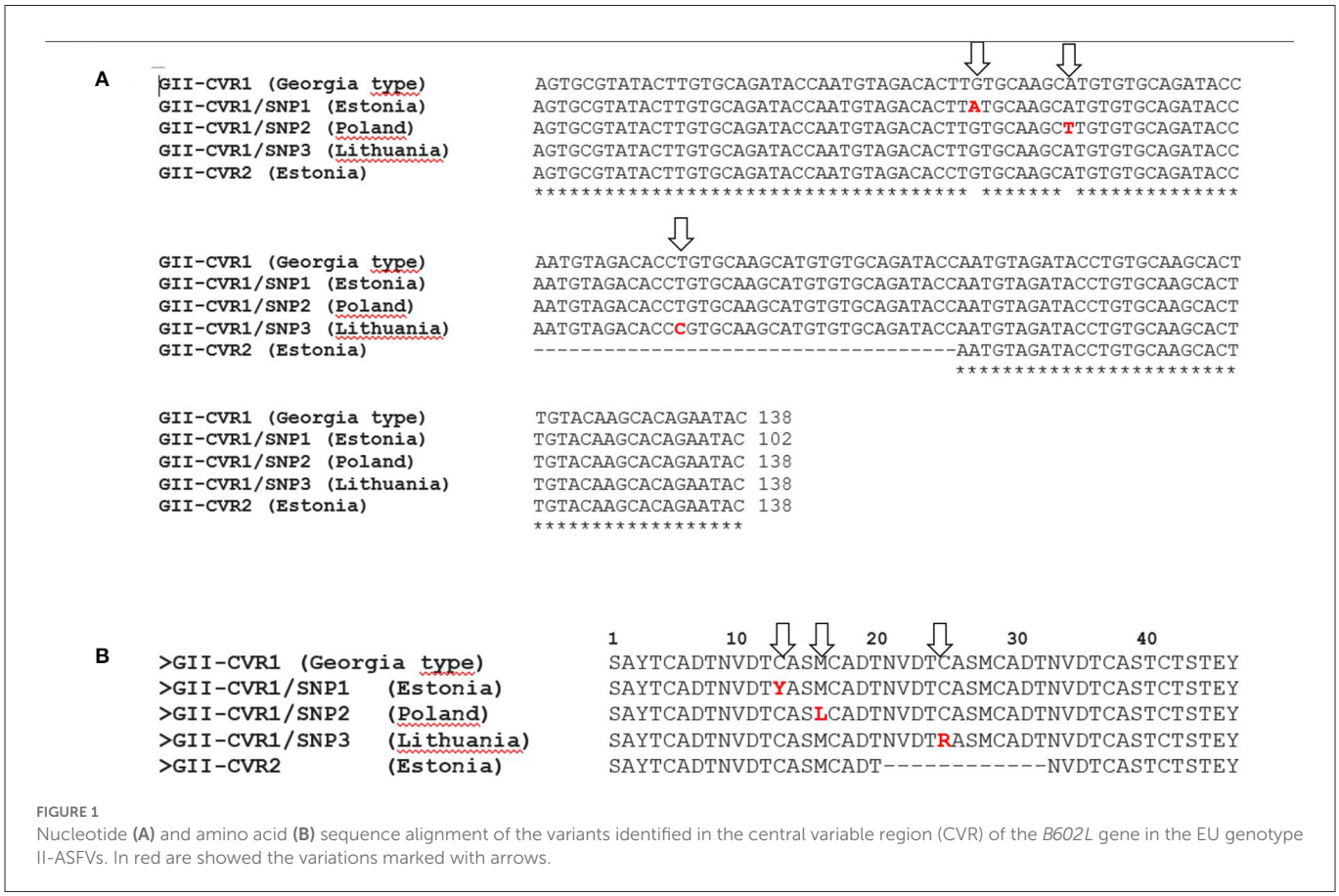


FIGURE 1
Nucleotide **(A)** and amino acid **(B)** sequence alignment of the variants identified in the central variable region (CVR) of the *B602L* gene in the EU genotype II-ASFVs. In red are showed the variations marked with arrows.

had not been previously identified in Europe or Asia, as demonstrated by comparison with 495 genotype II-ASFV homologous sequences available in the GenBank.

The details of the obtained results are included in the Supplementary Table S1 and the geographical location of the CVR variants identified in this study in Figure 2A.

3.2. Analysis of the TRS of the IGR *I73R/I329L* (IGR region)

For IGR *I73R/I329L* classification, 367 samples from ASF wild boar cases and pigs outbreaks collected between 2012 and 2022 were sequenced within this study. Data were supplemented

10.3389/fvets.2023.1112850

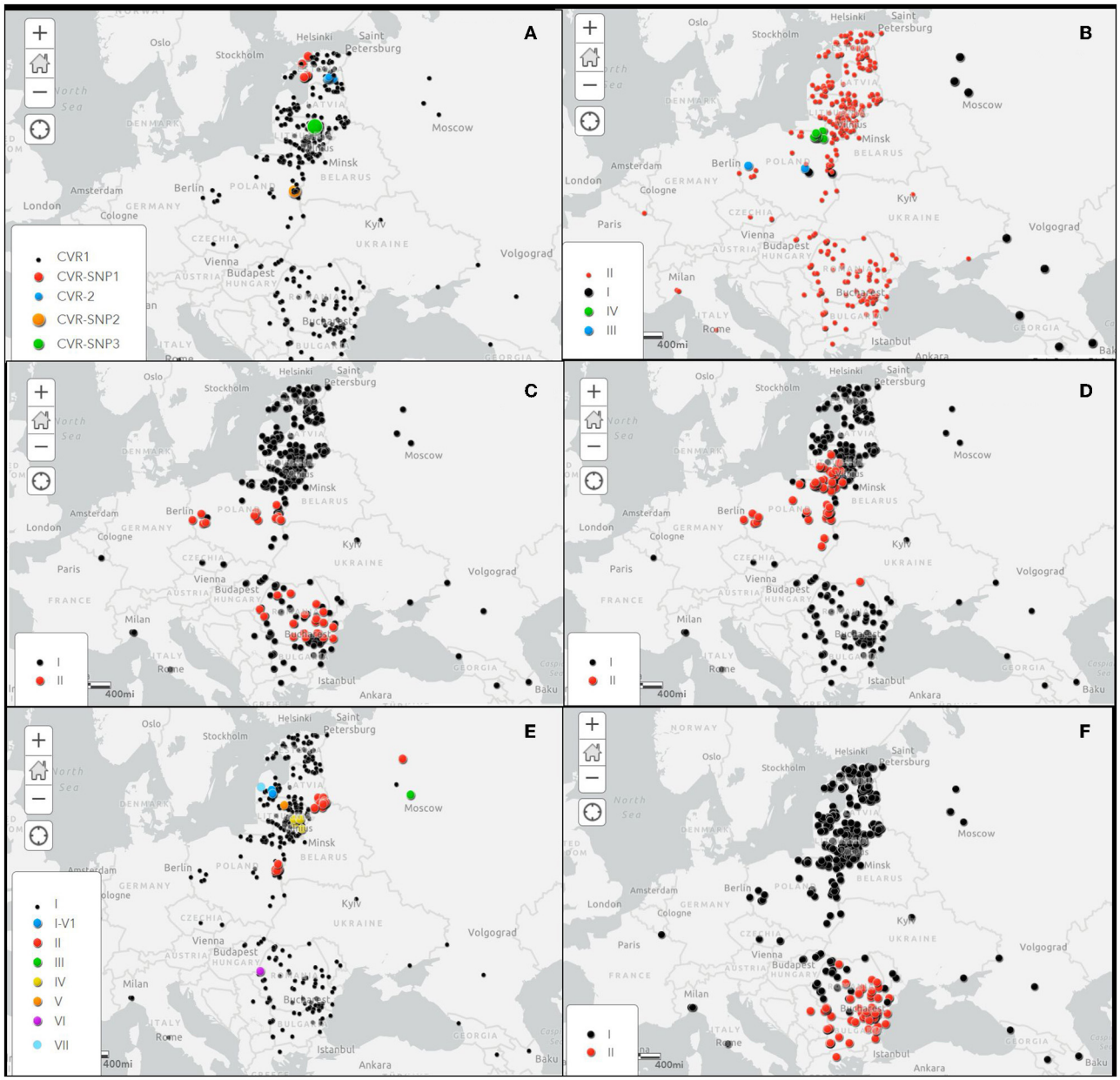


FIGURE 2
Spatial distribution of investigated gene variants of ASFVs in Europe during the period 2007–2022. **(A)** Central variable region (CVR) within the *B602L* gene, **(B)** Intergenic region (IGR) between *I73R*/*I329L* genes, **(C)** *O174L* gene, **(D)** *K145R* gene, **(E)** IGR between the *9R*/*10R* multigene family 505 (MGF505), and **(F)** ECO2 variants in the IGR between the *I329L*/*I215L* and partial *I215L* gene. Black dots shows the variant 1 100% homologous to the Georgia2007/1 reference strain.

with IGR sequences of 14 genotype II ASFVs from a previous study (8).

As expected, the IGR-II variant was the most frequent (352/367, 95.91% strains), followed by IGR-IV (10/367 strains, 2.72%), IGR-III (2/367 strains, 0.54%) and IGR-I (strain 2/367, 0.54%). While the IGR-II variant was present in all sampled regions, variants I, III and IV were only detected in Poland (Figure 2A). The IGR-I variant was identified in a sample taken from a wild boar at a distance of about 20 km from the border with Belarus, in Lublin voivodship in March 2017. The same variant was identified in a wild boar hunted in Masovian voivodship in November of the same year, about 160 km from the first detection of the IGR-I variant. The IGR-III variant was linked to an outbreak in domestic pigs in May 2017 in Lublin voivodship, western Poland, and to a wild boar case in Masovian voivodship, east-central Poland in January 2018. The IGR-IV variant was detected in 10 samples collected from wild boar cases in 2018 and 2019 in the eastern Warmian-Masurian voivodship, located in the north-eastern part of Poland, adjacent to Kaliningrad Oblast, Russian Federation. The details of the obtained results are included in the Supplementary Table S1 and the geographical location of the IGR variants in Figure 2B.

In order to improve the picture of ASFV epidemiology in time, the IGR sequences were compared with 540 genotype II ASFV homologous sequences available in the GenBank, resulting in a

final data set of 910 IGR sequences. As shown in Table 2, IGR-II was the predominant (92.6%) present in all affected countries of Europe and Asia. The IGR-I variant (frequency 3.8%) has been detected in China, the Russian Federation, Poland, South Korea, and Vietnam, while the IGR-III (frequency 1.6%) in China, South Korea and Vietnam. The IGR-IV variant has only been detected in Poland.

3.3. Analysis of the TRS of the *O174L* gene (*O174L* region)

In this study, 382 samples from ASF cases in wild boar and outbreaks in pigs from 2007 to 2022 were used. Most of the samples analyzed (351/382, 91.88%) showed 100% sequence identity with the Georgia strain 2007/1, thus presenting variant I of the *O174L* gene. A 14 nt insertion of CAGTAGTGATTTTT representing variant II of the *O174L* gene (15) was observed in 31 samples (8.11%), including 19 from Romania, 11 from Poland and one from Germany. Consistent with previous studies (15, 16), *O174L* variant II was identified in all samples taken in Lubusz and Masovian voivodeships in central and western Poland, and in four of the six from Lublin, located in the southeast of the country. In contrast, variant II was only identified in one of the 17 samples submitted from Podlaskie, in the northeast of Poland, whilst no variant II samples were observed in the north of Poland. In Romania, *O174L* variant II was detected for the first time in January 2019 in domestic pigs in western Romania, subsequently causing outbreaks in the rest of the country. All viruses tested during this year, except two of wild boar, were grouped within this variant. In contrast, the 2017, 2018 and 2021 Romanian viruses belong to variant I. Except in Germany (27) the *O174L* variant II has not been described in other European and Asian countries as it was determined by comparing 254 homologous sequences available at the GenBank.

The details of these results are included in the Supplementary Table S1 and the geographical location of the *O174L* variants in Figure 2C.

3.4. Analysis of the SNP of the *K145R* Gene (*K145R* region)

Conventional sequencing of the *K145R* gene was performed using the same panel of samples as that used for sequencing of the *O174L* gene. The vast majority of the investigated samples (320 of 382, 83.7%) showed 100% identity to the reference strain Georgia 2007/1 representing KP145R variant I. Variant II, characterized by the presence of one SNP (transversion C65167A, referring to Georgia 2007/1) (17) was found in Poland, Lithuania and Romania and in Germany.

In Poland, the *K145R*-II variant was dominant (40/47, 85.1%), while the *K145R*-I variant was identified in only seven ASFVs, two from 2014 index cases and five from wild boar hunted in 2018. It is interesting to note that, except for one hunted wild boar in central Poland, the remaining *K145R*-I variants were collected from areas close to the borders with Belarus, Russian Federation (Kaliningrad Oblast) and Ukraine. In Lithuania, 82 of the 102 ASFVs (80.39%) belonged to variant *K145R*-I and 20 (19.6%) to variant *K145R*-II with strains from Poland and Germany. The first identification of variant II in Lithuania dates from July 2017, when it was detected in domestic pigs on a farm located in Alytus County, on the border with Belarus. Since then, the *K145R*-II variant has been circulating in the wild boar population, mainly in the south of Lithuania, with a sporadic presence in the north of the country. The last cluster of variant II was detected near the border with Kaliningrad Oblast in late 2021-early 2022. Only one of the 42 viruses sequenced from Romania belonged to the *K145R*-II variant group. The sample was collected in 2019 from a wild boar found dead about 6 km of Ukraine border in the Botosani County, in the northern part of Romania. The geographical distribution of the *K145R* variants is showed in Figure 2D.

These sequences were compared to 158 sequences available from GenBank, representing additional genotype II ASFVs obtained from Europe and Asia. The *K145R* variant II was identified in Ukraine in 2016 in Kiev (GenBank Accession No. MN194591) and in 2018 in the Kaliningrad Oblast in seven wild boar ASFVs (GenBank Accession No. OM966714–OM966718, OM966720, OM96672, and OM799941).

3.5. Analysis of the TRS of the IGR between the MGF505 *9R* and *10R* genes (MGF region)

Amplicons ranging from around 530 to 590 bp were obtained from the 382 genotype II-ASFV isolates sequenced in this study. The molecular basis of this variation involved alterations in the number and type of TRS identified between ORFs *9R*/*10R* (Figure 3B). Two sets of serially repeated DNA sequences could be seen. The first one at residues 45,217–45,302 of the Georgia 2007/1 strain, proximal to the *9R* gene, consisted of five units of 17 nts (AGTAGTTCAGTTAAGAT) with the structure ABBCD. The second set of repeats at residues 45,365–45,467 contains six repetitions of a 17 nts repeat sequence (AGTTCATTTAAGTCAAT) with the structure EFGHHH. The conserved core sequences of the TRS varies in one or two nts (Figure 3B).

The 382 ASFV sequenced were divided into eight different groups based on the number and type of TRS found (Table 3). The largest group, called the MGF-1 group, was 100% homologous to the Georgia 2007/1 reference strain and included 341 of the 382 ASFVs (89.26%) from all sampled countries. The PCR products amplified in this group generated a 551 bp amplicon characterized by the presence of 11 TRS type ABBCD-EFGHHH. The isolates from group 2 (MGF-2) presented a larger amplicon (569 bp) due to the presence of an additional TRS (type ABBBCD-EFGHHH). Twenty-six (6.80%) ASFVs from Russia, Poland, and Latvia were grouped into this group. The MGF-2 variant was initially identified in Russian Federation, in 2012 in a sample taken from an outbreak of domestic pigs. In November 2016, the same variant was found in a wild boar in northeastern Poland, 10 km from the Belarusian border. This was the variant responsible for all but one wild boar cases and domestic pig outbreaks in the same region until June 2017. No further MGF-2 variants has been detected in Poland. In Latvia, the MGF-2 variant was detected for the first time in wild boar in July 2017 in the easternmost region of Latvia that borders Belarus. All wild boar cases that have occurred in this region since then (2017–2021) have been caused by the MGF-2 variant. Interestingly, the only domestic pig

10.3389/fvets.2023.1112850

TABLE 2 IGR variants identified in 910 genotype II ASFVs from Europe and Asia.

Region	Country	Year	No. sequences		No. IGR variants			
			GenBank	This study	IGR_I	IGR-II	IGR-III	IGR-IV
Europe	Georgia	2007–2008	3	–	3	–	–	–
	Armenia	2007	1	–	1	–	–	–
	Azerbaijan	2008	2	–	2	–	–	–
	Russia Federation	2009–2019	38	1	24	14	–	–
	Ukraine	2012- 2016	4	2	–	4	–	–
	Belarus	2013	1	–	–	1	–	–
	Lithuania	2014–2022	102	100	–	102	–	–
	Poland	2014–2020	201	46	2	180	2	17
	Latvia	2014–2021	41	41	–	41	–	–
	Estonia	2014–2022	72	71	–	72	–	–
	Moldova	2016–2018	4	3	–	4	–	–
	Czech Republic	2017–2018	3	2	–	3	–	–
	Romania	2017–2021	46	46	–	46	–	–
	Hungary	2018–2019	5	4	–	5	–	–
	Bulgaria	2018–2020	23	23	–	23	–	–
	Belgium	2018	2	1	–	2	–	–
	Slovakia	2019	1	1	–	1	–	–
	Serbia	2019–2020	14	14	–	14	–	–
	Germany	2020	2	1	–	1	–	–
	Greece	2020	1	1	–	1	–	–
	North Macedonia	2022	6	6	–	6	–	–
	Italy	2022	5	4	–	5	–	–
Asia	China	2018–2020	84	–	1	75	8	–
	Vietnam	2018–2021	174	–	1	170	4	–
	Indonesia	2019–2020	2	–	–	2	–	–
	Mongolia	2019	5	–	–	5	–	–
	Timor-Leste	2019	1	–	–	1	–	–
	South Korea	2019–2020	58	–	1	56	1	–
	India	2020	4	–	–	4	–	–
	Malasya	2021	4	–	–	4	–	–
	Philippines	2021	1	–	–	1	–	–
Total number			**910**	**367**	**35**	**843**	**15**	**17**
Frequency					**3.8%**	**92.6%**	**1.6%**	**1.9%**

outbreak analyzed in this study that occurred in June 2018 in the same region was caused by the MFG-1 variant rather than variant MGF-2. The MGF group's 3, 5, 6, and 7 were represented by a single ASFV from Russia Federation, Lithuania, Romania and Latvia, respectively. The MGF group 4 contained five ASFVs from Lithuania, four from wild boar and one from domestic pig, all of them taken in the eastern counties of Vilna and Utena, border with Belarus. The number of TRS in these groups varied between 10 TRS of groups 4 and 6–13 TRS found in groups 3 and 7 (Table 3).

Group 8, with three Latvian ASFVs, showed a three nucleotide deletion at nucleotide position 477 compared to the MGF1 variant (data not shown). This variant, named MGF-1V, was initially detected in November 2017 in a wild boar ASFV in the Brocenu region in the south of the country. Interestingly, the

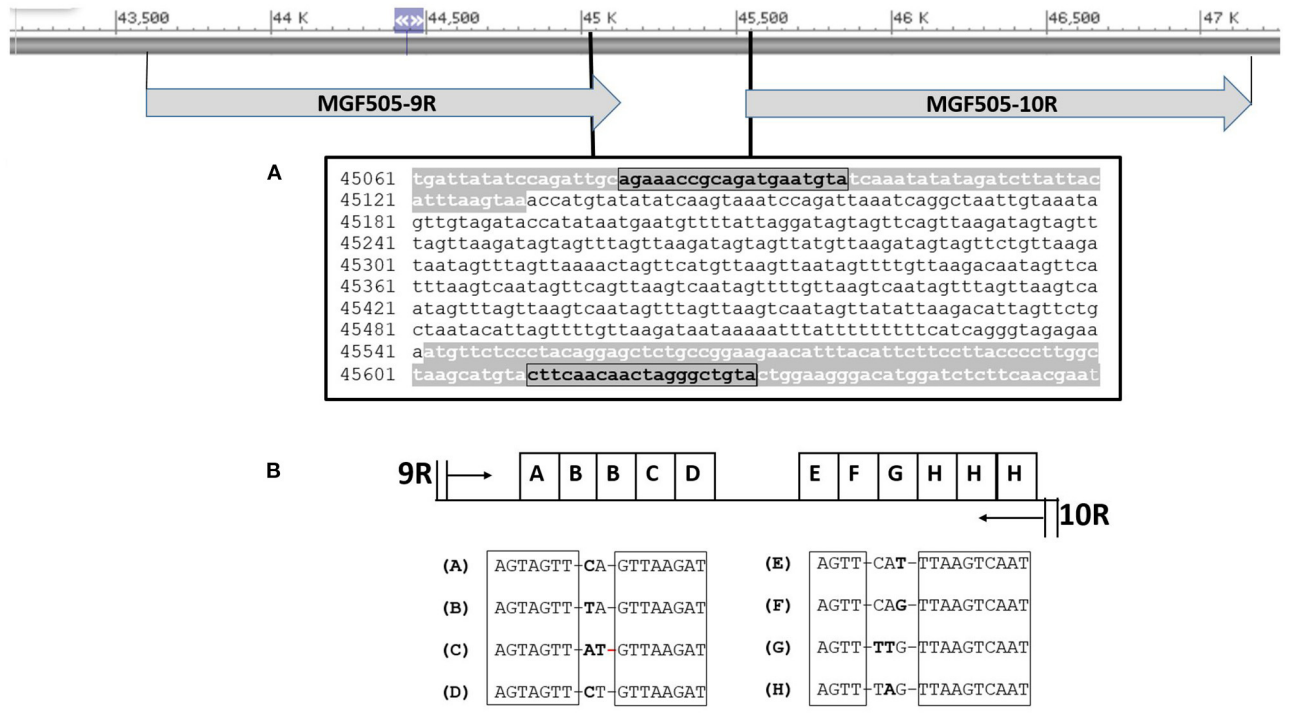


FIGURE 3
Map showing the location of the IGR between the *9R-10R* genes of MGF505 in reference to Georgia ASFV 2007/1. **(A)** Nucleotide sequence of the product amplified with the 505U/505L primers marked in bold. In gray the sequence of the *9R* and *10R* genes is shown. **(B)** Arrangement and identification of intergenic tandem repeat sequences (TRS). The conserved nucleotides in all TRS are boxed.

TABLE 3 Groups based on the analysis of the TRS in the IGR between the *9R/10R* genes of the multigene family (MGF) 505. Dashes indicate gaps introduced to enable similarities between sequences to be more easily visualized.

MGF group	Geographical distribution	No ASFVs (frequency)	Tandem repeat sequences	No repeats	Amplicon size (bp)
MGF-1	Armenia, Azerbaijan, Belarus, Belgium, Bulgaria, Czech Republic, Estonia, Georgia, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Moldova, North Macedonia, Poland, Romania, Russia Federation, Serbia, Slovakia, and Ukraine	341/382 (89.26%)	A-BB-CD__EFGHHH	11	551
MGF.2	Latvia, Poland, and Russia Federation	26/382 (6.80%)	A-BBBCD__EFGHHH	12	569
MGF-3	Russia Federation	1/382 (0.26%)	AABBBCD__EFGHHH	13	586
MGF-4	Lithuania	5/382 (1.30%)	A-BB—CD__EFGHH	10	535
MGF-5	Lithuania	1/382 (0.26%)	A-BB—CD__EFGHHHH	12	567
MGF-6	Romania	1/382 (0.26%)	A-B—-CD__EFGHHH	10	535
MGF-7	Latvia	1/382 (0.26%)	A-BBCBCD__EFGHHH	13	586
MGF-8	Latvia	3/382 (3.65%)	A-BB-CD__EFGHHH	11	549

Dashes indicate gaps introduced to enable similarities between sequences to be more easily visualized.

same variant was responsible for the outbreak that occurred on July 10, 2018 in a backyard farm located about 20 km from the initial case.

The details of these results are included in the Supplementary Table S1 and the geographical location of the MGF variants in Figure 2E.

3.6. Analysis of the IGR between the *I329L* and *I215L* genes and partial sequencing of the *I215L* gene (ECO2 region)

Amplicons of around 600 bp were obtained in the 382 ASFVs that were directly sequenced. Sequences were compared to 60 genotype II ASFVs from Europe and Asia retrieved from the GenBank, giving a final data set of 442 ASFVs. Based on the SNPs found in the *I215L* gene, four variants were identified. The vast majority of ASFVs from Europe and Asia (368/442, 83.25%) were 100% similar to the reference strain Georgia 2007/1 (ECO-I variant). However, in 67/442 (15.1%) of the ASFVs, the presence of a non-synonymous (C/T) SNP (ECO-II variant) was identified at nucleotide position 412 of the amplified fragment, that is at nucleotide 62 of the *I215L* gene (Figure 4A). The SNP resulted in an exchange of glutamic acid (E) for glycine (G) at amino acid position 192 of the *I215L* gene (Figure 4B). The ECO2-II variant was specific to ASFVs from Europe and was initially identified in the ASFVs responsible for the June 2018 domestic pig outbreak in Tulcea County, eastern Romania, on the border with Ukraine. Since then, the ECO2-I and ECO2-II variants circulate simultaneously in Romania, with variant I dominating in the west, while variant 2 predominates in the east, except for a single reported outbreak in Tulcea in 2019 (Supplementary Table S1). All ASFVs from Bulgaria (2018–2020), Serbia (2019–2020), Greece (2020) and North Macedonia (2022) showed a homogeneous nucleotide pattern, with the ECO2-II variant being the only one in circulation. Finally, ECO2-III (SNPA498G) and ECO2-IV (SNPG466A) were identified in two and one ASFVs from China, respectively. The SNP found in the variant 4 resulted in an amino acid exchange of alanine (A) for valine (V) in *I215L* (Figure 4B).

The details of the obtained results are included in the Supplementary Table S1 and the geographical location of the ECO2 variants in Figure 2F.

3.7. Genetic group classification of ASFV

Based on the concatenated nucleotide sequences of CVR, IGR *I73R*/*I329L*, *O174L*, *K145R*, IGR MGF505*9R*/*10R*, and IGR*I329L*-*I215L*, twenty four genetic groups may be distinguished in Europe (Table 4). The Georgia2007/1 type group, group 1 (CVR-1, IGR-I, *O174L*-I, *K145R*-I, MGF-I, and ECO-I), contains eight non-EU ASFVs (2.1%), collected from 2007 to 2012 in Georgia, Armenia, Azerbaijan, and the Russian Federation. The largest group was group 3 (CVR-1, IGR-II, *O174L*-I, *K145R*-I, MGF-I, and ECO-I) with 192 of the 382 (50.3%) ASFVs collected since 2012 to 2022 and originating from Ukraine, Belarus, Lithuania, Poland, Latvia, Estonia, the Czech Republic, Romania, Moldova, Hungary, Belgium, Slovakia, and the recent isolates from Italy in 2022. The second most frequent cluster was group 19 (CVR-1, IGR-II, *O174L*-I, *K145R*-I, MGF-I, and ECO-II) with 55 ASFVs (14.4%) obtained since 2018 to 2022 from Romania, Bulgaria, Serbia, Greece, and North Macedonia. Group 7 (CVR-1, IGR-II, *O174L*-I, *K145R*-II, MGF-I, and ECO-II) with 29 viruses (7.59%) was identified in ASFVs from Romania, Poland and Lithuania. With the exception of group 6 containing eight ASFVs from Poland and Germany, the remaining clusters were unique to a single country.

Per country, the greatest variability was found in the ASFVs from Poland distributed in nine different groups (3, 6, 7, 8, 10, 11, 13, 18, and 20), followed by six groups from Romania (3, 7, 19, 21, 22, and 24) and five groups from Lithuania (3, 7, 14, 15, and 16). In Estonia and Latvia 3 and 4 groups have been respectively identified, with group 3 being common to both countries, while groups 5, 9 (Estonia) and 12, 17, and 23 (Latvia) were exclusive to each country. Except in the Russian Federation with three identified groups (1, 2, and 4), in the remaining countries, circulating ASFVs were grouped into a single group. Regarding the host, most of the groups (11/24) included viruses obtained from both domestic pigs and wild boar (11/24), 10 were specific to wild boar and only three from domestic pigs.

The details of the obtained results are included in the Supplementary Table S1 and Table 4.

4. Discussion

One of the main gaps in the epidemiology of the ASFV is to differentiate the origin of the outbreaks or to trace the dynamics of the infection in the affected populations once a certain genotype is circulating in a country or region. This is mainly due to the large size of the ASFV genome that ranges from 170 to 193 kbp, which makes sequencing the entire genome difficult, and its low mutation rate as it is a DNA virus (2). In this study, we have investigated six variable regions of the ASFV genome and examined their use in distinguishing between closely related genotype II ASFVs circulating in Europe since 2007 up to 2022. Two approaches were utilized to study genome variability. First, the sequencing of four previously described regions of the genome that contain TRS or SNPs; the CVR (13), the IGR between the *I73R*-*I329L* genes (8), the *O174L* gene (15) and the partial sequencing of the *K145R* gene (16). Second, we designed two new PCRs to amplify two additional ASFV variable regions characterized by the presence of TRS located in the IGR between the *9R*−*10R* genes of MGF 505 (20, 21) and in the IGR of the *I329L*–*I215L* genes (22). The aim was to obtain additional markers capable of differentiating between the viruses circulating in Europe, especially those associated with the recent outbreaks that occurred in Italy and North Macedonia in 2022.

The CVR region is frequently the subject of sequence analysis due to single mutations capable of resolving phylogenies at the regional level (13, 28–34). Within genotype II-ASFVs, and despite the large number of outbreaks in Europe and Asia, CVR variants have only been described in Estonia, classifying the isolates into three different groups; CVR1 (Georgia2007/1 type), CVR2 characterized by an amino acid deletion of CASMCADTNVDT, and CVR1-SNP1 having a nucleotide mutation (14). Even though the resolving power of this gene region in isolates from Africa (28–34) and the polymorphisms observed in isolates from Estonia (14), only two SNP-based CVR variants were detected in this study in wild boar viruses from Poland (CVR1-SNP2) and Lithuania (CVR1-SNP3). Recent studies describe the presence of additional CVR variants in isolates from Russia (35) and China (19) based on the presence of SNPs. However, the variants described in this study were unique and were not further identified in any of the ASFVs sequenced or available in the GenBank.

The IGR *I73R*-*I329L* is characterized by the presence of 10-nucleotide tandem repeats of "TATATAGGAA" (8) that allow genotype II-ASFVs to be classified into one of the four variants

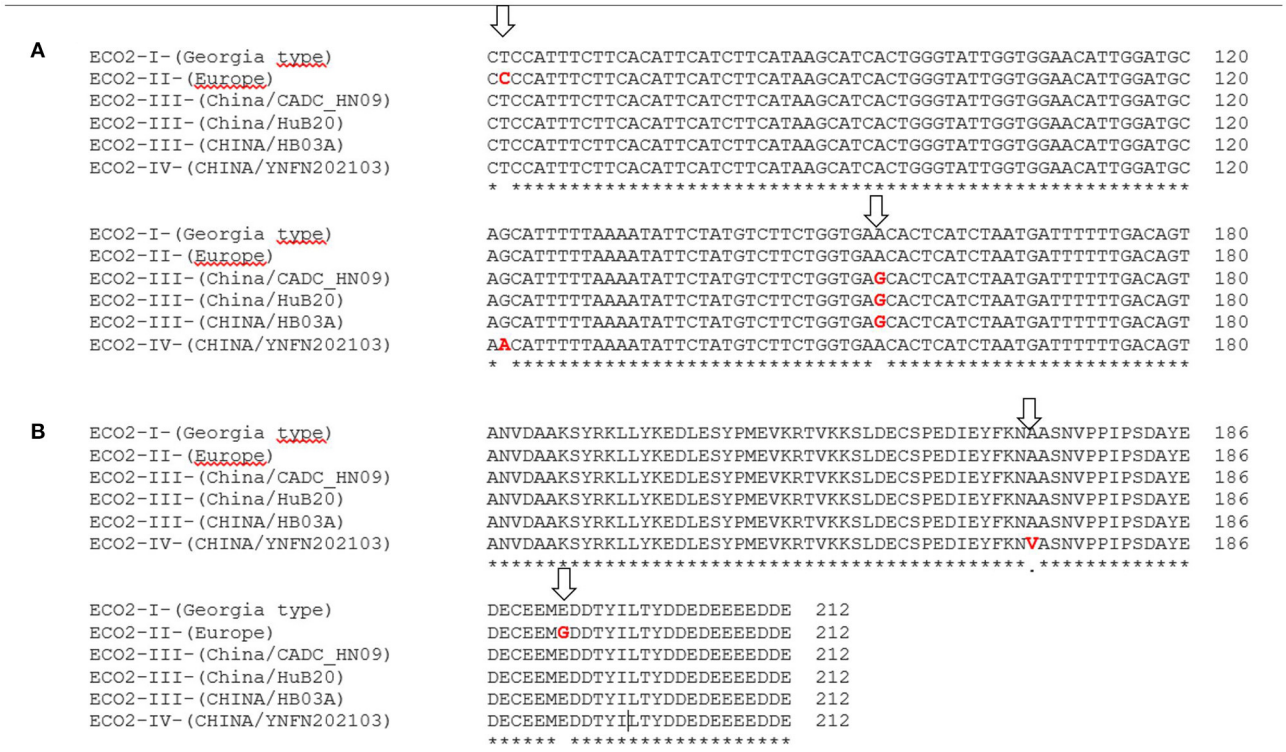


FIGURE 4
(A) Partial nucleotide sequence alignment and **(B)** amino acid alignment of the SNPs identified in the *I215L* gene amplified using the primers ECO2A/2B. In red are showed the variations marked with arrows.

identified so far; IGR-I (two copies), IGR-II (three copies), IGR-III (four copies) and IGR-IV (five copies) (18). According to the results of this study, the IGR-II variant, which emerged in 2012 or even earlier in the Russian Federation (8, 36), is the most frequently identified in the EU-ASFVs. This variant was responsible for the first cases in 2014 in Lithuania and Poland on the border with Belarus and has spread widely throughout all affected regions (8). The four IGR variants were only detected in Poland, being the first description in the EU for the IGR-I and IGR-III variants. Outside the EU these variants were circulating in China, Russia Federation, South Korea and Vietnam (17–19, 37–39). Consistent with what was described by Mazur-Panasiuk et al. (16), IGR-IV was exclusively identified in northern Poland on the border with Kaliningrad Oblast, Russia.

Previous sequencing studies of the *O174L* gene classified the isolates from Poland into two different variants, with variant II being the predominant one in the western part of the country and responsible for the ASF outbreaks in Germany (15, 16, 27). In our study, sequencing of the *O174L* gene revealed the presence of the *O174L*-II variant in Poland, as expected, but also in Romania, both in wild boar and domestic pigs. Similarly, the *K145R* variant II, described in Poland, Germany and Russia Federation (Kaliningrad Oblast) (15, 16), was found in Poland, Lithuania, and Romania.

The results obtained after the analysis of the four published variable regions did not allow to identify the origin of the outbreaks that occurred in 2022 in Italy and North Macedonia. For this, the IGR between MGF 505 *9R* and *10R* (MGF 505 *9R*/*10R*) characterized by the presence of TRS (20, 21) was sequenced. Eight MGF variants were identified based on the number and type of TRS. The majority (about 89%) of the ASFVs were 100% homologous to the Georgia 2007/1 strain and were classified as the MGF-1 variant, including those from the 2022 outbreaks. The MGF-2 variant, originating from the Russia Federation in 2012, formed a heterogeneous group with viruses from Poland and Latvia collected near the Belarusian border. Groups 3 to 8 were country specific. Despite its low utility in determining the origin of ASF 2022 epidemics, the IGR between the MGF*9R*/*10R* genes showed higher resolving power than the CVR, IGR, *O174L*, and *K145R* regions. This fact gives this region great potential as a molecular marker, especially in endemic regions where ASF has established within the wild boar population. To identify the origin of the 2022 outbreaks, a new PCR was designed that amplifies the IGR between the *I329L* and *I215L* genes and the carboxy terminus of the *I215L* gene (22). Partial sequencing of the *I215L* gene identified a SNP within the *I215L* gene, differentiating two geographically distinct genetic variants circulating in Europe. The ECO2-II variant, characterized by the presence of the SNP and initially identified in Romania in 2018, was responsible for the 2022 outbreaks in North Macedonia. This variant is circulating in Romania, Bulgaria, Serbia, Greece, and North Macedonia, forming a geographically distinct genetic group in Europe.

The combined analysis of the results obtained in the six ASFV genome regions sequenced enabled 382 ASFV isolates from Europe to be divided into 24 genetic groups. Group 1 is the oldest containing the isolate from Georgia in 2007 and is classified as the reference group, and includes historical genotype II-ASFVs from Europe. Group 3, the largest with 50.3% of ASFV isolates, includes the IGR-II variant identified in the EU "index" cases in 2014 in the vicinity of the

TABLE 4 Groups based on the variants identified in the CVR, IGR *I73R*/*I329L* (IGR), *O174L*, *K145R*, IGR MGF505*9R*/*10R* (MGF), and IGR*I329L*-*I215L* (ECO2) ASFV variable regions.

Genetic group	Geographical distribution (year)	No ASFVs (frequency)	HOST	Genetic variants					
				CVR	IGR	*O174L*	*K145R*	MGF	ECO2
1	Georgia (2007), Armenia (2007, 2008), Azerbaijan (2008), Russia Federation (2009, 2012)	8/382 (2.1%)	EWB, DP	I	I	I	I	I	I
2	Russia Federation (2012)	1/382 (0.26%)	DP	I	I	I	I	II	I
3	Ukraine (2012–2019), Belarus (2013), Lithuania (2014–2020), Poland (2014, 2018), Latvia (2014–2021), Estonia (2014–2022), Czech RP (2017, 2018), Romania (2017–2021), Moldova (2017–2018), Hungary (2018–2019), Slovakia (2019), Italy (2022)	192/382 (50.3%)	EWB, DP	I	II	I	I	I	I
4	Russia Federation (2012)	1/382 (0.26%)	EWB	I	I	I	I	III	I
5	Estonia (2015)	7/382 (1.83%)	EWB	II	II	I	I	I	I
6	Poland (2016, 2019), Germany (2020)	8/382 (2.09%)	EWB, DP	I	II	II	II	I	I
7	Poland (2016–2019), Lithuania (2017–2022), Romania (2019)	29/382 (7.59%)	EWB, DP	I	II	I	II	I	I
8	Poland (2016, 2017)	11/382 (2.88%)	EWB, DP	I	II	I	II	II	I
9	Estonia (2017)	10/382 (2.62%)	EWB, DP	I-SNP1	II	I	I	I	I
10	Poland (2017)	2/382 (0.52%)	EWB	I	I	II	II	I	I
11	Poland (2017)	1/382 (0.26%)	EWB	I-SNP2	II	I	II	I	I
12	Latvia (2017, 2018, 2021)	14/382 (3.66%)	EWB	I	II	I	I	II	I
13	Poland (2017)	1/382 (0.26%)	DP	I	III	II	II	I	I
14	Lithuania (2017)	1/382 (0.26%)	EWB	I-SNP3	II	I	I	I	I
15	Lithuania (2017)	1/382 (0.26%)	EWB	I	II	I	I	V	I
16	Lithuania (2017, 2018)	5/382 (1.31%)	EWB, DP	I	II	I	I	IV	I
17	Latvia (2017, 2018)	3/382 (0.79%)	EWB, DP	I	II	I	I	I-V1	I
18	Poland (2018)	1/382 (0.26%)	EWB	I	III	II	I	I	I
19	Romania (2018, 2021), Bulgaria (2018–2020), Serbia (2019, 2020), Greece (2020), North Macedonia (2022)	55/382 (14.40%)	EWB, DP	I	II	I	I	I	II
20	Poland (2018, 2019)	10/382 (2.62%)	EWB	I	IV	I	II	I	I
21	Romania (2019)	7/382 (1.83%)	EWB, DP	I	II	II	I	I	I
22	Romania (2019)	12/382 (3.14%)	EWB, DP	I	II	II	I	I	II
23	Lithuania (2020)	1/382 (0.26%)	EWB	I	II	I	I	VII	I
24	Romania (2021)	1/382 (0.26%)	DP	I	II	I	I	VI	I

EWB, European wild boar; DP, domestic pig.

Belarusian border. This group comprises ASFV from 12 countries, nine from the EU, including recent ASFV isolated in Italy in 2022. The second largest group (14.4%), group 19, contains the IGR II and ECO2-II variants and comprises viruses from Southeast Europe. This group was initially identified in 2018 in an outbreak on a backyard farm located in Tulcea, southeastern Romania, on the border with Odessa, Ukraine. Regarding the fact that this variant was completely absent in Romania, where group 3 was unique, it could be assumed that the disease probably jumped to Tulcea from neighboring countries such as Ukraine, but its exact origin cannot be determined based on the molecular data available from Ukraine. It was not further identified in Romania until January 2021, spreading throughout the country. This genetic group was subsequently identified in Bulgaria in August 2018, in Serbia in 2019,

in Greece in 2020 and in North Macedonia in 2022. In these countries it is the only group identified.

In contrast, the ASFVs from Romania were grouped into 6 distinct groups (3, 7, 19, 21, 23, 24). Interestingly, the 2019 epidemic wave, despite taking place in a short space of time (40), was caused by two distinct genetic groups specific only to Romania. Group 21, containing the IGR-II and *O174L*-II variants, was initially detected in January 2019 in the western regions and had a wide geographic spread, being identified in the north and east. Group 22 with the IGRII, *O174L*-II and ECO2-II variants formed a differentiated cluster in the southeast of the country. A similar unexpected introduction was observed in 2021 on a backyard farm in Arad, in the westernmost part of Romania, caused by a new genetic group, group 24 possessing the MGF-6 variant, never before identified. The origin of these genetic groups is unknown mainly due to the lack of information from neighboring non-EU countries. The epidemiological analysis carried out by Andraud et al. (40) clearly identified human activity and the distribution of the pig population as the main risk factors for the spread of ASF in epidemic waves in Romania. However, from the molecular data obtained in this study, we cannot exclude the role of wild boars in the infection of domestic pigs, probably due to the low level of biosecurity in backyard farms, predominant in the country (40). Although the situation in Romania cannot be directly translated to intensive pig farming countries, the results of this study highlight the need for strict biosecurity measures on farms and during transport to prevent the transmission of ASF on a large scale. Cluster 7 (IGR-II and *K145R*-II), also circulating in Romania, was associated with a wild boar case that occurred in the north of Botosani County, about 2 km from the Ukrainian border, assumed to have been introduced through the wild boar migration. This group, described by Mazur-Panasiuk et al. (16) as the most typical for Poland since 2016, was also identified in Lithuania in 2017 in an outbreak of domestic pigs on the Ukrainian border in Alytus County, southern Lithuania. Subsequently, two temporal a spacial differentiated groups were formed; the 2017–2018 group originating from Alytus that spread to central and northern Lithuania, and the 2021–2022 group identified in Taurage County, on the border with Kaliningrad Oblast. This same genetic group was identified in a wild boar hunted in the Kaliningrad Oblast in 2018 (ASFV/Kaliningrad_18/WB-12516; GenBank OM966720.1), so reaffirming that wild boar is playing a key role in the reintroduction of the ASFVs in affected countries. Three additional groups, 14 (IGRII and CVR-SNP1), 15 (IGR-II and MGF-5), and 16 (IGR-II, MGF-4) were sporadically detected in Lithuania in 2017 and 2018. It should be noted that group 17, detected on the border with Belarus, made it possible to genetically link isolates from domestic pigs and wild boars in Utena County. Indeed, 11 of the 24 groups clustered both wild boar and domestic pigs ASFVs.

The higher variability found in isolates from Poland with nine of 24 groups identified could be explained by the long persistence of the disease since 2014 within the wild boar population. However, in Lithuania, Estonia, and Latvia, where the disease has also been endemic since 2014, group 3 detected in the initial outbreaks was clearly the predominant one, with additional spatially and geographically restricted groups. Forth et al. (27) has recently described an increased mutation rate in the affected region of Germany, where the disease was introduced in 2020. Twenty-two ASFV isolates from Germany diverged into five clearly distinguishable lineages with at least 10 different variants characterized by high-impact mutations. Notably, all of the new variants share a 3′-end frameshift mutation of the PolX DNA polymerase *O174L* gene, initially described in Poland by Mazur-Panasiuk et al. (16), suggesting a causal role as a possible mutator gene. This could explain the greater variability found in countries where the *O174L*-II variant circulates, such as Poland and Romania, compared to countries where variant I is predominant.

Our results confirmed that genomic regions containing TRS generally occur in regions where the disease has long persisted within the wild boar population, appearing sporadically in domestic pigs. These regions are of particular interest in terms of standard genotyping procedures because of the difference in the length of the PCR product, which is convenient to observe during regular agarose electrophoresis. In addition to tandem repeats, SNPs represent an attractive molecular tool that allows discrimination of closely related ASFV genotype II strains into clearly distinct groups obtained for different space and time. Whole genome sequencing, while essential to identify new genetic markers, is not a feasible routine genotyping method due to the complexity of the required sequence analysis vs. the final results, the time consuming, high cost, and specialized personnel required. In this study, we describe a new multi-gene approach sequencing method that distinguishes European II-ASFV genotypes into 24 distinct groups by sequencing six independent ASFV genomic regions, including the never-before-described IGR 1329L–*I215L*, named ECO2 region. The introduction of a subtyping method into routine diagnosis within affected areas worldwide may help to identify potential origins of the disease and provide a deeper understanding of the spatial and temporal trajectories and routes of the disease. Additional sequencing of other genetic markers could cluster genotype II ASFV isolates at higher resolution and cannot be ruled out.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CG was the principal investigator, analyzed the sequences, wrote the manuscript, and designed the experiments. NC performed sequencing analysis. ASo and RN tested the clinical samples for ASF diagnostic processing the samples and did the PCR. ID and LK provided the clinical samples from North Macedonia and Latvia. EM purified the sequences and did the SANGER sequencing. CP and ASi did the conventional PCR for sequencing. EI, DD, VM, EC, IN, MF, FF, PV, and SP provided the clinical samples from affected ASF areas of Bulgaria, Romania, Serbia, Greece, Estonia, Poland, Italy, the Czech Republic, and Lithuania. AM is the EURL Director and overseeing all experimental work. All authors have read and approved the manuscript.

Funding

This study was supported by the European Union Reference laboratory for ASF (grant no. UE- LR PPA/03).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1112850/full#supplementary-material

References

1. Alonso C, Borca M, Dixon L, Revilla Y, Rodriguez F, Escribano JM. ICTV report consortium ICTV virus taxonomy profile: Asfarviridae. *J Gen Virol.* (2018) 99:613–4. doi: 10.1099/jgv.0.001049

2. Dixon LK, Chapman DA, Netherton CL, Upton C. African swine fever virus replication and genomics. *Virus Res.* (2013) 173:3–14. doi: 10.1016/j.virusres.2012.10.020

3. Achenbach JE, Gallardo C, Nieto-Pelegrín E, Rivera-Arroyo B, Degefa-Negi T, Arias M, et al. Identification of a new genotype of African swine fever virus in domestic pigs from Ethiopia. *Transbound Emerg Dis.* (2017) 64:1393–404. doi: 10.1111/tbed.12511

4. Quembo CJ, Jori F, Vosloo W, Heath L. Genetic characterization of African swine fever virus isolates from soft ticks at the wildlife/domestic interface in Mozambique and identification of a novel genotype. *Transbound Emerg Dis.* (2018) 65:420–31. doi: 10.1111/tbed.12700

5. Njau EP, Machuka EM, Cleaveland S, Shirima GM, Kusiluka LJ, Okoth EA, et al. African swine fever virus (ASFV): biology, genomics and genotypes circulating in Sub-Saharan Africa. *Viruses.* (2021) 13:2285. doi: 10.3390/v13112285

6. Sun E, Huang L, Zhang X, Zhang J, Shen D, Zhang Z, et al. Genotype I African swine fever viruses emerged in domestic pigs in China and caused chronic infection. *Emerg Microbes Infect.* (2021) 10:2183–93. doi: 10.1080/22221751.2021.1999779

7. Rowlands RJ, Michaud V, Heath L, Hutchings G, Oura C, Vosloo W, et al. African swine fever virus isolate, Georgia, 2007. *Emerg Infect Dis.* (2008) 14:1870–4. doi: 10.3201/eid1412.080591

8. Gallardo C, Fernández-Pinero J, Pelayo V, Gazaev I, Markowska-Daniel I, Pridotkas G, et al. Genetic variation among African swine fever genotype II viruses, eastern and central Europe. *Emerg Infect Dis.* (2014) 20:1544–7. doi: 10.3201/eid2009.140554

9. OIE (World Organisation for Animal Health). *African Swine Fever (ASF)—Situation Report 6.* (2022). Available online at: https://www.woah.org/app/uploads/2022/02/asf-report6.pdf (accessed November 21, 2022).

10. Ge S, Li J, Fan X, Liu F, Li L, Wang Q, et al. Molecular characterization of African swine fever virus, China, 2018. *Emerg Infect Dis.* (2018) 24:2131–3. doi: 10.3201/eid2411.181274

11. Gonzales W, Moreno C, Duran U, Henao N, Bencosme M, Lora P, et al. African swine fever in the Dominican Republic. *Transbound Emerg Dis.* (2021) 68:3018–9. doi: 10.1111/tbed.14341

12. Lubisi BA, Bastos AD, Dwarka RM, Vosloo W. Molecular epidemiology of African swine fever in East Africa. *Arch Virol.* (2005) 150:2439–52. doi: 10.1007/s00705-005-0602-1

13. Nix RJ, Gallardo C, Hutchings G, Blanco E, Dixon LK. Molecular epidemiology of African swine fever virus studied by analysis of four variable genome regions. *Arch Virol.* (2006) 151:2475–94. doi: 10.1007/s00705-006-0794-z

14. Vilem A, Nurmoja I, Niine T, Riit T, Nieto R, Viltrop A, et al. Molecular characterization of African swine fever virus isolates in Estonia in 2014-2019. *Pathogens.* (2020) 9:582. doi: 10.3390/pathogens9070582

15. Mazur-Panasiuk N, Wozniakowski G. The unique genetic variation within the *O174L* gene of polish strains of African swine fever virus facilitates tracking virus origin. *Arch Virol.* (2019) 164:1667–72. doi: 10.1007/s00705-019-04224-x

16. Mazur-Panasiuk N, Walczak M, Juszkiewicz M, Wozniakowski G. The spillover of African swine fever in Western Poland revealed its estimated origin on the basis of *O174L*, *K145R*, MGF 505-5R and IGR *I73R/I329L* genomic sequences. *Viruses.* (2020) 12:1094. doi: 10.3390/v12101094

17. Tran HTT, Truong AD, Dang AK, Ly DV, Nguyen CT, Chu NT, et al. Circulation of two different variants of intergenic region (IGR) located between the *I73R* and *I329L* genes of African swine fever virus strains in Vietnam. *Transbound Emerg Dis.* (2021) 68:2693–5. doi: 10.1111/tbed.13996

18. Nguyen VT, Cho KH, Mai NTA, Park JY, Trinh TBN, Jang MK, et al. Multiple variants of African swine fever virus circulating in Vietnam. *Arch Virol.* (2022) 167:1137–40. doi: 10.1007/s00705-022-05363-4

19. Shi K, Liu H, Yin Y, Si H, Long F, Feng S. Molecular characterization of African swine fever virus from 2019-2020 outbreaks in Guangxi Province, Southern China. *Front Vet Sci.* (2022) 9:912224. doi: 10.3389/fvets.2022.912224

20. Gallardo C, Elsukova A, Wozniakowski G, Nieto R, Soler A, Sánchez-Vizcaíno J-M, e al. Sequencing of the tandem repeat sequences (TRS) within the intergenic region between the multigene family 505 *9R—10R* genes: additional tool for subtyping genotype II African swine fever (ASF) isolates. In: *Proceedings of the 11th International Congress for Veterinary Virology 12th Annual Meeting of EPIZONE*, Vienna, Austria (2018), p. 71.

21. Mazloum A, Igolkin AS, Vlasova NN, Romenskaya DV. African swine fever virus: use of genetic markers in analysis of its routes of spread. *Vet Sci Today.* (2019) 3–14. doi: 10.29326/2304-196X-2019-3-30-3-8

22. Rodriguez JM, Salas ML, Viñuela E. Genes homologous to ubiquitin-conjugating proteins and eukaryotic transcription factor SII in African swine fever virus. *Virology.* (1992) 186:40–52. doi: 10.1016/0042-6822(92)90059-X

23. Fernández-Pinero J, Gallardo C, Elizalde M, Robles A, Gómez C, Bishop R, et al. Molecular diagnosis of African Swine Fever by a new real-time PCR using universal probe library. *Transbound Emerg Dis.* (2013) 60:48–58. doi: 10.1111/j.1865-1682.2012.01317.x

24. World Organisation for Animal Health (WOAH): 2021. African swine fever. In: *Manual of Diagnostic tests and Vaccines for Terrestrial Animals*, Vol. 2, Chapter 3.9.1. (2021). Available online at: https://www.woah.org/fileadmin/Home/eng/Health_standards/tahm/3.09.01_ASF.pdf (accessed November 20, 2022).

25. Gallardo C, Anchuelo R, Pelayo V, Poudevigne F, Leon T, Nzoussi J, et al. African swine fever virus p72 genotype IX in domestic pigs, Congo, 2009. *Emerg Infect Dis.* (2011) 17:1556–8. doi: 10.3201/eid1708.101877

26. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* (2021) 38:3022–7. doi: 10.1093/molbev/msab120

27. Forth JH, Calvelage S, Fischer M, Hellert J, Sehl-Ewert J, Roszyk H, et al. African swine fever virus - variants on the rise. *Emerg Microbes Infect.* (2023) 12:2146537. doi: 10.1080/22221751.2022.2146537

28. Alkhamis MA, Gallardo C, Jurado C, Soler A, Arias M, Sánchez-Vizcaíno JM. Phylodynamics and evolutionary epidemiology of African swine fever p72-CVR genes in Eurasia and Africa. *PLoS ONE.* (2018) 13:e0192565. doi: 10.1371/journal.pone.0192565

29. Sidi M, Zerbo HL, Ouoba BL, Settypalli TBK, Bazimo G, Ouandaogo HS, et al. Molecular characterization of African swine fever viruses from Burkina Faso, 2018. *BMC Vet Res.* (2022) 18:69. doi: 10.1186/s12917-022-03166-y

30. Minoungou GL, Diop M, Dakouo M, Ouattara AK, Settypalli TBK, Lo MM, et al. Molecular characterization of African Swine fever viruses in Burkina Faso, Mali, and Senegal 1989-2016: genetic diversity of ASFV in West Africa. *Transbound Emerg Dis.* (2021) 68:2842–52. doi: 10.1111/tbed.14240

31. Hakizimana JN, Nyabongo L, Ntirandekura JB, Yona C, Ntakirutimana D, Kamana O, et al. Genetic analysis of African swine fever virus from the 2018 outbreak in South-Eastern Burundi. *Front Vet Sci.* (2020) 7:578474. doi: 10.3389/fvets.2020.578474

32. Wade A, Achenbach JE, Gallardo C, Settypalli TBK, Souley A, Djonwe G, et al. Genetic characterization of African swine fever virus in Cameroon, 2010-2018. *J Microbiol.* (2019) 57:316–24. doi: 10.1007/s12275-019-8457-4

33. Gallardo C, Okoth E, Pelayo V, Anchuelo R, Martín E, Simón A, et al. African swine fever viruses with two different genotypes, both of which occur in domestic pigs, are associated with ticks and adult warthogs, respectively, at a single geographical site. *J Gen Virol.* (2011) 92(Pt 2):432–44. doi: 10.1099/vir.0.025874-0

34. Gallardo C, Mwaengo DM, Macharia JM, Arias M, Taracha EA, Soler A, et al. Enhanced discrimination of African swine fever virus isolates through nucleotide

sequencing of the p54, p72, and pB602L (CVR) genes. *Virus Genes.* (2009) 38:85–95. doi: 10.1007/s11262-008-0293-2

35. Mazloum A, Van Schalkwyk A, Chernyshev R, Shotin A, Korennoy FI, Igolkin A, et al. Genetic characterization of the central variable region in African swine fever virus isolates in the Russian Federation from 2013 to 2017. *Pathogens.* (2022) 11:919. doi: 10.3390/pathogens11080919

36. Goller KV, Malogolovkin AS, Katorkin S, Kolbasov D, Titov I, Höper D, et al. Tandem repeat insertion in African swine fever virus, Russia, 2012. *Emerg Infect Dis.* (2015) 21:731–2. doi: 10.3201/eid2104.141792

37. Kolbasov D, Titov I, Tsybanov S, Gogin A, Malogolovkin A. African swine fever virus, Siberia, Russia, 2017. *Emerg Infect Dis.* (2018) 24:796–8. doi: 10.3201/eid2404.171238

38. Kim SH, Lee SI, Jeong HG, Yoo J, Jeong H, Choi Y, et al. Rapid emergence of African swine fever virus variants with different numbers of a tandem repeat sequence in South Korea. *Transbound Emerg Dis.* (2021) 68:1726–30. doi: 10.1111/tbed.13867

39. Mazloum A, van Schalkwyk A, Shotin A, Igolkin A, Shevchenko I, Gruzdev KN, et al. Comparative analysis of full genome sequences of African swine fever virus isolates taken from wild boars in Russia in 2019. *Pathogens.* (2021) 10:521. doi: 10.3390/pathogens10050521

40. Andraud M, Bougeard S, Chesnoiu T, Rose N. Spatiotemporal clustering and random forest models to identify risk factors of African swine fever outbreak in Romania in 2018-2019. *Sci Rep.* (2021) 11:2098. doi: 10.1038/s41598-021-81329-x

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fvets.2023.1107059

OPEN ACCESS

EDITED BY
Christina Leyson,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Henry Muriuki Kariithi,
Agricultural Research Service (USDA),
United States
ANa Da Silva,
University of California, Davis, United States

*CORRESPONDENCE
Dong-Hun Lee
✉ donghunlee@konkuk.ac.kr
Chang-Seon Song
✉ songcs@konkuk.ac.kr

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 24 November 2022
ACCEPTED 11 January 2023
PUBLISHED 01 February 2023

CITATION
Kim H-J, Lee H-C, Cho AY, Choi Y-J, Lee H, Lee D-H and Song C-S (2023) Novel recombinant avian infectious bronchitis viruses from chickens in Korea, 2019–2021. *Front. Vet. Sci.* 10:1107059. doi: 10.3389/fvets.2023.1107059

Novel recombinant avian infectious bronchitis viruses from chickens in Korea, 2019–2021

Hyun-Jin Kim[1], Hyuk-Chae Lee[2], Andrew Y. Cho[1], Yun-Jeong Choi[1], Heesu Lee[1], Dong-Hun Lee[3]* and Chang-Seon Song[1,2]*

[1]Avian Diseases Laboratory, College of Veterinary Medicine, Konkuk University, Seoul, South Korea, [2]KHAV Co., Ltd., Seoul, South Korea, [3]Wildlife Health Laboratory, College of Veterinary Medicine, Konkuk University, Seoul, South Korea

Infectious bronchitis virus (IBV) has evolved through various mutation mechanisms, including antigenic drift and recombination. Four genotypic lineages of IBVs including GI-15, GI-16, GI-19, and GVI-1 have been reported in Korea. In this study, we isolated two IBVs from chicken farms, designated IBV/Korea/289/2019 (K289/19) and IBV/Korea/163/2021 (K163/21), which are two distinct natural recombinant viruses most likely produced by genetic reassortment between the S1 gene of K40/09 strain (GI-19 lineage) and IBV/Korea/48/2020 (GI-15 lineage) in co-infected commercial chickens. Comparative sequence analysis of hypervariable regions (HVRs) revealed that the K289/19 virus had similar HVR I and II with the K40/09 virus (100% and 99.2% nucleotide sequence identity, respectively), and HVR III with the IBV/Korea/48/2020 virus (100% nucleotide sequence identity). In contrast, the K163/21 virus had HVR I and II similar to the IBV/Korea/48/2020 virus (99.1% and 99.3% nucleotide sequence identity, respectively), and HVR III to the K40/09 virus (96.6% nucleotide sequence identity). The K289/19 virus exhibited similar histopathologic lesions, tissue tropism in trachea and kidney, and antigenicity with the parental K40/09 virus. The K163/21 exhibited similar pathogenicity and tissue tropism with the K40/09 virus, which were similar results with the isolate K289/19. However, it showed a lower antigenic relatedness with both parental strains, exhibiting R-value of 25 and 42, respectively. The continued emergence of the novel reassortant IBVs suggests that multiple recombination events have occurred between different genotypes within Korea. These results suggest that antigenic profiles could be altered through natural recombination in the field, complicating the antigenic match of vaccine strains to field strains. Enhanced surveillance and research into the characteristics of newly emerging IBVs should be carried out to establish effective countermeasures.

KEYWORDS

infectious bronchitis virus, recombination, pathogenicity, antigenicity, tissue tropism

1. Introduction

Infectious bronchitis virus (IBV) is a pathogen associated with acute respiratory tract diseases in chickens. The IBV is highly contagious and can also affect multiple organs according to their pathotypes, such as the respiratory tract, kidneys, and reproductive tract. Chickens infected with IBV show respiratory signs, reduced egg production, weight loss, and decreased weight gain. Mortality may vary depending on the IBV strain, secondary bacterial infection, or coinfection with other viruses. It is economically important to control this infection because of its detrimental effects on poultry production (1, 2).

High antigenic diversity of IBVs and the extensive emergence of variants are problematic, resulting in poor cross-protection by the available vaccine strains (3–5). The spike (S) protein of

IBV, comprising ~3.4 kb, is a major inducer of virus-neutralizing antibodies and an important factor in determining tissue tropism (6), which is post-translationally cleaved to S1 and S2 subunits. The S2 subunit is associated with membrane fusion (6, 7). The S1 subunit, responsible for host cell attachment, has a receptor binding domain and hypervariable regions (HVR). HVRs (HVR I, II, and III) are associated with neutralizing antibodies, virulence of the virus, and tissue tropism (8–12). The HVR I is a major epitope inducing neutralizing antibodies and associated with tissue tropism for respiratory tract (12, 13). In addition, a previous study reported that the HVR II is related to kidney affinity (14). Because of these characteristics of the S1 subunit, IBV genotypes have been classified based on the S1 gene (15–17).

According to the classification system of IBV lineages defined by Valestro et al. (15), IBVs isolated in Korea are classified into four genotypic lineages: GI-15, GI-16, GI-19, and GVI-1 (16). The GI-15, previously designated as Korean group I (K-I), is associated with respiratory diseases in chickens (18). The GI-19 includes nephropathogenic viruses that can be divided into three subgroups: KM91-like, QX-like, and K40/09-like (17). The KM91-like subgroup, also known as Korean group II a (K-IIa), have been detected from 1990's with nephropathogenicity in chickens (19, 20). The QX-like subgroup, also known as Korean group II b (K-IIb), causes outbreaks in chickens globally, including South Korea (15, 17). The K40/09-like subgroup was reported as the Korean new cluster I in our previous study, which is produced by recombination between the KM91-like virus and QX-like virus (21). The GI-16 strains were isolated in Korea during 2003–2006, designated as Korean group III (K-III) (20), which have caused respiratory syndrome and nephropathogenic diseases in chicken farms (22, 23). The GVI-1, reported as the Korean new cluster II causes clinical signs mainly in the respiratory tract of chickens (24, 25). Diverse recombinant genotypes of IBVs in Korea suggests that multiple recombination have occurred between different genotypes in Korea.

In this study, we isolated two novel IBVs which are natural recombinant between the GI-15 and GI-19 lineages. We investigated their pathobiological characteristics, such as pathogenicity, antigenicity, and tissue tropism in chickens.

2. Materials and methods

2.1. Virus isolation and propagation

The IBVs used in this study were K40/09, IBV/Korea/48/2020, K289/19, and K163/21. The K40/09 (K40/09-like subgroup of GI-19) and IBV/Korea/48/2020 (B4-like subgroup of GI-15) viruses were isolated from chickens in Korea (17, 21). The isolate K289/19 was isolated from 46-day-old broiler breeders showing respiratory signs and nephritis. The isolate K163/21 was isolated from a 77-day-old layer that showed respiratory signs. All IBVs used in this study were propagated in 10-day-old specific-pathogen free (SPF) embryonated chicken eggs for 48 h (21). The allantoic fluid was harvested from the inoculated eggs and stored at −70°C until use. Before using the allantoic fluid, quantification of viruses was conducted by titration to calculate 50% embryo infectious dose (EID_{50}) (26).

2.2. PCR, sequencing, and phylogenetic analysis

Viral RNA was extracted from harvested allantoic fluid using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The S1 gene of each IBV isolate was amplified using a OneStep RT-PCR Kit (Qiagen) according to the manufacturer's instructions and using two pairs of primers (S1 forward F: CGGAACAAAAGACMGACTTAGT and S1 forward R: CWGTACCATTAACAAARTAAGCMAG; S1 rear F: TGTGTATTTTAAAGCAGGTGGACC and S1 rear R: GTTTGTATGTACTCATCTGTAAC). The reaction was conducted in a ProFlex PCR System (Applied Biosystems, Forest City, CA, USA) at 50°C for 30 min; 95°C for 15 min, and 35 cycles of 94°C for 60 s, 53°C for 60 s, 72°C for 120 s, and a final extension step at 72°C for 7 min. The PCR products were purified using a GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced using Sanger sequencing (Macrogen Co., Ltd., Seoul, South Korea). The obtained nucleotide sequences of the S1 gene were assembled and aligned with the prototype of different lineages of IBV and reference strains (15, 17) shown in Supplementary Table S1 using Geneious Prime® 2022.1.1 software (https://www.geneious.com/). A phylogenetic tree was constructed in MEGA version 10.2.5, using the neighbor-joining method with 1000 bootstrap replicates.

2.3. Recombination analysis

The Recombination Detection Program (RDP) 4 software (v. 4.39) was used to identify putative recombination events using several detection methods: RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, Phylpro, LARD, and 3Seq (27). Recombination events where at least five detection methods showed a p-value $< 1 \times 10^{-14}$ were accepted (28). After identifying their putative parental strains from 59 reference sequences, the S1 gene sequences of the IBV isolates were compared with the sequences of their putative parental strains. Nucleotide sequences from the S1 gene were used to generate similarity plot using the Simplot software (v. 3.5.1), with a window size of 200 bp and a step size of 20 bp (29).

2.4. Pathogenicity and tissue tropism investigation

One-day-old SPF chickens ($n = 150$) were randomly divided into five groups (Group 1: K40/09; Group 2: IBV/Korea/48/2020; Group 3: K289/19; Group 4: K163/21; Group 5: Negative control; n=30 in each group) and maintained separately in isolation cabinets (Three Shine, Daejeon, Korea). All chickens used in this study were obtained from Namduk SPF (Incheon, Republic of Korea). Chickens in the treated group were inoculated with a virus with a 10^5 EID_{50} per chick *via* the ocular route. Chickens from the negative control group were inoculated with sterile phosphate-buffered saline (PBS) for the same volume as the virus inoculated groups. At 5 days post-inoculation (5 dpi), twenty chicks were sacrificed from each group. The upper, middle, and lower regions of tracheal and kidney tissues were collected from ten necropsied chickens from each group for histopathological examination. Ciliary

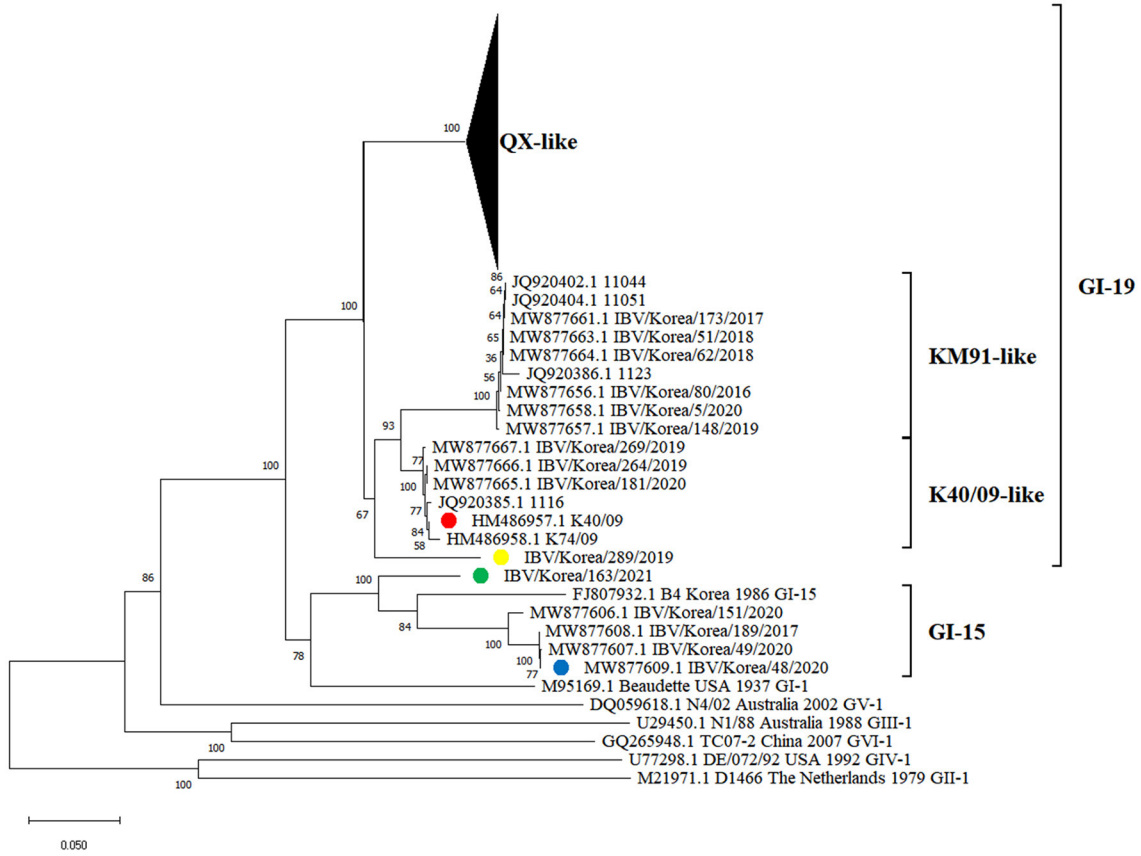


FIGURE 1
Phylogenetic tree based on the S1 gene nucleotide sequences of IBVs. The IBVs newly isolated in this study are highlighted in yellow and green circles (plural). Strains highlighted in red and blue circles (plural) are putative parental strains of isolates, which are identified as progenitors of recombination events using RDP4 software in this study.

TABLE 1 Recombinant detection of S1 gene in IBV isolated in this study.

Recombinant strain	Major parental strain	Minor parental strain	*p*-values of the detection methods[a]				
			RDP	GENECONV	MaxChi	Chimera	3Seq
IBV/Korea/289/2019	K40/09 (99.6%)	IBV/Korea/48/2020 (100%)	5.56×10^{-49}	3.97×10^{-39}	4.71×10^{-30}	4.68×10^{-30}	4.46×10^{-81}
IBV/Korea/163/2021	IBV/Korea/48/2020 (99.0%)	K40/09 (98.9%)	7.43×10^{-44}	2.29×10^{-37}	9.56×10^{-28}	1.04×10^{-27}	1.79×10^{-71}

[a]Recombination events were confirmed when P-values were $< 1 \times 10^{-14}$ from at least five detection methods. The major and minor parental strain is the virus contributing the larger fraction of the recombinant sequences and the smaller fraction of the recombinant sequences to the generation of recombinant IBV, respectively.

loss, inflammatory response of the trachea, and severity of kidney inflammation were scored as previously described (30). Tracheal and kidney tissues were collected from remaining 10 sacrificed chickens from each group to isolate virus. Tracheal and kidney tissues were homogenized and diluted to 10% (w/v) in PBS containing 400 mg/mL gentamicin. The supernatants of the homogenized tissue samples were clarified by centrifugation at 3,000 rpm (107×g) for 10 min, filtered using a 0.45 μm Minisart syringe filter (Sartorius, Göttingen, Germany), followed by propagationin 10-day-old SPF embryonated chicken eggs at 37°C for 72 h. After incubation, the allantoic fluid was harvested and viral RNA was extracted. Viral replication in trachea and kidney samples was confirmed using real-time reverse transcription-PCR (rRT-PCR) as previously described (31).

Ten of thirty chicks from each group were observed for clinical signs, morbidity, and mortality for 2 weeks after experimental infection. Clinical signs were monitored twice daily for clinical signs, including rales, nasal discharge, coughing, eye irritation, depression, and watery diarrhea. The morbidity and mortality rates were monitored daily. If mortality occurred, the dead chickens were necropsied to observe gross lesions.

2.5. Antigenicity study using cross-neutralization test

Allantoic fluids of the K40/09, IBV/Korea/48/2020, K289/19, and K163/21 viruses were inactivated with 0.1% formaldehyde for 24 h

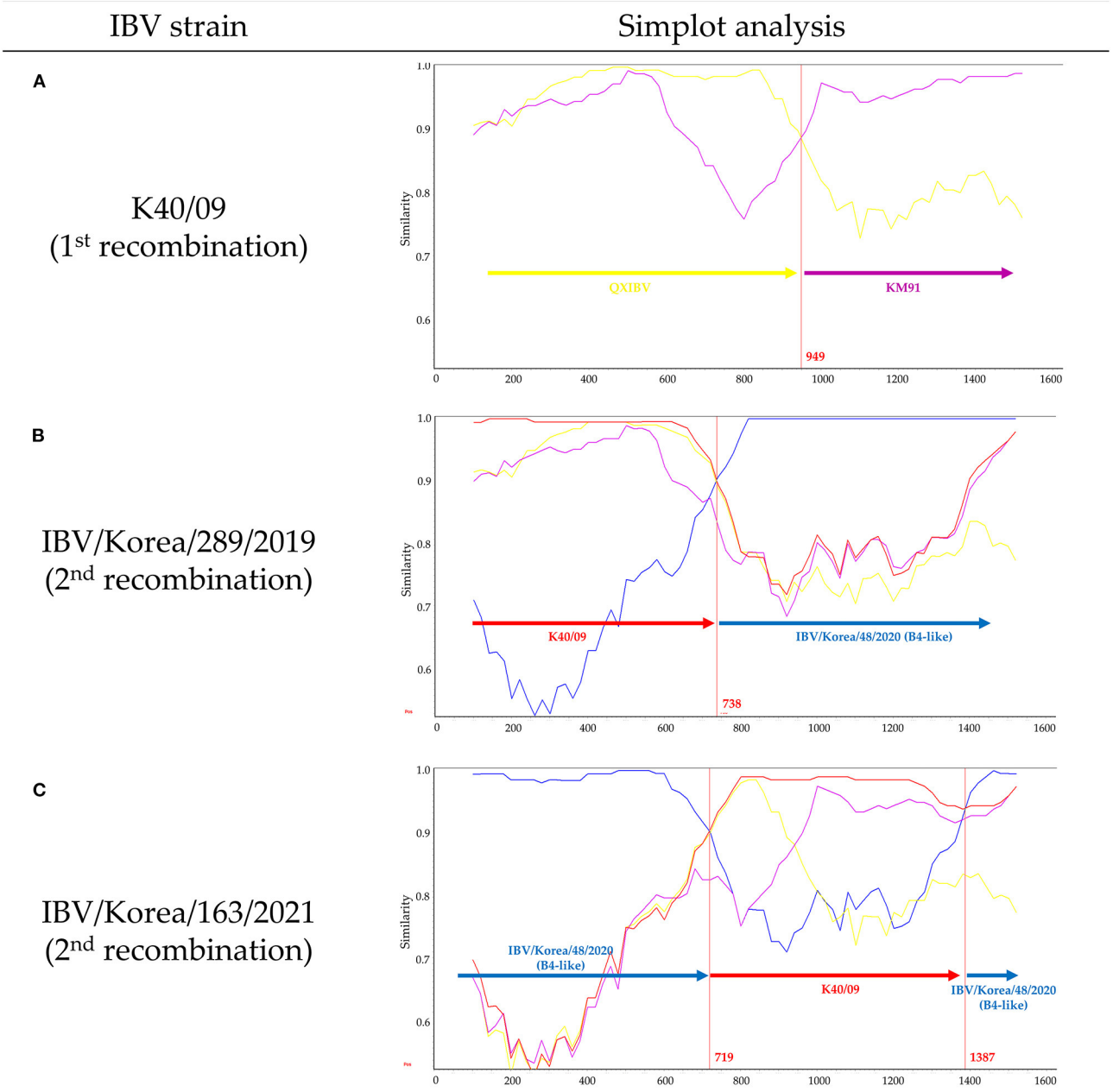


FIGURE 2
Evolutionary history of novel IBVs produced by multiple recombination events in Korea. Simplot analysis was performed with a window size of 200 bp and a step size of 20 bp. **(A)** Similarity plot of K40/09 strain, which was reported as recombinant between QXIBV (yellow) and KM91 (purple). **(B)** Similarity plot of K289/19. **(C)** Similarity plot of K163/21. Putative parental strains are highlighted in red (K40/09 strain) and blue (IBV/Korea/48/2020 strain). Red vertical lines and numbers indicate the crossover breakpoints and their nucleotide positions.

at room temperature (20–22°C). Inactivated allantoic fluids were emulsified with Montanide ISA70 (Seppic, Paris, France) at ratio of 3:7 (v/v) to be injected into SPF chickens to produce antisera (32).

Two-week-old SPF chickens ($n = 20$) were randomly separated into four groups and each oil-emulsified inactivated virus (0.5 mL) was intramuscularly inoculated into each group. Two weeks after inoculation, all groups were given a booster of the same inactivated virus. Two weeks after boosting, hyperimmune antisera were collected from each group and inactivated at 56°C for 30 min. Virus cross-neutralization tests were performed using SPF embryonated eggs with the Alpha method according to the OIE Terrestrial Manual (33). Briefly antisera and the ten-fold dilutions of the allantoic fluids, which began from the titer of 10^7 EID_{50}, were mixed and incubated at 37°C for 30 min. After incubation, the mixtures were inoculated into five SPF embryonated chicken eggs at each dilution. Eggs were inoculated with the corresponding titers of the virus with PBS in parallel. Endpoints were calculated using the Reed and Muench method (34). The neutralization index (NI) and antigenic relatedness (R-value) were calculated as previously described (19).

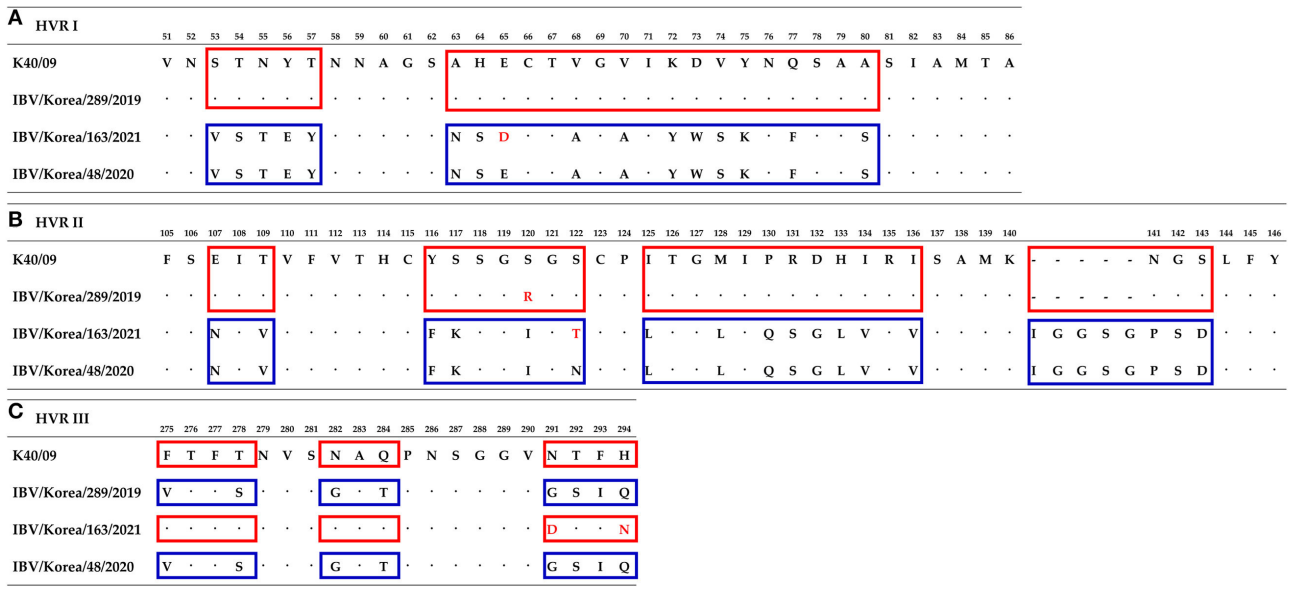


FIGURE 3
Comparison of amino acids sequences of hypervariable region at S1 gene between isolates and their parental strains. Numbering is based on isolate K289/19. **(A)** Hypervariable region I. **(B)** Hypervariable region II. **(C)** Hypervariable region III. Similar amino acids sequences with the K40/09 strain are highlighted in red boxes. Blue boxes indicate amino acids sequences similar to the IBV/Korea/48/2020 strain. Substitutions of amino acids are highlighted with red font.

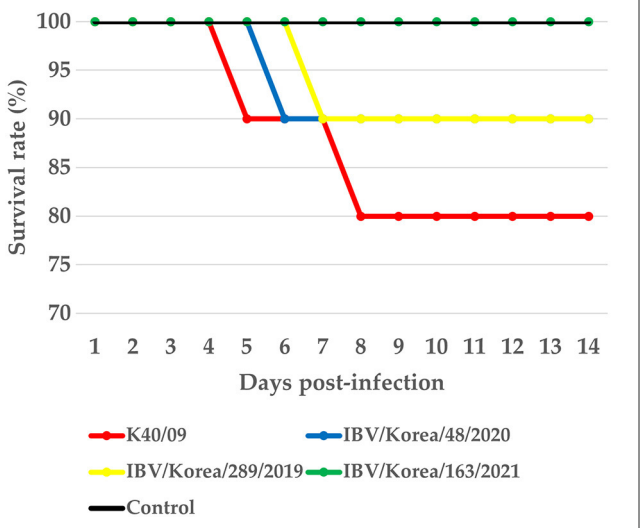


FIGURE 4
Survival rate of chickens infected with IBV isolates.

2.6. Statistical analyses

The ciliary loss and inflammatory response scores of test groups were compared with the score of negative group using Dunnett's multiple comparisons test. Re-isolation rate of the inoculated virus among the groups was analyzed using one-tailed Fisher's exact test. Statistical significance was set at $p < 0.05$. All statistical analyses were performed using GraphPad Prism version 9.2.0 software (San Diego, CA, USA).

2.7. Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of Konkuk University of South Korea (permission number KU22113, July 18, 2022).

3. Results

3.1. Phylogenetic, recombination and genomic analysis

Phylogenetic analysis of the S1 gene revealed that the K289/19 and K163/21 viruses were genetically distinct from previously identified IBVs in South Korea. These viruses did not belong to the QX-like, KM91-like, or K40/09-like and GI-15 subgroups in the phylogeny (Figure 1).

Recombination events in K289/19 and K163/21 were identified using RDP4 software (Table 1). Five different methods in RDP4 and Simplot analysis showed that their putative parental strains are most likely the K40/09 strain of GI-19 and the IBV/Korea/48/2020 strain of GI-15 (p value $<1 \times 10^{-14}$) (Figure 2). Other putative recombination events were excluded, because they did not satisfy the criteria for recombination event described above ($p < 1 \times 10^{-14}$ at least five detection methods).

The K40/09 strain is a recombinant strain derived from QXIBV and KM91 (21) (Figure 2A). Single crossover breakpoint in the S1 gene of K289/19 virus was identified at the nucleotide position 738 bp; 1–738 bp was similar to the K40/09 strain, followed by IBV/Korea/48/2020 strain (Figure 2B). However, two crossover breakpoints were identified in the S1 gene of K163/21 virus, at nucleotide positions 719 bp and 1387 bp; the 719–1387 bp region in the S1 gene of IBV/Korea/48/2020 strain was substituted by the

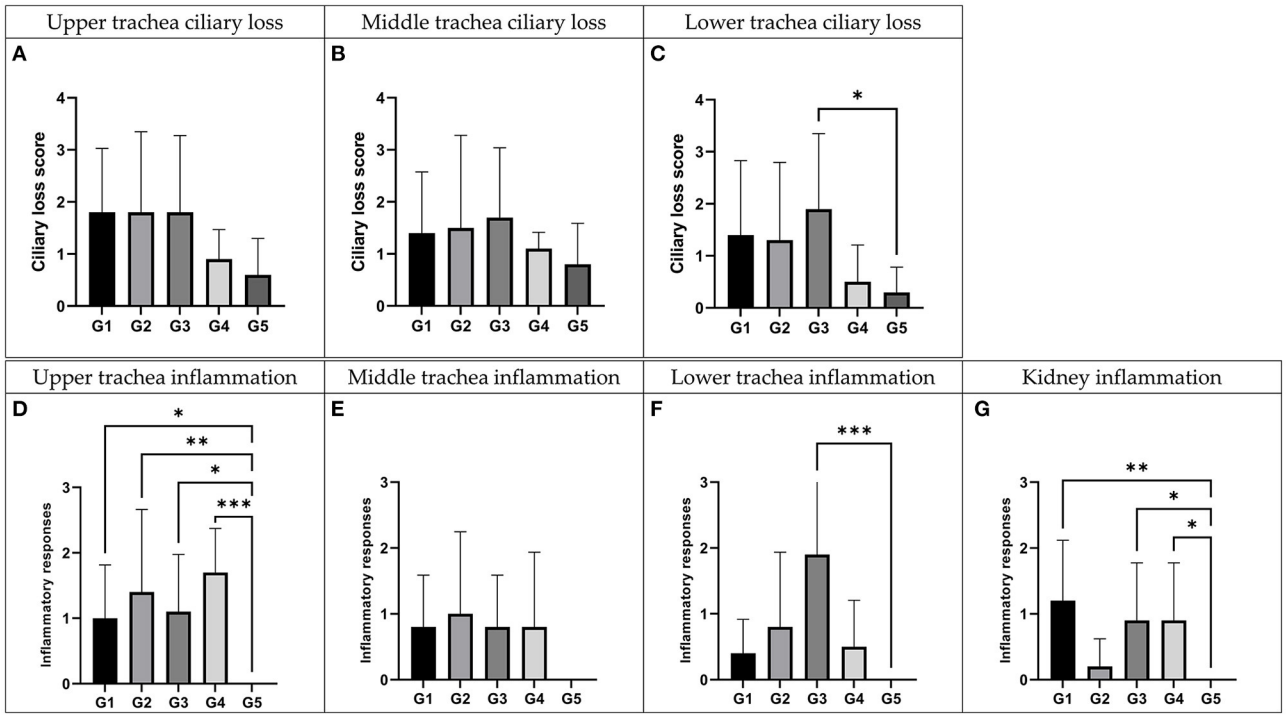


FIGURE 5
Histopathologic examination of recombinant strains and their parent strains. Trachea and kidney samples were collected at 5 days post-inoculation. G1: K40/09 challenged group, G2: IBV/Korea/48/2020 challenged group, G3: IBV/Korea/289/2019 challenged group, G4: IBV/Korea/163/2021 challenged and G5: negative control group. **(A–C)** Ciliary damages were compared with negative control. **(D–G)** Inflammation responses were estimated comparing with negative control. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$, compared to the negative control group as determined using Dunnett's multiple comparisons test.

TABLE 2 Virus re-isolation rate in the trachea and kidney.

Inoculation strain	No. of inoculation virus isolated/no. of inoculated	
	Trachea	Kidney
K40/09	10/10****	10/10****
IBV/Korea/48/2020	10/10****	0/10
IBV/Korea/289/2019	10/10****	9/10****
IBV/Korea/163/2021	10/10****	9/10****
PBS(Negative control)	0/10	0/10

****P < 0.0001 compared to the negative control group as determined using one-tailed Fisher's exact test.

partial S1 gene of K40/09 strain (Figure 2C). Briefly, the S1 genes of two isolates were produced by recombination between same parental strains, the K40/09 strain and IBV/Korea/48/2020, but in different patterns.

The nucleotide sequence identity of the S1 gene of each recombinant was compared with the GI-15 and GI-19 IBVs identified previously in Korea. Nucleotide sequence identities of K289/19 were 81.9–87.8% with the GI-15 IBVs, and 85.9–91.6% with the GI-19 IBVs. Nucleotide sequence identities of K163/21 were 84.6–91.6% with the GI-15 IBVs, and the 78.7–86.8% with GI-19 IBVs (data not shown). Sequence analysis of the HVR I, II, and III in the S1 gene showed that the HVR I and II of K289/19 were similar to the K40/09 strain, except one substitution at 120th amino acid (S120R), but HVR III was similar to the IBV/Korea/48/2020 strain (Figure 3). In contrast, the HVR I and II of the isolate K163/21 was similar to the IBV/Korea/48/2020 strain, except two substitutions at 65th and 122th amino acids (E65D and N122T). The HVR III of the isolate K163/21 was similar to the K40/09 strain except two substitutions at 291th and 294th amino acids (N291D and H294N).

3.2. Pathogenicity investigation and tissue tropism

Mortality was not observed in the K163/21 challenged group or the negative control group (Figure 4). The chickens challenged with the IBV/Korea/48/2020 or K289/19 viruses exhibited 10% mortality (1/10) at 6 dpi and 7 dpi, respectively. The chickens inoculated with K40/09 strain showed 20% mortality (2/10). Mild exudates and petechial hemorrhage in the trachea were commonly observed in the dead chickens. Nephritis and urate deposition were only observed in the dead chickens challenged with the K289/19 and the K40/09 viruses, indicating nephropathogenicity of these viruses.

Histopathological examination (Figure 5) and virus re-isolation test (Table 2) were performed at 5 dpi to evaluate the pathogenicity and tissue tropism of each virus. For the ciliary loss score of upper trachea, the score of the K40/09, IBV/Korea/48/2020, K289/19 group were higher than the negative control group, followed by the score of the K163/21 group. However, there was no statistical difference compared to the negative control group ($p > 0.05$). Likewise, for

TABLE 3 Antigenic relatedness (R-value) between the IBVs used in this study.

Virus	Antisera			
	K40/09 (GI-19 K40/09-like)	IBV/Korea/48/2020 (GI-15 B4-like)	IBV/Korea/289/2020 (recombinant)	IBV/Korea/163/2021 (recombinant)
K40/09 (GI-19 K40/09-like)	100[a]	-	-	-
IBV/Korea/48/2020 (GI-15 B4-like)	7	100	-	-
IBV/Korea/289/2019 (recombinant)	76	10	100	-
IBV/Korea/163/2021 (recombinant)	25	42	36	100

[a]R > 70%: little or no difference in antigenic identity; 33–70%: minor subtype difference; 11–32%: major subtype difference; 0–10%: different serotypes.

the ciliary loss score of middle trachea, the scores of inoculated groups were higher than that of the negative control group, but there was no significant difference compared with negative control group. However, for the ciliary loss score of lower trachea, the score of K289/19 challenged group was the highest ($p<0.05$), followed by the K40/09 challenged group and the IBV/Korea/48/2020 challenged group. The K163/21 had a slightly higher score compared to the negative control group. Based on the results of the scoring of ciliary loss, it seems that the K289/19 was more pathogenic than its parental strains, but the K163/21 was less pathogenic than its parental strains.

All the groups inoculated with each virus showed significantly higher inflammation score than the negative control group in the upper trachea ($P < 0.05$). Chickens inoculated with each virus showed higher inflammation score than the negative control group in the middle trachea, but there was no statistically significant difference ($P > 0.05$). For the inflammation score in the lower trachea, the score of K289/19 challenged group was only significantly higher than the negative control group ($P < 0.001$). Collectively, all inoculated viruses were identified to be pathogenic in trachea, although virulence of each virus was slightly different. The inflammation score in kidney tissues of IBV/Korea/48/2020 challenged group was similar to the negative control group. However, the K40/09 group and two novel recombinant viruses showed statistically higher inflammation scores in kidney compared to the negative control group ($P < 0.05$).

Similar results were also identified from the virus re-isolation test (Table 2). The positive rate of IBV was 100% in trachea of all groups, suggesting efficient infection and robust replication of all IBVs in the upper respiratory tract of chickens. However, the positive rates for the kidney tissues varied among the groups. Chickens inoculated with K40/09 showed a 100% re-isolation rate from kidney tissue samples. Chickens infected with K289/19 or K163/21 exhibited 90% positive rate from kidney tissues. However, the IBV/Korea/48/2020 group and the negative control group showed 0% positive rate from kidney tissues.

3.3. Antigenicity study using cross-neutralization test

Cross-neutralization tests were performed using the recombinant viruses and their putative parental strains. NI values retrieved from the cross-neutralization tests were converted to R-values using the calculation method described by Archetti and Horsfall (35). R-values implicate antigenic relatedness between two viruses. R-values >70% indicate the same serotype, R-values between 33 and 70% indicate subtype with a minor difference, between 11 and 32% indicate a subtype with a major difference, and R-values <11% indicate a different serotype (19). Recombinant viruses showed different antigenic properties as shown in Table 3. The two parental strains, K40/09 and IBV/Korea/48/2020, were antigenically different (7%). The isolate K289/19 was antigenically close with the K40/09 strain (76%), but different with the IBV/Korea/48/2020 strain (10%). The antigenic identity of K163/21 was identified to be major subtype difference from K40/09 strain (25%) and minor subtype difference from K289/19 (36%) and IBV/Korea/48/2020 (42%).

4. Discussion

The IBV evolves through a variety of mutation mechanisms, not only accumulations of point mutations but also recombination events, which makes it difficult to perfectly prevent the infection of IBVs (3–5). In particular, the recombination between field strains and vaccine strains contributes to the sudden emergence of various strains (36). In this study, we isolated two novel recombinant IBVs (K289/19 and K163/21) and investigated the genetic characteristics based on the S1 gene and pathobiological features in chickens. Our data showed that these viruses were produced by recombination between same progenitors, the GI-19-like virus and GI-15-like virus in different patterns. Since two recombinants were identified as progenies of identical parental strains based on the S1 gene, we had assumed that these viruses may have similar pathobiological characteristics to one of their parental strains. However, the pathogenicity, tropism, and antigenicity were not identical to those of parental strains. Although the S1 subunit is a major determinant of pathogenicity, tissue tropism, and antigenicity of IBV (8–12), genetic changes in S1 gene could not explain molecular mechanism of these phenotypical changes since other genes such as S2 gene and replicase gene can also alter their pathobiological characteristics (37, 38). Whole genome sequencing would be required to better understand the molecular mechanism for these findings.

The HVRs of S1 gene are highly diverse and often associated with antigenicity, pathogenicity, and tissue tropism (13, 39). In particular, it has been reported that the HVR II is associated with nephropathogenicity (14). The K289/19, harboring HVR II derived from the nephropathogenic IBV, exhibited kidney affinity. Unexpectedly, the K163/21, harboring HVR II derived from the non-nephropathogenic IBV, also showed nephropathgenicity in chickens.

It implies that the HVR II of S1 gene is not the only factor that determine the tissue tropism of IBV. Furthermore, other studies have suggested that genetic changes in other structural or non-structural protein genes could contribute to the pathogenicity and tissue tropism of coronaviruses (14, 37, 40–42). Therefore, further studies are needed to clarify the determinants of pathogenicity and tissue tropism.

Live attenuated vaccines derived from GI-15 and GI-19-like viruses have been used in Korea (30, 43). The antigenicity of the recombinant viruses was identified to be different, despite of being derived from the same parental strains. Since IBVs generally show poor cross-protective efficacy between different serotypes (44), vaccine strains should be carefully selected based on the epidemiology and antigenic properties of IBVs. It should be also noted that the pathogenicity and antigenicity of the recombinant virus in field condition could be different from laboratory animal study and insufficient efficacy of vaccines could accelerate viral evolution (45). Recently, an IBV that has highly similar recombinant pattern with the K163/21 virus was reported in Korea, which has the partial S1 gene of QX-like IBV(GI-19) between 724 and 1,102 bp in the S1 gene of GI-15 IBV (29). The emergence of such a diverse range of natural recombinant IBVs raises a concern that novel recombinants could possibly escape vaccine induced immunity and extensively spread in Korea possibly due to insufficient cross-neutralization between commercial vaccines and novel variants. Therefore, enhanced surveillance for newly emerged recombinants should be carried out, and IBV characteristics such as whole genome sequences, pathogenicity, and antigenicity should be quickly analyzed to prepare control measures.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was reviewed and approved by Konkuk University IACUC.

Author contributions

C-SS: conceptualization, validation, and project administration. D-HL, H-JK, H-CL, and AC: methodology. H-JK, Y-JC, and HL: formal analysis. D-HL, H-JK, Y-JC, and HL: investigation. H-JK and H-CL: resources. H-JK and AC: data curation. H-JK: writing—original draft preparation and visualization. D-HL, H-CL, AC, and C-SS: writing—review and editing. D-HL and C-SS: supervision and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Animal Disease Management Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant number: 122057-2).

Acknowledgments

We thank Ye-Ram Seo from Konkuk University for technical support.

Conflict of interest

H-CL and C-SS were employed by KHAV Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1107059/full#supplementary-material

References

1. Jackwood MW, de Wit S. Infectious bronchitis. In: Swayne DE, Boulianne M, Logue CM, McDougald LR, Nair V, Suarez DL, Wit S, Grimes T, Johnson D, Kromm M, Prajitno TY, Rubinoff I, Zavala G, editors. *Diseases of Poultry*. (2020). doi: 10.1002/9781119371199.ch4

2. Abdel-Moneim AS. Coronaviridae: Infectious Bronchitis Virus. In: Bayry J, editor. *Emerging and Re-emerging Infectious Diseases of Livestock*. Cham: Springer International Publishing. (2017). p. 133–66.

3. Cavanagh D, Casais R, Armesto M, Hodgson T, Izadkhasti S, Davies M, et al. Manipulation of the infectious bronchitis coronavirus genome for vaccine development and analysis of the accessory proteins. *Vaccine*. (2007) 25:5558–62. doi: 10.1016/j.vaccine.2007.02.046

4. Liu XL, Su JL, Zhao JX, Zhang GZ. Complete genome sequence analysis of a predominant infectious bronchitis virus (IBV) strain in China. *Virus Genes*. (2009) 38:56–65. doi: 10.1007/s11262-008-0282-5

5. Legnardi M, Tucciarone CM, Franzo G, Cecchinato M. Infectious bronchitis virus evolution, diagnosis and control. *Vet Sci*. (2020) 7:79. doi: 10.3390/vetsci7020079

6. Cavanagh D. Coronavirus avian infectious bronchitis virus. *Vet Res.* (2007) 38:281–97. doi: 10.1051/vetres:2006055

7. Bande F, Arshad SS, Bejo MH, Moeini H, Omar AR. Progress and challenges toward the development of vaccines against avian infectious bronchitis. *J Immunol Res.* (2015) 2015:424860. doi: 10.1155/2015/424860

8. Moore KM, Jackwood MW, Hilt DA. Identification of amino acids involved in a serotype and neutralization specific epitope with in the s1 subunit of avian infectious bronchitis virus. *Arch Virol.* (1997) 142:2249–56. doi: 10.1007/s007050050239

9. Wang CH, Huang YC. Relationship between serotypes and genotypes based on the hypervariable region of the S1 gene of infectious bronchitis virus. *Arch Virol.* (2000) 145:291–300. doi: 10.1007/s007050050024

10. Alvarado IR, Villegas P, Mossos N, Jackwood MW. Molecular characterization of avian infectious bronchitis virus strains isolated in Colombia during 2003. *Avian Dis.* (2005) 49:494–9. doi: 10.1637/7202-050304R.1

11. Zanaty A, Arafa AS, Hagag N, El-Kady M. Genotyping and pathotyping of diversified strains of infectious bronchitis viruses circulating in Egypt. *World J Virol.* (2016) 5:125–34. doi: 10.5501/wjv.v5.i3.125

12. Cavanagh D, Davis PJ, Mockett APA. Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Res.* (1988) 11:141–50. doi: 10.1016/0168-1702(88)90039-1

13. Promkuntod N, van Eijndhoven RE, de Vrieze G, Grone A, Verheije MH. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology.* (2014) 448:26–32. doi: 10.1016/j.virol.2013.09.018

14. Bouwman KM, Parsons LM, Berends AJ, Vries RPd, Cipollo JF, Verheije MH. Three amino acid changes in avian coronavirus spike protein allow binding to kidney tissue. *J Virol.* (2020) 94:e01363–19. doi: 10.1128/JVI.01363-19

15. Valastro V, Holmes EC, Britton P, Fusaro A, Jackwood MW, Cattoli G, et al. S1 gene-based phylogeny of infectious bronchitis virus: an attempt to harmonize virus classification. *Infect Genet Evol.* (2016) 39:349–64. doi: 10.1016/j.meegid.2016.02.015

16. Feng K, Wang F, Xue Y, Zhou Q, Chen F, Bi Y, et al. Epidemiology and characterization of avian infectious bronchitis virus strains circulating in southern China during the period from 2013-2015. *Sci Rep.* (2017) 7:6576. doi: 10.1038/s41598-017-06987-2

17. Lee HC, Jeong S, Cho AY, Kim KJ, Kim JY, Park DH, et al. Genomic analysis of avian infectious bronchitis viruses recently isolated in South Korea reveals multiple introductions of GI-19 Lineage (QX Genotype). *Viruses.* (2021) 13:1045. doi: 10.3390/v13061045

18. Lee HJ, Youn HN, Kwon JS, Lee YJ, Kim JH, Lee JB, et al. Characterization of a novel live attenuated infectious bronchitis virus vaccine candidate derived from a Korean nephropathogenic strain. *Vaccine.* (2010) 28:2887–94. doi: 10.1016/j.vaccine.2010.01.062

19. Choi KS, Lee EK, Jeon WJ, Park MJ, Kim JW, Kwon JH. Pathogenicity and antigenicity of a new variant of Korean nephropathogenic infectious bronchitis virus. *J Vet Sci.* (2009) 10:357–9. doi: 10.4142/jvs.2009.10.4.357

20. Lee EK, Jeon WJ, Lee YJ, Jeong OM, Choi JG, Kwon JH, et al. Genetic diversity of avian infectious bronchitis virus isolates in Korea between 2003 and 2006. *Avian Dis.* (2008) 52:332–7. doi: 10.1637/8117-092707-ResNote.1

21. Lim TH, Lee HJ, Lee DH, Lee YN, Park JK, Youn HN, et al. An emerging recombinant cluster of nephropathogenic strains of avian infectious bronchitis virus in Korea. *Infect Genet Evol.* (2011) 11:678–85. doi: 10.1016/j.meegid.2011.01.007

22. Yu L, Jiang Y, Low S, Wang Z, Nam SJ, Liu W, et al. Characterization of three infectious bronchitis virus isolates from China Associated with Proventriculus in Vaccinated Chickens. *Avian Dis.* (2001) 45:416–24. doi: 10.2307/1592981

23. Toffan A, Bonci M, Bano L, Bano L, Valastro V, Vascellari M, et al. Diagnostic and clinical observation on the infectious bronchitis virus strain Q1 in Italy. *Vet Ital.* (2013) 49:347–55. doi: 10.12834/VetIt.1303.01

24. Lim TH, Kim MS, Jang JH, Lee DH, Park JK, Youn HN, et al. Live attenuated nephropathogenic infectious bronchitis virus vaccine provides broad cross protection against new variant strains. *Poult Sci.* (2012) 91:89–94. doi: 10.3382/ps.2011-01739

25. Li L, Xue C, Chen F, Qin J, Xie Q, Bi Y, et al. Isolation and genetic analysis revealed no predominant new strains of avian infectious bronchitis virus circulating in South China during 2004-2008. *Vet Microbiol.* (2010) 143:145–54. doi: 10.1016/j.vetmic.2009.11.022

26. Kint J, Maier HJ, Jagt E. Quantification of infectious bronchitis coronavirus by titration *in vitro* and *in ovo*. *Methods Mol Biol.* (2015) 1282:89–98. doi: 10.1007/978-1-4939-2438-7_9

27. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* (2015) 1:vev003. doi: 10.1093/ve/vev003

28. Yan W, Qiu R, Wang F, Fu X, Li H, Cui P, et al. Genetic and pathogenic characterization of a novel recombinant avian infectious bronchitis virus derived from GI-1, GI-13, GI-28, and GI-19 strains in Southwestern China. *Poult Sci.* (2021) 100:101210. doi: 10.1016/j.psj.2021.101210

29. Youn SY, Lee JY, Bae YC, Kwon YK, Kim HR. Genetic and Pathogenic characterization of QX(GI-19)-recombinant infectious bronchitis viruses in South Korea. *Viruses*. (2021) 13:1163. doi: 10.3390/v13061163

30. Lim TH, Youn HN, Yuk SS, Kwon JH, Hong WT, Gwon GB, et al. Successful cross-protective efficacy induced by heat-adapted live attenuated nephropathogenic infectious bronchitis virus derived from a natural recombinant strain. *Vaccine.* (2015) 33:7370–4. doi: 10.1016/j.vaccine.2015.07.043

31. Callison SA, Hilt DA, Boynton TO, Sample BF, Robison R, Swayne DE, et al. Development and evaluation of a real-time Taqman RT-PCR assay for the detection of infectious bronchitis virus from infected chickens. *J Virol Methods.* (2006) 138:60–5. doi: 10.1016/j.jviromet.2006.07.018

32. Radmehri M, Talebi A, Ameghi Roudsari A, Mousaviyan SM, Gholipour MAJ, Taghizadeh M. Comparative study on the efficacy of MF 59, ISA70 VG, and nano-aluminum hydroxide adjuvants, alone and with nano-selenium on humoral immunity induced by a bivalent Newcastle+Avian influenza vaccine in chickens. *Arch Razi Inst.* (2021) 76:1213–20. doi: 10.22092/ari.2021.356666.1887

33. OIE. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals chapter 3. 3. 2. Avian infectious bronchitis virus. 2018. *World Organisation for Animal Health fouded as OIEAvian Infectious Bronchitis Virus*. [801]. Available online at: https://www.woah.org/en/what-we-do/standards/codes-and-manuals/terrestrial-manual-online-access/ (accessed May, 2018).

34. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints12. *Am J Epidemiol.* (1938) 27:493–7. doi: 10.1093/oxfordjournals.aje.a118408

35. Archetti I, Horsfall FL Jr. Persistent antigenic variation of influenza a viruses after incomplete neutralization in ovo with heterologous immune serum. *J Exp Med.* (1950) 92:441–62. doi: 10.1084/jem.92.5.441

36. Bande F, Arshad SS, Omar AR, Hair-Bejo M, Mahmuda A, Nair V. Global distributions and strain diversity of avian infectious bronchitis virus: a review. *Anim Health Res Rev.* (2017) 18:70–83. doi: 10.1017/S1466252317000044

37. Armesto M, Cavanagh D, Britton P. The replicase gene of avian coronavirus infectious bronchitis virus is a determinant of pathogenicity. *PLoS ONE.* (2009) 4:e7384. doi: 10.1371/journal.pone.0007384

38. Cheng J, Zhao Y, Xu G, Zhang K, Jia W, Sun Y, et al. The S2 Subunit of QX-type infectious bronchitis coronavirus spike protein is an essential determinant of neurotropism. *Viruses*. (2019) 11:972. doi: 10.3390/v11100972

39. Shan D, Fang S, Han Z, Ai H, Zhao W, Chen Y, et al. Effects of hypervariable regions in spike protein on pathogenicity, tropism, and serotypes of infectious bronchitis virus. *Virus Res.* (2018) 250:104–13. doi: 10.1016/j.virusres.2018.04.013

40. Liu S, Zhang X, Gong L, Yan B, Li C, Han Z, et al. Altered pathogenicity, immunogenicity, tissue tropism and 3'-7kb region sequence of an avian infectious bronchitis coronavirus strain after serial passage in embryos. *Vaccine.* (2009) 27:4630–40. doi: 10.1016/j.vaccine.2009.05.072

41. Phillips JE, Jackwood MW, McKinley ET, Thor SW, Hilt DA, Acevedol ND, et al. Changes in nonstructural protein 3 are associated with attenuation in avian coronavirus infectious bronchitis virus. *Virus Genes.* (2012) 44:63–74. doi: 10.1007/s11262-011-0668-7

42. Lin SY, Li YT, Chen YT, Chen TC, Hu CJ, Chen HW. Identification of an infectious bronchitis coronavirus strain exhibiting a classical genotype but altered antigenicity, pathogenicity, and innate immunity profile. *Sci Rep.* (2016) 6:37725. doi: 10.1038/srep37725

43. Choi KS, Jeon WJ, Lee EK, Kye SJ, Park MJ, Kwon JH. Development of an attenuated vaccine strain from a korean respiratory type infectious bronchitis virus. *Korean J Vet Res.* (2011) 51:193–201. doi: 10.14405/kjvr.2011.51.3.193

44. Kamble NM, Pillai AS, Gaikwad SS, Shukla SK, Khulape SA, Dey S, et al. Evolutionary and bioinformatic analysis of the spike glycoprotein gene of H120 vaccine strain protectotype of infectious bronchitis virus from India. *Biotechnol Appl Biochem.* (2016) 63:106–12. doi: 10.1002/bab.1298

45. Wickramasinghe IN, van Beurden SJ, Weerts EA, Verheije MH. The avian coronavirus spike protein. *Virus Res.* (2014) 194:37–48. doi: 10.1016/j.virusres.2014.10.009

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fvets.2023.1058133

OPEN ACCESS

EDITED BY
Christina Leyson,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Pavulraj Selvaraj,
Louisiana State University, United States
Laura Hoon-Hanks,
Inotiv, United States

*CORRESPONDENCE
Krisztián Bányai
✉ bkrota@hotmail.com

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 30 September 2022
ACCEPTED 06 January 2023
PUBLISHED 02 February 2023

CITATION
Varga-Kugler R, Ihász K, Marton S, Kaszab E, Marschang RE, Farkas S and Bányai K (2023) Genetic diversity among reptilian orthoreoviruses isolated from pet snakes and lizards. *Front. Vet. Sci.* 10:1058133. doi: 10.3389/fvets.2023.1058133

Genetic diversity among reptilian orthoreoviruses isolated from pet snakes and lizards

Renáta Varga-Kugler[1,2], Katalin Ihász[1], Szilvia Marton[1,2], Eszter Kaszab[1,2], Rachel E. Marschang[3], Szilvia Farkas[4] and Krisztián Bányai[1,2,5]*

[1]Veterinary Medical Research Institute, Budapest, Hungary, [2]National Laboratory for Infectious Animal Diseases, Antimicrobial Resistance, Veterinary Public Health and Food Chain Safety, Budapest, Hungary, [3]Laboklin GmbH & Co. KG, Bad Kissingen, Germany, [4]Department of Obstetrics and Food Animal Medicine Clinic, University of Veterinary Medicine, Budapest, Hungary, [5]Department of Pharmacology and Toxicology, University of Veterinary Medicine, Budapest, Hungary

Reovirus infections in reptiles are frequently detected and associated with various clinical diseases; yet, our knowledge about their genetic diversity and evolutionary relationships remains limited. In this study, we characterize at the genomic level five reptile origin orthoreovirus strains isolated from exotic snakes and lizards in Hungary and Germany. The genomic organization of the study strains was similar to that of the representative strains of reptile origin reoviruses belonging to species *Reptilian orthoreovirus* and *Testudine orthoreovirus*. Additionally, all five study strains clustered with the bush viper origin reference *Reptilian orthoreovirus* strain, 47/02. The nucleotide sequence divergence among strains fell from 56.64 to 99.36%. Based on genome segment constellations two well separated groups were observed, which may represent two genetic lineages of reptilian orthoreoviruses we tentatively referred here as genogroups, classifying two squamata origin strains with available whole genome sequences into genogroup I (GGI) and four strains into genogroup II (GGII). The representative GGI and GGII *Reptilian orthoreovirus* strains are characterized by moderate-to-high nucleotide and amino acid similarities within genogroups (range, 69.45 to 99.36% and 74.64 to 100.00%), whereas lower nucleotide and amino acid similarities (range, 56.64 to 77.24% and 54.53 to 93.85%) and different structures of the bicistronic S1 segment were found between genogroups. Further studies are needed to explore the genomic diversity among reptilian reoviruses of squamata origin; this would be critical to establish a robust classification system for these viruses and to see if interaction among members of distinct lineages may result in viable progenies with novel genetic features.

KEYWORDS

***Reptilian orthoreovirus*, phylogeny, next generation sequencing, snake, lizard**

1. Introduction

Members of the order *Reovirales* are double-stranded RNA viruses infecting a wide range of host species, including plants, fungi, protists and animals. The order is divided into two families, *Sedoreoviridae* and *Spinareoviridae* containing 6 and 9 genera, respectively. According to the International Committee on Taxonomy of Viruses (ICTV) the genus *Orthoreovirus* belonging to the family *Spinareoviridae*, is currently divided into ten species: *Avian orthoreovirus* (ARV), *Baboon orthoreovirus* (BRV), *Broome orthoreovirus* (BroRV), *Mahlapitsi orthoreovirus* (MAHLV), *Mammalian orthoreovirus* (MRV), *Nelson Bay orthoreovirus* (NBV), *Neoavian orthoreovirus* (NeARV), *Piscine orthoreovirus* (PRV), *Reptilian orthoreovirus* (RRV) and *Testudine orthoreovirus* (TRV) (1, 2). Orthoreoviruses are non-enveloped viruses with an

icosahedral capsid 70–80 nm in diameter. The 23 kilobasepairs (kbp) viral genome consists of 10 segments grouped into three categories based on their size: three large segments (L1-L3), three medium segments (M1-M3), and four small segments (S1-S4). With the exception of the S1 or S4 segment, which might be bi- or tricistronic, each genome segment encodes a single protein (3).

The genome size of reptile origin orthoreoviruses is about 24 kbp (range, 23,957 to 24,043 bp, based on two fully sequenced viral genomes) (4, 5). Genetic analysis of a short fragment of the RNA-dependent RNA polymerase (RdRp) gene identified two or three genetic clades; however more recent data indicate that reoviruses of reference snake and tortoise origin strains belong to different *Orthoreovirus* species (4–8). At present, the few known isolates of reptilian reoviruses are officially classified into two distinct *Orthoreovirus* species: *Reptilian* and *Testudine orthoreovirus*, latter represented by only a single strain isolated from a spur-thighed tortoise (5, 7). Reovirus infections in reptiles may be asymptomatic, but have been associated with various clinical diseases and experimental infection induced severe respiratory disease in snakes (9, 10). Reptilian orthoreoviruses induce giant cells *in vivo* and *in vitro*, a feature linked to the expression of fusion associated small transmembrane (FAST) protein encoded by the S1 genome segment (6, 10, 11). Although reptilian reoviruses are easily isolated and frequently detected, our knowledge about their genetic diversity and evolutionary relationships remained limited.

Genetic diversity within orthoreoviruses stems from accumulation of point mutations generated by the viral RdRp that lacks proofreading activity and genetic reassortment of cognate genomic segments, which may result in new combinations of gene constellations (12, 13). In addition to reassortment events among members of a particular virus species, exchange of homologous genomic segment has been hypothesized to occur among viruses belonging to different *Orthoreovirus* species (14, 15). Moreover, intrasegmental recombination involving cognate genomic segments may also occur among homologous viruses and this mechanism may further increase the genetic diversity within orthoreoviruses (15).

In the present study we performed whole genome sequencing of five reptilian orthoreovirus strains isolated from different exotic species in Hungary and Germany. Because of the segmented nature of their genome and the evidence of various reoviral strategies to increase genetic diversity, all genome segments were subjected to phylogenetic calculations in order to uncover their evolutionary relationships.

2. Materials and methods

2.1. Virus detection and isolation

The strains analyzed in the current study are listed in Table 1. Strain IBD26/00 was isolated from the liver of a *Boa constictor* that died with inclusion body disease as described elsewhere (16). For strains 2013/12, 2013/54, 2013/47, and KP3, organ samples (heart, tongue, throttle, lung, esophagus, stomach, intestine, liver, and kidney) were collected in sterile PBS solution from succumbed animals. We have no information about the health, body condition and the circumstances of death of the animals. A pan-reovirus-specific reverse transcription nested polymerase chain reaction targeting a conservative region of the orthoreoviral RdRp gene was used for detection of reptilian orthoreoviruses from the supernatant of the pooled organ samples (8). In order to isolate the orthoreoviruses, the supernatants of the pooled organ samples were propagated in a 6 well culture dish on viper heart (VH2) or iguana heart (IgH2) continuous cell lines. After 4 days when cytopathic effect appeared, isolates were used for VH2 or IgH2 cell lines applying 75-cm^2 flasks. On day 4 post infection when syncytium formation appeared (Figure 1) the virus isolates were harvested by freezing and thawing. Cell culture supernatants containing the virus isolates were stored at −80°C and periodically passaged on VH2 or IgH2 cells to maintain their viability.

2.2. Whole genome sequencing

The viral RNA was extracted from polyethylene glycol concentrated cell culture supernatant using TRIzol Reagent (Sigma-Aldrich) according to the manufacturer's recommendations. Extracted RNA was processed for whole genome sequencing as described in details previously (4). Random primed reverse transcription was used to generate complementary DNA (cDNA) from viral RNA. cDNA libraries were prepared using the NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent (New England Biolabs, Beverly, MA, USA) using the Ion Torrent Xpress barcode adapters (Life Technologies, Carlsbad, CA, USA) according to the instructions recommended by the manufacturers. The emulsion PCR and subsequent templated bead enrichment were performed with a OneTouch v2 instrument and Ion OneTouchTMES, respectively. Sequencing was carried out on a 316 chip using the Ion Torrent Personal Genome Machine® (Life Technologies). To obtain the 5' and 3' terminal sequences of the segments, DNA oligonucleotides were ligated to each end of the dsRNA as described in detail elsewhere (17, 18). The confirmation of next generation sequencing (NGS) data and the completion of missing parts of the sequences was carried out with additional oligonucleotide primers in PCR and Sanger sequencing reactions (not shown).

2.3. Computer analysis

Sequence reads generated by IonTorrent sequencing were assembled using the CLC Genomics Workbench version 7 (http://www.clcbio.com). Contigs were aligned with Sanger sequencing reads and were edited using Geneious Prime and AliView softwares (19, 20). BLASTn and BLASTx algorithms were used to identify homologous genes among sequences deposited in GenBank (21). Codon-based multiple sequence alignments were generated using the Muscle algorithm within the TranslatorX (22). Phylogenetic analysis was performed and sequence identity values were calculated using the MEGA6 and MEGAX package (23, 24). Gene-specific substitution models were evaluated and the best-fit models were selected based on the Bayesian information criterion. Maximum-likelihood trees were generated, and tree topologies were validated by bootstrap analysis (100) as implemented in MEGAX. A short fragment of the RdRp gene was analyzed by neighbor-joining method. In aim to discover potential recombination

TABLE 1 List of virus strains analyzed in the current study.

Sample	Species		Isolation		Source
			Location	Year	
IBD26/00	Boa constrictor	*Boa constrictor*	Germany	2000	Pet owner
KP3	Ball python	*Python regius*	Budapest, Hungary	2013	Pet shop*
2013/12	Schneider's skink	*Eumeces schneideri*	Budapest, Hungary	2013	Pet shop*
2013/54	Green iguana	*Iguana iguana*	Budapest, Hungary	2013	Pet shop*
2013/47	Unknown snake		Hungary	2013	Pet owner

*Originated from the same pet shop.

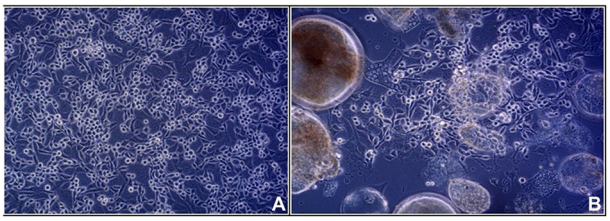


FIGURE 1
Viper heart (VH2) cell line: **(A)** control **(B)** syncytium formation induced by infection with KP3 reptilian orthoreovirus strain.

events Recombination Detection Program v4.100 (RDP4) was used applying the default parameters for the embedded methods BootScan, Chimera, GENECONV, MaxChi, RDP, SiScan, and 3Seq (25).

2.4. Sequence data availability

The genome sequences of strains 2013/12, 2013/54, 2013/47, IBD26/00, and KP3 were deposited in the GenBank database under the accession numbers MN313188–MN313197, MN313198–MN313207, MN313218–MN313227, MN313228–MN313237, and MN313238–MN313247, respectively. Additional sequence data were obtained from GenBank. The accession numbers of the corresponding segments of the reference reovirus strains are shown on the phylogenetic trees (Figure 2).

3. Results and discussion

3.1. Genomic features

Five reovirus strains were isolated in 2000 in Germany and around a decade later in Hungary (Table 1). All of the detected viruses possessed the ability to cause giant cell formation in cell culture (Figure 1), similar to other reptilian orthoreoviruses (6, 10, 11). The complete genomes of the strains ranged from 23,990 to 24,033 bp in length and the size of the individual genome segments ranged from 1,209 bp (S4) to 3,970 bp (L1) (Table 2). The G+C content was 45.9–47.6%. The genomic organization of the strains was similar to that of the two previously described reptile origin reovirus strains, 47/02 and CH1197/96 (4, 5). With the exception of S1, which was bicistronic in the studied strains, all genome segments were monocistronic. Using the Geneious Prime software and the BLASTx algorithm the following orthoreoviral proteins were identified by ORF prediction and sequence comparison: λA (core shell), λB (core RdRp), λC (core turret), μA (core NTPase), μB (outer shell), μNS (NS factory), σA (core clamp), σB (outer clamp), σNS (NS RNAb), σC (outer fiber), and p14 (FAST) (4, 26). The 3′ untranslated regions were 41-104 nt long and included the UCAUC 3′ terminal consensus sequence conserved between *Orthoreovirus* species. The 5′ untranslated regions were 12–31 nt long and found to be highly conserved at the termini with the exception of the fourth position which can hold an adenine or a cytosine (Table 2).

3.2. Comparison with representative members of other *Orthoreovirus* species

Nucleotide (nt) sequences of each genome segment were analyzed separately. With the exception of the ORFs encoding the proteins μA and σA, in which case BRV and BroRV clustered separately, phylogenetic analysis revealed four well separated lineages within the genus *Orthoreovirus*, reaffirming previous findings: (i) viruses representing the three serotypes of the classic MRV (Jones, Lang, Dearing); (ii) ARV and NeARV with those of bat origin: NBV and Pulau reoviruses; (iii) RRV and TRV along with BRV, BrRV, and MAHLV; and (iv) PRV (Figure 2) (4, 27, 28). The sequence identity values of homologous genes and the proteins they encode ranged between 29.86–63.98% and 10.84–66.72%, respectively, when study strains were compared with representative members of the non-reptilian origin *Orthoreovirus* species (data not shown). These identities mostly remained below the cut-off value currently used as demarcation criteria for distinct species within the genus *Orthoreovirus* (<60% for nt, <65% for the core protein aa and <35% for the outer capsid proteins aa sequence), but in the comparison of the μB protein we found higher identities that were above this cut-off value. The μB protein tended to be conserved to a similar degree as the core proteins as observed previously for other orthoreoviruses (5, 29). Furthermore, in the comparison with MAHLV we found identity values above the demarcation criteria defined for distinct species for three additional sequences (λB 65.76–66.72% aa; λA 60.37–62.23% nt; and λB 62.37–63.98% nt identity) confirming the tree topologies where MAHLV appeared to be the closest relative of reptile origin orthoreoviruses (27, 30, 31).

FIGURE 2
Unrooted phylogenetic trees based on the nucleotide sequences of the corresponding genes of different orthoreoviruses. Phylogenetic trees were calculated by the maximum-likelihood method based on the best-fit model selected for each gene. The scale bar is proportional to the genetic distance. Bootstrap values >60 are shown at the branch nodes. ARV, *Avian orthoreovirus*; BRV, *Baboon orthoreovirus*; BroRV, *Broome reovirus*; MAHLV, *Mahlapitsi orthoreovirus*; MRV, *Mammalian orthoreovirus*; NBV, *Nelson Bay orthoreovirus*; PRV, *Piscine orthoreovirus*; RRV, *Reptilian orthoreovirus* and TRV, *Testudine orthoreovirus*.

3.3. Comparison with reference reptilian and testudines RVs

Orthoreoviruses isolated from reptile hosts presented in this study composed one monophyletic clade together with the two previously described orthoreoviruses from reptiles, one isolated from a bush viper (*Atheris squamigera*) and the other from a spur-thighed tortoise (*Testudo graeca*), suggesting a common evolutionary relationship of these viruses (4, 5). All of the newly described reoviruses originating from different squamata species composed a common group with 47/02, the reference strain of the species *Reptilian orthoreovirus*, which was also isolated from a squamate host. On the other hand, CH1197/96, the reference strain of *Testudine orthoreovirus* clustered separately in all phylogenetic trees indicating differences in orthoreoviruses of phylogenetically distant reptilian hosts (Figure 2).

Confirming our previous findings on the existence of different reptile origin reovirus species, comparison of each of the newly described squamate reoviruses with sequences from CH1197/96 showed that none of the identity values reached the cut-off defined for identical species (>75% for nt, >55% for the core protein aa and >85% for the outer capsid proteins aa sequence) (5). Furthermore, identity values remained below the species demarcation criteria for distinct *Orthoreovirus* species in the case of μNS (48.49–49.28%), σB (50.17–51.30%), and σC (48.99–49.66%) nucleotide and λC (62.14–62.76%) protein sequences (Figure 3).

TABLE 2 General genomic features of the studied strains.

Genome segment	Size (nt)	Length (nt) of			Sequence at the termini 5′ end/ 3′ end	Encoded protein	Protein size (aa)
		5′ UTR	ORF	3′ UTR			
L1	3,967–3,970	13	3,870–3,873	84	GUUCUU/UUCAUC	λA (core shell)	1,289–1,290
L2	3,901–3,933	14	3,846–3,870	41–49	GUUCUU/UUCAUC	λC (core turret)	1,281–1,289
L3	3,847–3,849	14–16	3,786	47	GUUCUU/UUCAUC[#]	λB (core RdRp)	1,261
M1	2,480–2,488	27	2,373–2,382	78–81	GUUAUU/UUCAUC	μNS (NS factory)	790–793
M2	2,339–2,349	13–14	2,232–2,283	51–104	GUUCUU/UUCAUC*	μA (core NTPase)	743–760
M3	2,130–2,134	26	2,031–2,034	73–74	GUUAUU/UUCAUC	μB (outer shell)	676–677
S1	1,499–1,503	22–23	360–381	66–70	GUUAUU/UUCAUC	p14 (FAST)	119–126
			1,050–1,053			σC (outer fiber)	349–350
S2	1,314	12	1,251	51	GUUAUU/UUCAUC	σA (core clamp)	416
S3	1,279–1,288	31	1,161–1,170	86–87	GUUAUU/UUCAUC	σB (outer clamp)	386–389
S4	1,209–1,212	24–25	1,110–1,113	74	GUUAUU/UUCAUC	σNS (NS RNAb)	369–370

[#]Except: IDB 26/00 GUUAUU. *Except: 2013/54 GUUAUU.

3.4. Phylogeny of the genome of squamata reoviruses

Among the newly described reptilian orthoreoviruses two different lineages can be observed: strain 2013/54 (green iguana, *Iguana iguana*) composed a clade with reference RRV strain 47/02 (bush viper, *Atheris squamigera*) while strains 2013/12 (Schneider's skink, *Eumeces schneideri*), 2013/47 (unknown snake species) and KP3 (ball python, *Python regius*) made up another clade together with IBD 26/00 (boa constrictor), the latter clustering on a separate branch but in close relationship with the other two isolates (Figure 2). We clustered these lineages into genogroups (GG) and designated 47/02-like strains as GGI and KP3-like strains as GGII. Most of the sequence identities between the strains within each of the two genogroups reached the demarcation criteria for identical species (Figure 3). The only exception was the σC nucleotide sequence, in which case we found slightly lower values (69.45–69.84%) between IBD26/00 and the three other strains within the same lineage (2013/47, 2013/12, KP3).

In the comparison between the two genogroups, the results were more diverse. In the pairwise comparison of the nucleotide sequence data only the λA (75.20–76.42%) and σNS (75.52–77.24%) nt identities reached the cut-off value defined for identical species. In addition, the identity values for the μNS (56.59–57.56%) and σC (58.50–60.52%) ORFs remained below the cut-off value for distinct species. In the case of the other six genes, the identities fell between the two cut-off values, which gives no recommendation for species classification: λB (74.76–75.13%), λC (70.00–70.83%), μA (69.56–71.23%), μB (74.05–76.13%), σA (73.32–75.00%), and σB (63.28–65.46%).

The identity values calculated for the amino acid sequences of the outer capsid proteins were all above the species demarcation criteria for identical species (μB 89.00–89.30%, σB 69.95–70.74%, and σC 60.23–62.54%) as were the values for the λA (92.92–93.85%) and λB (89.69–90.17%) inner core proteins. However, the analysis of the more conserved inner core proteins revealed somewhat lower amino acid identities, falling under the cut-off value for identical species, in the case of λC (80.48–80.87%), μA (80.87–82.77%), and σA (84.38–85.34%) proteins.

The pairwise identity values affirmed the lineage diversification seen on the phylogenetic trees; strains grouping on different branches showed only moderate or low similarity. On the other hand, strains belonging to the same lineage showed high similarity. Also, the structure of the bicistronic S1 segment was slightly different in the two groups; GGII strains (2013/12, 2013/47, IBD26/00, and KP3) possessed two overlapping ORFs on the S1 segment, while the ORFs on the S1 segment of GGI strains (47/02 and 2013/54) were non-overlapping. Based on RDP4 analysis, recombination event was not detected in the protein coding genes of the squamata origin orthoreoviruses including the reference strain of RRVs and strains analyzed in this study.

4. Concluding remarks

Data about the genetic diversity among reptilian orthoreoviruses is scarce and most data obtained so far comes from animals held in captivity. Infection of wild reptiles with orthoreovirus has not been thoroughly investigated and thus the epizootiology of this virus in nature remains largely unclear; nonetheless, the few serological investigations performed to date confirmed the circulation of reoviruses under natural circumstances (32, 33). In this study, we described and analyzed five reptile origin orthoreovirus strains at the genomic level. The study strains originated from snakes and lizards, whose carcasses were obtained from pet owners and pet shops in two European countries, Germany and Hungary. Although the sample size was small and the strains were collected only from two countries within a narrow timeframe, this study enabled the comparison of orthoreoviruses with diverse host species origins. Apparently, the strict regulations on the trade of reptiles and the limited sources of animals may, at some extent, limit the diversity of known

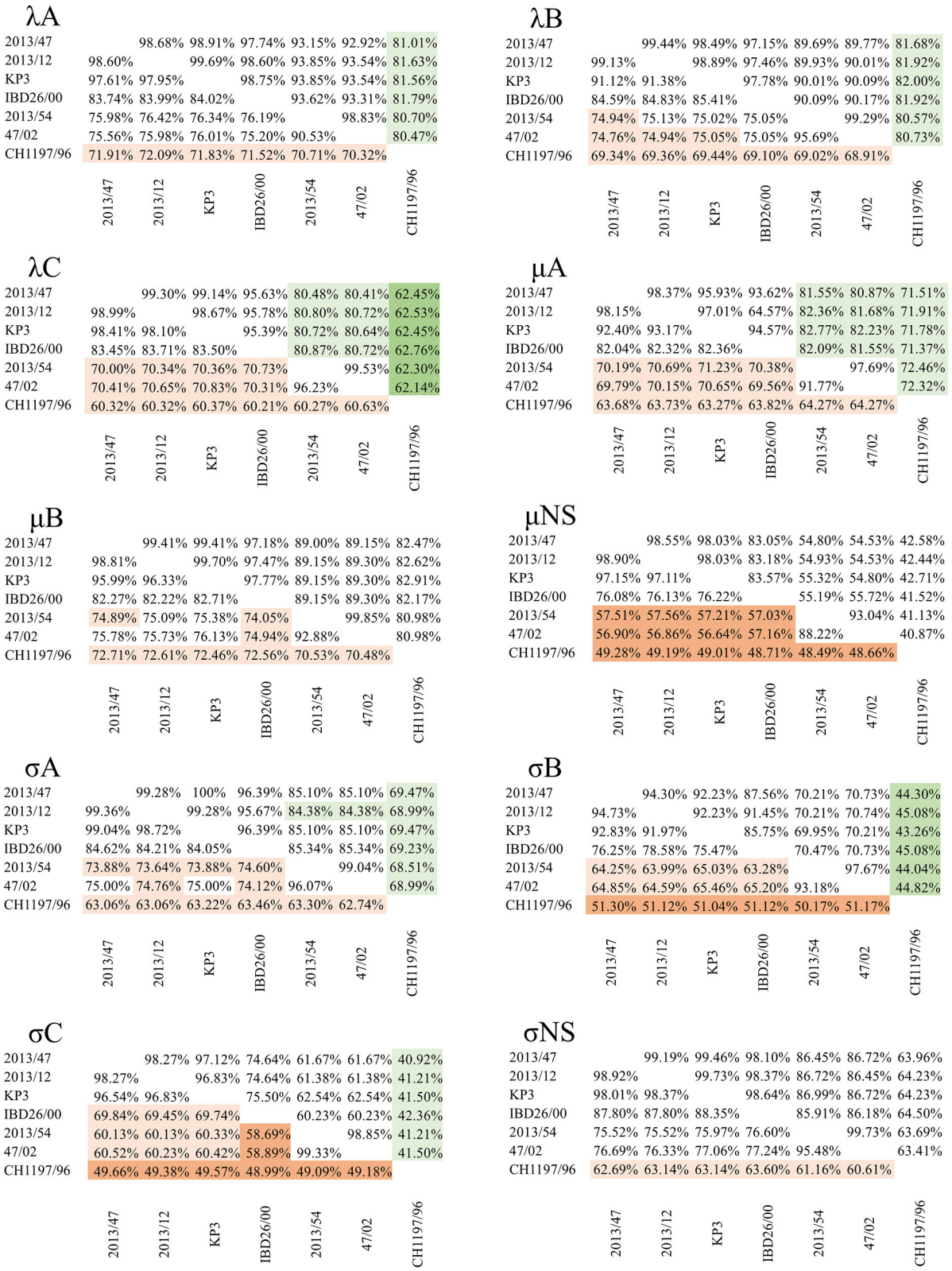


FIGURE 3
Pairwise identity values of nucleotide **(lower left triangle)** and amino acid **(upper right triangle)** sequences of the corresponding genes from reptilian origin orthoreoviruses. Values under the demarcation criteria for different species are indicated by dark orange (nt) and dark green (aa); values between the two cut-off values (i.e., above the cut-off value for different but under the value for identical species) are indicated by light orange (nt) and light green (aa) colors.

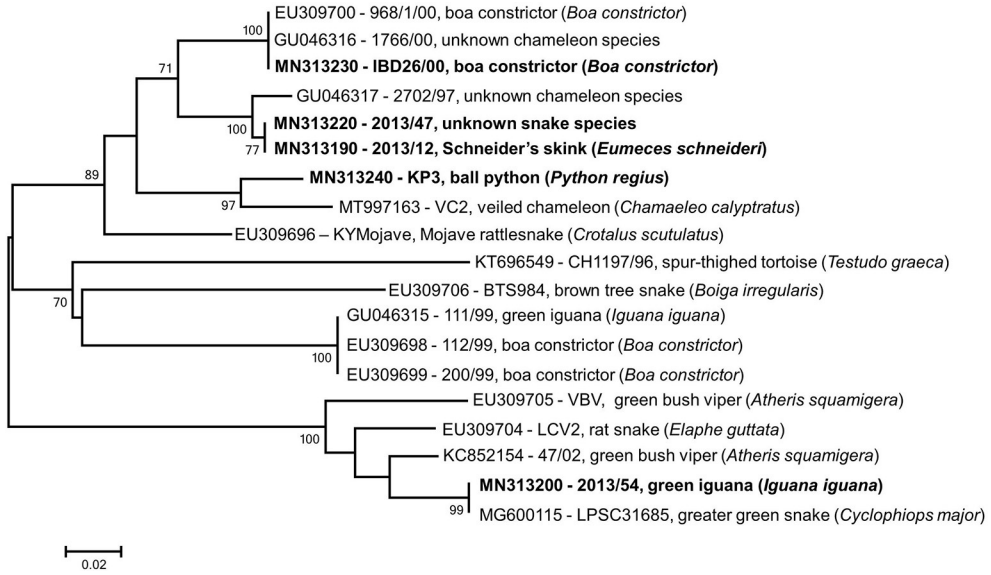


FIGURE 4
Phylogenetic tree based on partial RdRp gene using overlapping regions of reference strains available in GenBank. Taxa are identified by accession number, strain name and host species; study strains are highlighted. The scale bar is proportional to the genetic distance. Bootstrap values greater than 60 are shown at the branch nodes.

virus variants circulating among captive snakes and lizards and the observed viral diversity may not necessarily reflect the true viral diversity which could be isolated from populations of wild reptiles. Despite these shortcomings, our analyses revealed some notable features of the genetic diversity within reptile origin reoviruses.

All lizard and snake origin reovirus strains we analyzed in this study comprised a common phylogenetic cluster (*Reptilian orthoreovirus*) and showed the closest relationship with the only representative of the *Testudine orthoreovirus* (4, 5). Using available GenBank entries we generated an alignment containing short overlapping fragments of the RdRp gene (λB); next, we performed phylogenetic analysis that revealed additional diversity among squamata origin orthoreoviruses (Figure 4). In this analysis, three and five reference strains could be classified into the proposed genetic lineages, GGI and GGII, respectively. In addition, another two putative lineages of squamata origin reptilian orthoreoviruses, represented by strains BTS984 and 111/99, respectively, were identified. Although these strains showed low sequence similarity along a ∼100 nt long fragment with the single known member of *Testudine orthoreovirus*, they seemed to share common evolutionary roots. In general, evidence shows that very short gene sequences might be suitable to perform taxonomic classification of very diverse viruses (8, 34, 35). However, experience with reptilian reoviruses is limited and further data are needed to determine whether these divergent unclassified reovirus strains are members of the newly established *Testudine orthoreovirus* species, or, they represent distantly related lineages of the *Reptilian orthoreovirus* species, or, they constitute completely new virus species. The extension of genome sequencing will be required to update and validate the current taxonomy of reptilian reoviruses and establish a robust classification system below the virus species level.

Data availability statement

The data presented in the study are deposited in the GenBank repository under the accession numbers MN313188-MN313197, MN313198-MN313207, MN313218-MN313227, MN313228-MN313237, and MN313238-MN313247.

Author contributions

KB and SF contributed to the study conception and design. Material preparation, data collection, and analysis were performed by RV-K, KI, SM, EK, and RM. The first draft of the manuscript was written by RV-K and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Scientific Research Fund of Hungary [K108727]. Additional support was provided by the National Laboratory for Infectious Animal Diseases, Antimicrobial Resistance, Veterinary Public Health, and Food Chain Safety, RRF-2.3.1-21-2022-00001.

Conflict of interest

RM is employed by Laboklin GmbH & Co. KG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Matthijnssens J, Attoui H, Bányai K, Brussaard CPD, Danthi P, Del Vas M, et al. ICTV virus taxonomy profile: Sedoreoviridae 2022. *J Gen Virol.* (2022) 103:001782. doi: 10.1099/jgv.0.001782

2. Matthijnssens J, Attoui H, Bányai K, Brussaard CPD, Danthi P, Del Vas M, et al. ICTV virus taxonomy profile: Spinareoviridae 2022. *J Gen Virol.* (2022) 103:001781. doi: 10.1099/jgv.0.001781

3. Day JM. The diversity of the orthoreoviruses: Molecular taxonomy and phylogentic divides. *Infect Genet Evol.* (2009) 9:390–400. doi: 10.1016/j.meegid.2009.01.011

4. Bányai K, Borzák R, Ihász K, Fehér E, Dán Á, Jakab F, et al. Whole-genome sequencing of a green bush viper reovirus reveals a shared evolutionary history between reptilian and unusual mammalian orthoreoviruses. *Arch Virol.* (2014) 159:153–8. doi: 10.1007/s00705-013-1796-2

5. Kugler R, Marschang RE, Ihász K, Lengyel G, Jakab F, Bányai K, et al. Whole genome characterization of a chelonian orthoreovirus strain identifies significant genetic diversity and may classify reptile orthoreoviruses into distinct species. *Virus Res.* (2016) 215:94–8. doi: 10.1016/j.virusres.2016.02.005

6. Ihász K, Farkas SL, Lengyel G, Bányai K, Gál J. Occurence and genetic diversity of orthoreoviruses in exotic reptile species in Hungary. *Magy Állatorv Lapja.* (2014) 136:247–52.

7. Marschang RE. Viruses infecting reptiles. *Viruses.* (2011) 3:2087–126. doi: 10.3390/v3112087

8. Wellehan JFX, Childress AL, Marschang RE, Johnson AJ, Lamirande EW, Roberts JF, et al. Consensus nested PCR amplification and sequencing of diverse reptilian, avian, and mammalian orthoreoviruses. *Vet Microbiol.* (2009) 133:34–42. doi: 10.1016/j.vetmic.2008.06.011

9. Dandár E, Borzák R, Bányai K, Farkas SL. Orthoreovirus-caused diseases in reptiles, birds and mammals. *Magy Állatorv Lapja.* (2012) 134:564–73.

10. Latney LV, Wellehan J. Selected emerging infectious diseases of squamata. *Vet Clin North Am Exot Anim Pract.* (2013) 16:319–38. doi: 10.1016/j.cvex.2013.01.003

11. Duncan R, Corcoran J, Shou J, Stoltz D. Reptilian reovirus: a new fusogenic orthoreovirus species. *Virology.* (2004) 319:131–40. doi: 10.1016/j.virol.2003.10.025

12. Farkas SL, Dandár E, Marton S, Fehér E, Oldal M, Jakab F, et al. Detection of shared genes among Asian and European waterfowl reoviruses in the whole genome constellations. *Infect Genet Evol.* (2014) 28:55–7. doi: 10.1016/j.meegid.2014.08.029

13. Kugler R, Dandár E, Fehér E, Jakab F, Mató T, Palya V, et al. Phylogenetic analysis of a novel reassortant orthoreovirus strain detected in partridge (Perdix perdix). *Virus Res.* (2016) 215:99–103. doi: 10.1016/j.virusres.2015.11.018

14. Dandár E, Farkas SL, Marton S, Oldal M, Jakab F, Mató T, et al. The complete genome sequence of a European goose reovirus strain. *Arch Virol.* (2014) 159:2165–9. doi: 10.1007/s00705-014-2003-9

15. Farkas SL, Marton S, Dandár E, Kugler R, Gál B, Jakab F, et al. Lineage diversification, homo- and heterologous reassortment and recombination shape the evolution of chicken orthoreoviruses. *Sci Rep.* (2016) 6:36960. doi: 10.1038/srep36960

16. Darke S, Marschang RE, Hetzel U, Reinacher M. Experimental infection of Boa constrictor with an orthoreovirus isolated from a snake with inclusion body disease. *J Zoo Wildl Med.* (2014) 45:433–6. doi: 10.1638/2013-0194R.1

17. Bányai K, Dandár E, Dorsey KM, Mató T, Palya V. The genomic constellation of a novel avian orthoreovirus strain associated with runting-stunting syndrome in broilers. *Virus Genes.* (2011) 42:82–9. doi: 10.1007/s11262-010-0550-z

18. Lambden PR, Cooke SJ, Caul EO, Clarke IN. Cloning of noncultivatable human rotavirus by single primer amplification. *J Virol.* (1992) 66:1817–22. doi: 10.1128/jvi.66.3.1817-1822.1992

19. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* (2012) 28:1647–9. doi: 10.1093/bioinformatics/bts199

20. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* (2014) 30:3276–8. doi: 10.1093/bioinformatics/btu531

21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

22. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* (2010) 38:W7–13. doi: 10.1093/nar/gkq291

23. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 60. *Mol Biol Evol.* (2013) 30:2725–9. doi: 10.1093/molbev/mst197

24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K, MEGA X. Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* (2018) 35:1547–9. doi: 10.1093/molbev/msy096

25. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* (2015) 1:vev003. doi: 10.1093/ve/vev003

26. Coombs KM. Reoviruses: Molecular biology. In: *Encyclopedia of virology.* 3rd ed. Mahy BWJ, Van Regenmortel MHV (eds). New York: Academic Press (2009). p. 390–9. doi: 10.1016/B978-012374410-4.00485-4

27. Jansen van Vuren P, Wiley M, Palacios G, Storm N, McCulloch S, Markotter W, et al. Isolation of a novel fusogenic orthoreovirus from Eucampsipoda africana bat flies in South Africa. *Viruses.* (2016) 8:65. doi: 10.3390/v8030065

28. Palacios G, Lovoll M, Tengs T, Hornig M, Hutchison S, Hui J, et al. Heart and skeletal muscle inflammation of farmed salmon is associated with infection with a novel reovirus. *PLoS ONE.* (2010) 5:e11487. doi: 10.1371/journal.pone.0011487

29. Farkas SL, Varga-Kugler R, Marton S, Lengyel G, Palya V, Bányai K. Genomic sequence and phylogenetic analyses of two novel orthoreovirus strains isolated from Pekin ducks in 2014 in Germany. *Virus Res.* (2018) 257:57–62. doi: 10.1016/j.virusres.2018.09.001

30. Markussen T, Dahle MK, Tengs T, Løvoll M, Finstad ØW, Wiik-Nielsen CR, et al. Sequence analysis of the genome of piscine orthoreovirus (PRV) associated with heart and skeletal muscle inflammation (HSMI) in Atlantic salmon (Salmo salar). *PLoS ONE.* (2013) 8:e70075. doi: 10.1371/journal.pone.0070075

31. Fukase Y, Minami F, Masuda T, Oi T, Takemae H, Ishida H, et al. Genetic diversity, reassortment, and recombination of mammalian orthoreoviruses from Japanese porcine fecal samples. *Arch Virol.* (2022) 167:2643–52. doi: 10.1007/s00705-022-05602-8

32. Gravendyck M, Ammermann P, Marschang RE, Kaleta EF. Paramyxoviral and reovirual infections of iguanas on Honduran islands. *J Wildl Dis.* (1998) 34:33–8. doi: 10.7589/0090-3558-34.1.33

33. Marschang RE, Donahoe S, Manvell R, Lemos-Espinal J. Paramyxovirus and reovirus infections in wild-caught Mexican lizards (Xenosaurus and Abronia spp). *J Zoo Wildl Med.* (2002) 33:317–21. doi: 10.1638/1042-7260(2002)033[0317:PARIIW]2.0.CO;2

34. Wellehan JF, Johnson AJ, Harrach B, Benkö M, Pessier AP, Johnson CM, et al. Detection and analysis of six lizard adenoviruses by consensus primer PCR provides further evidence of a reptilian origin for the atadenoviruses. *J Virol.* (2004) 78:13366–9. doi: 10.1128/JVI.78.23.13366-13369.2004

35. Holbrook MG, Anthony SJ, Navarrete-Macias I, Bestebroer T, Munster VJ, van Doremalen N. Updated and validated pan-coronavirus PCR assay to detect all coronavirus genera. *Viruses.* (2021) 13:599. doi: 10.3390/v13040599

frontiers | Frontiers in Veterinary Science

TYPE Brief Research Report
PUBLISHED 20 February 2023
DOI 10.3389/fvets.2023.1094052

OPEN ACCESS

EDITED BY
Iryna Goraichuk,
Agricultural Research Service (USDA), United States

REVIEWED BY
Oliver Lung,
Canadian Food Inspection Agency (CFIA), Canada
Christopher Lewis Netherton,
The Pirbright Institute, United Kingdom
Denis V. Kolbasov,
Federal Research Center of Virology and Microbiology, Russia

*CORRESPONDENCE
Dong-Hun Lee
✉ donghunlee@konkuk.ac.kr

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 09 November 2022
ACCEPTED 23 January 2023
PUBLISHED 20 February 2023

CITATION
Hyeon J-Y, Tseren-Ochir E-O, Lee D-H, Nahm S-S, Gladue DP, Borca MV, Song C-S and Risatti GR (2023) Whole genome sequencing and phylogenetic analysis of African swine fever virus detected in a backyard pig in Mongolia, 2019. *Front. Vet. Sci.* 10:1094052. doi: 10.3389/fvets.2023.1094052

Whole genome sequencing and phylogenetic analysis of African swine fever virus detected in a backyard pig in Mongolia, 2019

Ji-Yeon Hyeon[1,2], Erdene-Ochir Tseren-Ochir[3], Dong-Hun Lee[2]*, Sang-Soep Nahm[2], Douglas P. Gladue[4], Manuel V. Borca[4], Chang-Seon Song[2,5] and Guillermo R. Risatti[1]

[1]Department of Pathobiology and Veterinary Science, University of Connecticut, Storrs, CT, United States, [2]College of Veterinary Medicine, Konkuk University, Seoul, Republic of Korea, [3]Department of Infectious Diseases and Microbiology, School of Veterinary Medicine, Mongolian University of Life Sciences, Ulaanbaatar, Mongolia, [4]Plum Island Animal Disease Center, Agriculture Research Service, US Department of Agriculture, Greenport, NY, United States, [5]KCAV Co., Ltd., Seoul, Republic of Korea

African swine fever (ASF) is a highly contagious and fatal disease affecting domestic and wild pigs caused by the African swine fever virus (ASFV). Since the first outbreak in China in August 2018, ASF has spread rapidly in Asia. and the first case in Mongolia was confirmed in January 2019. In this study, we report the first whole genome sequence of an ASFV (ASFV SS-3/Mongolia/2019) detected from a backyard pig in Mongolia in February 2019 using whole genome sequencing. We analyzed their phylogenetic relationship with other genotype II ASFVs from Eurasia. The ASFV SS-3/Mongolia/2019 belonged to genotype II (p72 and p54), serogroup 8 (CD2v), Tet-10a variant (pB602L), and IGRIII variant (intergenic region between the I73R/I329L genes). A total of five amino acid substitutions were observed in MGF 360-10L, MGF 505-4R, MGF 505-9R, NP419L, and I267L genes compared to the ASFV Georgia 2007/1 virus. ML phylogenetic analysis of the whole genome sequence showed that the virus shares a high nucleotide sequence identity with ASFVs recently identified in Eastern Europe and Asia and clustered with the ASFV/Zabaykali/WB5314/2020|Russia|2020 virus which was identified at the border between the Russian Federation and Mongolia in 2020. Our results suggest that trans boundary spread of ASF occurred through close geographic proximity.

KEYWORDS

African swine fever (ASF), whole genome sequencing (WGS), Mongolia, backyard pig, phylogenetic analysis

Introduction

African swine fever (ASF) is caused by the African swine fever virus (ASFV), a virus within the genus *Asfivirus* of the family *Asfarviridae*. It is a highly contagious and high-mortality disease affecting domestic pigs and wild boars and a notifiable disease to the World Organization for Animal Health (WOAH) (1, 2). ASF was first described in Kenya in 1921; it subsequently re-emerged from Africa into Georgia and rapidly spread into Eastern Europe and Asia (1). In December 2007, ASF was reported in Russia, with subsequent outbreaks occurring between 2008 and 2009 and affecting domestic pigs and wild boars (1). In August 2018, the disease was reported in China; it spread rapidly across that country and Mongolia, and then to Vietnam, Cambodia, North Korea, Laos, the Philippines, Myanmar, South Korea, Timor-Leste, and Indonesia (2).

On 10 January 2019, Mongolia's State Central Veterinary Laboratory (SCVL) confirmed the first ASF outbreak in Bulgan province, Mongolia (3). After this first detection, 11 additional

ASF outbreaks were recorded in Mongolia, involving seven provinces (4). A total of 105 farms and holdings were affected by the disease, resulting in the death or elimination of more than 3,000 exposed pigs (> 10 % of the total pig population in Mongolia) (4). While complete genome sequences of ASFV have been reported from several countries affected by ASF (5–12), a complete genome sequence of ASFV from Mongolia has not been published. However, several characterizations of the ASFV strain from the 2019 outbreak in Bulgan province, Mongolia have been reported in a previous study: partial p72, full p54, partial pB602L, and partial CD2v genes, as well as a 356-bp fragment between the I73R and I329L genes (2).

In this study, we report the first whole genome sequence of an ASFV detected in a backyard pig in Mongolia in February 2019 using next-generation sequencing (NGS). We also analyze its phylogenetic relationship with other genotype II ASFVs from Eurasia.

2. Materials and methods

Spleen tissue from a backyard pig carcass found at a dumping ground near Ulaanbaatar was collected during passive surveillance on 14 February 2019, and it was confirmed to be ASFV-positive using quantitative reverse transcription real-time PCR (RT-qPCR) assay (13, 14). DNA was extracted from spleen tissue from the ASF-positive backyard pig carcass using a DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) in accordance with the manufacturer's instructions and eluted in distilled water. DNA concentrations were determined using a Qubit BR dsDNA assay kit (Invitrogen, Carlsbad, CA). DNA samples were diluted to 0.2 ng/μl and libraries were prepared using the Illumina Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA). The concentration of sample libraries was determined using the Qubit dsDNA HS assay kit, and libraries were diluted to a 2 nM concentration and combined in equal volumes to form the pooled library. Subsequently, 600 μl of the 10 pM libraries were submitted for pair-end sequencing using the MiSeq reagent kit v2 (500 cycles) (Illumina, San Diego, CA).

The raw reads were adapter-trimmed for known Illumina adapters and quality-trimmed ($Q > 20$ and minimum length > 50) with Bbduk (https://sourceforge.net/projects/bbmap). Trimmed reads were mapped to the reference sequence, the ASFV Georgia 2007/1 (GenBank acc. FR682468.2), using minimap2 (15) with the default settings in Geneious Prime 10 Software (https://www.geneious.com/), and the consensus genome sequences (hereafter referred to as ASFV SS-3/Mongolia/2019) were called using Geneious Prime 10 with default parameter settings. To demonstrate the phylogenetic organization of the ASFV, ASFV genotype I 56/Ca/1978 (GenBank acc. MN270969) and all available full-length genome sequences (>150,000 bp) of genotype II ASFVs ($n = 70$) were downloaded from the NCBI GenBank database. The online multiple alignment server MAFFT, version 7 (https://mafft.cbrc.jp/alignment/software/), was used for sequence alignment of whole genome sequences with the default settings (16). Maximum likelihood (ML) phylogenies were constructed using RAxML-HPC v.8 using the general time-reversible (GTR) nucleotide substitution model with gamma distribution and with 1,000 rapid bootstrap

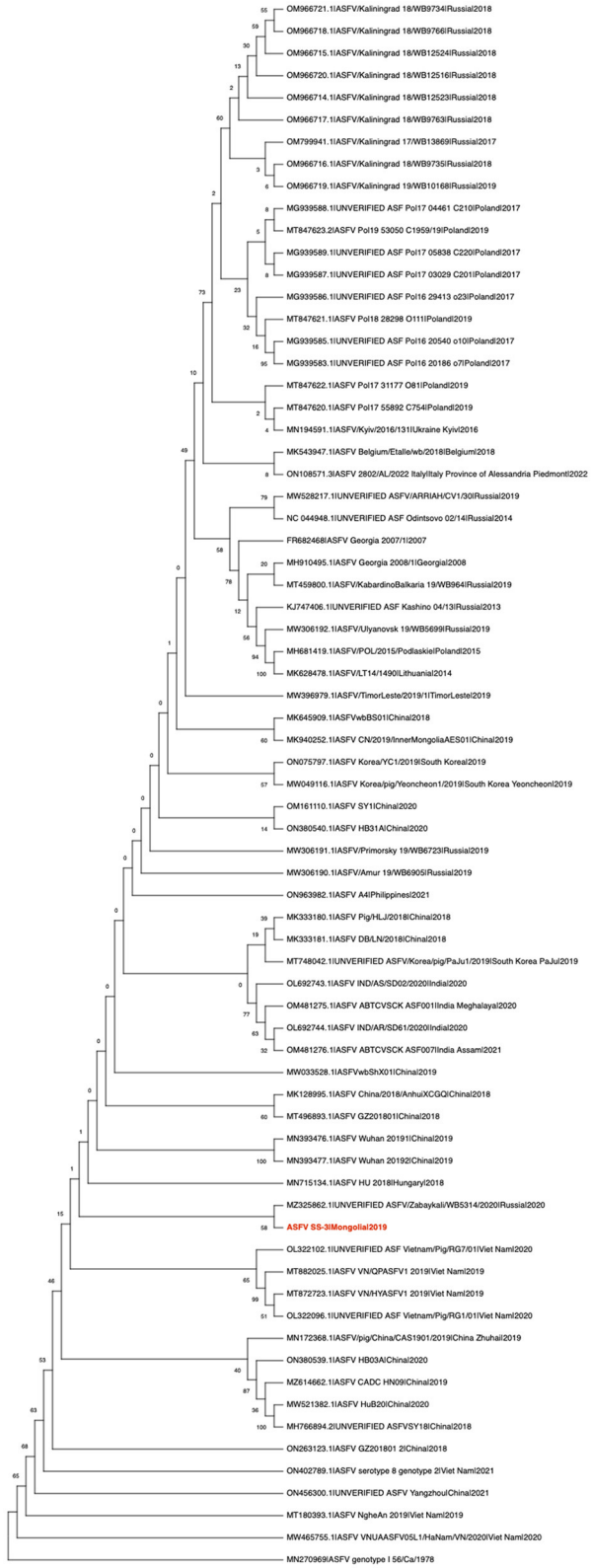


FIGURE 1
Maximum likelihood analysis of 70 complete coding sequences of ASFV, including the ASFV SS-3|Mongolia|2019 sequenced in this study (in red). The phylogeny was rooted to the ASFV genotype I 56/Ca/1978 as an outgroup and converted to cladogram form for better visualization of the genetic relationships. Numerical values represent 1,000 bootstrap replicate values expressed as percentages.

replicates (17). Phylogenetic trees were rooted to the ASFV genotype I 56/Ca/1978 as an outgroup and converted to cladogram form for better visualization of the genetic relationships.

3. Results and discussion

From 14,542,882 total raw reads, 111,352 reads were mapped to the reference ASF sequence. The length of the assembled genome was 190,591 bp, with a mean coverage depth of 162.3 reads (minimum: 43; maximum: 2,638). The molecular characteristics of ASFV SS-3/Mongolia/2019 were consistent with previous findings on ASFVs from the 2019 outbreaks in Mongolia (2), specifically genotype II (*p72* and *p54*), serogroup 8 (*CD2v*), Tet-10a (*pB602L*), IGRIII variant (intergenic region between the *I73R/I329L* genes). It was found to have a high degree of nucleotide identity (99.98%) with ASFV Georgia 2007/1(data not shown). A total of five amino acid substitutions were observed in genes MGF 360-10L, MGF 505-4R, MGF 505-9R, NP419L, and I267L relative to ASFV Georgia 2007/1 (Supplementary Table 1). MGF360 and MGF505 gene products have been reported to be related to suppression of a type I interferon response, as analyzed using a swine cDNA microarray (18). NP419L encodes DNA ligase involved in DNA replication, repair, nucleotide metabolism, transcription, and other enzymatic activities or host defense evasion (19). Finally, I267L is an important virulence factor that operates by impairing innate immune responses mediated by the RNA Pol-III-RIG-I axis (20).

ML phylogenetic analysis of the whole genome sequence suggested that the virus belongs to the Georgia-07-like genotype II ASF virus (Figure 1). It has a high degree of nucleotide identity (>97%) with ASFVs recently detected in Eastern Europe and Asia (data not shown). It was found to cluster with the ASFV/Zabaykali/WB5314/2020|Russia|2020 virus, which was identified at the border between Russia and Mongolia in 2020 (Figure 1) (5), suggesting that transboundary spread of ASF occurred through close geographic proximity. In late 2020, ASF-related mass mortality of wild boar was observed in MinJiin Khangai mountain, Mungunmorit soum of Tuv province, and Yeroo soum of Selenge province, which are remote areas of Mongolia bordering with Russia. These areas are far from domestic animal farms. We assume that the virus might have been disseminated by the movement of wild boar carriers between Russia and Mongolia.

Data availability statement

The data presented in the study are deposited in GenBank under the accession number OP787478 and Bioproject accession number PRJNA924538.

Ethics statement

Ethical review and approval was not required for the study on animals in accordance with the local legislation and institutional requirements.

Author contributions

J-YH: data analysis and manuscript writing. D-HL: study design, data analysis, and manuscript editing. E-OT-O: data collection and manuscript editing. DG, MB, and GR: study design and manuscript editing. S-SN and C-SS: data analysis, study design, and manuscript editing. All authors contributed to the article and approved the submitted version.

Funding

J-YH and S-SN are supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through the Animal Disease Management Technology Development Program, funded by the Ministry of Agriculture, Food, and Rural Affairs (MAFRA) (122016-02).

Conflict of interest

C-SS was employed by KCAV Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1094052/full#supplementary-material

References

1. Mazur-Panasiuk N, Wozniakowski G, Niemczuk K. The first complete genomic sequences of african swine fever virus isolated in Poland. *Sci Rep.* (2019) 9:4556. doi: 10.1038/s41598-018-36823-0

2. Ankhanbaatar U, Sainnokhoi T, Khanui B, Ulziibat G, Jargalsaikhan T, Purevtseren D, et al. African Swine Fever Virus Genotype Ii in Mongolia, 2019. *Transbound Emerg Dis.* (2021) 68:2787–94. doi: 10.1111/tbed.14095

3. Heilmann M, Lkhagvasuren A, Adyasuren T, Khishgee B, Bold B, Ankhanbaatar U, et al. African Swine Fever in Mongolia: Course of the Epidemic and Applied Control Measures. *Vet Sci*. (2020) 7:24. doi: 10.3390/vetsci7010024

4. Food and Agriculture Organization of the United Nations. *South Asia Animal Disease Outbreaks and News* 61 ed. Rome, Italy: Food and Agriculture Organization of the United Nations (2020).

5. Mazloum A, Igolkin AS, Shotin AR, Zinyakov NG, Vlasova NN, Aronova EV, et al. [Analysis of the Whole-Genome Sequence of an Asf Virus (Asfarviridae: Asfivirus: African Swine Fever Virus) Isolated from a Wild Boar (Sus Scrofa) at the Border between Russian Federation and Mongolia]. *Vopr Virusol.* (2022) 67:153–64. doi: 10.36233/0507-4088-104

6. Mileto P, da Conceicao F, Stevens V, Cummins D, Certoma A, Neave MJ, et al. Complete genome sequence of african swine fever virus isolated from a domestic pig in timor-leste, 2019. *Microbiol Resour Announc.* (2021) 10:e0026321. doi: 10.1128/MRA.00263-21

7. Hien ND, Nguyen LT, Hoang LT, Bich NN, Quyen TM, Isoda N, et al. First Report of a Complete Genome Sequence of a Variant African Swine Fever Virus in the Mekong Delta, Vietnam. *Pathogens*. (2022) 11:797. doi: 10.3390/pathogens11070797

8. Granberg F, Torresi C, Oggiano A, Malmberg M, Iscaro C, De Mia GM, et al. Complete Genome Sequence of an African Swine Fever Virus Isolate from Sardinia, Italy. *Genome Announc*. (2016) 4:e01220–16. doi: 10.1128/genomeA.01220-16

9. Hakizimana JN, Ntirandekura JB, Yona C, Nyabongo L, Kamwendo G, Chulu JLC, et al. Complete Genome Analysis of African Swine Fever Virus Responsible for Outbreaks in Domestic Pigs in 2018 in Burundi and 2019 in Malawi. *Trop Anim Health Prod.* (2021) 53:438. doi: 10.1007/s11250-021-02877-y

10. Koltsov A, Tulman ER, Namsrayn S, Kutish GF, Koltsova G. Complete genome sequence of virulent genotype I African swine fever virus strain K49 from the democratic Republic of the Congo, Isolated from a Domestic Pig (Sus Scrofa Domesticus). *Arch Virol.* (2022) 167:2377–80. doi: 10.1007/s00705-022-05543-2

11. Kovalenko G, Ducluzeau AL, Ishchenko L, Sushko M, Sapachova M, Rudova N, et al. Complete genome sequence of a virulent african swine fever virus from a domestic pig in Ukraine. *Microbiol Resour Announc*. (2019) 8:e00883–19. doi: 10.1128/MRA.00883-19

12. Senthilkumar D, Rajukumar K, Venkatesh G, Singh F, Tosh C, Kombiah S, et al. Complete genome analysis of african swine fever virus isolated from domestic pigs during the first asf outbreaks in India. *Transbound Emerg Dis.* (2022) 69:e2020–e7. doi: 10.1111/tbed.14536

13. Fernandez-Pinero J, Gallardo C, Elizalde M, Robles A, Gomez C, Bishop R, et al. Molecular diagnosis of African swine fever by a new real-time pcr using universal probe library. *Transbound Emerg Dis.* (2013) 60:48–58. doi: 10.1111/j.1865-1682.2012.01317.x

14. King DP, Reid SM, Hutchings GH, Grierson SS, Wilkinson PJ, Dixon LK, et al. Development of a Taqman Pcr assay with internal amplification control for the detection of african swine fever virus. *J Virol Methods.* (2003) 107:53–61. doi: 10.1016/S0166-0934(02)00189-1

15. Li H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics.* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191

16. Katoh K, Rozewicki J, Yamada KD. Mafft online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* (2019) 20:1160–6. doi: 10.1093/bib/bbx108

17. Stamatakis A. Raxml Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033

18. Afonso CL, Piccone ME, Zaffuto KM, Neilan J, Kutish GF, Lu Z, et al. African swine fever virus multigene family 360 and 530 genes affect host interferon response. *J Virol.* (2004) 78:1858–64. doi: 10.1128/JVI.78.4.1858-1864.2004

19. Dixon LK, Chapman DA, Netherton CL, Upton C. African Swine Fever Virus Replication and Genomics. *Virus Res.* (2013) 173:3–14. doi: 10.1016/j.virusres.2012.10.020

20. Ran Y, Li D, Xiong MG, Liu HN, Feng T, Shi ZW, et al. African Swine Fever Virus I267l Acts as an important virulence factor by inhibiting Rna polymerase Iii-Rig-I-mediated innate immunity. *PLoS Pathog.* (2022) 18:e1010270. doi: 10.1371/journal.ppat.1010270

frontiers | Frontiers in Veterinary Science

TYPE Data Report
PUBLISHED 02 March 2023
DOI 10.3389/fvets.2023.1112552

OPEN ACCESS

EDITED BY
Iryna Goraichuk,
Agricultural Research Service (USDA),
United States

REVIEWED BY
John M. Ngunjiri,
Targan Inc., United States
Haizhou Liu,
Wuhan Institute of Virology (CAS), China

*CORRESPONDENCE
Chang-Seon Song
songcs@konkuk.ac.kr
Ji-Yeon Hyeon
jyhyeon1205@gmail.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

RECEIVED 30 November 2022
ACCEPTED 15 February 2023
PUBLISHED 02 March 2023

CITATION
Cho AY, Kim T-H, Lee S-H, Lee H, Choi Y-J, Seo Y-R, Lee D-H, Hyeon J-Y and Song C-S (2023) Whole genome sequencing of Avian metapneumovirus type B genomes directly from clinical samples collected from chickens in live bird markets using multiplex tiling RT-PCR method. *Front. Vet. Sci.* 10:1112552. doi: 10.3389/fvets.2023.1112552

Whole genome sequencing of Avian metapneumovirus type B genomes directly from clinical samples collected from chickens in live bird markets using multiplex tiling RT-PCR method

Andrew Y. Cho[1†], Tae-Hyeon Kim[1†], Sun-Hak Lee[1], Heesu Lee[1], Yun-Jeong Choi[1], Ye-Ram Seo[1], Dong-Hun Lee[2], Ji-Yeon Hyeon[1*] and Chang-Seon Song[1,3*]

[1]Avian Disease Laboratory, College of Veterinary Medicine, Konkuk University, Gwangjin-gu, Seoul, Republic of Korea, [2]Wildlife Health Laboratory, College of Veterinary Medicine, Konkuk University, Seoul, Republic of Korea, [3]KHAV Co. Ltd., New Millennium Hall, Gwangjin-gu, Seoul, Republic of Korea

KEYWORDS

Avian metapneumovirus (aMPV), chickens, live bird market, clinical samples, whole genome sequencing

Introduction

Avian metapneumovirus (AMPV), a member of the family *Pneumoviridae*, genus *Metapneumovirus* possesses a non-segmented negative-sense RNA genome of approximately 13–15 kb with eight genes (3′-N-P-M-F-M2-SH-G-L-5′) (1–3). AMPV causes rhinotracheitis in turkeys and swollen head syndrome (SHS) in chickens, mostly contributing to secondary bacterial infections leading to more severe symptoms in chickens (4).

Different types of AMPVs have been classified based on the nucleotide sequence divergence of the attachment glycoprotein (G) and antigenic differences between strains (5). Types A and B are found all around the world, and type C was reported in North America, China, and South Korea, and in a retrospective study in France in 1990 (6). Type D was reported only once in a turkey flock in France in 1985 (5, 7). AMPV types A, B, and C have been detected in South Korea; the type A and B in poultry farms and the type C from pheasants in live bird markets (6, 8, 9).

The study of AMPV infection is particularly difficult due to the transient nature of viral shedding in the host before the symptoms develop. In most cases, multiple blind passages of AMPV from the clinical samples are required for isolation and identification (4, 5). Therefore, a limited number of AMPV sequences are available in the NCBI GenBank database. As of November 29, 2022, there are only three complete genome sequences of AMPV type B (LN16, VCO3/60616, and Hungary/657/4; MH745147, AB548428, and MN729604 respectively) (3, 10, 11).

The tiling amplicon method has been proven to be efficient and prolific in producing whole genome sequences directly from clinical samples during the Covid-19 pandemic (12). In a previous study, the multiplex tiling RT-PCR method was applied to enrich the genetic material of Zika virus and Rabies lyssavirus directly from brain tissue samples, and it enables to sequence of the full genome of low titer samples containing as low as 50 genome copies in a reaction (13, 14).

In this study, we detected 6 AMPVs from chickens in live bird markets (LBM) in South Korea during 2019–2022. For whole genome sequencing of AMPVs, we developed a PCR primer panel to efficiently amplify the complete coding region of AMPV type B using the multiplex tiling RT-PCR method. We successfully obtained the complete coding region of AMPVs using Illumina next-generation sequencing (NGS) and conducted comparative a phylogenetic analysis to analyze the genetic relatedness of AMPVs from LBMs in Korea with other AMPVs.

Materials and methods

Sample collection

A total of 138 slaughtered chickens were purchased from poultry meat vendors in LBMs in Korea during the period of 2019 to 2022. The nasal turbinate or whole beaks of the chickens were collected and washed with phosphate-buffered saline (PBS) using enough to fully immerse the samples. Total RNA was extracted from the nasal turbinate wash samples using the Qiagen RNeasy mini kit (Hilden, Germany) and used to detect both Avian metapneumovirus type A and B using the real-time qRT-PCR as previously described (8). All positive nasal turbinate wash samples were inoculated to Vero cells for virus isolation.

Primer design and tiling amplicon PCR

To design a tiling amplicon PCR primer panel, full genome sequences of AMPV type B available in the NCBI GenBank (LN16, VCO3/60616, and Hungary/657/4; Accession no. MH745147, AB548428, and MN729604 respectively) were downloaded and aligned using the MAFFT 1.4.0 program on Geneious Prime software (https://www.geneious.com) (3, 10, 11). Primers were designed to amplify 380–420 bp region with about 100 bp overlap, and 43 sets of designed primers were pooled into two pools according to the primer design output by Primal Scheme (http://primalscheme.com) (Supplementary Table 2) (13). Three of six positive samples with lower cycle threshold (Ct) values (ranging from 25.77 to 30.6) were selected for subsequent multiplex RT-PCR (Table 1). cDNA was synthesized from the RNA using the LunaScript RT SuperMix kit (NEB, Massachusetts, United States) following the manufacturer's instructions. For the multiplex PCR assay, an equal volume of each 100 uM primer stock was pooled together as designated as pool 1 and 2 (Supplementary Table 2). The PCR mixture was prepared by mixing 12.5 ul Q5 Hotstart 2X Master Mix (NEB, Massachusetts, United states), 0.015 uM of each primer pool, 5 ul of template cDNA, and nuclease-free water up to 25 ul. The reaction mixture was prepared for each pool as it induces consistency between reactions and PCR assay was performed. PCR amplification conditions were: 98°C for 15 s, followed by 35 cycles of 95°C for 15 s and 63°C for 5 min (12). The PCR products were then visualized by electrophoresis on 2% agarose gels showing around 400 bp amplicons (13). PCR products of both primer pools were combined and used for library preparation after purification using the Qiaquick PCR purification kit (Qiagen, Hilden, Germany). Library was prepared using the TruSeq Nano DNA kit (Illumina, California, United States) to produce 2 × 151 bp paired-end reads. Library preparation and sequencing on NextSeq 500 sequencing system (Illumina, California, United States) was done by LAS (Gimpo, Republic of Korea).

Assembly and phylogenetic analysis

Raw reads were trimmed of adapters and low-quality bases using BBDuk version 38.84 by setting the minimum quality to 20 (15). *De novo* and reference-based assemblies of genome sequences were performed. For reference-based assembly, trimmed reads were mapped to the LN16 virus genome (GenBank accession number: MH745147) using the Minimap 2.24 (https://github.com/lh3/minimap2) with default options and visualized on Geneious Prime software. Trimmed reads were assembled *de novo* using the SPAdes assembler 3.15.5. The assembled genome sequences produced by reference mapping and *de novo* assembly approaches were combined to generate the final consensus genome sequences. A total of 48 G sequences of AMPV type B were downloaded from the GenBank database and aligned using the MAFFT

TABLE 1 Genome sequencing and assembly results of Avian metapneumovirus type B isolates from this study.

Isolates	Real-time PCR (Ct value[a])	Total NGS reads	Trimmed reads (>Q20)	Genome assembly results		
				Number of assembled reads	Coverage	Complete CDS
AMPV/B/Korea/N19-29/2019	28.67	9,016,084	8,632,844	7,925,758 (91.8%)[b]	99.3% (13,414 of 13,513)	Yes
AMPV/B/Korea/N21-41/2021	25.77	8,427,438	8,178,950	7,563,434 (92.5%)	99.2% (13,411 of 13,513)	Yes
AMPV/B/Korea/N21-83/2021	30.6	7,867,590	7,609,232	6,967,759 (91.6%)	98.3% (13,279) of (13,513)	No [Gaps in F (5 bp) and L (100 bp)][c]

[a]Cycle threshold values of real-time RT-PCR.
[b]Percentage of mapped reads/trimmed reads.
[c]Gaps were filled with two sanger sequencing reads produced with sample PCR primers used for multiplex RT-PCR.

Multiple Sequence Alignment software v7.450 for phylogenetic analysis (16). A maximum-likelihood (ML) phylogenetic tree was reconstructed using the RAxML GUI 2.0 (https://antonellilab.github.io/raxmlGUI/) with a rapid bootstrap option set to 1,000 (17).

Descriptive results

A total of 6 out of 138 (4.34%) chicken nasal turbinate wash samples tested positive for AMPV type B by real-time qRT-PCR (Supplementary Table 1 and Supplementary Figure 1). Sequencing results confirmed that the positive samples from the surveillance were not a result of contamination as the detected APMV sequences differ from each other and the rest of the APMV genomes found in the GenBank database. The inoculated Vero cells did not exhibit any cytopathic effect, suggesting it may require additional passages for virus isolation (data not shown). We successfully obtained complete coding genome sequences (CDS) of all three nasal turbinate samples (AMPV/B/Korea/N19-29/2019, AMPV/B/Korea/N21-83/2021, and AMPV/B/Korea/N21-41/2021; hereafter N19-29, N21-83, and N21-41, respectively) (Table 1). Our

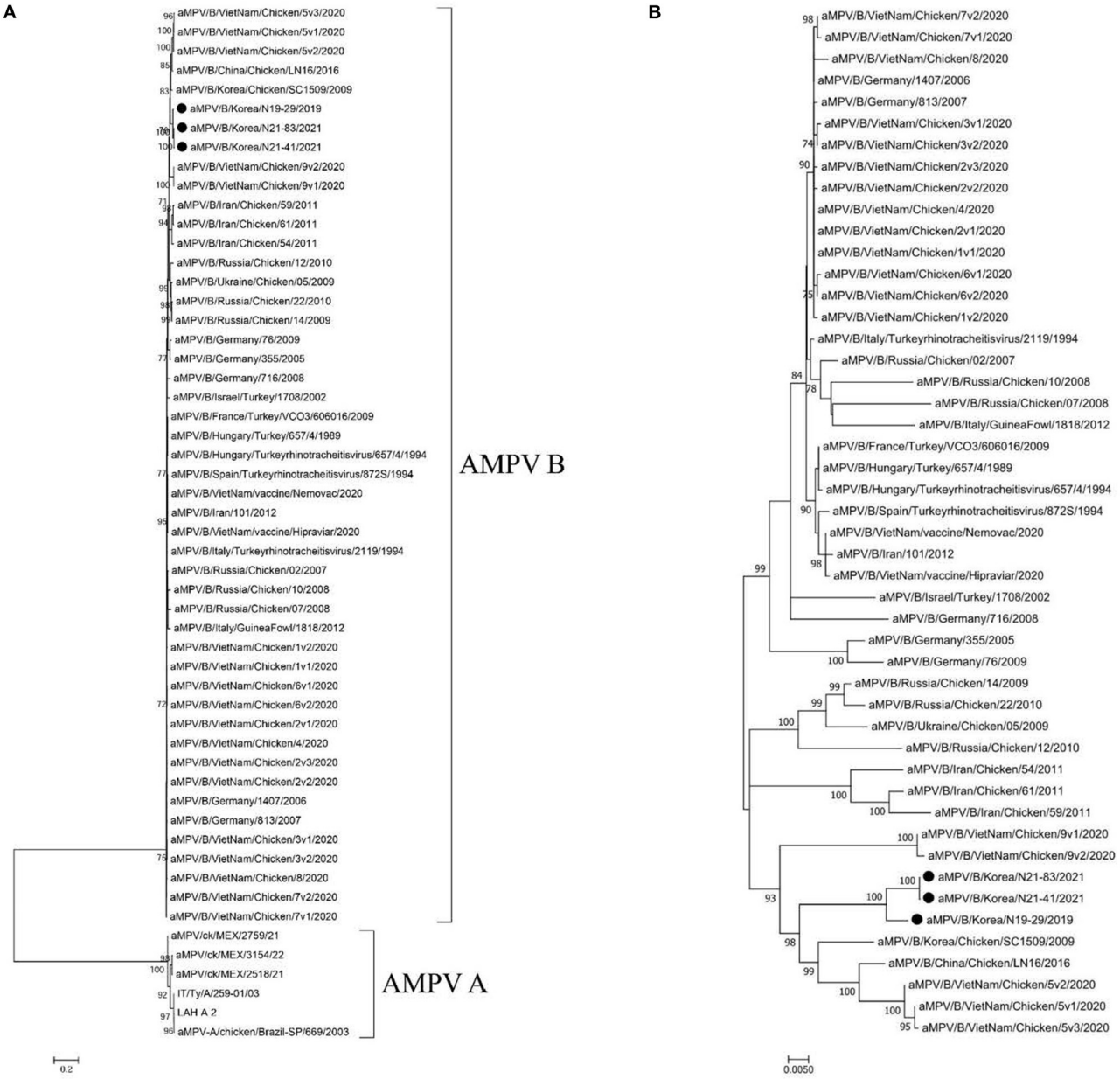


FIGURE 1
(A) Maximum-likelihood analysis of 54 glycoprotein (G) sequences of AMPV type A and B. It is rooted to the AMPV type A sequences and **(B)** Maximum-likelihood analysis of 48 glycoprotein (G) sequences of AMPVs type B downloaded from GenBank database, rooted to midpoint. Closed circles (•) indicate the viruses sequenced in this study. The scale bars show the number of substitutions per site. The numerical values represent 1,000 bootstrap replicate values expressed as a percentage.

tiling amplicon PCR reactions showed a fairly small amplification bias when clinical samples with relatively low Ct values, 25.77 and 28.67, were used. We were able to assemble complete CDS of N19-29 and N21-41 viruses using the tiling amplicon PCR primer panel designed in this study coupled with illumina NGS. However, the N21-83 virus which showed a higher Ct value (30.6) had two short gaps in the initial genome assembly results, including 5 bp in F gene and 100 bp in L gene. The gaps were covered with two sanger sequencing reads using the same primers used in this study. Since the two fragments were amplified using the same primers, the outcome of low amplification efficiency was not likely due to primer-template mismatch. We assume that competitive inhibition between primers may have decreased PCR efficiency of the fragments.

Phylogenetic analysis of G gene sequences revealed that the viruses sequenced in this study belong to the AMPV type B and showed the closest genetic relationship with the SC1509 (GenBank Accession no. DI187010.1) virus (nucleotide sequence identity: 95.74–96.14%) from Korea in 2009 and the LN16 virus (nucleotide sequence identity: 95.42–95.82%) identified from China in 2016 (Figure 1). The N19-29, N21-83, and N21-41 viruses did not cluster with the commercial live attenuated AMPV type B vaccine strains in phylogenies (Figure 1). We were not able to present phylogenetic analysis based on whole-genome sequences as currently there are only 3 complete genome sequences of AMPV type B available in the NCBI GenBank database.

In conclusion, we developed the tiling amplicon PCR method for genome sequencing and successfully sequenced three AMPVs directly from clinical samples. The multiplex tiling RT-PCR and NGS approach developed in this study has the potential to be implemented in a diagnostic setting, providing a rapid and reliable method for complete genome sequencing and molecular epidemiological study of AMPV from clinical samples. In addition, the complete CDS of AMPVs established in this study would be useful as reference data for future investigations on AMPVs. The relative ease of acquiring complete CDS directly from clinical samples without a labor-intensive adaptation process in cell culture will help further the diversity of the AMPV genome database. Full genome sequencing of AMPV have suggested subpopulation present in vaccine strain could be selected for better replication during *in vivo* replication in field conditions (18). Our method could also be used to monitor mutations and subpopulations of field strains without prior adaptation, allowing for more accurate representation of AMPV quasi-species. In addition, since the live bird markets have been recognized as a reservoir, amplifier, and source of Avian viruses (19), genomic surveillance of AMPVs in LBMs should be enhanced for monitoring of further evolution and spread of the AMPVs.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, OP924005, OP924006, and OP924007. https://www.ncbi.nlm.nih.gov/sra, SRR22993455, SRR22993456, and SRR22993457.

Author contributions

Conceptualization and data analysis: AC. Methodology, data curation, and writing—original draft preparation: AC and T-HK. Sample preparation: AC, T-HK, S-HL, HL, Y-JC, and Y-RS. Writing—review and editing: D-HL, J-YH, and C-SS. Supervision and funding acquisition: J-YH and C-SS. All authors have read and agreed to the published version of the manuscript.

Funding

This study is supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Animal Disease Management Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant number: 122057-2).

Acknowledgments

We thank Sang-Han Kim, Ye-Jeong Park, and Sang-Ah Park for their excellent technical support.

Conflict of interest

C-SS was employed by KHAV Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1112552/full#supplementary-material

References

1. Rima B, Collins P, Easton A, Fouchier R, Kurath G, Lamb RA, et al. ICTV virus taxonomy profile: pneumoviridae. *J General Virol.* (2017) 98:2912–3. doi: 10.1099/jgv.0.000959

2. Franzo G, Legnardi M, Mescolini G, Tucciarone CM, Lupini C, Quaglia G, et al. Avian Metapneumovirus subtype B around Europe: a phylodynamic reconstruction. *Vet Res.* (2020) 51:88. doi: 10.1186/s13567-020-00817-6

3. Yu M, Xing L, Chang F, Bao Y, Wang S, He X, et al. Genomic sequence and pathogenicity of the first avian metapneumovirus subtype B isolated from chicken in China. *Vet Microbiol.* (2019) 228:32–8. doi: 10.1016/j.vetmic.2018.11.009

4. Suarez DL, Miller PJ, Koch G, Mundt E, Rautenschlein S. Newcastle disease, other Avian paramyxoviruses, and avian metapneumovirus infections. *Dis Poultry*. (2020) 109–66. doi: 10.1002/9781119371199.ch3

5. Jesse ST, Ludlow M, Osterhaus ADME. Zoonotic origins of human metapneumovirus: a journey from birds to humans. *Viruses.* (2022) 14:677. doi: 10.3390/v14040677

6. Lee E, Song MS, Shin JY, Lee YM, Kim CJ, Lee YS, et al. Genetic characterization of avian metapneumovirus subtype C isolated from pheasants in a live bird market. *Virus Res.* (2007) 128:18–25. doi: 10.1016/j.virusres.2007.03.029

7. Bayon-Auboyer MH, Arnauld C, Toquin D, Eterradossi N. Nucleotide sequences of the F, L and G protein genes of two non-A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup. *J Gen Virol.* (2000) 81:2723–33. doi: 10.1099/0022-1317-81-11-2723

8. Kwon J-S, Lee H-J, Jeong S-H, Park J-Y, Hong Y-H, Lee Y-J, et al. Isolation and characterization of avian metapneumovirus from chickens in Korea. *J Vet Sci.* (2010) 11:59. doi: 10.4142/jvs.2010.11.1.59

9. Choi K-S, Jeon W-J, Park M-J, Lee E-K, Kwon J-H. Isolation and characterization of Avian Metapneumovirus from broiler breeder chickens in Korea. *J Bacteriol Virol.* (2009) 39:373. doi: 10.4167/jbv.2009.39.4.373

10. Goraichuk IV, Kapczynski DR, Seal BS, Suarez DL. Complete genome sequence of an avian metapneumovirus subtype B strain from Hungary. *Microbiol Resour Announc*. (2020) 9:e00177–20. doi: 10.1128/MRA.00177-20

11. Sugiyama M, Ito H, Hata Y, Ono E, Ito T. Complete nucleotide sequences of avian metapneumovirus subtype B genome. *Virus Genes.* (2010) 41:389–95. doi: 10.1007/s11262-010-0518-z

12. Quick J. *nCoV-2019 sequencing protocol v3 (LoCost) V.3*. (2020). Available online at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye. doi: 10.17504/protocols.io.bbmuik6w (accessed November 12, 2022).

13. 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) 12:1261–76. doi: 10.1038/nprot.2017.066

14. Bashashati M, Chung DH, Fallah Mehrabadi MH, Lee DH. Evolution of H9N2 avian influenza viruses in Iran, 2017–2019. *Transbound Emerg Dis.* (2021) 68:3405–14. doi: 10.1111/tbed.13944

15. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. In *Conference: 9th Annual Genomics of Energy and Environment Meeting.* Walnut Creek, CA, March 17–20 (2014) United States. DE-AC02-05CH11231 2016-04-082014. p. Medium: ED.

16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* (2013) 30:772–80. doi: 10.1093/molbev/mst010

17. Edler D, Klein J, Antonelli A, Silvestro D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol Evolut.* (2021) 12:373–7. doi: 10.1111/2041-210X.13512

18. Franzo G, Naylor CJ, Drigo M, Croville G, Ducatez MF, Catelli E, et al. Subpopulations in aMPV vaccines are unlikely to be the only cause of reversion to virulence. *Vaccine.* (2015) 33:2438–41. doi: 10.1016/j.vaccine.2015.03.092

19. Lee DH, Song CS. H9N2 avian influenza virus in Korea: evolution and vaccination. *Clin Exp Vaccine Res.* (2013) 2:26–33. doi: 10.7774/cevr.2013.2.1.26

frontiers | Frontiers in Veterinary Science

TYPE Brief Research Report
PUBLISHED 28 April 2023
DOI 10.3389/fvets.2023.1085554

OPEN ACCESS

EDITED BY
Christina Leyson,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Ruth H. Nissly,
The Pennsylvania State University, United States
Giulia Mencattelli,
Fondazione Edmund Mach, Italy

*CORRESPONDENCE
Dong-Hun Lee
✉ donghunlee@konkuk.ac.kr

RECEIVED 31 October 2022
ACCEPTED 06 January 2023
PUBLISHED 28 April 2023

CITATION
Hyeon J-Y, Helal ZH, Appel A, Tocco N, Hunt A, Lee D-H and Risatti GR (2023) Whole genome sequencing and phylogenetic analysis of West Nile viruses from animals in New England, United States, 2021. *Front. Vet. Sci.* 10:1085554. doi: 10.3389/fvets.2023.1085554

Whole genome sequencing and phylogenetic analysis of West Nile viruses from animals in New England, United States, 2021

Ji-Yeon Hyeon[1,2,3], Zeinab H. Helal[1,2], Allison Appel[1], Natalie Tocco[1], Amelia Hunt[1,2], Dong-Hun Lee[3]* and Guillermo R. Risatti[1,2]

[1]Department of Pathobiology and Veterinary Science, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, United States, [2]Connecticut Veterinary Medical Diagnostic Laboratory, Department of Pathobiology and Veterinary Science, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, United States, [3]College of Veterinary Medicine, Konkuk University, Seoul, Republic of Korea

West Nile virus is a mosquito-borne Flavivirus which is the leading cause of global arboviral encephalitis. We sequenced WNVs from an American crow found in Connecticut and an alpaca found in Massachusetts which were submitted to the Connecticut Veterinary Medical Diagnostic Laboratory (CVMDL). We report here the complete protein-coding sequences (CDS) of the WNVs (WNV 21-3957/USA CT/Crow/2021 and WNV 21-3782/USA MA/Alpaca/2021) and their phylogenetic relationship with other WNVs recovered from across the United States. In the phylogenetic analysis, the WNVs from this study belonged to the WNV lineage 1. The WNV 21-3957/USA CT/Crow/2021 clustered with WNVs from a mosquito and birds in New York during 2007–2013. Interestingly, the virus detected in the alpaca, WNV 21-3782/USA MA/Alpaca/2021 clustered with WNVs from mosquitos in New York, Texas, and Arizona during 2012–2016. The genetic differences between the viruses detected during the same season in an American crow and an alpaca suggest that vector-host feeding preferences are most likely driving viral transmission. The CDS of the WNVs and their phylogenetic relationships with other WNVs established in this study would be useful as reference data for future investigations on WNVs. Seasonal surveillance of WNV in birds and mammals and the genetic characterization of detected viruses are necessary to monitor patterns of disease presentations and viral evolution within a geographical area.

KEYWORDS

West Nile virus, genome sequencing, phylogenetic analysis, surveillance, epidemiology

1. Introduction

West Nile virus (WNV) is a neurotropic mosquito-borne Flavivirus genus within the Flaviviridae family (1). Its transmission cycle involves mosquitoes belonging to *Culex* spp. as vectors and birds as amplifying hosts or reservoirs. Some mammalian species including humans and horses are accidental dead-end hosts (1, 2). It was first isolated in Uganda in 1937 and is currently the most widespread arbovirus geographically worldwide due to its spread throughout two continents within 2 years (3, 4).

Partial sequencing of the gene encoding for the envelope protein (E) of WNV led to the classification of the virus into five distinct phylogenetic lineages, and the WNV lineage 1 and WNV lineage 2 have been associated with outbreaks in humans (5, 6). Lineage 1

encompasses viruses from Africa, the Middle East, Eastern Europe, the United States, and Australia. Lineage 2 comprises viruses from sub-Saharan Africa and Madagascar (5).

In the United States, WNV was first detected in New York City in 1999 and spread rapidly across the United States within only a couple of years; New York (1999), Connecticut (2000), Florida (2001), Rocky Mountains and Washington state (2002), and Southern California (2003) (3, 7). It has been suggested that there are multiple possible origins of WNV in the United States, but the most likely explanation is that it was due to the human transportation of birds and/or mosquitoes (7).

Although WNVs are seasonally detected in birds and less frequently in mammals in most of the United States, there is limited phylogenetic data based on whole genome sequences (WGS), impeding a more detailed understanding of WNV evolution. Regionally, like in the Northeastern region of the United States, where WNV was first introduced, there are no reports regarding complete genomes of the virus detected in birds and mammals during the last two decades. WGS-based phylogenetic analysis would be a useful tool to understand the spread and evolution of WNV.

In this study, we report the complete protein-coding sequences (CDS) of WNVs detected at the Connecticut Veterinary Medical Diagnostic Laboratory (CVMDL) in an American crow (*Corvus brachyrhynchos*) from Connecticut and an alpaca (*Vicugna pacos*) from Massachusetts during 2021 using next-generation sequencing (NGS). We analyzed their phylogenetic relationship with other WNVs recovered from across the United States to reconstruct the origin of these viruses.

2. Materials and methods

2.1. Samples

We found two WNVs from animals submitted to the Connecticut Veterinary Medical Diagnostic Laboratory (CVMDL). An American crow (*Corvus brachyrhynchos*) found dead in Branford, CT, and a female alpaca (*Vicugna pacos*) from the state of Massachusetts were confirmed WNV positive using the quantitative reverse transcription real-time PCR (RT-qPCR) assay (8) at the CVMDL, Department of Pathobiology and Veterinary Science, the University of Connecticut in 2021.

2.2. RT-qPCR and whole genome sequencing

Total RNA was extracted from brain tissue samples using the TRIzol reagent (ThermoFisher Scientific, USA) according to the manufacturer's instructions for RT-qPCR. Ct values were 14.31 and 17.18 for the American crow and the alpaca samples, respectively. The RNA samples were then used for whole genome sequencing. Sequence-Independent, Single-Primer-Amplification (SISPA) was performed to amplify viral RNA as described in the previous study (9). The Swift 2S Turbo DNA Library Kits (Swift Biosciences, Coralville, IA) were used according to the manufacturer's instructions to generate multiplexed paired-end sequencing libraries. The dsDNA was fragmented and tagged with adapters by Nextera transposase (Illumina, San Diego, CA). Sequencing libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) and analyzed on a High Sensitivity DNA Chip on the Bioanalyzer (Agilent Technologies, Santa Clara, CA). The libraries were adjusted to 1 nM concentration and equal volumes of 5 μl of each library were pooled. The pool was denatured with sodium hydroxide (0.2 N final concentration) and further diluted to 100 pM. Five percent of PhiX control library (Illumina) was added to the pool. The library pool was loaded in the flow cell of the MiSeq Reagent Kit V3 (Illumina). The barcoded multiplexed library sequencing (2 × 300 bp) was performed on an Illumina MiSeq platform (Illumina).

2.3. Assembly of sequencing reads

Residual adapters, SISPA primer K (GACCATCTAGCGACCTCCAC), and bases with low-quality scores ($Q < 20$) were removed from fastq files using BBduk. Then, reference-guided genome assemblies against reference genome sequences (GenBank accession number: KX547196 and KX547200) were performed using the Minimap2 in Geneious Prime 10 Software (https://www.geneious.com/) and the consensus genome sequences were called using the Geneious Prime 10 with default parameter settings, hereafter referred to as WNV 21-3782/USA MA/Alpaca/2021 and WNV 21-3957/USA CT/Crow/2021 virus.

2.4. Phylogenetic analysis

The CDS of WNVs identified in the United States which have metadata including host and collection date ($n = 902$), WNV lineage 1 reference sequence (NC 009942), and WNV lineage 2 reference sequence (NC 001563) were downloaded from NCBI GenBank database. The ElimDupes software (http://hcv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html) was used to down-sample the data set of 902 WNVs with 99.5% sequence similarity cutoff level to 140 sequences. Two reference sequences and our sequences were added to the datasets for phylogenetic analysis.

The MAFFT multiple alignment software v1.4.0 in Geneious was used for multiple sequence alignment of complete CDS of the WNV genomes. Maximum likelihood (ML) phylogenies were constructed using RAxML-HPC v.8 using the general time-reversible (GTR) nucleotide substitution model and discrete gamma distribution with 1,000 rapid bootstrap replicates, and TempEst v1.5.3 was used to identify potential outliers that substantially deviated from the linear regression of root-to-tip genetic distance against time, and the outliers were removed from this study (10, 11). Phylogenetic trees were rooted to the WNV lineage 2 reference sequence (NC 001563) as an outgroup. Subtrees including the WNVs from this study were extracted from ML phylogenies to better visualize the genetic relationships.

To investigate amino acid changes, the CDS of WNVs were annotated using the "find annotations" feature in the Geneious prime by comparing with the WNV strain HNY1999 polyprotein

TABLE 1 Nucleotide sequence identities between the WNV 21-3782/USA MA/Alpaca/2021 virus and nearest virus homologs in the GenBank database (as of 15 May 2022).

Accession no.	Virus name	Collection date	Location	Host	% identity
KX547196.1	West Nile virus strain WNV-1/Culex/USA/13290644/2013	17-Sep-2013	USA: Nassau Co., NY	*Culex sp.*	99.52%
KY782106.1	West Nile virus isolate B3	25-Jul-2012	USA: Illinois, West Chicago suburbs	*Spinus tristis*	99.48%
KX547167.1	West Nile virus strain WNV-1/Culiseta sp./USA/13330613/2013	04-Sep-2013	USA: Onondaga Co., NY	*Culiseta sp.*	99.47%
KX547565.1	West Nile virus strain WNV-1/Culiseta sp./USA/13330653/2013	24-Sep-2013	USA: Onondaga Co., NY	*Culiseta sp.*	99.46%
KM012188.1	West Nile virus isolate ARC13-12	2012	USA: IL	*Homo sapiens*	99.38%
KX547391.1	West Nile virus strain WNV-1/Culex/USA/10140626/2010	25-Aug-2010	USA: Erie Co., NY	*Culex sp.*	99.38%
KX547485.1	West Nile virus strain WNV-1/Culex/USA/14430021/2014	13-Jun-2014	USA: Rockland Co., NY	*Culex sp.*	99.37%
KX547556.1	West Nile virus strain WNV-1/Culiseta sp./USA/12370596/2012	06-Sep-2012	USA: Oswego Co., NY	*Culiseta sp.*	99.35%
KX547473.1	West Nile virus strain WNV-1/Culex/USA/13141233/2013	17-Jul-2013	USA: Erie Co., NY	*Cultex sp.*	99.35%
KC736498.1	West Nile virus isolate AVA1204579	2012	USA: Texas	*Culex quinquefasciatus*	99.35%

gene (Accession no. AF202541) and translated. Amino acid mutations in the molecular markers for virulence determinants in mammalian and avian hosts reported in a previous study (12) were investigated.

3. Results

The length of sequenced WNV genomes was 10,533 bp, and nucleotide pairwise identity among the two sequences was 98.4% (data not shown). The NCBI BLAST searches revealed that WNV 21-3782/USA MA/Alpaca/2021 virus shared high nucleotide identity (>99.0%) with WNVs identified from mosquitos, birds, and humans in the United States between 2010 and 2014 (Table 1), whereas WNV 21-3957/USA CT/Crow/2021 virus shared nucleotide identity with WNVs identified from mosquitos and a bird in the United States between 2012 and 2015 (Table 2).

The phylogenetic analysis revealed that WNV 21-3782/USA MA/Alpaca/2021 and WNV 21-3957/USA CT/Crow/2021 belonged to the WNV lineage 1 (Supplementary Figures 1, 2). The WNV 21-3957/USA CT/Crow/2021 clustered with WNVs detected in mosquitoes in New York during 2013-2015 (Figure 1A). This virus was not genetically related to the WNVs detected in American crows in Connecticut during 1999–2002 (Supplementary Figures 1, 2). The WNV 21-3782/USA MA/Alpaca/2021 clustered with WNVs detected in mosquitos in New York from 2013 to 2014 (Figure 1B).

We investigated if predicted amino acid substitutions observed in the CDS of WNV 21-3957/USA CT/Crow/2021 and WNV 21-3782/USA MA/Alpaca/2021 sequences encompass previously identified molecular markers of virulence in WNV (12). The CDS of WNV 21-3782/USA CT/Alpaca/2021 harbors amino acid substitutions, including E-159A and NS4A-46E/47L/50A, but the amino acid sequences were different from the previous study (Table 3). A T249P substitution in non-structural protein 3 (NSP3) associated with decreased virulence in the avian model of WNV virulence was detected in WNV 21-3957/USA CT/Crow/2021 (Table 4).

4. Discussion

Since WNV was first detected in the United States in 1999 in New York City, WNV has spread from the initial point of entry in the northeastern United States to the rest of the United States, Mexico, Canada, and the Caribbean (13, 14). Most complete genome sequences of WNVs available in the NCBI GenBank are WNVs from mosquitoes in the northeastern United States and only a few phylogenetic studies have been performed based on full-length genome sequences (13, 15, 16). In this study, we detected WNVs via RT-qPCR in a mammalian host (alpaca) and an avian host (American crow). These animals originated from the New England region (Connecticut and Massachusetts) and were submitted in 2021 to the Pathology section of CVMDL for post-mortem diagnostics. Detected viruses were then genetically characterized via next generation sequencing considering that only a few phylogenetic studies have been based on the WGS of WNV (13, 15, 16). Here, NGS in combination with sequence enrichment steps via SISPA allowed the obtention of complete CDS of WNV directly from clinical samples.

In the ML phylogenetic trees, the long tree branch length between the WNVs sequenced in this study and their closest relatives suggest that the virus had been circulating before being detected in the crow and alpaca in 2021. We assume that the

TABLE 2 Nucleotide sequence identities between the WNV 21-3957/USA CT/Crow/2021 virus and nearest virus homologs in the GenBank database (as of 15 May 2022).

Accession no.	Virus name	Collection date	Location	Host	% identity
KX547235.1	West Nile virus strain WNV-1/Culex sp./USA/15350183/2015	13-Aug-2015	USA: Orange Co., NY	*Culex sp.*	99.60%
KX547482.1	West Nile virus strain WNV-1/Culiseta sp./USA/13370456/2013	07-Aug-2013	USA: Oswego Co., NY	*Culiseta sp.*	99.44%
KX547254.1	West Nile virus strain WNV-1/Culex/USA/13510557/2013	17-Jul-2013	USA: Suffolk Co., NY	*Culex sp.*	99.43%
KC333376.1	West Nile virus isolate TX8546	14-Jun-2012	USA: Katy, Harris Co., Texas	*blue jay*	99.43%
KX547337.1	West Nile virus strain WNV-1/Culex/USA/12430655/2012	28-Sep-2012	USA: Rockland Co., NY	*Culex sp.*	99.42%
KX547333.1	West Nile virus strain WNV-1/Culex/USA/13511447/2013	09-Oct-2013	USA: Suffolk Co., NY	*Culex sp.*	99.42%
KX547201.1	West Nile virus strain WNV-1/Aedes sp./USA/12290391/2012	20-Jul-2012	USA: Nassau Co., NY	*Aedes sp.*	99.42%
KY216153.1	West Nile virus isolate 557	25-Jul-2012	USA: Not further specified	*Culex pipiens*	99.41%
KY229073.1	West Nile virus isolate 875	21-Aug-2012	USA: Not further specified	*Culex pipiens*	99.39%
KX547612.1	West Nile virus strain WNV-1/Culex/USA/12510875/2012	21-Aug-2012	USA: Suffolk Co., NY	*Culex sp.*	99.39%

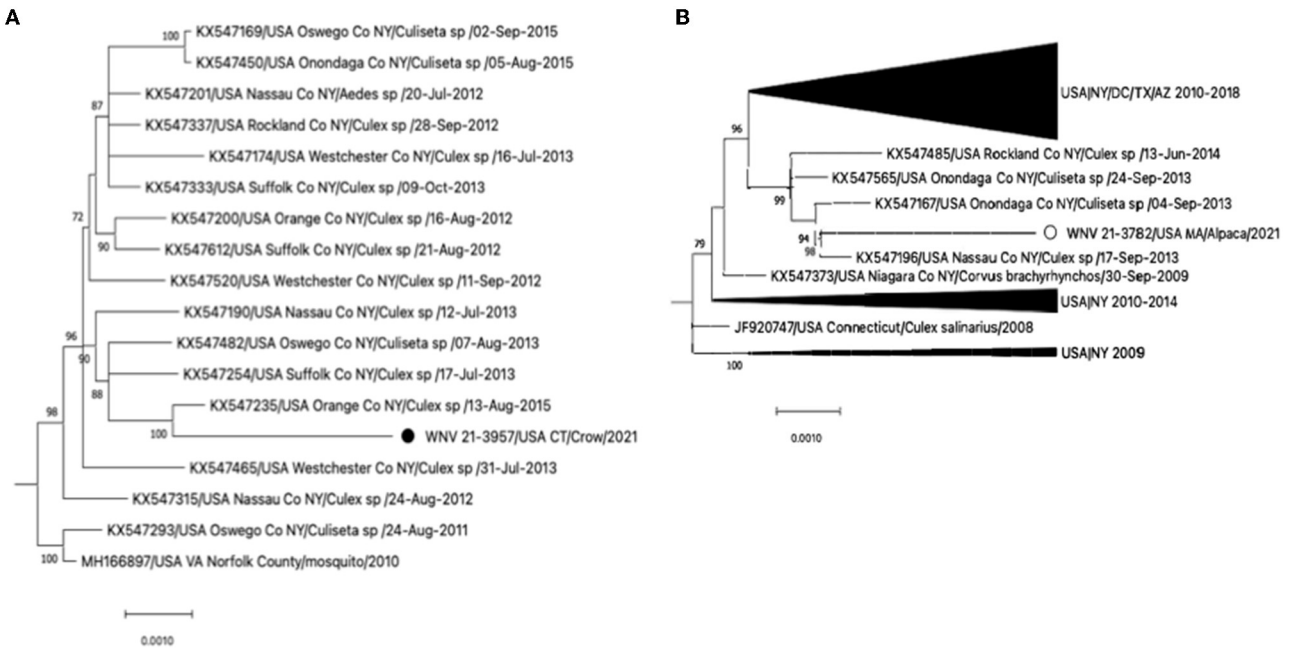


FIGURE 1
(A) Expansion of the clade of WNV 21-3957/USA CT/Crow/2021 and **(B)** the clade of WNV 21-3957/USA CT/Crow/2021 from the full Maximum-likelihood tree in Supplementary Figure 2. The black circle and open circle identify WNV 21-3957/USA CT/Crow/2021 and WNV 21-3957/USA CT/Crow/2021 virus, respectively. The scale bars show the number of substitutions per site.

WNVs have been maintained in mosquitoes and were transmitted to these animals. In addition, the root-to-tip regression analysis of ML phylogeny using 904 complete genome sequences of WNVs showed a positive correlation between time and genetic distance with a high correlation coefficient value (0.93) and R square value (0.87), indicating gradual genetic evolution of WNVs in North America at an estimated evolutionary rate of 4.59×10^{-4} substitutions/site/year (data not shown). However, the details of the transmission routes of these viruses remain uncertain due to the lack of recent genome sequences and surveillance data.

Both the American crow and the alpaca had a history of neurological signs preceding death. Considering the seasonality

TABLE 3 Amino acid substitutions of WNV 21-3782/USA CT/Alpaca/2021 virus in virulence determinants in the WNV genome found in mammalian host.

Virulence determinants (protein-amino acid substitution)		WNV 21-3782/USA CT/Alpaca/2021
Mutations involved in the attenuated viral phenotype	PrM-N15Q	N
	M-I36F	I
	M-A43G	A
	E-L107F	L
	E-A316V/E	A
	E-K440R	K
	NS1-N130A	N
	NS1-N175A	N
	NS1-N207A	N
	NS1-P250L	P
	NS2A-A30P	A
	NS3-249	P
	NS3-483	D
	NS4A-Q46K	E
	NS4A-Q47K	L
	NS4A-D50K	A
	NS4B-P38G	P
	NS4B-C102S	C
	NS4B-E249G	E
	NS5-K61A	K
	NS5-D146A	D
	NS5-K182A	K
	NS5-E218A	E
	NS5-A804V	A
Mutations involved in the more virulent viral phenotype	E-154	N
	E-156	S
	E-I159V	A

TABLE 4 Amino acid substitutions of WNV 21-3957/USA CT/Crow/2021 virus in virulence determinants in the WNV genome found in avian host.

Virulence determinants (protein-amino acid substitution)		WNV 21-3957/USA CT/Crow/2021
Mutations involved in the attenuated viral phenotype	PrM-I141T	I
	E-S156P	S
	NS3-483	D
Mutations involved in the more virulent viral phenotype	NS1-K110N	K
	NS3-T249P	P
	NS4A-F92L	F

of the disease, WNV was considered one of the diagnostics differentials for both animals. WNV is considered a major public and animal health problem, causing diverse pathologies ranging from mild febrile to severe neurological damage and death. WNV pathotypes in birds and mammals have been associated with specific genotypes (12). For instance, the NS3-T249P mutation observed in WNV 21-3957/USA CT/Crow/2021 is a critical determinant of WNV virulence in American crows and present in many WNV strains that caused major outbreaks in humans such as in Egypt (1950), Romania (1996), Russia (1996), New York (1999), and Israel (1997–1998) (12, 17, 18). In Brault et al.'s studies (17, 18), the mutant WNV NY99-P249T and KN3829 (attenuated strain) gave rise to a low level, delayed viremia at day 3 pi, compared to high titers observed in WNV NY99 and KN3829-T249P. In addition, the E-159A observed in the WNVs sequenced in this study was found in many WNVs recovered after 2001, suggesting a possible link with the enhanced WNV spread and pathology in America after the year 2000 (19). Mutations at the NS4A sites have been associated with increased WNV virulence in mammals (12). The identification of virulence determinants and mutants as determined here via WGS is a crucial step in understanding WNV epidemiology, transmission, and pathogenesis. The study conducted here highlights the need for enhanced genomic surveillance of WNVs.

In this study, we report the complete CDS of WNVs identified from a crow and an alpaca in New England in 2021. NGS in combination with SISPA approach enabled WGS of WNVs directly from clinical samples. The use of the described NGS approach will allow efficient complete genome sequencing of circulating WNVs that can provide abundant information to understand the evolution and spread of WNVs. Additionally, the complete genome sequences and their phylogenetic relationships with other WNVs established in this study would be useful as reference data for future genomic surveillance of WNVs. Continued surveillance and genome sequencing of WNVs from animals as well as mosquitos would be needed to monitor virus evolution and transmission and to assess the emergence of genetic mutations that may be relevant for public health.

Data availability statement

The data presented in the study are deposited in GenBank under the accession number ON994909 and ON994910.

Ethics statement

Ethical review and approval was not required for the study on animals in accordance with the local legislation and institutional requirements.

Author contributions

J-YH: WGS data analysis and manuscript writing. ZH, AA, NT, and AH: sample preparation and data collection. D-HL: study design, WGS data analysis, and manuscript editing. GR: supervision and data collection. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by projects University of Connecticut SPS #180229 and University of Connecticut SPS #181033. D-HL was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Animal Disease Management Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA; grant number: 122057-2).

Acknowledgments

We thank the unconditional support of staff and faculty from the Connecticut Veterinary Medical Diagnostic Laboratory (CVMDL), Department of Pathobiology and Veterinary Science, CAHNR, University of Connecticut.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1085554/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Maximum-likelihood analysis of WNV lineage 1 and 2 reference sequences and 140 complete genome sequences of WNVs identified in United States including two WNVs of this study. The scale bars show the number of substitutions per site. The numerical values represent 1,000 bootstrap replicate values expressed as a percentage. The WNVs sequenced in this study were highlighted in red. Phylogenetic tree was rooted to the WNV lineage 2 reference sequence (NC 001563) as an outgroup.

SUPPLEMENTARY FIGURE 2
Maximum-likelihood analysis of WNV lineage 1 and 2 reference sequences and 906 complete genome sequences of WNVs identified in United States including two WNVs of this study. The scale bars show the number of substitutions per site. The numerical values represent 1,000 bootstrap replicate values expressed as a percentage. The WNVs sequenced in this study were highlighted in red. Phylogenetic tree was rooted to the WNV lineage 2 reference sequence (NC 001563) as an outgroup.

References

1. Richter J, Tryfonos C, Tourvas A, Floridou D, Paphitou NI, Christodoulou C. Complete genome sequence of west nile virus (Wnv) from the first human case of neuroinvasive Wnv infection in cyprus. *Genome Announc.* (2017) 5:17. doi: 10.1128/genomeA.01110-17

2. Kramer LD, Li J, Shi PY. West nile virus. *Lancet Neurol.* (2007) 6:171–81. doi: 10.1016/S1474-4422(07)70030-3

3. Hadfield J, Brito AF, Swetnam DM, Vogels CBF, Tokarz RE, Andersen KG, et al. Twenty years of west nile virus spread and evolution in the americas visualized by nextstrain. *PLoS Pathog.* (2019) 15:e1008042. doi: 10.1371/journal.ppat.1008042

4. Ciota AT. West nile virus and its vectors. *Curr Opin Insect Sci.* (2017) 22:28–36. doi: 10.1016/j.cois.2017.05.002

5. Lanciotti RS, Ebel GD, Deubel V, Kerst AJ, Murri S, Meyer R, et al. Complete genome sequences and phylogenetic analysis of west nile virus strains isolated from the United States, Europe, and the Middle East. *Virology.* (2002) 298:96–105. doi: 10.1006/viro.2002.1449

6. Berthet FX, Zeller HG, Drouet MT, Rauzier J, Digoutte JP, Deubel V. Extensive nucleotide changes and deletions within the envelope glycoprotein gene of Euro-African west nile viruses. *J Gen Virol.* (1997) 78:2293–7. doi: 10.1099/0022-1317-78-9-2293

7. Roehrig JT. West nile virus in the United States—A historical perspective. *Viruses.* (2013) 5:3088–108. doi: 10.3390/v5123088

8. Lanciotti RS, Kerst AJ. Nucleic acid sequence-based amplification assays for rapid detection of west nile and St. Louis encephalitis viruses. *J Clin Microbiol.* (2001) 39:4506–13. doi: 10.1128/JCM.39.12.4506-4513.2001

9. Chrzastek K, Lee DH, Smith D, Sharma P, Suarez DL, Pantin-Jackwood M, et al. Use of sequence-independent, single-primer-amplification (sispa) for rapid detection, identification, and characterization of avian RNA viruses. *Virology.* (2017) 509:159–66. doi: 10.1016/j.virol.2017.06.019

10. Stamatakis A. Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033

11. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using tempest (formerly path-O-gen). *Virus Evol.* (2016) 2:vew007. doi: 10.1093/ve/vew007

12. Fiacre L, Pages N, Albina E, Richardson J, Lecollinet S, Gonzalez G. Molecular determinants of west nile virus virulence and pathogenesis in vertebrate and invertebrate hosts. *Int J Mol Sci.* (2020) 21:239117. doi: 10.3390/ijms21239117

13. Davis CT, Ebel GD, Lanciotti RS, Brault AC, Guzman H, Siirin M, et al. Phylogenetic analysis of North American west nile virus isolates, 2001-2004: Evidence for the emergence of a dominant genotype. *Virology.* (2005) 342:252–65. doi: 10.1016/j.virol.2005.07.022

14. Gray RR, Veras NM, Santos LA, Salemi M. Evolutionary characterization of the west nile virus complete genome. *Mol Phylogenet Evol.* (2010) 56:195–200. doi: 10.1016/j.ympev.2010.01.019

15. Grinev A, Daniel S, Stramer S, Rossmann S, Caglioti S, Rios M. Genetic variability of west nile virus in us blood donors, 2002-2005. *Emerg Infect Dis.* (2008) 14:436–44. doi: 10.3201/eid1403.070463

16. Herring BL, Bernardin F, Caglioti S, Stramer S, Tobler L, Andrews W, et al. Phylogenetic analysis of WNV in North American blood donors during the 2003-2004 epidemic seasons. *Virology.* (2007) 363:220–8. doi: 10.1016/j.virol.2007.01.019

17. Brault AC, Huang CY, Langevin SA, Kinney RM, Bowen RA, Ramey WN, et al. A single positively selected west nile viral mutation confers increased virogenesis in American crows. *Nat Genet.* (2007) 39:1162–6. doi: 10.1038/ng2097

18. Brault AC, Langevin SA, Bowen RA, Panella NA, Biggerstaff BJ, Miller BR, et al. Differential virulence of west nile strains for American crows. *Emerg Infect Dis.* (2004) 10:2161–8. doi: 10.3201/eid1012.040486

19. Moudy RM, Meola MA, Morin LL, Ebel GD, Kramer LD. A newly emergent genotype of west nile virus is transmitted earlier and more efficiently by culex mosquitoes. *Am J Trop Med Hyg.* (2007) 77:365–70. doi: 10.4269/ajtmh.2007.77.365

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 17 May 2023
DOI 10.3389/fvets.2023.1086001

OPEN ACCESS

EDITED BY
Christina Leyson,
U.S. National Poultry Research Center, Agricultural Research Service (USDA), United States

REVIEWED BY
Cristina Kraemer Zimpel,
Michigan State University, United States
Hind Yahyaoui Azami,
University of Georgia, United States
Laetitia Canini,
Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES), France

*CORRESPONDENCE
Gianluigi Rossi
✉ g.rossi@ed.ac.uk

RECEIVED 31 October 2022
ACCEPTED 14 April 2023
PUBLISHED 17 May 2023

CITATION
Rossi G, Shih BB-J, Egbe NF, Motta P, Duchatel F, Kelly RF, Ndip L, Sander M, Tanya VN, Lycett SJ, Bronsvoort BM and Muwonge A (2023) Unraveling the epidemiology of *Mycobacterium bovis* using whole-genome sequencing combined with environmental and demographic data. *Front. Vet. Sci.* 10:1086001. doi: 10.3389/fvets.2023.1086001

Unraveling the epidemiology of *Mycobacterium bovis* using whole-genome sequencing combined with environmental and demographic data

Gianluigi Rossi[1,2]*, Barbara Bo-Ju Shih[1], Nkongho Franklyn Egbe[3], Paolo Motta[4], Florian Duchatel[1], Robert Francis Kelly[5], Lucy Ndip[6,7], Melissa Sander[8], Vincent Ngwang Tanya[9], Samantha J. Lycett[1,2], Barend Mark Bronsvoort[1,2] and Adrian Muwonge[1]

[1]The Roslin Institute, R(D)SVS, University of Edinburgh – Easter Bush Campus, Midlothian, United Kingdom, [2]Centre of Expertise on Animal Diseases Outbreaks, EPIC, Edinburgh, United Kingdom, [3]School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln, United Kingdom, [4]The Food and Agriculture Organization of the United Nations, Regional Office for Asia and the Pacific, Bangkok, Thailand, [5]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, United Kingdom, [6]Laboratory for Emerging Infectious Diseases, University of Buea, Buea, Cameroon, [7]Department of Biomedical Sciences, Faculty of Health Sciences, University of Buea, Buea, Cameroon, [8]Tuberculosis Reference Laboratory, Bamenda, Cameroon, [9]Cameroon Academy of Sciences, Yaoundé, Cameroon

When studying the dynamics of a pathogen in a host population, one crucial question is whether it transitioned from an epidemic (i.e., the pathogen population and the number of infected hosts are increasing) to an endemic stable state (i.e., the pathogen population reached an equilibrium). For slow-growing and slow-evolving clonal pathogens such as *Mycobacterium bovis*, the causative agent of bovine (or animal) and zoonotic tuberculosis, it can be challenging to discriminate between these two states. This is a result of the combination of suboptimal detection tests so that the actual extent of the pathogen prevalence is often unknown, as well as of the low genetic diversity, which can hide the temporal signal provided by the accumulation of mutations in the bacterial DNA. In recent years, the increased availability, efficiency, and reliability of genomic reading techniques, such as whole-genome sequencing (WGS), have significantly increased the amount of information we can use to study infectious diseases, and therefore, it has improved the precision of epidemiological inferences for pathogens such as *M. bovis*. In this study, we use WGS to gain insights into the epidemiology of *M. bovis* in Cameroon, a developing country where the pathogen has been reported for decades. A total of 91 high-quality sequences were obtained from tissue samples collected in four abattoirs, 64 of which were with complete metadata. We combined these with environmental, demographic, ecological, and cattle movement data to generate inferences using phylodynamic models. Our findings suggest *M. bovis* in Cameroon is slowly expanding its epidemiological range over time; therefore, endemic stability is unlikely. This suggests that animal movement plays an important role in transmission. The simultaneous prevalence of *M. bovis* in co-located cattle and humans highlights the risk of such transmission being zoonotic. Therefore, using genomic tools as part of surveillance would vastly improve our understanding of disease ecology and control strategies.

KEYWORDS

***Mycobacterium bovis*, whole genome sequencing (WGS), genomic surveillance, zoonotic tuberculosis, phylodynamics, phylogeography, multi-host system, one health**

1. Introduction

In the last two decades, the increased availability, efficiency, and reliability of genomic reading techniques, such as whole-genome sequencing (WGS) techniques, have ignited a profound transformation in understanding disease ecology and epidemiology. This, coupled with improved statistical methodologies and high-performance computing, has enhanced our understanding of pathogen dynamics and evolution (1).

Techniques such as WGS can identify polymorphisms in the genetic material, which is generated by transcription errors that can occur to the pathogen while replicating within their host (2). As the pathogen is transmitted through the host population, the accumulation of polymorphisms in its DNA/RNA can be used as a "transmission signature". Therefore, by tracking these mutations across bacterial genomes sampled in a host population, we are now able to infer transmission events between individual hosts, sub-populations, geographical areas, or species, while at the same time, we are able to gather insights about the evolutionary trajectory of a pathogen (2). Furthermore, when accurate spatial information on the sampled isolates is available, we can combine it with pathogen genetic data to disentangle the spatiotemporal dynamics of outbreaks, particularly in natural or other scarcely sampled animal populations (3).

Despite these advances, many challenges still exist, including the reconciliation between the temporal signal of outbreaks with pathogen mutations (4). *Mycobacterium tuberculosis* Complex (MTBC) members are clonal species, and therefore, recombination has been considered rare [although a recent publication showed otherwise (5)]. A few mutations are expected to occur for these species per year, generating little diversity during outbreaks in host populations. Consequently, there is inherent uncertainty in establishing infection patterns within the infected population and their associated infections. Therefore, combining genomic information with metadata is essential for accurate transmission chain estimation (4).

Mycobacterium bovis, a member of the MTBC group, is the etiological agent of animal or bovine tuberculosis (bTB) in bovids and other mammalians and of zoonotic tuberculosis (TB) in humans (6). Its infections are characterized by chronic disease, with or without a latent period, where infected cattle are hard to identify, making it hard to quantify potential infectious contacts (7). The estimation of *M. bovis* prevalence is often affected by several factors, including the inaccuracy of diagnostic tests (8), and the potential co-infection with other pathogens (9). Such challenges explain why *M. bovis* has only been successfully eliminated or controlled in a few countries. However, it still represents a significant threat to cattle industries and human health in many other countries. For example, zoonotic tuberculosis due to *M. bovis* is a major public health problem in low- and medium-income countries (LMICs), where close interaction between people and livestock is common and the access to pasteurized milk is limited (6, 10). Indeed, the magnitude of this burden is likely underestimated since human–animal transmission is predominantly via ingestion of infected products and it presents with a range of non-specific symptoms (11).

In Cameroon, *M. bovis* is circulating in the cattle population, both in the southern areas (12) and, in particular, in the northern regions, where a previous study on cattle sampled at four regional abattoirs showed a sampled population prevalence ranging from 2.75% (31 positive over 1,129 cattle inspected, Northwest) to 21.25% [34 over 160, North (13)]. Abattoirs surveillance, where carcasses are inspected for TB-like lesions, is the only surveillance strategy regularly implemented in the country; in Bamenda (Northwest region), Awah-Ndukum et al. (14) showed that the TB-like lesion in cattle increased in the period from 1994 to 2010.

Commonly to many LMICs, bTB control in Cameroon is also made difficult by the absence of detailed records on cattle population, by local rearing practices such as pastoralism which expose animals to contact with other herds and potential reservoir wildlife species, and by the transhumance cattle movements westward toward Nigeria, where the demand of meat is driven by a fast human population increase (15).

In a previous study, Egbe et al. (15) employed two molecular typing techniques to understand the relatedness of *M. bovis* strains circulating in the region. These are spoligotyping and MIRU-VNTR typing as follows: the former is based on the presence of multiple spacer oligonucleotides in the genome direct repeat region, while the latter is based on 12 loci containing variable numbers of tandem repeats of mycobacterial interspersed repetitive units (16, 17). Compared with WGS, these techniques consider a limited genome region and can be more subject to homoplasy (18). The results reported by Egbe (15) showed that most of the isolates belonged to the Af1 clonal complex ($n = 250$/total $n = 255$), while the remaining ones had an unidentified clonal complex. They also highlighted an unexpectedly high genetic diversity, as showed by the 37 sampled spoligotypes, of which, 19 were newly observed, and a total of 97 genotypes were obtained by combining spoligotypes with MIRU-VNTR (15).

While those techniques are instrumental to investigating potential infection clusters at a broader level, they can be limited for a more in-depth understanding of the spatiotemporal dynamics of the disease. This study aimed to fill these gaps and enhance our understanding of the *M. bovis* epidemiology and spatial dynamics in Cameroon using WGS. We applied novel phylogenetic techniques to determine whether there was endemic stability across Cameroon's cattle-rearing regions while examining the role of environmental and ecological variables and animal movements in the pathogen spread.

We used 91 high-quality *M. bovis* sequences obtained from cattle tissues sampled at regional abattoirs as described by Egbe et al. (13). After determining the single nucleotide polymorphisms (SNPs), we built a tree by joining the Cameroonian WGSs with other African sequences obtained from publicly available repositories, in order to understand how the sampled population fit in the continent context. Then, we ran a continuous space phylogeographical analysis with *BEAST* (19) on the Cameroonian sequences while testing different random walk diffusion models (20). This was possible because the origin village of the cattle tested at the abattoir was known for 64 *M. bovis* cattle isolates, allowing us to associate spatial coordinates with these sequences. Furthermore, we tested the association between the spatial pathogen distribution obtained with the georeferenced

phylogenetic tree and environmental, anthropic, and ecological factors (21), and we finally ran a machine learning analysis to test whether the empirical cattle movement network (22) or other variables could explain the genetic diversity across isolates.

Our findings strengthen the call for improved *M. bovis* molecular surveillance in underrepresented regions and countries, to gather insights into potential patterns that can be missed when limiting the studies to areas of low genetic diversity, the consequence of strict control practices such as test-and-cull.

2. Materials and methods

2.1. Data collection

Four regional abattoirs were sampled between 2012 and 2013, in the Northwest (Bamenda), Adamawa (Ngaoundere), North (Garoua), and Extreme North (Maroua) regions of Cameroon (Supplementary Figure 1). As part of the regular operations, cattle carcasses were inspected for the presence of TB-like lesions. The tissues, including lymph nodes, of all animals with lesions and of some randomly chosen ones without lesions were collected to be cultured, and information about the animal (age, breed, and village of provenance, among others) was taken. A detailed description of the data collection and bacterial isolation can be found in the study by Egbe et al. (13). The DNA extraction was conducted in BSL 3 facilities (Tuberculosis Reference Laboratories in Bamenda, Cameroon), and the procedure is fully described in the study by Egbe et al. (15). Sequencing was also attempted for *M. bovis* isolates sampled in human hosts at the Bamenda hospital (Northwest region) during a cross-sectional study within the wider project. We reported a summary of the number of sampled animals and the number of *M. bovis*-positive ones in Supplementary Table 1.

2.2. Whole-genome sequencing processing

The sequencing was carried out at Edinburgh Genomic Facilities (University of Edinburgh). Samples were prepared with 1 TruSeq Nano 550 bp insert, 76 Pippin selected library from the supplied genomic DNA, while MiSeq v2 (Illumina) was used to generate 250 base paired-end sequences from the library to yield at least 11M + 11M reads (1 run) at 30x coverage. The output was read from a 4-lane Miseq. A total of 124 *M. bovis* WGSs were obtained (two from human hosts), while for nine isolates (one from human hosts), the sequencing failed.

We used an adapted *BovTB*-nf pipeline (23) for quality control. Reads were deduplicated using *fastuniq*, trimmed using *Trimmomatics* (24) (-phred33 ILLUMINACLIP:$adapters:2:30:10 SLIDINGWINDOW:10:20 MINLEN:36), and mapped to the reference genome using *bwa-mem2* (25). The mapped reads were filtered (-ShuF 2308 -) and sorted using *Samtools* (26), and then classified using *Kraken2* (27) (–quick) against a prebuilt Kraken 2 database [*Minikraken* v2 (27)]. The Kraken2 output was summarized with Bracken (28) (-r 150 -l S), and the top 20 species from the Bracken output were used to determine if the sample was contaminated with other microorganisms. Variants were called using *bcftools* (29) (–IndelGap 5 -e 'DP<5 && AF<0.8'), and strain-specific SNPs were used for classifying whether the samples were *M. bovis* or not [custom script and differentiating SNPs taken from the study mentioned in MMMO (23)]. The percentage of coverage (>60%) on the reference, read depth (>10), and number of reads (> 600,000) were used to identify and remove samples with insufficient data. To curate aligned core-variants for the downstream phylogenetic analysis, variants were called and filtered using *Snippy* v4.6.0 (30), using the default settings (minimum coverage = 10, minimum VCF variant call quality = 100), with the *M. bovis* AF2122/97 genome (GenBank: LT708304.1) as the reference genome. Variants from repeated regions were removed [mask for repetitive regions taken from the study mentioned in MMMO (23)]. Core-SNPs were determined by the snippy-core function within *Snippy*, where a genomic position was considered to be a core-site when present in all samples. We defined "high-quality" sequences as the ones with genome coverage > 90% and reading depth > 10 (31), and we renamed the sequences with a string composed of the following information: host species, location (administrative subdivision or country, see Section 2.3), sequential number, and date.

For each sequence, the spoligotype and the clonal complex were retrieved from the study by Egbe et al. (15). In one case, a sequence was missing the spoligotype number; however, it was assessed with the *vSNP* pipeline (32). For all bioinformatics tools, we used the default settings unless stated otherwise.

We checked if divergent sequences belonged to other mycobacteria species. We tested the presence of regions of difference (RDs) 1, 4, 9, and 12 patterns (33) in the outlier samples, and raw reads from each sample were aligned to *M. tuberculosis* (NC_000962.3) with *Burrows-Wheeler Aligner* v0.7.17 (34) and sorted and indexed with *SAMtools* v1.10 (35). Primer flanking regions for the RDs on *M. tuberculosis* were determined by querying the sequences using NCBI web nucleotide BLAST with the default parameters (36), while the presence of RDs was manually determined by examining the read alignment in *Integrative Genomics Viewer* v2.14.1 (37).

2.3. Cameroonian *M. bovis* sequences in the African context

We obtained other *M. bovis* genomes from online repositories as follows: first, from the *Patric* (now, BV-BRC) dataset (38), and second, selecting the appropriate genomes among the ones listed by Loiseau et al. (39) and obtained from the EBI dataset (for details and references, see Supplementary Table 2). We selected all the available sequences sampled in Africa, in order to qualitatively detect potential genetic similarities between the sampled Cameroonian *M. bovis* population and other isolates from the African continent, and thus provide a broader context to our analysis.

When analyzing sequences from *Patric*, genomes were shredded into pseudo by *Snippy,* followed by the process of alignment and SNP identification described above. The core-SNP alignments were made with and without the other African genomes. We used the *iqtree* web server (40, 41) to compute a phylogenetic tree ($n = 212$), which included all the Cameroonian

high-quality sequences ($n = 91$) and the other African ones plus the 1997 reference from the UK ($n = 121$).

2.4. Cameroonian sequences phylogenetic analysis

The quantitative analyses were performed on a subsample of the Cameroonian sequences, obtained after removing the non-cattle ones, the ones missing the geographical coordinates and potential outliers, i.e., isolates not clustering within the main Cameroonian clade. We initially joined the remaining sequences ($n = 64$) of the tree using the TN93 genetic distance model and the neighbor-joining (NJ) algorithm *ape* package (42) in *R* v4.0.5 (43),

with the sole purpose of estimating a temporal signal within the sample in *TempEst* v1.5.3 (44). We, then, used the sequences SNP alignment completed with sampling dates, to infer time-scaled phylogenetic trees using *BEAST* v1.10.4 (19) with the *BEAGLE* library (45) and evaluated the results with *Tracer* v1.7.2 (46). Since the sequences had associated geographical location metadata, we included latitudes and longitudes as an additional continuous space variable for phylogeographic inference.

To select the best model, we ran a series of exploratory models using an HKY (47) substitution model, similar to other studies (48–50) and a strict molecular clock. First, we sequentially selected the best continuous trait model and then the best bacterial population size model (tree prior). We tested the Brownian random walk, Cauchy Relaxed Random Walk (RRW), lognormal RRW, and Gamma RRW for the former, and constant population, exponential growth, and Bayesian *Skygrid* (51, 52) for the latter. In the exploratory *BEAST* runs, we chose a truncated (between 0 and 0.1) normally distributed clock rate prior, with mean and standard deviation set as the slope in the root-to-tip obtained in *Tempest*; the chain length was set to 10^8 and sampled every 10^4 steps. The models were compared using marginal likelihood estimation (MLE), with path sampling (PS) and stepping-stone sampling (SS), if they reached a satisfactory effective sample size (>200). Once the model features were selected, we ran a final one setting the chain length to 10^9 steps and sampled every 10^5 steps. In this case, we used the clock rate posterior of the selected exploratory model as a prior for the final model. The maximum clade credibility (MCC) tree was extracted with *TreeAnnotator* v1.10.4 (part of the *BEAST* suite), and clades were visually defined within the MCC tree branches. The MCC tree was plotted against the sequence spoligotype and MIRU-VNTR typing to visually assess the correspondence between molecular typing and clades.

2.5. Spatial statistics and environmental factors analysis

From the final *BEAST* run, we extracted a set of 100 trees from the posterior distribution and further analyzed using *seraphim* v1.0 (21, 53), to obtain the spatial spread statistics: branch velocity and epidemic wavefront. The former was calculated for each branch dividing the geographical distance from the origin to the destination nodes by the time branch time duration. The epidemic wavefront shows the geographical range of the epidemic over time: at each time, it is calculated as the geographical distance between the positions of the tree's estimated root and the most distant node (spatial distance wavefront) or accounting for the distance of nodes closer to the root (patristic distance wavefront).

Additionally, *seraphim* allows to statistically test hypothesis on the effect of environmental layers on the epidemic dynamics; the effect can either be of "conductance", when the layer favors the pathogen diffusion, or "resistance", when the layer hampers it. We tested nine layers as follows: elevation, cattle population density, human population density, two describing the road infrastructure (number of intersections and total road length), and four land cover types (waterbodies, forest, grassland, and grazeland, and other vegetation types, such as mosaic, shrub, and sparse vegetation). The original raster layers were downloaded from online repositories (see Supplementary Table 3 for the sources) and adapted to a 5 km × 5 km grid using *QGIS* v3.26.1. For each cell, elevation, cattle, and human populations were averaged for the 5 × 5 km grids, while road intersections were counted, and road lengths were measured starting from the same road original raster. For the land cover, each value represents the percentage of that cell covered by each land cover type. The original land cover raster included 38 different cover types. To ease computation, we selected the most relevant for the study and merged them into four layers as follows: waterbodies, forest, cropland/grassland, and other vegetation, including mosaic, shrub, and partial cover (Supplementary Table 4).

First, we ran a preliminary analysis on each variable, to determine if it could have played a role as conductance or resistance in the pathogen spread. For each of the 100 extracted trees, we estimated the correlation between dispersal duration and environmental distance. The results are summarized by two statistics as follows: the number of positive variable's coefficient of determination out of the 100 trees, and the number of positive Q statistic, calculated as $Q = R^2_{var} - R^2_{null}$, that is, the difference between the correlation R^2 for the variable's raster and for a null raster, again calculated for each tree (53). For the analysis, we used two path models as follows: straight line (where the branch "weight" is calculated by summing the cell values through which the straight line passes) and least-cost path (where the branch "weight" is calculated by summing the values between adjacent cells along the least-cost path).

Once we identified the potential resistance or conductance factors, we performed 10 tree randomizations and calculated the statistics again. In this case, we used the Bayes Factor (BF_e), calculated as $BF_e = p_e /(1 - p_e)$, where p_e is the probability that $Q_{observed} > Q_{randomized}$. We used two criteria for tree randomizations as follows: (1) randomizations of node positions while maintaining the branches' lengths, the tree topology, and the location of the most ancestral node; and (2) randomizations of node positions while maintaining only the branches' lengths.

2.6. Genetic distance regression and role of the cattle movements

We, finally, tested which variables can better explain the genetic distances between the sampled *M. bovis* isolates, to understand the signatures of temporal, spatial, and demographic factors (54, 55). We ran this analysis using a Boosted Regression Tree (BRT)

regression model (56) in *R* software [packages *dismo* (57) and *gbm* (58)], a very flexible tool that combines decision trees and boosting techniques (59). In this model, the dependent variable was the genetic distance between *M. bovis* strains, expressed as SNPs. We tested a total of 28 relational variables, calculated for each pair of isolates (Supplementary Table 5). Except for the temporal and spatial distance (which were calculated from the original isolate metadata) and a binary variable indicating whether two sequences have the same spoligotype, MIRU-VNTR, and clade (yes/no), the other variables are associated with the *M. bovis* isolates administrative subdivision.

We built two subdivision-level contact networks. The first one is a spatial network where nodes represent subdivisions, and edges between them are positive if they share a border. This network is undirected (edges are not directional) and unweighted (all edge values are set to 1). For this network, we computed six variables to be associated with each pair of isolates: degree and betweenness centrality (60) of both isolates' subdivisions; shortest path and a binary variable indicating whether the two subdivisions belonged to the same network's community.

The second network represented the cattle movements, and edges correspond to the number of animals moved between subdivisions over a year. We built this network by aggregating the empirical data collected by Motta et al. (22), which originally reported the monthly number of cattle exchanged between markets. For this network, we computed eight variables as follows: degree, strength, and between centrality of both isolates' subdivisions; shortest path and the same community binary variable. The degree counts the number of each subdivision's connections, while the strength is the sum of the number of cattle moved to and from each subdivision. All networks' metrics were computed using the *R* package *igraph* (61).

Once we computed all the variables (the full list is presented in Supplementary Table 5), we trained the BRT model using 75% of the observations, while the remaining 25% were used for testing. We evaluated the models based on pseudo-R^2 and Root Mean Squared Error (RMSE) on the test dataset. These were both calculated using the package *caret* (62). For BRT, the relative influence of the variables is determined by the times each variable is selected to split the data in a decision tree, which, in turn, is weighted by the improvement in the model fit that resulted from the variable being used at each split (56). All models were fitted with a 10-fold cross-validation. The BRT algorithm has two main parameters as follows: the learning rate, which controls the contribution of each tree to the final model, and the tree complexity, which corresponds to the number of nodes in the tree. We ran some preliminary tests to tune the BRT in order to improve the predictions. Finally, we set the learning rate to 0.05 and the tree complexity to 8.

3. Results

3.1. Cameroonian sequences in the African context

We analyzed 124 *M. bovis* sequences (nine of the original 133 failed), with 91 having enough read depth and genome coverage to allow further analyses (see Supplementary Table 6 for further details). Two of these sequences came from isolates sampled humans, while for a third, the sequencing failed. One of the excluded sequences was marked as not *M. bovis*, and based on the presence of the four RDs 1, 4, 9, and 12 patterns (33), it was likely *M. tuberculosis*. All the high-quality *M. bovis* sequences were merged into a tree, with other 22 obtained from the *Patric* dataset, 99 from EBI, and the 1997 UK Reference to provide a continental context. The qualitative phylogenetic tree, in Figure 1, shows that most of the Cameroonian sequences (two of which obtained from human tissue samples) cluster with the Ghanaian human samples, and two Nigerians recovered from unreported hosts. All human samples from West Africa cluster with cattle sequences, except for the Malian human sequence. Most sequences ($n = 89$) belonged to Af1 clonal complex, and except for one, the spoligotypes were already known; for the other, we identified a new pattern (hex code: 6F-1F-5F-7F-BF-40). Being characterized by the absence of spacer 30, this spoligotype was considered Af1 (63). The dominant spoligotype was SB0944 ($n = 32/89$).

Two outlier sequences did not cluster with the rest of the sampled Cameroonian population. Their average distance from the rest of the Cameroonian population (respectively, 235 and 231 SNPs) was slightly higher than the average distance of the 1997 UK Reference from the Cameroonian isolates (222 SNPs), and they did not cluster with any other WGS sequence sampled in Africa (Figure 1). For both outlier sequences, the spoligotype was SB2332, found for the first time in Cameroon and submitted for classification at www.Mbovis.org by Egbe et al. (15). Following Warren et al. (33), we tested the presence of RDs 1, 4, 9, and 12 patterns, finding only the first one, confirming that they are likely *M. bovis*. We compared this spoligotype pattern with all the others from the *www.Mbovis.org* database, and we identified four patterns differing by two spacers as follows: SB0858 sampled in Spain (64) (different in spacers 20 and 22), SB1102 sampled in Chad (63) and Cameroon (12) (different in spacers 33 and 34), SB2333 reported by Egbe et al. (15) (different in spacers 22 and 34), and SB2691 sampled in France (not found in publications, different in spacers 20 and 34). We also identified 11 patterns differing by three spacers, sampled in France (65), Portugal (66), and Spain (64).

3.2. *M. bovis* evolutionary time scale in Cameroon

A total of 1,540 SNPs were determined from the *Snippy* core-SNP analysis on Cameroonian *M. bovis* genomes (Supplementary Figure 2). This was reduced to 1,106 SNPs when the dataset was reduced to 64 samples with complete metadata and excluding the non-cattle ones (two sampled in humans), which were used for the downstream quantitative analysis. The median SNP distance among the remaining high-quality sequences was 70 SNPs (mean 68, range from 0 to 144, 2.5th and 97.5th quantiles 14 and 118). For two cattle (one from Bibemi and the other from Touboro), two *M. bovis* isolates sequenced were available (obtained from different tissues). In both cases, the two strains were identical (Bibemi, 3 and 4; Touboro, 7 and 8, Figure 2), which suggests a

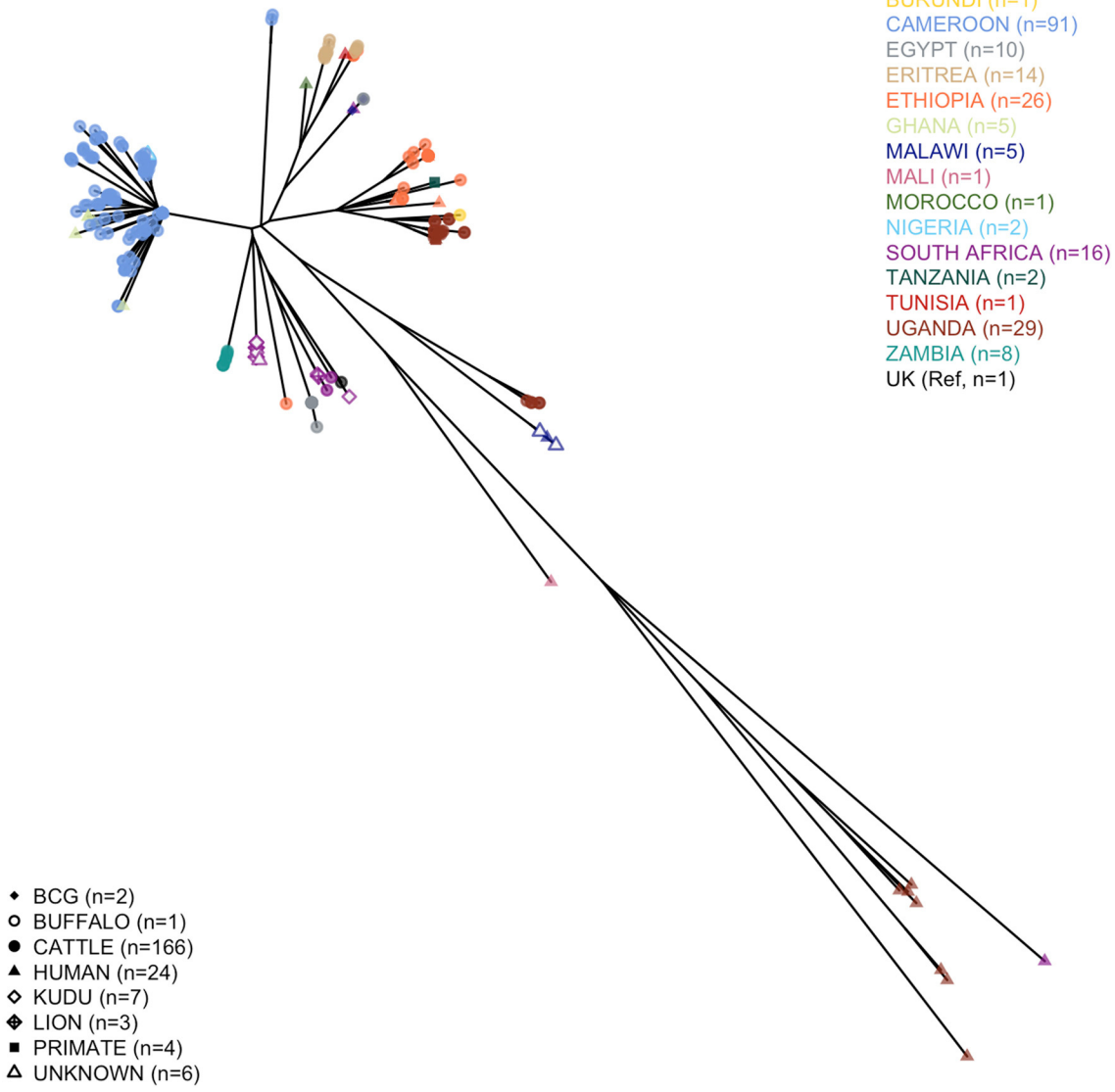


FIGURE 1
Phylogenetic tree of the African *Mycobacterium bovis* whole-genome sequences considered in the study. The tree includes 91 high-quality Cameroonian sequences, 101 from the EBI dataset, 20 from Patric, and the 1997 UK *M. bovis* reference.

single infection disseminated in different organs rather than two separate infections.

The analysis in *Tempest* showed a slightly positive temporal signal (coefficient of determination 0.11 and correlation coefficient 0.33) and a slope of 1.267×10^{-2} (Supplementary Figure 3). We used a sequential approach in *BEAST* to select the best spatial model and bacterial population models. Based on the MLE estimation of the exploratory models (Supplementary Table 7), we determined the best model included a Gamma Relaxed Random Walk (RRW) spatial model (first step of the sequential analysis) and the SkygGrid population model (second step). The final *BEAST* model was run with 10 bins and a cutoff of 400 years. The population trend is shown in Supplementary Figure 4. The model estimates suggest that the mean age of the root was in July 1950 (95th high-posterior density, HPD, April 1938–August 1961), while the average clock rate was 1.32×10^{-7} substitution/site/year (95th HPD 1.20×10^{-7} – 1.44×10^{-7}). The maximum clade credibility (MCC) tree is presented in Figure 2, which also shows the division in four clades as follows: clade 1 (green, 22 isolates), clade 2 (blue, 17 isolates), clade 3 (purple, 19 isolates), and clade 4 (red, five isolates). One sequence was excluded from all clades (Belel, 4; Figure 2, reported as "no clade" in the figures). The geographical distribution of the clades is presented in Figure 3, showing the number of *M. bovis* isolates per administrative subdivision, which ranged from 1 to 17 (see Supplementary Table 8 for the number of isolates per clade by regional abattoir). In Figure 4, we superimposed the MCC tree with spoligotypes; the most prevalent spoligotype, SB0944, occurred 26 times (out of 64 sequences) and was present in three of the four clades. The second most prevalent spoligotypes were SB0953 and SB2312, the first occurring five times in two clades and the latter occurring five times in one clade only (clade 2). We also superimposed the MIRU-VNTR types, as shown in

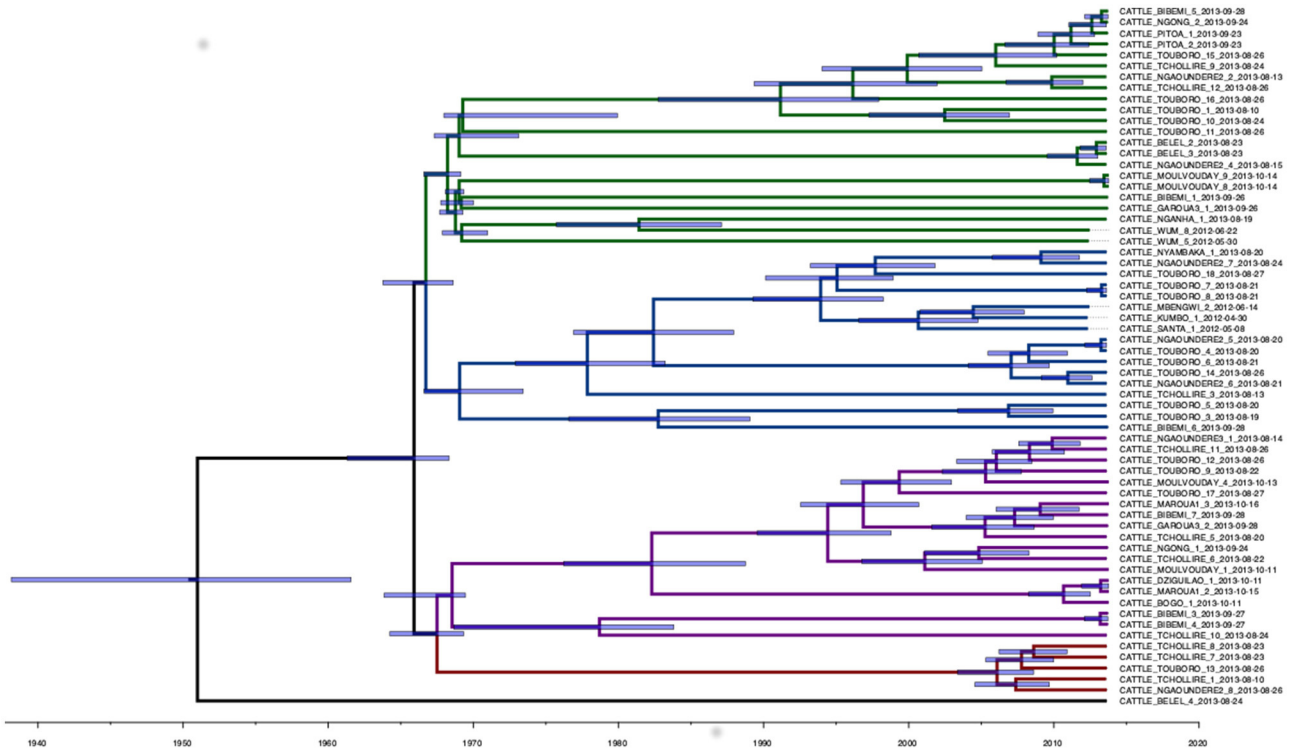


FIGURE 2
Phylogenetic time-scaled MCC tree of the 64 high-quality *M. bovis* whole-genome sequences sampled in Cameroon in 2012 and 2013. The blue transparent bars represent the 95th HPD of the internal node dates, while the branch colors represent different clades: 1 (green), 2 (blue), 3 (purple), and 4 (red). A non-time scaled tree showing the genetic distance between the 64 sequences is presented in Supplementary Figure 6.

Supplementary Figure 5. The most prevalent MIRU-VNTR type in the sampled population was V89, which occurred nine times; V82 and V37, respectively, occurred six and four times; and V81, V76, and V100 all occurred three times. Seven MIRU-VNTR types occurred two times, while 39 types occurred only once.

3.3. Spatiotemporal pathogen expansion

The estimated mean branch velocity was 53.1 km/year (95th CI 18.4–219.0, temporal trend presented in Supplementary Figure 7). The wavefront statistics in Figure 5 suggest that the pathogen expansion was slow until the mid-1960s but accelerated thereafter to reach the entire study area, with a slow but constant expansion in the following period. This is reflected in an increase in the branch velocity at the same time (Supplementary Figure 7), which is approximately the period when the branches formed the observed clades (Figure 2). The timing of the different branches in space is presented in Figure 6 (95th HPD in Supplementary Figure 8, with nodes colored by estimated/observed date).

We tested the association between nine geographical variables with the dispersal duration. Table 1 shows the results obtained using the straight line and the least-cost path models, and the latter run considering the variables as potential conductance or resistance factor. A total of six variables resulted in a significant association (positive coefficients for all at least 95 out of 100 trees and above 75% of positive *Q*) as follows: mosaic, shrub, and other vegetation covers (with both path models as resistance in the least-cost one); forest cover, elevation, and waterbodies cover (all as conductance); and cattle density (as resistance). However, when their statistical significance was tested through randomization, only forest cover and elevation (both as conductance) showed a Bayes factor significance [$\geq$ 3 (67)]. The result was robust against two different tree randomization algorithms for the forest layer, while for the elevation, this was true only when maintaining the branches' length and excluding the other tree topological characteristics.

3.4. Factors associated with genetic distance

The RMSE of the boosted regression tree BRT model ran using all 28 variables was 20.23, while the R^2 was 0.450. We simplified the model using the *dismo* package, which tests the performance of the model by dropping the less important variables with a procedure similar to backward selection in regression (56). The algorithm brought to eliminating 12 variables (see Supplementary Table 5); nonetheless, the model run using the remaining 16 variables performed very similarly to the original one (RMSE = 20.22 and R^2 = 0.452). Therefore, we used the latter to calculate the variable importance (Figure 7).

As expected, the most relevant variables were the temporal distance between the samples (first), and the binary variable indicating whether the two *M. bovis* isolates belonged to the

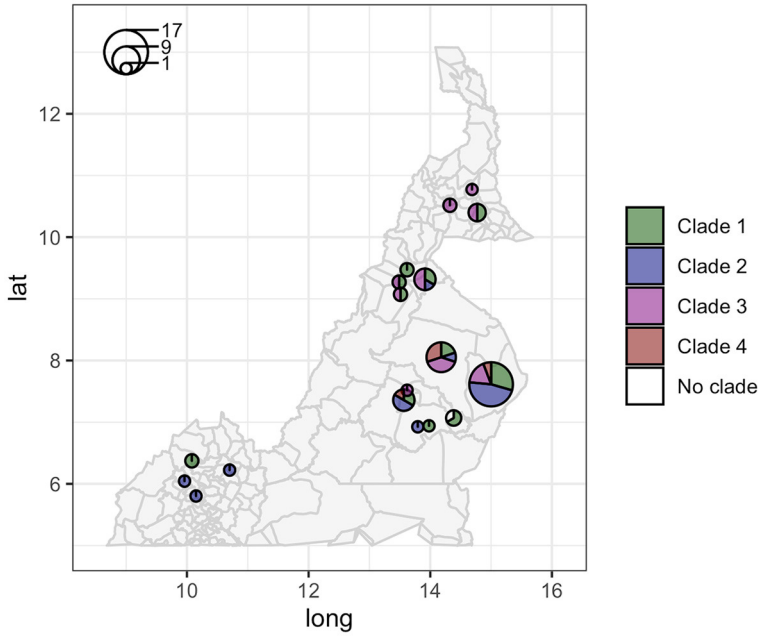


FIGURE 3
Geographic distribution of the 64 high-quality *M. bovis* whole-genome sequences in Cameroon. Circle sizes correspond to the number of sequences per administrative subdivision, and colors represent different clades (clade 1 green, clade 2 blue, clade 3 purple, and clade 4 red).

same clade in the MCC tree (second). The variables describing the subdivisions' population were also relevant in the model (population.y, third, and population.x, fifth), as well as whether the two isolates shared the same MIRU-VNTR (fourth). This was more relevant than the two isolates shared the same spoligotypes (11th), suggesting the former was more useful to discriminate closer *M. bovis* strains. The market movement network strength (i.e., the number of cattle moved from/to a subdivision) was the most important (sixth and ninth) among network-related variables, while the betweenness (eighth and 10th) was the only spatial network variable retained in the simplified model. Interestingly, when both variables were selected for the same metric, the one related to the youngest isolate (marked by *y*) was always preferred to the one related to the oldest isolate (marked by *x*). The partial dependency plots, showing the relationship between SNP distance and variables, are presented in Supplementary Figure 9.

4. Discussion

We sought to unravel the characteristics of the spread of a pathogen with zoonotic potential in time and space to improve our understanding and inform control and preparedness strategies. Our basic premise is that the accumulation of mutations in the pathogen's genome can be used as signatures of transmission events from host to host across time and space. Within space, the environment can create barriers that influence the population dynamics of diseases, i.e., altering host-to-host and pathogen–host interactions has direct effects on the genetic structure of the pathogen (68). The availability of high-throughput genomic techniques means we can interrogate the structural changes linked to the environment over time to gain critical insights into how the epidemic has evolved. In this study, we aimed to characterize *M. bovis* samples from cattle in Cameroon using genetic and demographic data to understand whether the pathogen is in a stable endemic state and the influence on the spread dynamic of environmental and ecological factors and cattle movements.

4.1. Evidence of dynamic endemicity

An important question was whether the *M. bovis* outbreak in North Cameroon was in a steady state, at an endemic equilibrium, or if it was expanding. Determining whether a pathogen is endemic has implications on risk perception and consequently on resource allocation. At the same time, the chances of zoonotic transmission are likely to be higher in the case of endemicity. In our analysis, the Bayesian model estimation with SkyGrid as a population model showed an increasing pathogen effective population size, corresponding to a constant increase in the disease velocity after the sudden jump during the mid-to-late 1960s. This suggests that the pathogen is not in a state of endemic stability, instead, it has been expanding at various rates over the years. This is in agreement with a previous publication using spoligotypes and MIRU-VNTR (15) and with the study by Awah-Ndukum et al. (14). The expansion of *M. bovis* might represent an issue for livestock and humans, particularly as we showed that the bacterium is circulating in both. At the moment, disease control in the area is absent, while, on the other hand, the dairy industry in Africa is generally expanding. A lack of widespread milk pasteurization could lead to an increase in zoonotic TB cases, which already represent a problematic issue in the region (11).

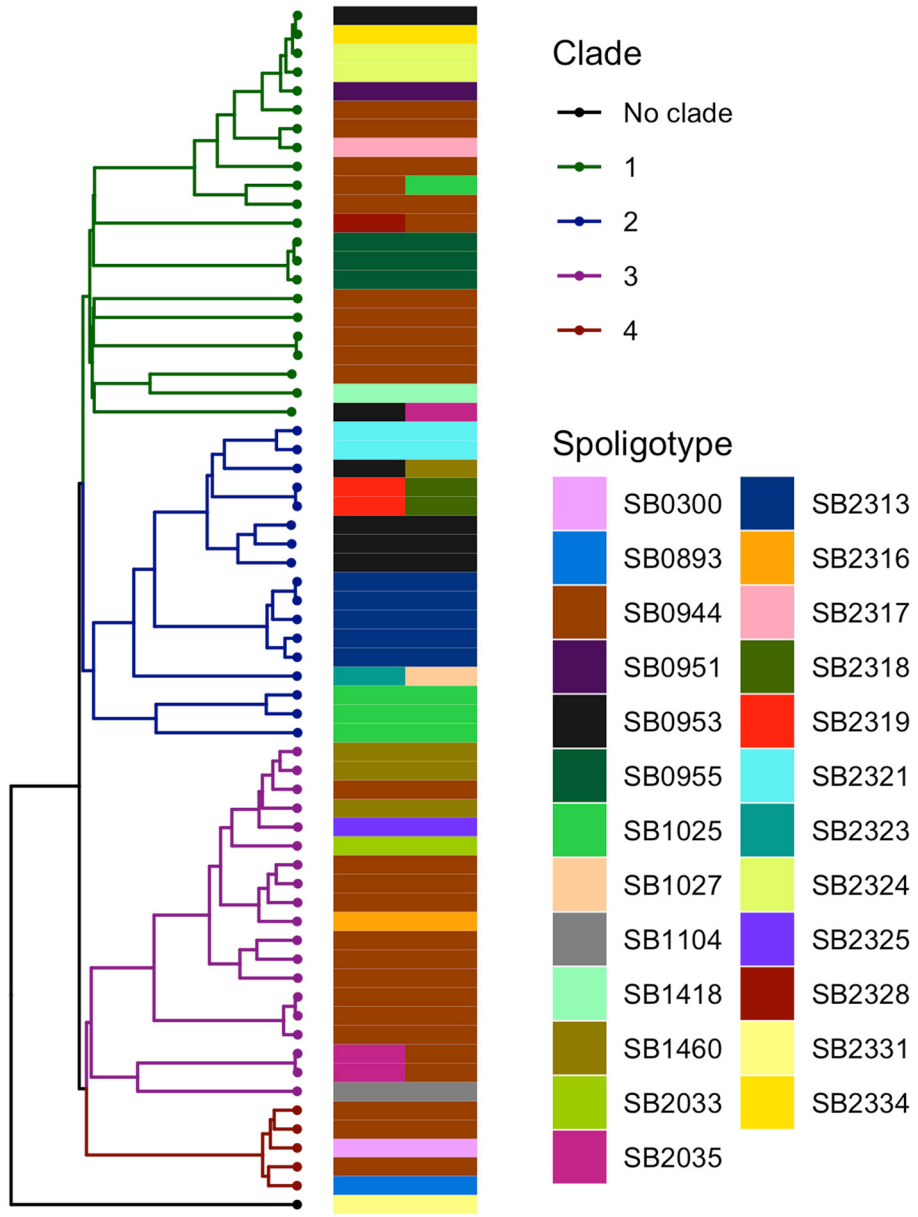


FIGURE 4
Visual comparison between the *M. bovis* phylogenetic time-scaled MCC tree and the spoligotypes obtained by Egbe et al. (15). A total of 10 sequences were associated with two spoligotypes because multiple samples from the same animal (up to three) were submitted for spoligotyping.

4.2. Genetic diversity of *M. bovis* in Cameroon

We observed a high diversity of *M. bovis*, confirming earlier observations with molecular typing techniques, providing less granular information (15), and considering the short time span of the sampling campaign and the small sample size. This contrasts with areas such as Great Britain and other European countries, where strict control measures, such as routine testing and stamping out of positive individuals, have been in place for decades. These measures could have resulted in a genetic bottleneck which hampered the pathogen's genetic variability, in particular by reducing the time a pathogen has to develop inside a domestic host and, therefore, the likelihood of substitutions in the DNA. As an example, Crispell et al. (55) reported a similar SNP distance range (0 to 150), albeit across a much bigger sample ($n = 230$), with a lower median (20 SNPs) and with isolates dating back two decades, while in a similar size monophyletic outbreak ($n = 64$), Rossi et al. reported a maximum SNP distance of only six SNPs (54). In Spain, Pozo et al. (69) found a similar SNP distance average and range (62 and 0 to 150) in a bigger *M. bovis* population, sampled in both cattle and wildlife over 13 years. It is noteworthy that high diversity can be associated with dynamic epidemiology and not with endemic stability.

All 64 core isolates belonged to the clonal complex Af1, which was observed in the region in previous studies (63). The

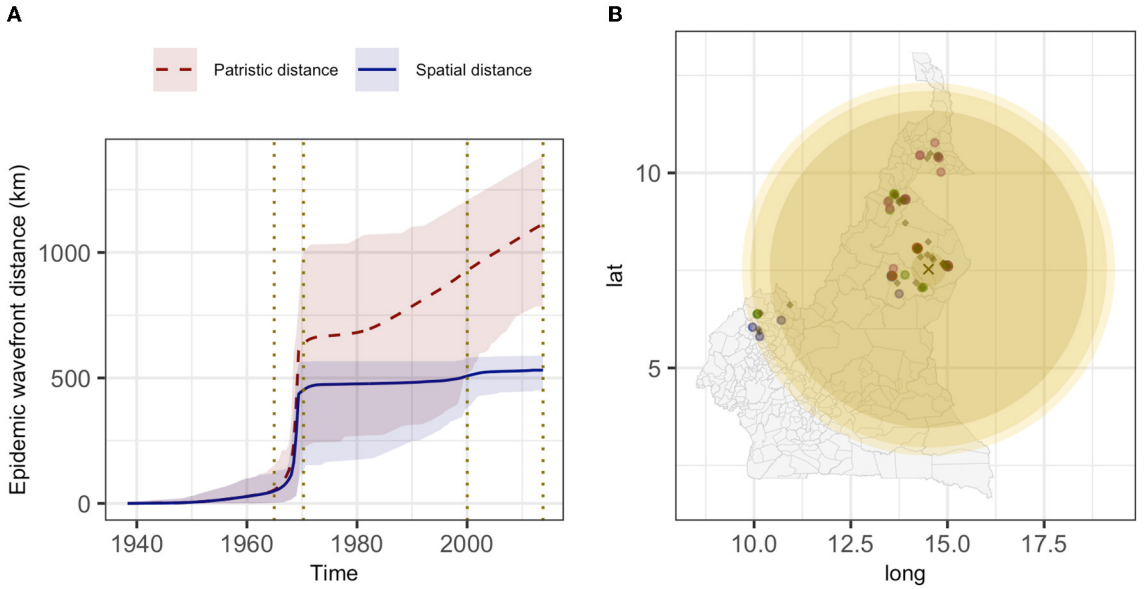


FIGURE 5
The estimated epidemic wavefront over time **(A)** and the expansion of the epidemic wavefront on the map **(B)**. **(A)** mean (lines) and 95th HPD (shades) of the epidemic wavefront spatial distance (blue) and patristic distance (red) over time. **(B)** different yellow shades represent the epidemic wavefront at a sequential point in time [marked by vertical dotted lines in **(A)**], and lighter shades of yellow correspond to more recent expansion; the estimated tree's root location is indicated by the black cross, diamonds represent the internal nodes estimated locations and circles the sampled isolates (colored by clade: 1, green; 2, blue; 3, purple; and 4, red).

most common spoligotype, SB0944, was found by Müller et al. (63) as the most prevalent in West Africa and considered the original of the Af1 clonal complex. Our findings also suggest zoonotic transmission in West Africa, as sequences recovered from humans in Cameroon and Ghana clustered with Cameroonian cattle *M. bovis* isolates (70). Because it is known that zoonotic TB represents a minoritarian but still crucial part of all TB cases in Africa, these results strengthen the case for One-Health approaches to control, which involve humans, livestock, wildlife, and environmental health (11, 71). Except for the one sequence in Mali and the two Cameroonian outliers, all the sequences from West Africa clustered together, hinting at high connectivity likely caused by cattle movements throughout the area, as previously shown by another study (72). Our results showed that the areas with the highest *M. bovis* diversity were in the Adamawa and North regions, both reporting all the clades identified by the maximum clade credibility (MCC) tree. All clades were also sampled in the towns of Touboro and Tchollire, both located in the North region but close to the Adamawa border. Previous studies reported that this area receive cattle from neighboring country as part of the transhumance migration, suggesting that cattle movements and markets play an important role in defining the dynamics of the pathogen, and therefore influencing its genetic diversity (15, 22). The Northwest region was underrepresented in the sample, with only five high-quality sequences on 31 infected cattle detected at the abattoir. This inherently reduces the level of diversity, which is far lower than reported using spoligotypes and MIRU-VNTR (15).

Despite covering a smaller portion of the genome and the higher occurrence of homoplasy with respect to WGS, in other contexts, spoligotypes have been used as a proxy cluster or to narrow down potential transmission within the study population (55, 73). Our results showed that this cannot be conducted for areas with high diversity such as the one we considered, as we observed little correspondence between the MCC tree branches and the spoligotypes. Similarly, other studies pointed out the limitations of such typing techniques (18), in the case of an expanding infection where transmission is steadily ongoing, compared with point-source ones (74). The high SNP distances among the sampled isolates also precluded the use of methods to infer direct transmission between hosts (7, 75).

When considering the entire sampled population, therefore including the sequences with incomplete metadata, we found two of the 91 sequences not belonging to the clonal complex Af1. In their spoligotype pattern (SB2332), we noted the absence of spacer 21 (76), and the closest relatives analyzed by Loiseau et al. (39) were identified as part of the clonal complex Eu2, including isolates sampled both in south-western Europe (SB0837, SB1090, and SB1308) and West Africa [SB1102, isolated in Cameroon as well (12)]. We can, then, speculate that these sequences likely belong to Eu2 as well, although we could not exclude one of the "unknown" clonal complexes identified by other studies (39, 77). Further development on this point was beyond the scope of this study, as we focused on the 64 core sequences to gather insights into the pathogen dynamics in the area.

4.3. Tracking the spread of *M. bovis* in Cameroon

We acknowledge that our estimates for the most recent common ancestor (MRCA) have a wide credible interval around it

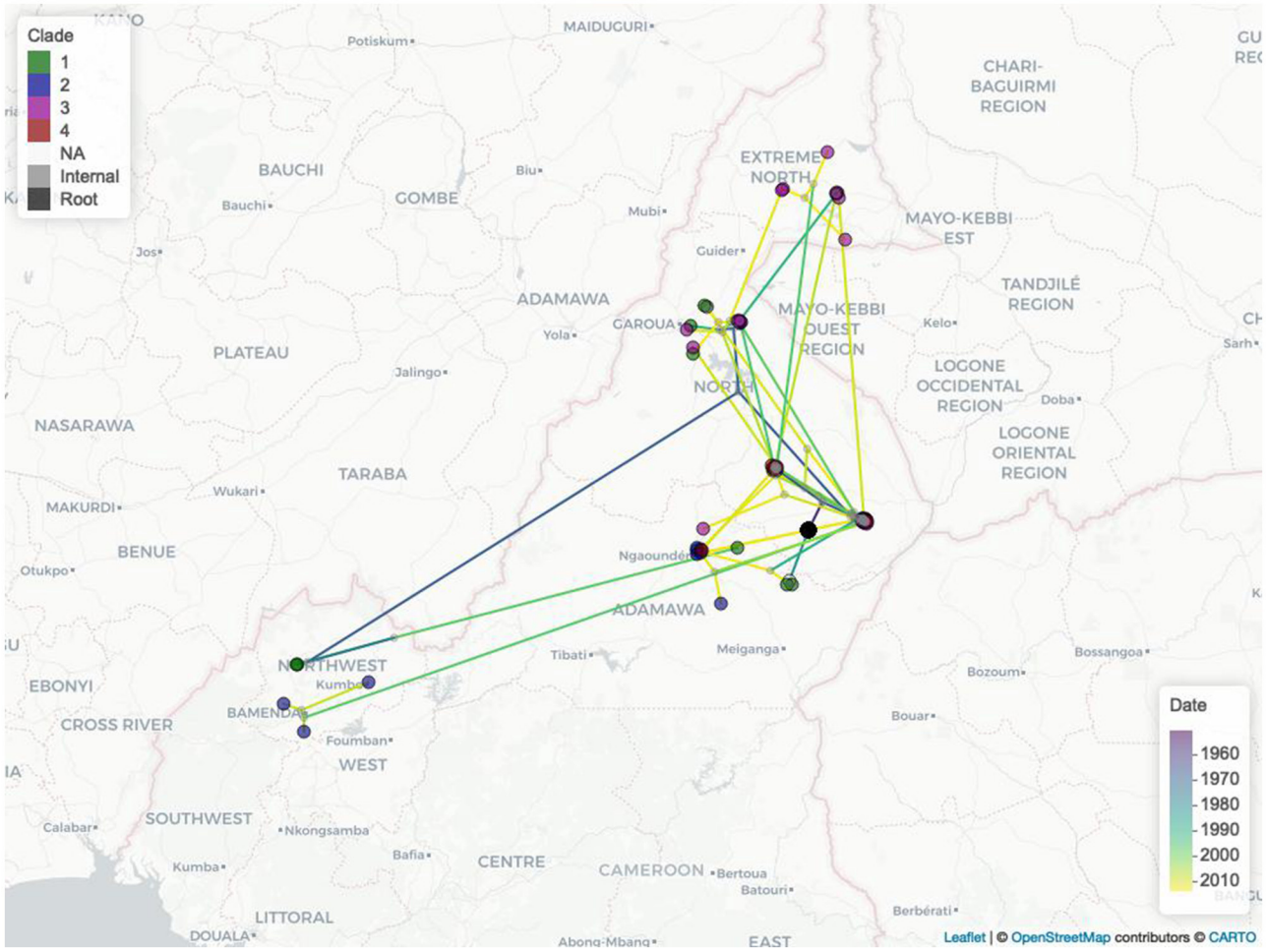


FIGURE 6
The Cameroonian *M. bovis* epidemic estimated expansions in space and time. Nodes are colored by clade (1, green; 2, blue; 3 purple; 4, red; no clade, light gray; internal nodes, dark gray; tree root, black), while the branches are colored by estimated movement date from 2007 (purple) to 2013 (yellow).

(23 years). This uncertainty is likely due to the short duration of the sample collection campaign, which also generated a weak temporal signal, although the coefficient of determination was similar to other *M. bovis* studies in highly sampled populations (54, 55). Nonetheless, our estimates coalesce around 1950, suggesting that the pathogen has been spreading in the area for at least six decades at the time of sampling. For the same reason, the estimated clock rate was higher than others in the literature but in the same order of magnitude ($0.67–1.26 \times 10^{-7}$, $n = 2625$ (39)].

The estimated MCC tree located at the most recent common ancestor (MRCA) in Touboro (North region), and from there, a rapid expansion of the outbreak reached most of the study area by the early 1970s. From the estimated origin, the pathogen likely spread first northward to Garoua (in the same region) and westward to the Northwest region, and later to the Extreme North and Adamawa regions and again to the Northwest.

The results of the spatial factor analysis showed that forest cover and elevation were the only significant ones, both acting as "conductance". Forest cover could be a proxy for potential wildlife interactions, as *M. bovis* is known to be quite effective in spreading at the wildlife–livestock (and humans) interface (71, 78). The elevation as conductance was counter-intuitive; however, this could be linked to cattle movements in pastoralist communities within the plateau located in the study area. This is important because, if confirmed, altitude could be used as a proxy for the missing pastoralist movements.

Our regression model performed reasonably well, although the amount of variability explained was below 50%. However, our objective was to understand which variables could better explain the genetic distance between *M. bovis* isolates, expressed as SNP distance. Except for the isolates between temporal distance and clade, the demographic variables were the most effective in explaining SNP distance, particularly the administrative subdivision human population size. These variables had a negative effect on the SNP distance, meaning that a smaller population was associated with a close relatedness of the *M. bovis* strains. This could be an effect of the population distribution on the country because the northern regions, where cattle are most

TABLE 1 The results of the analysis on nine spatial variables, assuming two path models, straight line and least cost, and for the least cost path, whether the variable worked as a conductance or resistance.

Variable	Type	Path model	Number of positive coefficients	Number of positive Q statistic	Mean Bayes factor	
					(Randomization #1)	(Randomization #2)
Mosaic_shrub_otherv	Resistance	Least cost	100	99	1.89	1.83
Forest	**Conductance**	**Least cost**	**100**	**96**	**3.66**	**3.66**
Mosaic_shrub_otherv	NA	Straight line	100	89	0.62	1.32
Elevation	**Conductance**	**Least cost**	**100**	**88**	**2.39**	**3.00**
Waterbodies	Conductance	Least cost	100	87	1.10	0.97
Cattle_density	Resistance	Least cost	99	77	2.49	2.88
Cattle_density	NA	Straight line	100	73	Not run	
Cattle_density	Conductance	Least cost	100	70	Not run	
Grassland_cropland	Resistance	Least cost	100	66	Not run	
Grassland_cropland	NA	Straight line	100	56	Not run	
Mosaic_shrub_otherv	Conductance	Least cost	100	44	Not run	
Roads_intersections	Conductance	Least cost	100	42	Not run	
Waterbodies	Resistance	Least cost	100	38	Not run	
Waterbodies	NA	Straight line	100	27	Not run	
Grassland_cropland	Conductance	Least cost	100	15	Not run	
Forest	Resistance	Least cost	100	12	Not run	
Elevation	NA	Straight line	100	7	Not run	
Forest	NA	Straight line	100	3	Not run	
Pop_density	Conductance	Least cost	99	40	Not run	
Elevation	Resistance	Least cost	99	15	Not run	
Roads_length	NA	Straight line	97	0	Not run	
Pop_density	NA	Straight line	96	16	Not run	
Pop_density	Resistance	Least cost	96	7	Not run	
Roads_length	Conductance	Least cost	92	11	Not run	
Roads_length	Resistance	Least cost	59	0	Not run	
Roads_intersections	NA	Straight line	56	1	Not run	
Roads_intersections	Resistance	Least cost	6	1	Not run	

The results show the number of positive coefficients for the 100 sampled trees, the number of positive Q statistics, and the mean Bayes factor calculated over 10 randomizations, testing two algorithms as follows: (1) randomizations of nodes positions while maintaining branches lengths, tree topology, and location of the most ancestral node; and (2) randomizations of nodes positions while maintaining only the branches lengths. The randomization algorithm was not run for spatial variables not significant in the preliminary analysis, variables significant in the randomization analysis were marked in bold.

concentrated, are less populated compared with the cities in the south. The simplified model performed similarly to the full model, suggesting some variables were not important in explaining the genetic distance. Beyond the human population size, also the other demographic variables (population and cattle density) were retained. Conversely, only five network-related variables were retained, three for the cattle movement network (out of eight) and two for the spatial network (out of six). All network-related variables had a positive effect on the SNP distance, with the number of cattle moved in or out of a subdivision (i.e., strength), having the higher predictive effect. Interestingly, this result was similar to other studies where cattle movements alone could not fully capture *M. bovis* genetic diversity (54, 55).

4.4. Limitations

The major limitation of this dataset was the short data collection time window, less than a year and a half, which resulted in uncertainty in the MRCA estimate and a weak temporal signal. While we can speculate the sampled bacterial population already reached the entire study area before the 1970s, a wider sampling time window would likely allow a stronger temporal signal and

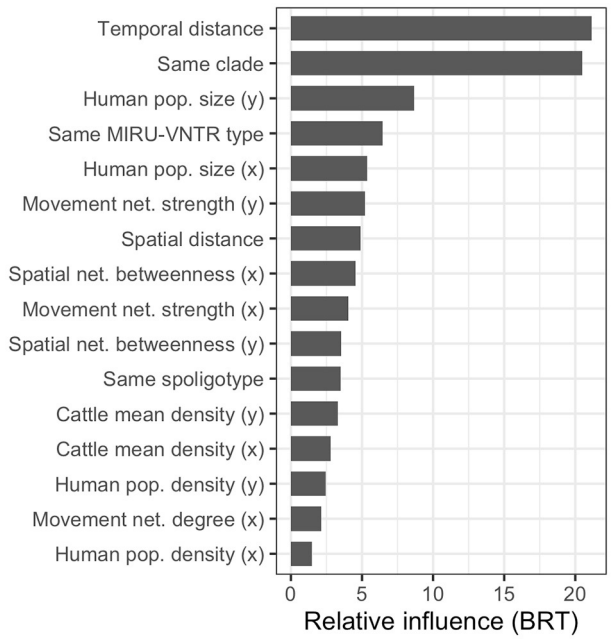


FIGURE 7
Relative influence of the most relevant variables in the simplified boosted regression tree (BRT) model used to explain the SNP distance between the 64 high-quality *M. bovis* isolates. Many variables are calculated between isolates pairs, *x* refers to the oldest isolate's subdivision, and *y* to the youngest one.

improve our estimate of the MRCA, which might be prior with respect to the current estimate. In turn, this affected the pathogen's expansion patterns, including the branch velocity and wavefront, which are also limited by the sampled area size. The spatial uncertainty might also be affected by the absence of dense cattle movement records, so the known spatial coordinates associated with each sequence correspond to the last village the animal lived in. The Adamawa and Northwest regions are home to 1.25 million and 450.000 cattle, respectively (22), and while this abattoir-based study provides a very informative snapshot of the *M. bovis* population in North Cameroon, it adds to the calls to improve cattle records and movement routine data collections in LMICs (79), as well as bTB detection efforts.

The low-quality WGSs disproportionally affected the Northwest region, as presented in Supplementary Table 7. This could have hampered the representativeness of the *M. bovis* diversity in that region, reducing the number of clades observed. The Adamawa region was the most represented, despite most of the sequences excluded from the quantitative analysis because of missing coordinates, which came from the Ngaoundere abattoir. The bacterium diversity in the Northwest might also be affected by the demographic of the slaughtered cattle in the region (13) because the region is highly populated by humans and more isolated in the trade network (22), and local animals of both sexes and at any age are slaughtered. Conversely, young male calves from the Adamawa, North, and Extreme North regions are often sent to richer southern regions to maximize their economic values, leaving the older cows to be slaughtered. By being exposed to the *M. bovis* for longer, the latter has more chances to develop lesions. On the contrary, these trends likely reduce the impact of missing information on the previous location of the animals because these animals have more chances of being reared locally.

In agreement with many studies and with the *vSNP* analysis result, we used AF2122/97 as a reference genome (49, 50, 54, 55, 77, 80, 81). To account for genes, absent in *M. bovis*, Loiseau et al. (39) used *M. tuberculosis* H37Rv, a choice driven by the different purpose of their study compared with ours (define the origin and the global population structure of *M. bovis*). Generally, the pipelines used to call the SNPs differed in many of the aforementioned studies, contributing to the uncertainty of the estimates and potentially generating biases in the analysis results and the clock rate calculations.

5. Conclusion

In conclusion, our study indicates endemic stability of *M. bovis* is unlikely in North Cameroon, but rather the disease is slowly expanding over time. Our findings highlight the importance of collecting data in underrepresented areas to enrich insights into the current body of literature, predominantly from developed countries. Moreover, our results pave the way for future research aimed to understand whether the observed *M. bovis* high-genetic diversity affects the spread dynamics.

Our findings underscore the need to adopt a One-Health surveillance strategy for *M. bovis* control (11). More studies on combining tools such as phylogeography, statistical modeling, landscape, and ecology will be beneficial to map spread patterns and effectively inform control and preparedness strategies (54).

Data availability statement

The Mycobacterium bovis sequences for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB61415.

Author contributions

BB, LN, and VT conceived the original project. BB, RK, LN, VT, MS, and NE designed the field study, databases, and survey instrument. RK, NE, VT, and BB developed the field SOPs and collected the data. PM collected the cattle movement data. GR, AM, FD, and FE cleaned the data. GR, BB, SL, and AM conceived the quantitative analysis. GR and SL run the phylogenetic analysis. BS implemented the bioinformatical pipelines. GR performed the machine learning analysis and responsible for writing the initial drafts. All authors contributed to comments for the final draft, contributed to the article, and approved the submitted version.

Funding

The primary data used in this study were generated with funding from Wellcome Trust (WT094945), with BCMB as the

principal investigator. BS, SL, MB, BB, and AM are supported by the Biotechnology and Biological Sciences Research Council (BBSRC) program grant to Roslin Institute (Award numbers BBS/E/D/20002172 and BBS/E/D/200021723), and AM later by his BBSRC Future leader Fellowship and current Chancellor's fellowship. GR, SL, and BB additionally received support from the Scottish Government Rural and Environment Science and Analytical Services Division as part of the Center of Expertise on Animal Disease Outbreaks (EPIC). We are also grateful to the staff of MENIPIA, especially the veterinarians and delegates who diligently supported the primary fieldwork that generated this data. BS was partially funded by a BBSRC Core Capability Grant BB/CCG1780/1 awarded to The Roslin Institute. *M. bovis* sequencing was carried out by Edinburgh Genomics, The University of Edinburgh, which is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1), and BBSRC (BB/J004243/1).

Acknowledgments

The authors would like to thank Dr. Isaac D. Otchere for providing the isolation year of the *M. bovis* sequences sampled in Ghana, Dr. Stella Mazeri and Dr. William Harvey, Roslin Institute University of Edinburgh, for data collection and analysis support, and Prof. Kim Vanderwaal, University of Minnesota for the useful comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1086001/full#supplementary-material

References

1. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet.* (2009) 10:540–50. doi: 10.1038/nrg2583

2. Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: How whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.* (2014) 22:282–91. doi: 10.1016/j.tim.2014.02.011

3. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci.* (2012) 109:15066–71. doi: 10.1073/pnas.1206598109

4. Campbell F, Strang C, Ferguson N, Cori A, Jombart T. When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* (2018) 14:1–21. doi: 10.1371/journal.ppat.1006885

5. Patané JSL, Martins J, Castelão AB, Nishibe C, Montera L, Bigi F, et al. Patterns and processes of *Mycobacterium bovis* evolution revealed by phylogenomic analyses. *Genome Biol Evol.* (2017) 9:521–35. doi: 10.1093/gbe/evx022

6. WHO. *Zoonotic Tuberculosis Factsheet*. (2017). Available online at: http://www.who.int/tb/areas-of-work/zoonotic-tb/en (accessed October 26, 2022).

7. Rossi G, Crispell J, Balaz D, Lycett S, Delahay R, Kao R. Identifying likely transmission pairs with pathogen sequence data using Kolmogorov Forward Equations; an application to *M. bovis* in cattle and badgers. *Sci Rep.* (2020) 8:1–13. doi: 10.1101/2020.06.11.146894

8. Bernitz N, Kerr TJ, Goosen WJ, Chileshe J, Higgitt RL, Roos EO, et al. Review of diagnostic tests for detection of *Mycobacterium bovis* infection in south African wildlife. *Front Vet Sci.* (2021) 8:1–11. doi: 10.3389/fvets.2021.588697

9. Kelly RF, Gordon LG, Egbe NF, Freeman EJ, Mazeri S, Ngwa VN et al. Bovine tuberculosis antemortem diagnostic test agreement and disagreement in a naturally infected african cattle population. *Front Vet Sci.* (2022) 9:877534. doi: 10.3389/fvets.2022.877534

10. Sichewo PR, Vander Kelen C, Thys S, Michel AL. Risk practices for bovine tuberculosis transmission to cattle and livestock farming communities living at wildlife-livestock-human interface in northern Kwazulu Natal, South Sfrica. *PLoS Negl Trop Dis.* (2020) 14:1–18. doi: 10.1371/journal.pntd.0007618

11. Olea-Popelka F, Muwonge A, Perera A, Dean AS, Mumford E, Erlacher-Vindel E, et al. Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis*—a call for action. *Lancet Infect Dis.* (2017) 17:e21–5. doi: 10.1016/S1473-3099(16)30139-6

12. Koro Koro F, Ngatchou AF, Portal JL, Gutierrez C, Etoa FX, Eyangoh SI. The genetic population structure of *Mycobacterium bovis* strains isolated from cattle slaughtered at the Yaoundé and Douala abattoirs in Cameroon. *OIE Rev Sci Tech.* (2015) 34:1001–10. doi: 10.20506/rst.34.3.2390

13. Egbe NF, Muwonge A, Ndip L, Kelly RF, Sander M, Tanya V, et al. Abattoir-based estimates of mycobacterial infections in Cameroon. *Sci Rep.* (2016) 6:1–14. doi: 10.1038/srep24320

14. Awah-Ndukum J, Kudi AC, Bradley G, Ane-Anyangwe I, Titanji VPK, Fon-Tebug S, et al. Prevalence of bovine tuberculosis in cattle in the highlands of Cameroon based on the detection of lesions in slaughtered cattle and tuberculin skin tests of live cattle. *Vet Med.* (2012) 57:59–76. doi: 10.17221/5252-VETMED

15. Egbe NF, Muwonge A, Ndip L, Kelly RF, Sander M, Tanya V, et al. Molecular epidemiology of *Mycobacterium bovis* in Cameroon. *Sci Rep.* (2017) 7:1–17. doi: 10.1038/s41598-017-04230-6

16. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol.* (1997) 35:907–14. doi: 10.1128/jcm.35.4.907-914.1997

17. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol.* (2006) 44:4498–510. doi: 10.1128/JCM.01392-06

18. Reis AC, Salvador LCM, Robbe-Austerman S, Tenreiro R, Botelho A, Albuquerque T. Article whole genome sequencing refines knowledge on the population structure of *Mycobacterium bovis* from a multi-host tuberculosis system. *Microorganisms.* (2021) 9:523. doi: 10.1101/2021.04.26.441523

19. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ. Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 110. *Virus Evol.* (2018) 4:1–5. doi: 10.1093/ve/vey016

20. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol.* (2010) 27:1877–85. doi: 10.1093/molbev/msq067

21. Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG, SERAPHIM. Studying environmental rasters and phylogenetically informed movements. *Bioinformatics.* (2016) 32:3204–6. doi: 10.1093/bioinformatics/btw384

22. Motta P, Porphyre T, Handel I, Hamman SM, Ngu Ngwa V, Tanya V, et al. Implications of the cattle trade network in Cameroon for regional disease prevention and control. *Sci Rep.* (2017) 7:1–13. doi: 10.1038/srep43932

23. MMMO. BovTB-nf-docker. (2022). Available online at: https://github.com/oxfordmmm/BovTB-nf-docker (accessed October 31, 2022).

24. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170

25. Md V, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *Proceedings - 2019 IEEE 33rd International Parallel and Distributed Processing Symposium, IPDPS 2019*. IEEE (2019). p. 314–324.

26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352

27. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* (2019) 20:1–13. doi: 10.1186/s13059-019-1891-0

28. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput Sci.* (2017) 2017:1–17. doi: 10.1101/051813

29. Li H, A. statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* (2011) 27:2987–93. doi: 10.1093/bioinformatics/btr509

30. Seemann T. *Snippy: Rapid Haploid Variant Calling and Core Genome Alignment*. (2022). Available online at: https://github.com/tseemann/snippy (accessed October 31, 2022).

31. Jiang Y, Jiang Y, Wang S, Zhang Q, Ding X. Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinformatics.* (2019) 20:1–12. doi: 10.1186/s12859-019-3164-z

32. USDA. vSNP. (2023). Available online at: https://github.com/USDA-VS/vSNP (accessed January 16, 2023).

33. Warren RM, Gey van Pittius NC, Barnard M, Hesseling A, Engelke E, de Kock M, et al. Differentiation of Mycobacterium tuberculosis complex by PCR amplification of genomic regions of difference. *Int J Tuberc Lung Dis*. (2006) 10:818–22.

34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324

35. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* (2021) 10:1–4. doi: 10.1093/gigascience/giab008

36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

37. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* (2011) 29:24–6. doi: 10.1038/nbt.1754

38. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* (2017) 45:D535–42. doi: 10.1093/nar/gkw1017

39. Loiseau C, Menardo F, Aseffa A, Hailu E, Gumi B, Ameni G, et al. An African origin for *Mycobacterium bovis*. *Evol Med Public Heal.* (2020) 2020:49–59. doi: 10.1093/emph/eoaa005

40. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* (2020) 37:1530–4. doi: 10.1093/molbev/msaa015

41. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ, W-IQ-TREE. a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* (2016) 44:W232–5. doi: 10.1093/nar/gkw256

42. Paradis E. Schliep K. Ape 50: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics.* (2018) 35:526–8. doi: 10.1093/bioinformatics/bty633

43. R Core Team. R: A Language and Environment for Statistical Computing. (2021). Available online at: https://www.r-project.org/ (accessed March 31, 2021).

44. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* (2016) 2:1–7. doi: 10.1093/ve/vew007

45. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol.* (2012) 61:170–3. doi: 10.1093/sysbio/syr100

46. Rambaut A, Drummond AJ, Xie D, Baele G. Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 17. *Syst Biol.* (2018) 67:901–4. doi: 10.1093/sysbio/syy032

47. Hasegawa M, Kishino H, Yano T. Aki Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* (1985) 22:160–74. doi: 10.1007/BF02101694

48. Crispell J, Zadoks RN, Harris SR, Paterson B, Collins DM, De-Lisle GW, et al. Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC Genomics.* (2017) 18:180. doi: 10.1186/s12864-017-3569-x

49. Duault H, Michelet L, Boschiroli M-L, Durand B, Canini L, A. Bayesian evolutionary model towards understanding wildlife contribution to F4-family *Mycobacterium bovis* transmission in the South-West of France. *Vet Res.* (2022) 53:1–12. doi: 10.1186/s13567-022-01044-x

50. Salvador LCM, O'Brien DJ, Cosgrove MK, Stuber TP, Schooley A, Crispell J. Disease management at the wildlife-livestock interface: using whole-genome sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan, USA. *Mol Ecol.* (2019) 10:1–14. doi: 10.1111/mec.15061

51. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol.* (2013) 30:713–24. doi: 10.1093/molbev/mss265

52. Hill V. Baele G. Bayesian estimation of past population dynamics in BEAST 110 Using the skygrid coalescent model. *Mol Biol Evol.* (2019) 36:2620–8. doi: 10.1093/molbev/msz172

53. Dellicour S, Rose R, Pybus OG. Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinf.* (2016) 17:1–12. doi: 10.1186/s12859-016-0924-x

54. Rossi G, Crispell J, Brough T, Lycett SJ, White PCL, Allen A, et al. Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections. *J Appl Ecol.* (2022) 59:210–22. doi: 10.1111/1365-2664.14046

55. Crispell J, Benton CHCH, Balaz D, De Maio N, Ahkmetova A, Allen A, et al. Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *Elife.* (2019) 8:1–36. doi: 10.7554/eLife.45833

56. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* (2008) 77:802–13. doi: 10.1111/j.1365-2656.2008.01390.x

57. Hijmans RJ, Phillips S, Leathwick J, Elith J. *Dismo: Species Distribution Modeling*. (2021) Available online at: https://cran.r-project.org/package=dismo (accessed September 19, 2022).

58. Greenwell B, Boehmke B, Cunningham J, Developers GBM. *Gbm: Generalized Boosted Regression Models*. (2022). Available online at: https://cran.r-project.org/package=gbm (accessed August 11, 2022).

59. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, Drakeley CJ, Ferguson HM, Kao RR. Predictive analysis across spatial scales links zoonotic malaria to deforestation. *Proc R Soc B Biol Sci.* (2019) 286:2351. doi: 10.1098/rspb.2018.2351

60. Newman M. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc. (2010).

61. Csárdi G, Nepusz T. The igraph software package for complex network research. *J Comput Appl.* (2014) 1695:1–17.

62. Kuhn M, Weston S, Keefer C, Engelhardt A, Cooper T, Mayer Z, et al. *Classification and Regression Training*. (2016). Available online at: https://cran.r-project.org/package=caret (accessed August 09, 2022).

63. Müller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, Boschiroli ML, et al. African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. *J Bacteriol.* (2009) 191:1951–60. doi: 10.1128/JB.01590-08

64. Rodríguez S, Romero B, Bezos J, de Juan L, Álvarez J, Castellanos E, et al. High spoligotype diversity within a *Mycobacterium bovis* population: Clues to understanding the demography of the pathogen in Europe. *Vet Microbiol.* (2010) 141:89–95. doi: 10.1016/j.vetmic.2009.08.007

65. Hauer A, De Cruz K, Cochard T, Godreuil S, Karoui C, Henault S, et al. Genetic evolution of *Mycobacterium bovis* causing tuberculosis in livestock and wildlife in France since 1978. *PLoS ONE.* (2015) 10:1–17. doi: 10.1371/journal.pone.0117103

66. Matos F, Cunha M V, Canto A, Albuquerque T, Amado A, Botelho A. Snapshot of *Mycobacterium bovis* and *Mycobacterium caprae* infections in livestock in an area with a low incidence of bovine tuberculosis. *J Clin Microbiol.* (2010) 48:4337–9. doi: 10.1128/JCM.01762-10

67. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc.* (1995) 90:773–95. doi: 10.1080/01621459.1995.10476572

68. Real LA, Biek R. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *J R Soc Interface.* (2007) 4:935–48. doi: 10.1098/rsif.2007.1041

69. Pozo P, Lorente-Leal V, Robbe-Austerman S, Hicks J, Stuber T, Bezos J. Use of Whole-Genome Sequencing to Unravel the Genetic Diversity of a Prevalent *Mycobacterium bovis* Spoligotype in a Multi-Host Scenario in Spain. *Front Microbiol.* (2022) 13:915843. doi: 10.3389/fmicb.2022.915843

70. Otchere ID, van Tonder AJ, Asante-Poku A, Sánchez-Busó L, Coscollá M, Osei-Wusu S, et al. Molecular epidemiology and whole genome sequencing analysis of clinical *Mycobacterium bovis* from Ghana. *PLoS ONE.* (2019) 14:1–13. doi: 10.1371/journal.pone.0209395

71. Mohamed A. Bovine tuberculosis at the human–livestock–wildlife interface and its control through one health approach in the Ethiopian Somali Pastoralists: a review. *One Heal.* (2020) 9:100113. doi: 10.1016/j.onehlt.2019.100113

72. Valerio VC, Walther OJ, Eilittä M, Cissé B, Muneepeerakul R, Kiker GA. Network analysis of regional livestock trade in West Africa. *PLoS ONE.* (2020) 15:1–20. doi: 10.1371/journal.pone.0232681

73. Trewby H, Wright DM, Skuce RA, McCormick C, Mallon TR, Presho EL, et al. Relative abundance of *Mycobacterium bovis* molecular types in cattle: a simulation study of potential epidemiological drivers. *BMC Vet Res.* (2017) 13:268. doi: 10.1186/s12917-017-1190-5

74. Rodriguez-Campos S, Aranaz A, De Juan L, Sáez-Llorente JL, Romero B, Bezos J, et al. Limitations of spoligotyping and variable-number tandem-repeat typing for molecular tracing of *Mycobacterium bovis* in a high-diversity setting. *J Clin Microbiol.* (2011) 49:3361–4. doi: 10.1128/JCM.00301-11

75. Campbell F, Cori A, Ferguson N, Jombart T. Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data. *PLoS Comput Biol.* (2019) 15:1–20. doi: 10.1371/journal.pcbi.1006930

76. Rodriguez-Campos S, Schürch AC, Dale J, Lohan AJ, Cunha M V, Botelho A, et al. European 2 - A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infect Genet Evol.* (2012) 12:866–72. doi: 10.1016/j.meegid.2011.09.004

77. Zimpel CK, Patané JSL, Guedes ACP, de Souza RF, Silva-Pereira TT, Camargo NCS, et al. Global Distribution and Evolution of *Mycobacterium bovis* Lineages. *Front Microbiol.* (2020) 11:1–19. doi: 10.3389/fmicb.2020.00843

78. Renwick AR, White PCL, Bengis RG. Bovine tuberculosis in southern African wildlife: a multi-species host-pathogen system. *Epidemiol Infect.* (2007) 135:529–40. doi: 10.1017/S0950268806007205

79. Chaters GL, Johnson PCD, Cleaveland S, Crispell J, De Glanville WA, Doherty T, et al. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Philos Trans R Soc B Biol Sci.* (2019) 374:264. doi: 10.1098/rstb.2018.0264

80. Reis AC, Cunha M V. The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. *Microb Genomics.* (2021) 7:664. doi: 10.1099/mgen.0.000664

81. Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H, et al. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. *Microb Genomics.* (2020) 6:1–8. doi: 10.1099/mgen.0.000388

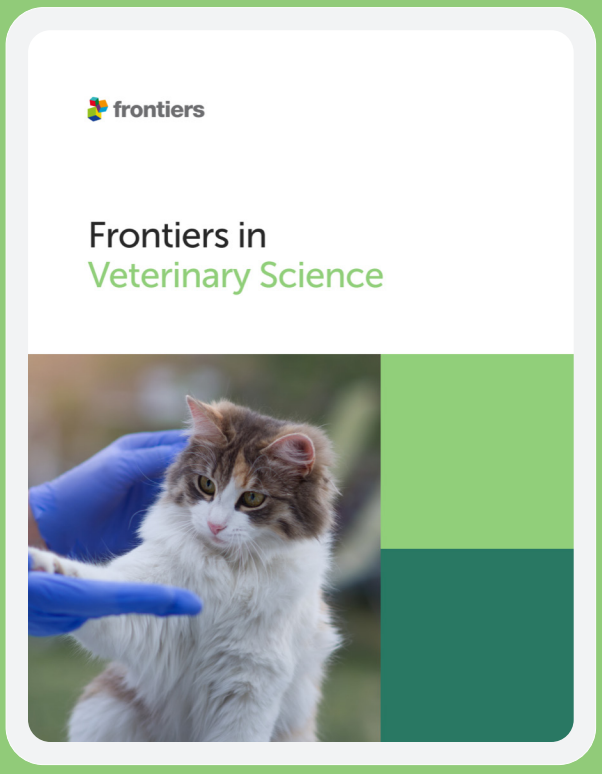
Frontiers in Veterinary Science

Transforms how we investigate and improve animal health

The third most-cited veterinary science journal, bridging animal and human health with a comparative approach to medical challenges. It explores innovative biotechnology and therapy for improved health outcomes.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact