Summary of PhD thesis

Molecular epidemiology of infectious bronchitis virus

dr. Hoitsyné Bali Krisztina

Supervisors: dr. Bányai Krisztián and dr. Bálint Ádám

UNIVERSITY OF VETERINARY MEDICINE

Doctoral School of Veterinary Science

Budapest, 2023.

Supervisors:

……

Dr. Bányai Krisztián

Veterinary Medical Research Institute

……

Dr. Bálint Ádám

National Food Chain Safety Office

Veterinary Diagnostic Directorate

……..

dr. Hoitsyné Bali Krisztina

Table of contents

1. Introduction

Members of the family *Coronaviridae* are single-stranded RNA viruses infecting a wide range of human and animal hosts. The genus *Gammacoronavirus* mainly affects birds and its economically most important species is *Avian coronavirus* (IBV). It causes infectious bronchitis, a highly contagious viral disease of chickens. Infection by *Avian coronavirus* causes respiratory disease, proventriculitis, nephritis, decreased egg production depending on the viral pathotype and the age of the affected flocks. Infectious bronchitis was first described in the 1930s and still remains a global problem for the poultry industry. Since its first report in Hungary in the 1960s, many variants of the virus have been identified in the country.

Phylogenetic analysis based on the S1 gene sequencing is used to identify genotypes and genetic lineages of IBV. The dominant genetic lineages are widely distributed but most lineages are confined to specific geographic regions. In addition to the high mutation rate of the virus, its extreme variability due to recombination between different strains makes vaccination and other control measures difficult.

Analysis of S1 sequences is extremely useful for identifying IBV genotypes and genetic lineages. However, the examination of complete genome sequences can help to reveal the entire structure and function of the genome. It also provides additional information for epidemiological studies, performing comprehensive recombination analyses, and identifying genetic markers related to pathogenicity and vaccine design.

In this context, the main objective of our studies was to collect and characterize complete genomic sequence data of different *Avian coronavirus* strains. By comparing the genomes of the virus strains, we would like to broaden our knowledge about the genetic diversity of infectious bronchitis virus, the importance of evolutionary mechanisms, such as recombination, and the phylogenic relationships of the studied strains.

2. Objectives

The aim of our study was to uncover the genetic diversity of *Avian coronavirus* strains causing infectious bronchitis:

1. Determining the complete genome sequence of a large number of strains by using next-generation sequencing (NGS) and adding the obtained sequence data to the relevant database;
2. Describing the genome structure, genetic diversity and phylogenetic relationships of the studied strains in order to gain information about the evolutionary mechanisms of *Avian coronavirus*;
3. Identifying new variants and genetic lineages and adding to the currently used S1 gene-based genotyping system;
4. Updating knowledge about the geographic distribution of individual genotypes with the help of newly generated data.

3. Materials and Methods

Virus strains

The analyzed virus strains were kindly provided by CEVA-Phylaxia and Hungarian National Food Chain Safety Office in the form of allantoic fluid.

Molecular methods

To enrich viral genomic RNA from the harvested allantoic fluid, samples were filtered and digested. Viral nucleic acid was extracted from the supernatant using the innuPREP Virus DNA/RNA Kit according to the manufacturer's instruction. Reverse transcription (RT) was performed using the oligonucleotide FR26RV-N containing a 3' random hexamer tag. cDNA was amplified in PCR using the FR20RV oligonucleotide. The PCR product was run in agarose gel, products between the size of 200-2000 base pairs were excised and extracted. Libraries were prepared and whole genome sequencing was performed on IonTorrent PGM or Illumina NextSeq 500 NGS platforms.

Bioinformatics

Geneious Prime software was used to assemble NGS data and to map raw sequence data to reference

sequences. Phylogenetic trees were generated and visualized using MEGA X software. 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 (1000). Nucleotide and amino acid sequence distances were calculated using p-distance method. Possible recombination events were detected with the RDP5 and SimPlot software.

Database analysis

All available complete or near complete S1 gene sequences of IBV were downloaded from GenBank. Following previously established criteria, sequences shorter than 1440 bp were removed. In the end, the analyzed 3590 S1 gene sequences originated from strains isolated between 1937 and 2022 from 43 countries.

4. Results and discussion

We determined the complete genome sequence of 102 *Avian coronavirus* strains using next generation sequencing methods. The complete genomes were ~27 kb in size and consisted of minimum 11, maximum 13 ORFs. A few strains lacked some accessory protein-coding genes such as 4b, 4c or 6b. The order of the 13 ORFs was identical (5′-UTR-1a-1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b-3′ UTR). The genome of the investigated strains encoded at least 24, but mostly 26 proteins.

Recombination analysis

We further analyzed the frequency and location of possible recombination sites using whole-genome sequence data of 80 strains. A total of 215 recombination events were identified within 51 strains. The majority of recombination events were detected among field strains and between field and vaccine strains. Simplot analysis identified at least 10 events when a transferred genomic fragment retained over 99% similarity with the putative vaccine strain.

The similarity values over 99% are thought to represent more recent recombination events between

vaccine and field strains with a greater likelihood that the event just preceded or was co-incidental with the actual isolation of the particular field strain. Some strains had multiple recombination breaking points with various vaccine strains.

Recombination hot spots were identified along the entire genome, mostly in genomic positions immediately upstream of the S gene, in the nsp2, nsp3, nsp12, nsp16 and envelope and membrane protein coding genes. In our study, recombination cold spots were found intermittently near the 5′ and 3′ ends of the genome and in a large fragment of the S gene. During our analyses, recombination was extremely frequent and affected roughly all regions of the genome. Recombination in the ORF 1ab region, encoding non-structural proteins essential for viral replication, may affect viral pathogenicity. Changes in envelope and membrane proteins essential for virion assembly can affect the efficiency of virus particle formation and virus transmission.

The scattered geographical and temporal origin of IBV genome sequence data from Europe prevented us from depicting a more structured landscape of the genomic evolution of IBVs. Thus, we demonstrated that

recombination among European strains involves different IBV genotypes and both field and vaccine strains.

During the recombination analysis of 600 complete S1 sequences a total of 39 recombination events were identified mostly between strains belonging to different genetic lineages (inter-lineage recombination). Changes in this region may result in the emergence of new genotypes and serotypes of IBV.

Phylogenetic relationship based on the S1 gene

Genetic classification based on the S1 coding region identified 11 genotypes among the 102 study strains. The most common was the GI-19 lineage. The remaining strains were classified as GI-23, GI-16, GI-13, GI-1, GI-21, GI-27, GI-9, GI-14, GI-11, and GI-25 lineages. In addition, four strains turned out to be recombinants (D1623/1/1/11/HU, D2584/12/1/13/PH, D3386/1/2/16/SA, 211776/Imrehegy/2011), and four strains were recognized as unique variants (D2334/11/2/13/CI, D2804/3/3/13/ID, D2586/4/6/13/PH, D3276/4/2/16/PH).

On the phylogenetic tree based on the complete S1 sequence the D1623/1/1/11/HU, D3386/1/2/16/SA and D2584/12/1/13/PH recombinant strains are located near to the strains belonging to the GI-19 lineage, but they are

grouped on a separate branch. During the S1 gene-based recombination analysis, a GI-19 parental sequence was involved in all recombination events involving the above-mentioned strains. Furthermore, they shared more than 90% nucleotide sequence identity with GenBank reference strains belonging to the GI-19 lineage. The recombinant 211776/Imrehegy/2011 strain is located close to the members of the GI-13 and GI-21 lineages on the phylogenetic tree. Moreover, it showed high nucleotide identity (87.88 and 88.56%) with reference GI-13 and GI-21 strains.

Based on the phylogenetic analyses, the unique variant strains did not cluster together with any previously determined genetic lineages. High S1 gene nucleotide identity (98.70%) was observed between the two unique variant strains originating from the Philippines (D2586/4/6/13/PH and D3276/4/2/16/PH). While they shared only 83.78 and 84.22% sequence identity compared to the Indonesian unique variant (D2804/3/3/13/ID) located closest to them on the phylogenetic tree. Based on our phylogenetic analyses, the two Philippine variants resemble each other more than any other strains, which may be evidence of a common evolutionary origin.

The variant from Ivory Coast (D2334/11/2/13/CI) showed the highest S1 sequence identity with another study strain belonging to the GI-21 lineage and a unique variant reference (RF/01/99) strain of Russian origin (78.11%). However, based on the phylogenetic analyses, the two unique variants did not form a monophyletic group. The unique variant strain from Ivory Coast clustered with members of the GI-26 lineage indigenous in Africa, but the S1 gene nucleotide identity was only 78%.

Comparative amino acid analysis of S protein

The role of the amino acid (aa) changes observed in the S1 region and their effects on virulence and pathogenicity are difficult to determine. The analyses of recombinantly expressed proteins support that the aa composition of the S protein, especially the receptor binding domain (RBD), effects tissue tropism. Since these regions are important, we analyzed the aa motifs found in the region.

Based on previous studies, the RBD region of S1 subunit is considered critical for binding of the M41 strain spike protein to chicken respiratory tract receptors. According to our analyses, the aa sequence of the strains

isolated from birds showing respiratory signs did not match the reference M41 strain. We also addressed the aa motifs of strains isolated from birds showing various symptoms. However, no correlation was found between the sequence in the studied aa positions and the assumed tissue tropism. No uniform pattern could be recognized in the sequences of strains isolated from birds showing different symptoms. Moreover, several strains originating from birds showing diverse symptoms shared identical aa sequences in the examined positions.

Previous studies with purified S1 proteins have shown that introduction of aa $_{110}$KIP$_{112}$ of the nephropathogenic QX spike protein into the RBD region of M41 was sufficient to extend its tropism toward the kidney. However, our data indicated that IBV strains isolated from animals with kidney lesions did not show the $_{110}$KIP$_{112}$ motif of the QX strain. Therefore, our findings were unable to confirm the correlation between the aa motif sequences and the observed clinical symptoms in the case of study strains.

In conclusion, the aa substitutions identified in particular positions were mostly linked to genetic lineages. No correlation was found between the motifs and the assumed tissue tropism of the strains included in the

study. Nevertheless, different aa sequences found in the given positions do not necessarily result in a decrease in virulence.

Geographic distribution of IBV lineages

According to the S1 gene-based classification system, there are 8 genotypes (GI to GVIII) of IBV. In our study, we have analyzed the strains belonging to the most diverse and widespread GI genotype.

The vast majority (66.1%) of the 3590 IBV S1 gene sequences available in the GenBank database originate from Asia and 88.4% of the Asian strains come from China. The distribution of the remaining sequences was the following: 12,7% North America, 1,9% Middle and South America, 6,5% Europe, 4,9% Middle East, 1,8% Africa, and 1,7% Australia and New Zealand. The sequences available in the GenBank database were completed with 102 sequences from our collection. Thus, the distribution of genotypes and genetic lineages was investigated with the addition of another 8 strains from Africa (Cameroon, Ghana, Ivory Coast, South Africa, Sudan), 21 strains from Asia (Far East, Indonesia, Philippines), 31 strains from Europe (Belarus, Greece, Hungary, Netherlands, Poland, Russia, Ukraine),

3 strains from North America (USA), 7 strains from Middle and South America (Argentina, Brazil, Mexico, Peru), and 32 strains originating from the Middle East (Egypt, Iran, Jordan, Lebanon, Morocco, Saudi Arabia, United Arab Emirates).

In brief, only 8.2% of GenBank strains and three out of the 102 strains determined in this study were classified into the widely distributed GI-1 lineage. Additionally, 9.3% of the GenBank strains and 7 strains from our collection (originating from North America, the Middle East and the Far East) could be classified into the GI-13 lineage. Only 0.9% of GenBank strains were members of the globally distributed GI-16 lineage. From our collection many strains of African, Middle Eastern, and South American origin strains belonged to this lineage (13/102). A significant proportion of the sequences available in the GenBank database came from strains belonging to the GI-19 lineage (42%). From our collection, many strains from Africa, Europe, the Far East and the Middle East were representatives of the GI-19 lineage (42/102). The dominant GI-23 lineage in Europe and in the Middle East includes 4.7% of the GenBank strains, and 15 mostly Middle Eastern strains from our collection. The fact that nearly 95% of the 164 complete S1 and 8 complete

genome sequences available in the GenBank database were isolated recently, indicates the expansion of the strains belonging to the GI-23 genetic lineage.

In addition to the globally distributed lineages, we also identified several regionally distributed indigenous genetic lineages in our strain collection. In Europe, three lineages are considered indigenous: GI-21, GII-1 and GII-2. Three GI-21 strains were identified in our collection from Morocco and Romania. Extant data show that 7 different lineages are geographically strictly confined to Asia: GI-7, GI-15, GI-18, GI-22, GI-24, GI-28, GI-29, GVI-1 and GVII-1. A large number of lineages, falling into two well distinct genotypes (GI and GIV), have been reported as indigenous to North America: GI-8, GI-9, GI-17, GI-20, GI-25, GI-27 and GIV-1. However, only some of these – GI-9, GI-27 and GIV-1 – are widespread. Most of the North American origin strains in our collection belong to the GI-27 lineage. We also identified two Mexican strains classified into GI-9 and GI-25 lineages as the first report from the country. The GI-11 lineage is a unique and dominant lineage in South America and the strain belonging to this lineage from our collection also originates from Brazil. The GI-26 lineage represents a unique African cluster of viruses that were

identified in Nigeria and Niger between 2006 and 2007. Likely due to their geographic isolation, Australia and New Zealand possess only unique indigenous variants. Five distinct IBV lineages are found in these localities (GI-5, GI-6, GI-10, GIII és GV), possessing large evolutionary distances compared to those found elsewhere. Even the globally widespread GI-1 lineage is absent here, probably as a result of the use of vaccine strains derived from local isolates instead of commercial vaccines.

5. New scientific results

1. We determined and analyzed the complete genome sequence of 102 *Avian coronavirus* strains. Our results highlighted the importance of recombination, which plays an important role in the evolutionary mechanisms of the *Avian coronavirus*. During our analyses, we identified recombination events affecting several regions of the genome, including the S1 gene.

2. We determined and analyzed the complete genome sequence of five novel Hungarian IBV strains belonging to the GI-1 and GI-19 lineages. Furthermore, we identified an S gene recombinant strain, originating from multiple recombination events, showing the highest identity with GI-13 and GI-21 strains.

3. In our study, new recombination hot spots were found in the nsp8 and nsp12 genes, and recombination cold spots were identified near the 5′ and 3′ ends of the genome and in the S gene.

4. We identified new unique variant strains from the Philippines, Ivory Coast, and Indonesia, which genome sequences differ greatly from other reference strains. Therefore, they could be the first representatives of new

genetic lineages. Our findings revealed genetic lineages previously not reported from Africa and Central America (Ivory Coast GI-16; Mexico GI-9, GI-25).

5. We revised the currently used S1 gene-based classification system using complete genome and complete S1 sequences available in the GenBank database along with our own strain collection. In addition, by integrating our own data, we updated the knowledge about the geographic distribution of IBV genotypes.

6. Regarding the examined peptide motifs, no correlation was found between the sequence observed in individual amino acid positions and the assumed tissue tropism. According to our results, the studied motifs were mostly arranged based on genetic lineages and not according to the caused symptoms. Therefore, the amino acid sequences responsible for tissue tropism could not be identified.

6. Publications in peer reviewed journals

1. Bali, K, Bálint, Á, Farsang, A, Marton, S, Nagy, B, Kaszab, E, Belák, S, Palya, V, Bányai, K, **Recombination events shape the genomic evolution of infectious bronchitis virus in europe.** Viruses 13(4), 535. 2021. https://doi.org/10.3390/v13040535
2. Bali K, Kaszab E, Marton Sz, Hamdiou SH, Bentaleb RK, Kiss I, Palya V, Bányai K, **Novel lineage of infectious bronchitis virus from Sub-Saharan Africa identified by random amplification and next-generation sequencing of viral genome.** Life 14(4), 475. 2022. https://doi.org/10.3390/life12040475
3. Bali K, Bálint Á, Bányai K, **Geographic distribution of IBV lineages.** Magyar Állatorvosok Lapja, 144. 673-690. 2022.