

APPROVED: 24 July 2024

# Analysis of zoonotic mutations of avian influenza viruses

Puglia Ilaria, Caporale Marialuigia, Mangone Iolanda, Petrova Tetyana, Castelli Pierluigi, Radomski Nicolas, De Ruvo Andrea, Di Pasquale Adriano, Lorusso Alessio


## Abstract

Avian influenza viruses (AIVs) pose a global health risk for animals and humans with outcomes ranging from asymptomatic cases to fatalities. In recent years, HPAI were detected in several mammal species globally including minks in Spain, marine mammals in South America, cats in Poland, and seals along the North Atlantic coast. The H5N1 outbreak in U.S. dairy cattle marked the first detection of HPAI in this species. Although human-to-human transmission has not been clearly demonstrated so far, this scenario raises epidemiological concerns, particularly due to the ability of AIVs to adapt and infect mammalian hosts. Close monitoring of genomic sequences and mutations is essential to quickly identify markers that affect host range or virulence, potentially leading to spillover in mammals with possible harmful effects. Efficient evaluation of AIV risks requires the analysis of mammalian adaptation markers and the screening of genomic sequences. To enhance pandemic preparedness measures and to understand the evolution of AIVs, a total of 592 potential zoonotic markers across eight viral segments were identified from a literature review and screened in over 1.4 million sequences from available public databases. From these inputs, mutation frequencies were observed within avian and mammalian hosts, including humans, per geographical area, and per period, generating a database composed of a total of 1,974 spreadsheets. The present report updates the list of potential zoonotic markers and provides an overview of their frequency in avian and mammalian populations, including humans. It serves as the basis for future in-depth investigations to assess the zoonotic risks associated with avian influenza viruses and to develop a comprehensive risk assessment.

© European Food Safety Authority, 2024

**Key words:** Avian Influenza, HPAI, spillover, mammals, mammalian adaptation, mutations

**Correspondence:** [biohaw@efsa.europa.eu](mailto:biohaw@efsa.europa.eu)



**Disclaimer:** The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Acknowledgements:** We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We additionally acknowledge the GitHub database for providing essential resources and data that significantly contributed to our study. Furthermore, we sincerely acknowledge Alice Fusaro and Isabella Monne from the European Union Reference Laboratory (EURL) for Avian Influenza and Newcastle Disease, as well as Alessandro Broglia from EFSA, for their constant and fundamental contribution and the EFSA/ECDC WG on AI for providing their comments and suggestions throughout the process.

**Suggested citation:** Puglia Ilaria, Caporale Marialuigia, Mangone Iolanda, Petrova Tetyana, Castelli Pierluigi, Radomski Nicolas, De Ruvo Andrea, Di Pasquale Adriano, Lorusso Alessio, 2024. Analysis of zoonotic mutations of avian influenza viruses. EFSA supporting publication 2024

**ISSN:** 2397-8325

© European Food Safety Authority, 2024

Reproduction is authorised provided the source is acknowledged.

Reproduction of the images listed below is prohibited and permission must be sought directly from the copyright holder.




## Summary

Avian influenza viruses (AIVs) are globally circulating pathogens with pandemic potential that pose a significant health risk due to their ability to infect a wide range of species. These pathogens primarily affect birds, including wild waterfowl and domestic poultry, but are also capable of infecting mammals and humans, ranging from asymptomatic cases to fatal outcomes. In recent years, the Highly Pathogenic Avian Influenza (HPAI) H5N1 subtype raised concern by infecting multiple mammalian species. In 2022, infections were registered in minks in Spain, and in marine mammals (sea lions and dolphins) from South America. The same year, a severe outbreak in harbour and gray seals along the North Atlantic coast in New England (USA) was observed, coinciding with mass avian infections and deaths. Similarly, in 2023, cases of H5N1 were observed in cats in Poland. More recently, in the spring of 2024, an H5N1 outbreak in U.S. dairy cattle was recorded, marking the first-ever detection of HPAI H5N1 affecting this species, although human-to-human transmission has not been demonstrated so far.

To efficiently evaluate the AIV risk, and thus enhance pandemic preparedness, prevention, and control measures, EFSA requested a detailed study of genetic markers associated with mammalian adaptation, along with a thorough AIV sequences screening on public databases. In agreement with the first Terms of Reference (ToR) of this work, a comprehensive literature review was conducted to identify potential zoonotic markers from the scientific literature. A search string composed of MeSH Terms and Boolean operators was used to search peer-reviewed articles on PubMed, resulting in a total of 113 records examined. Additional relevant records not initially included by the search string and the periodic EFSA reports on avian influenza were also considered. From the selected records, 592 potential zoonotic markers were identified and listed in eight different spreadsheets (one for each viral segment PB2, PB1, PA, HA, NP, NA, MP, NS) summarizing the point mutations at the amino acid level, the subtype in which the marker was initially observed and its related phenotypic effect. In addition, a total of 103 multi-basic cleavage sites sequence stretches in the hemagglutinin (HA) protein, which play a crucial role in the pathogenicity of influenza A in avian species, were also collected and classified as HPAI, LPAI, and unusual.

To answer the second ToR, to screen all the available AIV sequences in all subtypes encompassing those from spillover events to other mammals (humans included), nucleotide sequences and related metadata were collected from public databases. Sequences of 46 selected AIV subtypes from avian and mammalian hosts, collected over the past 24 years (2000-2024), were downloaded from GISAID and GitHub. Sequences from human seasonal influenza viruses, and endemic viruses among swine, dog, and equine species, were excluded through clustering analysis by the CD-HIT tool. Additionally, to maximize the dataset quality and thus reduce bias in downstream analysis, low-coverage sequences, laboratory-derived, and duplicates were also removed.

The sequences were analyzed to generate a comprehensive database in the form of spreadsheets (.tsv format). These report all the potential zoonotic markers found in each specific virus and associated metadata (subtype, host, collection date, and geographical area). Each output was meticulously organized both by subtype, classifying the mutations per segment, and for the overall protein segments (PB2, PB1, PB1-F2, PA, PA-X, HA, NP, NA,



MP1, MP2, NS1, NS2). This double classification ensures a detailed and accessible format to analyze the point mutations and understand their distribution across different virus segments and subtypes. Mutation frequency outputs were generated as well. In these cases, spreadsheets (organized by subtype and segments) report phenotypic characteristics of each marker and frequencies of mutations across the screened sequences. The frequency data were determined for different host classes (avian and mammalian), with separate data for humans and mammalian subgroups (e.g., *Canidae*, *Felidae*, *Mustelidae*). Furthermore, these frequency data were organized by geographical area (continent) and collection date (five-year intervals).

In summary, from the literature review, eight different datasets were produced identifying a sum of 592 potential zoonotic markers. More than 1,400,000 amino acid sequences, released in the period 2000-2024, were screened. This analysis resulted in the creation of a database consisting of 1,974 spreadsheets.

This analysis is crucial to monitor the genetic evolution of AIVs. The identification of potential zoonotic markers, the screening of sequences, and the determination of marker frequencies across mammalian species, have provided a comprehensive overview of the current situation. The analysis can represent the basis for future in-depth investigations to assess the risk associated with specific genome constellations and to develop a comprehensive risk assessment.

## Table of contents

Abstract.....	1
Summary.....	3
1 Introduction .....	6
1.1 Background and terms of reference as provided by the requestor .....	7
2 Methodology.....	8
2.1 Identification of avian influenza zoonotic markers .....	8
2.2 Multi-basic cleavage site sequences .....	9
2.3 AIV sequences download.....	10
2.4 Dataset filtering and sequences exclusion .....	10
2.4.1 Endemic strains .....	10
2.4.2 Poor-quality sequences and duplicates.....	11
2.5 Screening of point mutations at the amino acid level .....	11
3 Results .....	11
3.1 Identification of AIV zoonotic markers .....	11
3.2 AIV sequences download.....	13
3.3 Dataset filtering and sequence exclusion .....	15
3.3.1 Endemic strains .....	15
3.3.2 Poor-quality sequences and duplicates.....	16
3.4 Screening AIV sequences presenting zoonotic markers .....	19
3.5 Mutation frequency .....	24
3.6 Multi-basic cleavage sites.....	25
3.7 Data availability.....	25
4 Conclusion .....	25
5 Acknowledgements.....	<b>Error! Bookmark not defined.</b>
References.....	27

# 1 Introduction

Avian influenza viruses (AIVs) represent a significant public health threat due to their ability to infect a wide range of mammalian hosts, including humans. Primarily maintained in aquatic bird populations, which act as natural reservoirs, AIVs can rapidly mutate, enhancing their ability to cross species barriers and adapt, causing spillover to other species.


AIVs belong to the family *Orthomyxoviridae*, class *Insthoviricetes*, and phylum *Negarnaviricota*. AIVs are enveloped and segmented viruses, composed of a single-stranded negative-sense RNA genome. The genomic sequence encodes for ten proteins essential for viral infection, namely PB1 (polymerase basic protein 1), PB2 (polymerase basic protein 2), PA (polymerase protein), NP (nucleoprotein), HA (haemagglutinin), NA (neuraminidase), M1 (matrix protein 1), M2 (matrix protein 2), NS1 (non-structural protein 1), and NEP/NS2 (non-structural protein 2), along with auxiliary proteins as PB1-F2 and PA-X. HA binds to host receptors and mediates membrane fusion; NA is involved in progeny virions releasing, while M2 acts as an ion channel for replication and cellular homeostasis. PB1/PB2/PA constitutes the polymerase complex for RNA synthesis, NP encapsidates viral RNAs, M1 is involved in virions assembly and budding, NS1 impairs antiviral responses, while NEP/NS2 aids in viral assembly and export (Liang, 2023).

The genome segmentation is one of the reasons for the high variability of influenza viruses. It undergoes reassortment, by which viral gene segments are exchanged between different strains infecting the same host, leading to new genotypes with potential novel phenotypic characteristics (Abdelwhab et Mettenleiter, 2023). Transmission to different hosts can occur directly from infected birds or through intermediary hosts (Shortridge et al., 1992). Another crucial factor increasing AIV variability is linked to the lack of proofreading ability of the viral polymerase. The resulting error-prone replication potentially generates a population of viruses with diverse mutations (Boivin et al., 2010).

Based on genetic and antigenic variations of HA and NA genes, AIVs are categorized into subtypes. Currently, different subtypes have been identified in wild bird populations, based on multiple combinations of 16 HA (H1-16) and 9 NA (N1-9) (Zhang et Lei, 2024). Additionally, recent studies have identified the presence of HA subtypes H17 and H18, as well as NA subtypes N10 and N11, in bats (Tong et al., 2012; 2013).

Furthermore, AIVs are categorized into highly and low pathogenic strains (HPAI and LPAI, respectively), depending on their pathogenicity in avian species. The HPAI strains cause severe disease and high mortality rates in poultry, potentially leading to significant economic losses in the industry and posing risks to food security and public health. On the other hand, LPAI viruses typically cause mild or asymptomatic infections in birds but can evolve into HPAI strains through mutation and reassortment events (Tan et al., 2021).

Several studies have shown that mutations in the HA and NA proteins as well as in internal genes, can promote the binding to human receptors, increase replication efficiency in human cells, and favor the evasion of the host immune response (Tan et al., 2021). H3N8, H5N1, H5N6, H5N8, H6N1, H7N1, H7N2, H7N3, H7N4, H7N7, H7N9, H9N2, H10N3, H10N7, and H10N8 have been recognized so far as subtypes capable of infecting humans (AbuBakar et al., 2023). This plethora of viruses increases the risk of a pandemic strain emerging from avian influenza viruses, potentially repeating historical occurrences such as the Spanish flu (H1N1, 1918–1919), the Asian flu (H2N2, 1957–1958), the Hong Kong flu (H3N2, 1968–1969), and the swine flu (H1N1, 2009) (Akin et Gözel, 2020).



The current global landscape of avian influenza is marked by ongoing outbreaks and continued inter-species transmission. Currently, the most worrisome AIV subtype is represented by the global and multi-species spread of HPAI H5N1 (Bonilla-Aldana et al., 2024). Indeed, H5N1 infections in minks in Spain (Agüero et al., 2023), in marine mammals and seabirds from South America (Leguia et al., 2023), in seals along the North Atlantic coast in New England (Puryear et al., 2023), as well as in cats in Poland (Szaluś-Jordanow et al., 2023) are just few examples which demonstrate the ability of this virus to infect a wide range of species. The potential for widespread ecological impacts is extremely concerning.

In March 2024, the USDA confirmed the first case of HPAI H5N1 in U.S. dairy cattle. The virus, identified as H5N1 Eurasian lineage goose/Guangdong clade 2.3.4.4b, has been detected in multiple states (CDC, <https://www.cdc.gov/bird-flu/situation-summary/index.html>). The strain has caused significant impacts on dairy production, with infected cattle showing reduced milk production and other clinical signs. The virus has been found in milk samples from affected herds. Up to September 6, 2024, the CDC reported 14 human infections linked to the ongoing dairy cow outbreak in the USA country, in Texas, Michigan, Colorado and Missouri (CDC, <https://www.cdc.gov/bird-flu/h5-monitoring/index.html>). Out of these cases, 13 occurred among farm workers exposed to infected cattle, showing symptoms from conjunctivitis to mild upper respiratory distress. Genetic analysis connected these cases to cattle AIV strains (EFSA, ECDC, EURL, 2024). In contrast, the 14th human H5N1 case in the U.S. for 2024, reported in Missouri, is the first without known occupational exposure to animals. Missouri has experienced H5N1 outbreaks in poultry, but there have been no reported outbreaks in cattle. Avian influenza viruses represent a persistent and evolving threat to global public health. The capacity to infect multiple species, accumulate mutations, and adapt to new hosts necessitates rigorous surveillance, research, and preparedness measures. By closely monitoring wild bird populations, domestic poultry farms, as well as mammals, health authorities can mitigate the spread and the risks posed by these viruses. Genomic surveillance is essential to understand virus transmission dynamics across species and regions, thus preventing and controlling future outbreaks. By sequencing and analyzing the circulating strains, researchers can detect mutations affecting virulence, transmissibility, and resistance to antiviral drugs. The real-time monitoring facilitates the early identification of emerging strains with pandemic potential, enabling timely public health responses and the development of vaccines and treatments.

## 1.1 Background and terms of reference as provided by the requestor

The contract entitled "Analysis of zoonotic mutations of avian influenza viruses" (framework contract number: PO/EFSA/BIOHAW/2024/03) was awarded to the Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale".

The evolution of the HPAI epidemiological situation with a high number of infected birds and newly affected mammalian species is prompting a response by animal and health authorities in EU Member States, indicating the need for enhanced preparedness and prevention. Given the escalating mammalian adaptation through the evidence of point mutations (e.g., single nucleotide polymorphisms and small deletions) and the potential of mammal-to-mammal transmission events, public health authorities are currently working on tools and solutions which can mitigate this issue. Thus, authorities are developing or adapting their tools for epidemiological investigations, strengthening collaboration, issuing new emergency national legislations, and updating their surveillance guidance and programs in light of the upcoming seasonal flu season.

The situation and actions described above have been taken in reaction to:



- ECDC's current risk assessment<sup>1</sup> that focuses on the immediate risk of avian influenza for human health fulfilling ECDC's new mandate and the Regulation EU 2022/2371 (OJ L 314, 6.12.2022, p. 26–63) aiming to prevent and prepare for cross-border health threats, including epidemics;
- EFSA's latest scientific opinions and reports<sup>2</sup> coupled by measures taken (e.g., surveillance, prevention and control measures) under the Animal Health Law (e.g., EU 2020/689 (OJ L 174, 3.6.2020, p. 21) and EU 2020/687 (OJ L 174, 3.6.2020, p. 64)).

By considering the One Health nature of this subject, and pursuant to Article 3 of Regulation (EU) 2022/2370 and Articles 29 and 31 of Regulation (EC) No 178/2002, the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Authority (EFSA) are requested to carry out a scientific opinion on preparedness, prevention, and control measures for zoonotic avian influenza in animals. This should be provided:

- a. By assessing the risk posed by H5N1 clade 2.3.4.4b, other HPAI viruses, as well as other AI viruses; if possible, by quantifying the risk of mutations of AI viruses and by identifying factors influencing the AIVs adaptation to birds and/or mammals;
- b. By implementing a comprehensive set of prevention, protection, and preparedness measures in response to HPAI H5N1 and other AIVs observed in birds and mammals. Key factors to consider are:
  1. Surveillance, identifying the most suitable avian influenza surveillance tools and optimizing surveillance programs to address zoonotic risks;
  2. The prevention, biosecurity, and control measures to reduce the risk of zoonotic avian influenza.

The scope of the present scientific report is to answer to the previous Terms of reference (ToR), which were translated into the following objectives:

- a. To identify zoonotic markers from the scientific literature, to extract part of the information from the review by Suttie et al. (2019); and to expand the list of mutations to include as much as possible relevant markers;
- b. To screen all the AIV sequences available, including the ones obtained from spillover events in mammals (also humans) for zoonotic markers.
  1. To generate a database in the form of spreadsheets, encompassing all the viruses presenting zoonotic markers along with the list of mutations found in each specific virus and associated metadata: virus subtype, clade (for H5 subtype), host species, location, and collection date;
  2. To provide a mutation frequency overview per mutation, per subtype, per species, and per geographical area.

## 2 Methodology


The methodology as explained below and the outcomes requirements were discussed and agreed during interim web meetings with EURL experts and EFSA, to ensure the best coordination and development of the project.

### 2.1 Identification of avian influenza zoonotic markers

<sup>1</sup> <https://www.ecdc.europa.eu/en/infectious-disease-topics/z-disease-list/avian-influenza/threats-and-outbreaks/risk-assessment-h5>

<sup>2</sup> <https://www.efsa.europa.eu/en/topics/topic/avian-influenza>





A comprehensive literature review was performed, to compile a robust list of zoonotic markers and highlight critical mutations of AIVs among mammals. Our approach included the recently published AIV cases in mammals to assess the current epidemiological context and to evaluate the transmission dynamics within the wild population. The list of cases from the periodic EFSA reports on AI monitoring was also integrated.

The literature review initially focused on the zoonotic marker inventory provided by Suttie et al. (2019). To encompass a broader list of point mutations, e.g., targeting those in the NA and NP genes, which were not covered in that review, and to cover the temporal gap between 2000 and May 28, 2024, we formulated a search string, including MeSH Terms and Boolean operators. We used the following terms to search for mutations in AIVs: *influenza A virus [MeSH Terms] AND (mutation OR mutagenesis OR virulence [MeSH Terms]) AND (mammalian adaptation)*. We performed this focused search on PubMed, by applying filters for the selected period (2020-2024), and by selecting only papers available in English for ease of translation. In May 2024, the search yielded 189 records, exported into a single dataset. The Rayyan software was used to organize the bibliography and facilitate the screening process. Screening by title and abstract was conducted manually. Subsequently, the remaining papers underwent full-text screening, enabling data collection from pertinent studies and facilitating snowballing.

A total of 113 records were included from articles derived from the search string. Additionally, other relevant papers, that were not initially selected by the search string, were also enrolled. After a deep analysis of the screened records, zoonotic markers were identified and used to expand the tables from the inventory by Suttie et al. (2019). Indeed, point mutations and molecular markers affecting phenotypic function from avian-origin influenza viruses, including those isolated from human cases, were used to generate a database for each segment as spreadsheets. For ease of understanding, HA mutations were numbered according to the H5 reference sequence A/Vietnam/1203/2004 (H5N1) (GISAID acc. n. EPI1990181) and to the H3 reference sequence A/Aichi/2/1968 (H3N2) (GISAID acc. n. EPI130007), while NA and internal proteins were numbered according to the genome segments of H5N1 A/Goose/Guangdong/1/1996 (NCBI Reference Sequence PB2: NC\_007357.1, PB1: NC\_007358.1, PA: NC\_007359.1, NP: NC\_007360.1, NA: NC\_007361.1, MP: NC\_007363.1, NS: NC\_007364.1).

We excluded computational studies due to the lack of experimental validation. Furthermore, to focus on AIV and zoonotic infection, specific markers of viruses that are endemic in swine (H1 and H3) (Bi et al., 2010), in dog and equine (H3) (Cui et al., 2016) were also excluded. Finally, as several publications report the same mutations causing similar biological characteristics, we excluded duplicate information.

The subtypes of avian influenza viruses (AIVs) included in the analysis were those that had reported spillover to mammals based on literature. In detail, a total of 46 subtypes were included: next to HPAI H5Nx and H7Nx viruses, H10 (mostly affecting marine mammals), H3N8 and H9N2 found in wild mammals and identified as causative agents in human cases (EFSA, ECDC and EURL, 2023) were considered. H17 and H18, as well as N10, and N11 affecting bats (Tong et al., 2012; 2013) were deliberately excluded since these subtypes have not been detected in avian species to date. All the subtypes included are reported in Table 1.

## 2.2 Multi-basic cleavage site sequences

To conduct a comprehensive analysis, we decided to examine the cleavage site sequences that play a crucial role in the pathogenicity of influenza A in all the selected subtypes (as from

the sequences downloaded, see Section 2.3). For this analysis, the inventory provided by OFFLU, version 4th January 2022 (<https://www.offlu.org/wp-content/uploads/2022/01/Influenza-A-Cleavage-Sites-Final-04-01-2022.pdf>) was used. In this inventory, cleavage sites are classified as multi-basic cleavage sites of H5Nx HPAI viruses, multi-basic cleavage sites of sporadic H7 HPAI viruses, and unusual, 2-3 basic residue multi-basic cleavage sites with variable LP and HP phenotypes. To this list, we added some recent HPAI sequences, such as the one belonging to the strain A/chicken/South Africa/SA2310/2023 (H7N6; HA GISAID acc. N. EPI2699047) (Monjane et al., 2024).

## 2.3 AIV sequences download

To screen the zoonotic markers identified in the literature review (see Section 2.1) within AIVs sequences, including those from spillover events to mammals, public sequence databases were consulted. Indeed, the sequences of the selected subtypes of AIV from avian, mammalian, and human hosts were downloaded from GISAID. Additionally, the sequences from the current H5N1 dairy cattle outbreak were obtained from GitHub, to include data from AIV-infected cows missing in GISAID.

GitHub sequences were directly obtained from <https://github.com/moncla-lab/avian-flu-USDA-cattle/tree/main>, while those from GISAID were collected from GISAID EpiFlu™ database (<https://platform.epicov.org/epi3/frontend#502094>) by applying specific filters, as detailed below.

All records were downloaded from the “avian-flu-USDA-cattle” GitHub section (FASTA files and specific metadata). The download was conducted, in the GISAID EpiFlu™ database, by selecting the influenza type A. Then, each subtype was selected based on the hemagglutinin (H) and the neuraminidase (N) fields. Host selection was performed as follows: human; avian hosts (chicken, curlew, duck, eagle, falcon, goose, grouse, guineafowl, gull, ostrich, other avian species, partridge, passerine birds, penguin, pheasant, pigeon, rails, sandpiper, shearwater, swan, turkey, turnstone, and US quail); and mammals (canine, dairy cow, equine, feline, other mammals, and swine). Those marked as “laboratory-derived” viruses were deliberately excluded. No filter was applied for the “Location”, the “Clades”, and the “Lineages” fields. The “Collection date from” and the “Submission date from” were set to January 2000 to May 2024. For each AIV segment, we conducted separate downloads for the viral segments PB2, PB1, PA, HA, NP, NA, MP, and NS, ensuring only complete sequences were included (by applying the filter “only complete” on GISAID). Otherwise, for the PB1 and PA segments, we also downloaded sequences marked as incomplete. This is because we detected that many of the sequences flagged by GISAID as incomplete, were complete. We selected all the available records, thus downloading the specific metadata (isolates as XLS with only virus metadata) and nucleotide sequences in FASTA format.

## 2.4 Dataset filtering and sequences exclusion

To improve the dataset quality and reduce potential biases, isolates were filtered. The aim was to remove sequences endemic in swine, canine, equine, and human species; poor-quality sequences; those laboratory-derived, and duplicates.

### 2.4.1 Endemic strains

The H1 and H3 endemic strains, circulating in swine, canine, and equine species, as well as sequences from human seasonal influenza viruses, were excluded according to the aims of the current opinion. To exclude these endemic strains, the HA segments from the following

subtypes including H1N1, H1N2, H3N1, H3N2, H3N3, and H3N8 were clustered at 95% of identity using the CD-HIT v4.8.1 software (<https://github.com/weizhongli/cdhit>).

#### 2.4.2 Poor-quality sequences and duplicates

Except for PB1 and PA segments, we used the GISAID filter “only complete” to remove incomplete and poor-quality sequences from our dataset. However, we observed samples with low coverage during the alignment process. Some sequences, identified by GISAID as complete, actually contain missing portions represented as multiple 'N's. Since these poor-quality sequences could adversely affect results, a threshold of 90% coverage was established. Therefore, all sequences that fell below this threshold were removed. Additionally, we noted that some nucleotide sequences contained extraneous elements within the sequences like anomalous letters, or even references to the host (e.g., the entire word “chicken”) and the geographical locations (e.g., the entire word “poland”). In these instances, the identified sequences were removed to maximize the overall quality of the dataset.

Furthermore, sequences marked as “laboratory-derived” in GISAID were originally excluded. Then we examined the “passage history” data associated with each sequence to ensure a more comprehensive filtering. This allowed us to identify and exclude sequences that underwent multiple passages on cell cultures or chicken embryos.

Finally, we also identified and removed duplicate sequences, preserving only those sequences from the first isolation passage. Any repetition (e.g., sequences available in both databases) was recognized by assessing the repetition/similarity of “isolate names” among sequences and sequentially removed.

### 2.5 Screening of point mutations at the amino acid level

Each remaining sequence was mapped to the previously identified references for each segment, using NCBI’s mapping tool BLAST v2.14.0 ([https://github.com/ncbi/blast\\_plus\\_docs](https://github.com/ncbi/blast_plus_docs)). Sequences carrying the point mutations of interest at the amino acid level were identified and, for each of them, a set of in-house scripts was executed to extract the presence/absence profiles of potential zoonotic markers, and associate related metadata.

From the matrices, the frequencies of each potential zoonotic marker were calculated using different metadata to group the sequences based on host, location, and collection date. Mutation frequencies were calculated:

- across subtypes and its segments;
- and for each segment, across all subtypes, as total frequencies.

## 3 Results

### 3.1 Identification of AIV zoonotic markers

From the screened records, potential AIV zoonotic markers were identified and used to expand the tables from the inventory by Suttie et al. (2019), resulting in a sum of 592 markers (Table 1). It must be noted that mutations reported as occurring in combination were enumerated as a single entity, meaning that multiple mutations occurring together were reported as a

separate entry in the dataset (e.g. in the PB2 segment, the combined mutation E627K-D701N was reported as a separate line in the dataset). A table was created for each viral segment (PB2, PB1, PA, HA, NP, NA, MP, NS), resulting in eight spreadsheets.

In these files, each entry provides the following details, organized into a series of columns:

- the specific mutation, as the exact amino acid substitution according to the references (HA mutations were numbered according to the H5 and H3 reference sequences A/Vietnam/1203/2004 (H5N1) and A/Aichi/2/1968 (H3N2), respectively; while NA and internal proteins were numbered according to the genome segments of A/Goose/Guangdong/1/1996 (H5N1));
- the subtype in which the mutation was originally observed;
- the phenotypic effect resulting from the mutation;
- the study type (*in vitro*, *in vivo* studies, or mutations exclusively observed through genomic characterization);
- the corresponding scientific reference (literature).

The phenotypic effect of the marker was initially described in the tables based on information extracted from the literature. Subsequently, this information was disaggregated into different columns, to discriminate effects on:

- evasion from mammalian restriction factors,
- increased mammalian specificity of the viral polymerase,
- increased mammalian specificity of virus attachment to receptor (receptor preference),
- increased HA stability in mammal's environment (decreased pH of fusion, increased thermal stability),
- disruption of the second sialic acid binding site (2SBS) in neuraminidase,
- increased antiviral resistance,
- increased virulence in mammals,
- increased transmission in mammals.

The presence/absence of mutation effects or study type is expressed using the binary code, for each field described above: "0" denotes absence, while "1" indicates presence. Mutations are categorized based on specific criteria. For better understanding, some examples are provided below:

- the "*In vitro*" field was marked with "1" when the mutation's effect was solely described *in vitro* (e.g., "increased virus binding to  $\alpha 2-6$ "),
- the "*In vivo*" field was marked with "1" when the mutation has experimental evidence, and the specific column "*In which species*" reflects the animal model(s) used,
- the "*Observed*" field was marked with "1" when the mutation's effects were observed through genomic characterization, with the column "*In which species*" indicating the species in which these mutations were observed,
- the column "*Evades mammal restriction factors*" was marked with "1" for mutations affecting evasion from Butyrophilin subfamily 3 member A3 (BTN3A3) or interferon response,
- the field "*Increases mammalian specificity of the viral polymerase*" was marked with "1" if the effect was observed in the polymerase complex,
- the field "*Increases mammalian specificity of virus attachment to receptor (receptor preference)*" was marked with "1" for effects involving the HA segment (e.g., increased virus binding to  $\alpha 2-6$ ),
- the "*Disruption of the second sialic acid binding site (2SBS) in neuraminidase*" field was marked with "1" for mutations affecting the 2SBS in the NA segment,

- the fields related to “*Increased virulence in mammals*” and “*Increased transmission to mammals*” were compiled based on effects reported from *in vivo* experiments. If the original study provides information about specific traits for virulence or transmission to mammals e.g. “Increased mammalian specificity of virus attachment to receptor” + “increases HA stability in mammal’s environment” + etc, a “1” will also be present in the corresponding columns of the specific phenotype trait.

The markers without a negative impact on mammals (e.g., mutations related to decreased virulence in mammals), as well as mutation with a phenotype exclusively described in avian hosts, were excluded, being not relevant for the potential zoonotic risk assessment. The corresponding fields in the tables were filled with “/”.

Table 1: Number of identified markers for each AIVs segment.

	Segments												
	PB2	PB1	PB1	PA	PA-X	HA	NP	NA	MP1	MP2	NS1	NS2	tot
		1	-F2										
n. marker	97	48	6	73	2	166	50	82	16	6	40	6	592

PB2, polymerase basic protein 2; PB1 and PB1-F2, polymerase basic protein 1; PA and PA-X, polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; MP1, matrix protein 1; MP2, matrix protein 2; NS1, non-structural protein 1; NS2, non-structural protein 2; tot, total number of markers identified from literature review.

The updated spreadsheets summarizing the identified AIV zoonotic markers can be accessed at the following link: <https://zenodo.org/records/14024971>.

## 3.2 AIV sequences download

A total of 46 subtypes were included based on their spillover to mammals and if they were described along the mutations identified in Section 3.1. The collected sequences were isolated from January 2000 to May 2024, from avian and mammalian hosts including humans. Overall, 1,091 species were included: 980 avian hosts, 110 mammalian species and humans. The undefined terms “mammals” and “other mammals” were also in the host list, related to two viral isolates. Table 2 reports the subtypes included and the number of sequences downloaded for each segment and each subtype. Sequences were downloaded in FASTA format, from the “avian-flu-USDA-cattle” GitHub section and from the GISAID EpiFlu™ database. A number of 1,482,574 nucleotide sequences were downloaded, corresponding to a total of 2,201,362 amino acidic sequences from 304,086 unique viruses, referring to viruses for which at least one segment sequence was available.

The downloaded metadata (.xlsx format) reported:

- Isolate ID (virus identification) and ID from each segment,
- Isolate name,
- Subtype, Lineage, and Clade,
- Location,
- Host,
- Collection date,



- and other minor information such as passage history, submitting lab, authors, etc.

Table 2: Number of sequences downloaded from GISAID and GitHub, for each segment and each subtype.

Segment										
Subtype	PB2	PB1	PA	HA	NP	NA	MP	NS	tot nt/sub	tot aa/sub
<b>H1N1</b>	59571	45125	62206	93871	58308	79124	70576	58657	527438	764002
<b>H1N2</b>	3778	3881	3958	6618	3879	6553	4735	3959	37361	53894
<b>H2N2</b>	41	48	48	44	43	43	45	45	357	543
<b>H3N1</b>	205	218	217	247	210	238	226	236	1797	2694
<b>H3N2</b>	61422	85504	79884	78221	55505	104794	87433	73928	626691	953440
<b>H3N3</b>	25	30	30	29	26	25	29	27	221	337
<b>H3N8</b>	1993	2135	2151	2259	2043	2286	2100	2197	17164	25747
<b>H4N1</b>	11	13	13	13	12	12	16	15	105	162
<b>H4N2</b>	164	193	191	160	172	169	177	185	1462	2157
<b>H4N5</b>	18	19	18	15	16	16	16	18	136	207
<b>H4N6</b>	1509	1607	1606	1543	1549	1520	1608	1610	12552	18983
<b>H5N1<sup>GISAID</sup></b>	10860	11926	11958	13739	11577	12136	10693	11568	94457	140646
<b>H5N1<sup>GitHub</sup></b>	239	239	239	239	239	239	239	239	1912	16298
<b>H5N2</b>	1355	1384	1372	1524	1306	1359	1309	1312	10921	2429
<b>H5N3</b>	190	206	209	278	195	202	171	196	1647	235
<b>H5N4</b>	20	19	20	20	20	19	19	20	157	2416
<b>H5N5</b>	196	204	204	235	188	195	188	205	1615	24213
<b>H5N6</b>	1875	2001	2012	2392	1937	2065	1979	1980	16241	183
<b>H5N7</b>	14	15	15	20	17	16	11	17	125	35776
<b>H5N8</b>	2891	3049	3065	3442	2911	3068	2714	2904	24044	488
<b>H5N9</b>	40	40	39	52	40	34	41	41	327	5089
<b>H6N1</b>	418	424	432	426	426	441	413	420	3400	8368
<b>H6N2</b>	662	715	715	726	671	763	659	684	5595	915
<b>H6N5</b>	73	77	75	77	74	75	80	76	607	8912
<b>H6N6</b>	705	763	734	836	714	753	718	737	5960	2189
<b>H6N8</b>	168	191	185	187	171	185	181	182	1450	1519
<b>H7N1</b>	108	147	148	112	121	112	119	119	986	740
<b>H7N2</b>	53	62	63	73	57	59	60	64	491	9526



<b>H7N3</b>	761	813	807	837	791	731	795	788	6323	568
<b>H7N4</b>	44	47	46	55	42	51	46	49	380	181
<b>H7N5</b>	16	17	10	15	16	16	15	17	122	917
<b>H7N6</b>	73	78	79	90	67	93	69	71	620	5356
<b>H7N7</b>	600	419	418	663	358	581	349	391	3779	386
<b>H7N8</b>	30	32	32	34	30	32	34	32	256	26904
<b>H7N9</b>	2190	2272	2250	2347	2187	2294	2206	2215	17961	64243
<b>H9N2</b>	4654	5136	5071	9419	4860	5222	4823	5014	44199	879
<b>H10N1</b>	71	76	76	72	73	71	71	73	583	2009
<b>H10N3</b>	164	169	168	166	166	165	170	167	1335	910
<b>H10N4</b>	75	81	80	76	76	73	73	71	605	1276
<b>H10N5</b>	107	106	107	102	106	105	108	107	848	8123
<b>H10N7</b>	644	685	687	636	678	663	686	693	5372	1535
<b>H10N8</b>	125	135	135	133	121	128	122	122	1021	195
<b>H11N6</b>	13	17	15	13	16	13	17	21	125	1037
<b>H13N2</b>	84	87	89	78	86	87	88	87	686	2985
<b>H13N6</b>	242	254	258	239	244	240	246	252	1975	1355
<b>H13N8</b>	111	114	113	112	113	113	113	113	902	395
<b>H13N9</b>	30	34	34	36	30	35	32	32	263	764002
<b>tot/seg</b>	15863 8	17083 2	18230 8	22252 1	15248 7	22721 4	19661 8	17195 6	148257 4	220136 2

PB2, polymerase basic protein 2; PB1, polymerase basic protein 1; PA polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; MP, matrix proteins; NS, non-structural proteins; tot/seg, total number of sequences for segment; tot nt/sub, total number of nucleotide sequences for subtype; tot aa/sub, total number of amino acid sequences for subtype. For H5N1, sequences were downloaded from GISAID (GIS) and GitHub (Git) to include genomic data from AIV-infected cows missing in GISAID.

All the nucleotide sequences (FASTA format) and related metadata (.xlsx format) can be available on request by EFSA.

### 3.3 Dataset filtering and sequence exclusion

Sequences were filtered through a multi-step process to remove endemic strains in swine, canine, equine, and humans; poor-quality sequences; laboratory-derived, and duplicates. The aim was to improve the dataset quality and minimize potential biases.

#### 3.3.1 Endemic strains

The H1 and H3 endemic strains, circulating in swine, canine, and equine species, as well as sequences from human seasonal influenza viruses, were excluded to focus only on avian influenza spillover into mammals. By grouping sequences at 95% of identity using the CD-HIT v4.8.1 software (<https://github.com/weizhongli/cdhit>), clusters mainly composed of

strains from swine (H1 and H3), human (H1 and H3), canine (H3), and equine species (H3), circulating in different geographical areas during multiple years, were manually excluded from downstream analyses. For clarity, examples of removed clusters are listed below:

- Pandemic H1N1-derived viruses (clusters composed of multiple swine isolates, and a limited number of human and/or few other species strains);
- Groups of human seasonal viruses (composed of several human isolates, and few strains from swine and some other animals);
- Circulating strains in swine species (H1 and H3) (composed mainly of viruses from swine origin);
- Circulating strains in equine and canine species (H3) (composed mainly of equine/canine strains and a low number of strains from viruses isolated from other animals);
- Groups composed of low numbers of avian species that may have caused spillover among several strains from swine/human samples.

The sequences of other segments were also removed for these strains. Conversely, clusters composed of sequences of viruses from avian hosts, groups formed by mammalian sequences, and strains from human, swine, canine, and equine species not included in the previously removed clusters, were manually included in the subsequent analysis. A number of 177,436 viruses were removed based on clustering analysis, indicating that more than half of the initial dataset (58.35%) was excluded from subsequent analysis. In the following table (Table 3) the numbers of H1 and H3 removed sequences are reported, for each subtype.

Table 3: Number of sequences belonging to H1 and H3 subtypes, removed based on clustering analysis on HA segment, for each subtype.

Subtypes	Downloaded HA	Discarded HA based on clustering	Analyzed HA
<b>H1N1</b>	764002	92893	671109
<b>H1N2</b>	53894	6350	47544
<b>H3N1</b>	2694	64	2630
<b>H3N2</b>	953440	77782	875658
<b>H3N3</b>	337	0	337
<b>H3N8</b>	25747	347	25400
<b>tot</b>	1036112	84543	951569

HA, hemagglutinin; tot, total number of sequences removed after clustering by the CD-HIT tool for the H1 and H3 subtypes.

### 3.3.2 Poor-quality sequences and duplicates

Similarly, to maintain the integrity and reliability of our dataset, sequences were filtered before the bioinformatic analysis for the zoonotic marker screening, to eliminate poor-quality sequences, as well as those derived from laboratory handling (multiple cell cultures or chicken embryo passages) and those observed as duplicates.

As for the horizontal coverage on reference sequences, by setting a threshold of 90%, we identified and removed 7,037 sequences with low coverage, out of a total of 1,482,574 downloaded FASTA files (0.47%). These sequences belonged to a total of 3,489 viruses. It is important to note that among these sequences, a number of 6,154 (87.45%) corresponds to PB1, PB1-F2, PA, and PA-X segments since we downloaded both complete and incomplete sequences for these two genes from GISAID. This decision was based on our observation that on GISAID, especially for the PA segment, many sequences are identified as incomplete even if these contain the entire length of the segment. Anyway, only the specific segment that exhibited low coverage was eliminated. The other segments from the same strain, with coverage higher than 90%, were preserved and included in the subsequent analysis.

Regarding sequences containing extraneous elements like anomalous letters, a total of 315 amino acid sequences out of 1,482,574 FASTA files (0.021%) were identified and discarded. In these instances, the complete viruses (69 isolates) were removed to ensure a higher integrity of the dataset.

Initially, we excluded samples identified as "laboratory derived" from GISAID. However, we also reviewed the "passage history" data associated with each sequence, to ensure more comprehensive filtering. Specifically, sequences indicating a "passage history" involving cell cultures or cell passaging were excluded. A total of 413 "passage history" descriptions were excluded (e.g., SIAT, MDCK2; passage details: SIAT1, MDCK1; MDCK-SIATX, SIAT3; CaCo-2 2 +hCK1; 4046\_Day\_4; 10 passages - MDCK cells; Passage Line 1, Clone 2; and so on), corresponding to the removal of a total of 15,301 unique isolates (5.03%). In these cases, the entire viruses were removed (all the segments) to ensure consistency across the different viral segments.

Finally, duplicates were identified by analyzing the "isolate name" among sequences. Doing that, sequences present in both databases (GISAID and GitHub) were considered as single entities. In addition, four identical sequences reported in GitHub were distinguished by minor variations in their isolate names, as reported in the example:


- >A/Cattle/USA/24-009110-023-**original**/2024
- >A/Cattle/USA/24-009110-023-**v**/2024

In these cases, sequences reporting the word "original" were retained, while those expressing "v" were removed. After duplicate removal, a total of 352 sequences deposited on GitHub but missing from GISAID, were included in the final dataset, corresponding to 44 unique viruses.

Tables 4 and 5 express the number of amino acidic sequences downloaded, discarded, and subjected to bioinformatic analysis for each subtype and each segment, respectively.

Table 4: Number of amino acid sequences downloaded, discarded, and subjected to bioinformatic analysis for each subtype.

Subtypes	Downloaded	Discarded	Analyzed
<b>H1N1</b>	764002	673963	90039
<b>H1N2</b>	53894	44280	9614
<b>H2N2</b>	543	8	535
<b>H3N1</b>	2694	650	2044



<b>H3N2</b>	953440	594623	358817
<b>H3N3</b>	337	3	334
<b>H3N8</b>	25747	2231	23516
<b>H4N1</b>	162	3	159
<b>H4N2</b>	2157	13	2144
<b>H4N5</b>	207	9	198
<b>H4N6</b>	18983	79	18904
<b>H5N1<sub>GIS+Git</sub></b>	140646	718	139928
<b>H5N2</b>	16298	54	16244
<b>H5N3</b>	2429	18	2411
<b>H5N4</b>	235	0	235
<b>H5N5</b>	2416	12	2404
<b>H5N6</b>	24213	33	24180
<b>H5N7</b>	183	1	182
<b>H5N8</b>	35776	157	35619
<b>H5N9</b>	488	10	478
<b>H6N1</b>	5089	34	5055
<b>H6N2</b>	8368	163	8205
<b>H6N5</b>	915	0	915
<b>H6N6</b>	8912	153	8759
<b>H6N8</b>	2189	16	2173
<b>H7N1</b>	1519	6	1513
<b>H7N2</b>	740	1	739
<b>H7N3</b>	9526	39	9487
<b>H7N4</b>	568	1	567
<b>H7N5</b>	181	0	181
<b>H7N6</b>	917	0	917
<b>H7N7</b>	5356	15	5341
<b>H7N8</b>	386	1	385
<b>H7N9</b>	26904	37	26867
<b>H9N2</b>	64243	715	63528
<b>H10N1</b>	879	10	869
<b>H10N3</b>	2009	2	2007
<b>H10N4</b>	910	14	896
<b>H10N5</b>	1276	5	1271
<b>H10N7</b>	8123	26	8097

<b>H10N8</b>	1535	3	1532
<b>H11N6</b>	195	0	195
<b>H13N2</b>	1037	14	1023
<b>H13N6</b>	2985	31	2954
<b>H13N8</b>	1355	2	1353
<b>H13N9</b>	395	1	394
<b>tot</b>	2201362	1318154	883208

Downloaded, sequences downloaded from public databases; Analyzed, sequences used for the screening of zoonotic markers; Discarded, sequences removed following filtering process; tot/seg, the total number of amino acid sequences downloaded, discarded, and subjected to bioinformatic analysis for each subtype. As for H5N1<sub>GIS+Git</sub>, sequences were downloaded from both databases (GISAID and GitHub) to include genomic data from AIV-infected cows missing in GISAID (duplicates from isolates reported in both databases were excluded).

Table 5: Number of amino acid sequences downloaded, discarded, and subjected to bioinformatic analysis for each segment.

<b>Subtypes</b>	<b>Downloaded</b>	<b>Discarded</b>	<b>Analyzed</b>
<b>PB2</b>	158399	92213	66186
<b>PB1</b>	170568	89230	81338
<b>PB1-F2</b>	170568	88763	81805
<b>PA</b>	182043	100761	81282
<b>PA-X</b>	182043	102380	79663
<b>HA</b>	222326	178203	44123
<b>NA</b>	226975	136802	90173
<b>NP</b>	152248	101125	51123
<b>MP1</b>	196379	117972	78407
<b>MP2</b>	196379	117989	78390
<b>NS1</b>	171717	96437	75280
<b>NS2</b>	171717	96279	75438
<b>tot/seg</b>	2201362	1318154	883208

PB2, polymerase basic protein 2; PB1, polymerase basic protein 1; PA polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; MP, matrix proteins; NS, non-structural proteins; Downloaded, sequences downloaded from public databases; Analyzed, sequences used for the screening of zoonotic markers; Discarded, sequences removed following filtering process; tot/seg, the total number of amino acid sequences downloaded, discarded, and subjected to bioinformatic analysis for each subtype.

### 3.4 Screening AIV sequences presenting zoonotic markers

By searching for the identified mutations in the AIV sequences from GitHub and GISAID, a set of twelve databases in the form of spreadsheets was generated for each subtype: PB2, PB1 and PB1-F2, PA and PA-X, HA, NP, NA, M1 and M2 from MP, NS1 and NS2 from NS. The coupled mutations, composed of more than one marker on different segments, have been reported in a separate file. Thus, considering the number of subtypes (46), a total of 598 spreadsheets were produced. In addition, a summarizing file was created for each segment and the coupled mutations (13 spreadsheets), including all the subtypes evaluated.

Each of these outcomes is grouped into a folder titled with the specific subtype (e.g. H1N1) and entitled with the name of the specific segment associated with the term “positive samples” (e.g., HA\_positive samples). Each file reports the following information:

- Isolate ID (virus identification),
- Segment ID (segment identification),
- Isolate name,
- Subtype,
- Clade,
- Location (divided into continent; state and when available specific location),
- Host,
- Host class (avian or mammal),
- Host family (mammalian subgroups),
- Collection date,
- along with the list of mutations related to the specific segment. For ease of understanding, mutations related to HA are expressed according to the H5 and H3 numbering (reference sequences H5N1 A/Vietnam/1203/2004 and H3N2 A/Aichi/2/1968, respectively); while NA and internal proteins were numbered according to genome segments of H5N1 A/Goose/Guangdong/1/1996.

The missing metadata from the database appear as empty fields. As for “isolate ID” which identifies the unique virus, codes from GISAID are formatted as EPI\_ISL\_XXXXX (e.g., EPI\_ISL\_19074717), whereas those from GitHub are presented in numeric format (e.g., 24-009027-004). The presence or absence of a specific mutation is indicated using binary coding, where “0” denotes absence and “1” indicates presence. The host class was manually extrapolated from metadata due to the observation that, during sequence downloading, hosts were often categorized as “other mammals” or “other avian”. To determine the precise number of hosts, a manual analysis was conducted to remove generic categories and extract real data from the “Isolate name” field, obtaining a sum of 1.091 species (980 avian hosts, humans, and 110 mammalian species).

Additionally, for subsequent frequency calculations, mammalian species have been grouped into subcategories, considering their large number, as follows:

- Canidae,
- Felidae,
- Mustelidae,
- Phocidae,
- Otariidae,
- Cetacea,
- Bovidae,
- Primate,
- Rodentia,
- Swine,
- Equidae,



- other mammals (attributed to only two sequences).


The spreadsheets (named as "*name of marker\_positive samples.tsv*") can be accessed at the following link: <https://zenodo.org/records/14033485>.

and those per each subtype can be available on request by EFSA.

Table 6, Figures 1 and 2 report the number of amino acid sequences screened for zoonotic markers for each subtype and segment. Overall, following filtering processes, a global of 883,208 amino acid sequences were analyzed. These correspond to a total of 126,480 isolates considering viruses for which at least a segment sequence was available, or a count of 27,875 unique viruses, considering isolates for which all the segments were accessible.

Table 6: Amino acid sequences screened for each amino acid segment and each subtype.

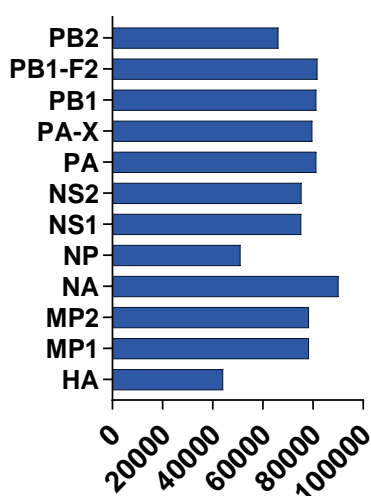
Subtypes	Segments												tot/su b
	PB2	PB1	PB1- F2	PA	PA- X	HA	NP	NA	MP 1	MP 2	NS1	NS2	
<b>H1N1</b>	7793	5559	5607	7317	7270	527	7605	12906	8407	8407	9322	9319	90039
<b>H1N2</b>	817	871	877	875	877	256	845	546	915	915	909	911	9614
<b>H2N2</b>	41	46	46	46	46	44	43	43	45	45	45	45	535
<b>H3N1</b>	154	157	161	157	158	180	165	172	177	177	193	193	2044
<b>H3N2</b>	24235	39772	39963	37879	36206	208	8345	40571	35668	35664	30072	30234	358817
<b>H3N3</b>	25	30	30	29	29	29	26	25	29	28	27	27	334
<b>H3N8</b>	1856	1937	1968	1941	1972	1908	1891	2035	1945	1945	2059	2059	23516
<b>H4N1</b>	11	12	12	12	13	13	12	12	16	16	15	15	159
<b>H4N2</b>	164	190	191	188	188	159	172	168	177	177	185	185	2144
<b>H4N5</b>	17	17	17	16	16	15	16	16	16	16	18	18	198
<b>H4N6</b>	1508	1580	1597	1582	1596	1539	1549	1517	1608	1608	1610	1610	18904
<b>H5N1<sub>GIS+Gi</sub></b>	10836	11731	11785	11800	11830	13760	11568	12127	10686	10684	11561	11560	139928
<b>H5N2</b>	1354	1375	1373	1357	1357	1523	1306	1359	1309	1308	1312	1311	16244
<b>H5N3</b>	190	204	205	199	206	277	195	201	171	171	196	196	2411
<b>H5N4</b>	20	19	19	20	20	20	20	19	19	19	20	20	235
<b>H5N5</b>	196	201	201	201	201	235	188	195	188	188	205	205	2404
<b>H5N6</b>	1872	1989	1995	2006	2007	2391	1937	2065	1979	1979	1980	1980	24180
<b>H5N7</b>	14	14	15	15	15	20	17	16	11	11	17	17	182
<b>H5N8</b>	2886	3000	3011	3041	3037	3439	2909	3066	2712	2712	2903	2903	35619
<b>H5N9</b>	40	38	38	36	36	52	40	34	41	41	41	41	478
<b>H6N1</b>	418	409	420	425	424	426	426	441	413	413	420	420	5055
<b>H6N2</b>	643	683	682	683	682	720	671	758	659	657	683	684	8205



<b>H6N5</b>	73	77	77	75	75	77	74	75	80	80	76	76	915
<b>H6N6</b>	694	716	727	717	719	828	707	750	716	715	735	735	8759
<b>H6N8</b>	168	188	189	181	181	186	171	183	181	181	182	182	2173
<b>H7N1</b>	108	145	145	147	147	112	121	112	119	119	119	119	1513
<b>H7N2</b>	53	62	62	62	63	73	57	59	60	60	64	64	739
<b>H7N3</b>	759	807	806	803	804	834	789	728	793	793	786	785	9487
<b>H7N4</b>	44	46	47	46	46	55	42	51	46	46	49	49	567
<b>H7N5</b>	16	17	17	10	10	15	16	16	15	15	17	17	181
<b>H7N6</b>	73	78	78	79	79	90	67	93	69	69	71	71	917
<b>H7N7</b>	600	416	417	413	414	663	358	581	349	348	391	391	5341
<b>H7N8</b>	30	32	32	32	32	34	30	31	34	34	32	32	385
<b>H7N9</b>	2190	2265	2267	2236	2240	2347	2187	2294	2206	2205	2215	2215	26867
<b>H9N2</b>	4626	4927	4990	4920	4928	9408	4849	5215	4822	4818	5013	5012	63528
<b>H10N1</b>	71	74	74	73	73	72	73	71	71	71	73	73	869
<b>H10N3</b>	164	168	169	167	168	166	166	165	170	170	167	167	2007
<b>H10N4</b>	75	79	79	75	75	76	76	73	73	73	71	71	896
<b>H10N5</b>	106	104	106	106	106	102	106	105	108	108	107	107	1271
<b>H10N7</b>	641	677	684	680	682	634	678	663	686	686	693	693	8097
<b>H10N8</b>	125	134	133	135	135	133	121	128	122	122	122	122	1532
<b>H11N6</b>	13	17	17	15	15	13	16	13	17	17	21	21	195
<b>H13N2</b>	84	82	82	87	87	78	86	87	88	88	87	87	1023
<b>H13N6</b>	242	247	247	251	251	238	244	240	246	246	251	251	2954
<b>H13N8</b>	111	113	113	113	113	112	113	113	113	113	113	113	1353
<b>H13N9</b>	30	33	34	34	34	36	30	35	32	32	32	32	394
<b>tot/seg</b>	66186	81338	81805	81282	79663	44123	51123	90173	78407	78390	75280	75438	883208

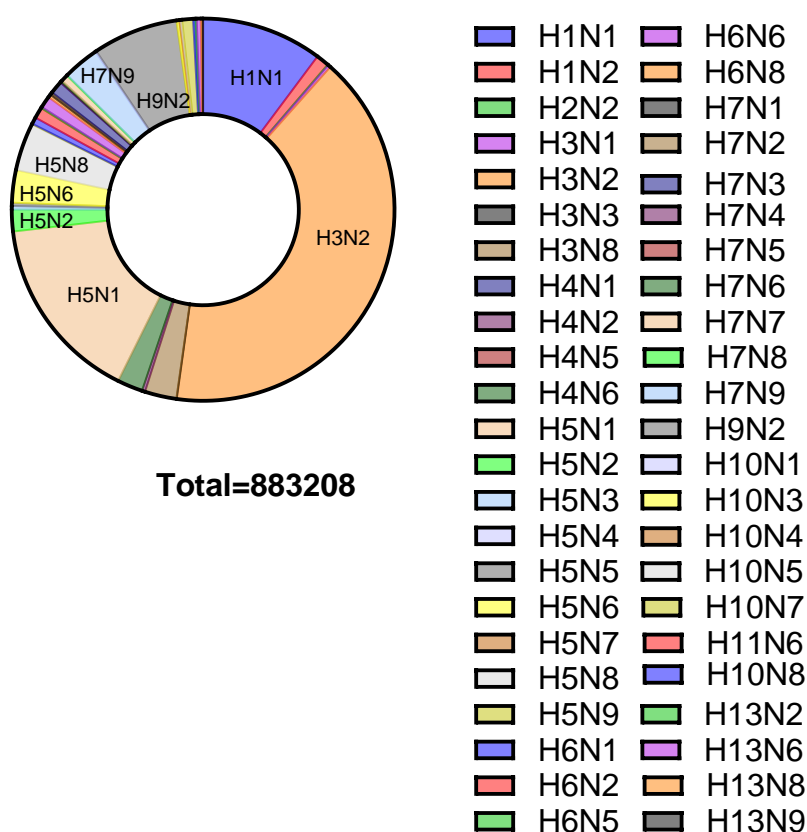
PB2, polymerase basic protein 2; PB1 and PB1-F2, polymerase basic protein 1; PA and PA-X polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; MP1, matrix protein 1; MP2, matrix protein 2; NS1, non-structural protein 1; NS2, non-structural protein 2; tot/seg, the total number of amino acidic sequences for segment; tot/sub, the total number of amino acidic sequences for subtype. As for H5N1<sub>GIS+Git</sub>, sequences were downloaded from both databases (GISAID and GitHub) to include genomic data from AIV-infected cows missing in GISAID (duplicates from isolates reported in both databases were excluded).

Figure 1: Amino acid sequences analyzed for viral segment.



In the Y-axis viral segments (PB2, polymerase basic protein 2; PB1 and PB1-F2, polymerase basic protein 1; PA and PA-X polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; MP1, matrix protein 1; MP2, matrix protein 2; NS1, non-structural protein 1; NS2, non-structural protein 2), in the X-axis the number of amino acid sequences analyzed.

Figure 2: Amino acid sequences analyzed for subtype included (46).



### 3.5 Mutation frequency

Data generated from the zoonotic markers screening in the AIV sequences (3.4 section) were combined with metadata and elaborated to answer the second ToR, thus providing a mutation frequency overview per mutation, subtype, species, and geographical area.

Similarly, for each subtype (46), a panel of thirteen spreadsheets (one for each segment, plus the one for combined mutations) was generated, for a sum of 611 files. In addition, thirteen summarizing files were created including all the subtypes evaluated.

Each of these is entitled with the name of the subtype and the specific segment (e.g., H2N2\_HA\_frequencies or HA\_TOT\_freq) and reports:

- the subtype analyzed,
- the viral protein segment,
- the specific mutation (mutations related to HA are expressed according to H5 and H3 numbering (reference sequences H5N1 A/Vietnam/1203/2004 and H3N2 A/Aichi/2/1968, respectively); while NA and internal proteins were numbered according to genome segments of H5N1 A/Goose/Guangdong/1/1996),
- its phenotypic function (*in vitro*, *in vivo* studies or mutations exclusively observed through genomic characterization; evasion of mammalian restriction factors, increase of specificity of the viral polymerase, mammalian receptor preference, increase of viral stability, disruption of the second sialic acid binding site in NA, increase of antiviral resistance, and increase of virulence and transmission),
- frequency in viruses from avian species,
- frequency in viruses from mammals,
- frequency in viruses from humans,
- frequencies, in viruses from different mammal subgroups (Canidae; Felidae; Mustelidae; Phocidae; Otariidae; Cetacea; Bovidae; Primate; Rodentia; Swine; and Equidae),
- frequencies of mutation based on geographical area (determined for continents: Europe, North America, South America, Asia, Africa, Antarctica, and Oceania),
- frequencies of mutation by collection date (five-year intervals were determined: 2000-2004; 2005-2009; 2010-2014; 2015-2019; 2020-2024).

Frequencies were defined as the number of sequences isolated from a class (e.g. humans) showing a mutation divided by the total number of sequences from that class. Furthermore, duplicate spreadsheets were generated, with frequencies expressed as absolute numbers rather than percentages. In both cases, headers indicate the number of sequences analyzed, while each row shows the percentage or number of positive samples for the specific mutation in that species/location/period. Results are indicated as NA (not applicable), when a mutation was not identified at all in a certain subgroup, differentiating it from mutations characterized by very low frequencies (e.g. 0.000001), expressed as 0.

It is important to underline that these percentages pertain specifically to sequenced cases, not to sampled cases or the overall total.

The spreadsheets of frequencies expressed as percentage or absolute numbers can be accessed at the following link: <https://zenodo.org/records/14033485> ("name of marker\_TOT\_freq.tsv" and "name of marker\_TOT\_freq\_absolute number.tsv"), while the spreadsheets per each subtype can be available on request by EFSA.

### 3.6 Multi-basic cleavage sites

The hemagglutinin sequences of all subtypes tested (a total number of 44,123) were also analyzed for the cleavage site evaluation, which is involved in HPAI avian influenza pathogenicity in birds. By considering the sequences listed in the OFFLU document and recent HPAI sequences, a total of 103 multi-basic cleavage site sequences (HPAI, LPAI, and unusual) were screened in the hemagglutinin sequences.

For each subtype, a spreadsheet was generated reporting:

- the investigated cleavage site sequence,
- the subtypes related (H5, H7, unusual),
- the associated pathogenicity to the subtype (HPAI, LPAI),
- and the corresponding frequencies across different categories (as specified in the previous section).

Similarly to the point mutations screening, a set of “positive results” outcomes, and two “mutation frequency” sets were generated. For the three sets, frequencies have been calculated for each subtype (46) and globally (one summarizing file), thus 141 spreadsheets were created (available on request from EFSA). Related spreadsheets of the summarising files:

- clivage\_TOT\_freq.tsv (%)
- clivage\_TOT\_freq.tsv (absolute numbers)
- clivage\_TOT\_positive\_samples.tsv

can be accessed also at the following link: <https://zenodo.org/records/14033485>.

### 3.7 Data availability


Data provided to the requestor were organized in folders, specifically:

- AIV zoonotic markers (eight tables, one for each viral segment, reporting identified zoonotic mutations),
- Downloaded sequences (FASTA format) and related metadata (.xlsx format),
- Outcomes, divided into subfolders:
  - Positive results (658 spreadsheets from the screening of zoonotic markers in AIV sequences, divided by subtypes and segments),
  - Mutation frequency (1,316 spreadsheets from frequencies analysis for hosts, locations, and periods; divided by subtypes and segments; expressed as percentages and absolute numbers).

The dataset produced are available in the tsv. format, are accessible at the links in the present document or can be accessed on request by EFSA.

## 4 Conclusion

AIVs pose a significant risk of zoonotic transmission, host switching, and the potential generation of pandemic strains. The continuous circulation in wild and domestic birds and the high genetic variability highlight the importance of understanding the potential risk of emerging new strains. Factors like deforestation, urbanization, and habitat changes further increase interactions between domestic and wild species, influencing the transmission



dynamics of AIVs. Global trade and traveling also contribute to the AIVs' spread, potentially introducing novel strains to different regions or host species. Active monitoring is crucial as it allows the early detection of outbreaks, thus enabling rapid response measures to prevent the spread to humans and other animals. Recently, an increase in AIV infections among mammals has been observed. Notably, H5N1 infections have been documented in minks in Spain, dolphins in South America, and seals in the USA, coinciding with significant avian outbreaks (Bonilla-Aldana et al., 2024; Agüero et al., 2023; Leguia et al., 2023; Puryear et al., 2023; Szaluś-Jordanow et al., 2023). The recent outbreak of H5N1 in dairy cattle has heightened public health concerns regarding the zoonotic potential of these viruses, emphasizing the critical need for a One Health approach (CDC, <https://www.cdc.gov/bird-flu/situation-summary/index.html>; EFSA, ECDC, EURL, 2024). It is fundamental to enhance preparedness and mitigation strategies, by integrating veterinary and human health sectors with environmental and wildlife management, ultimately reducing the risks posed by zoonotic diseases and promoting global health security.

This research aimed to identify potential zoonotic markers documented in the literature that can contribute to mammalian adaptation in avian influenza viruses and to screen sequences deposited in public databases over the past 24 years. These were analyzed by collecting mutation frequencies per host (avian, mammals, humans, and mammalian subgroups), per geographical area (by continent), and per period (five-year ranges). From a literature review comprising approximately 120 records published between 2020 and 2024, a total of 592 potential zoonotic markers were identified and listed in eight different datasets (one for each viral segment: PB2, PB-1, PA, HA, NA, NP, MP, NS). Precisely, a number of 883,208 amino acid sequences deposited in GISAID and Github, collected worldwide between January 2000 and May 2024 were screened. A database composed of a total of 1,974 spreadsheets was created. In detail, 658 positive result outcomes, and 1,316 frequency mutation outcomes were produced.

However, we acknowledge that our analyses are subject to biases. A primary concern is related to the bias associated with observed mutation frequencies. This is because frequencies are intricately linked to the surveillance activity of particular species. Specifically, the sequencing may not fully reflect the sampling, as not all sampled cases are sequenced. Moreover, the sampling may not cover all cases, as not all cases are tested. Consequently, our data predominantly represent sick or deceased cases, potentially overlooking those asymptomatic. Similarly, species subjected to more frequent monitoring might be over-represented in the dataset, potentially distorting the results. A similar consideration regarding the frequency related to collection dates. The advances in sequencing technologies, the broader surveillance efforts, and the increase of data sharing have significantly enhanced the accuracy and efficiency of monitoring operations. As a result, recent sequences may dominate the dataset, potentially overlooking older data.

Furthermore, it is important to note that from GISAID database download, sequences containing non-nucleotide characters have been identified. The immediately identifiable data (315 sequences from 69 unique viruses) have been promptly removed. However, due to the challenge of systematically detecting and excluding all potentially problematic sequences, some sequences with non-amino acidic characters may still be in our dataset.

The GISAID filter “only complete” for the data download may represent another bias. Originally, the filter was applied to remove incomplete and poor-quality sequences. However, we noted that the majority of PB1 and PA sequences marked by GISAID as incomplete were complete. To overcome this issue, we re-downloaded and re-analyzed genomic data for these two segments, thus including incomplete sequences. However, due to time constraints, a



similar detailed assessment was not conducted for the other viral segments, representing a limitation in the comprehensiveness of the analysis.

Similarly, the short deadline prevented the phylogenetic tree building for the clustering analysis of H1 and H3 sequences. Indeed, this method requires significant computational resources and time. After a first attempt with unsuccessful results, the clustering analysis was conducted based on sequence similarity using the CD-HIT tool. Anyway, it is important to emphasize that observed clusters strictly depend on our dataset, thus the final dataset may not accurately reflect reality. This is because some endemic strains, that could have generated distinct clusters, have probably been excluded during the downloading and filtering processes.

Finally, two limitations must be highlighted, regarding the neuraminidase and the accessory proteins PB1-F2 and PA-X. As for the NA, it should be noted that the screening regarded point mutations and some deletions in the second sialic acid binding site (2SBS) described in the literature. These mutations, adjacent to the catalytic site, contribute to sialidase activity against multivalent substrates (Du et al., 2019). However, it is important to clarify that the bioinformatics analysis did not identify any deletions different from those specified in the input. Therefore, some deletions in the 2SBS of the neuraminidase may have been potentially missed. Regarding the auxiliary proteins PB1-F2 and PA-X, it should be remarked that, due to the application of the 90% coverage filter (as with all other segments), the analysis may have overlooked the identification of truncated PB1-F2 and PA-X. The PB1-F2 truncation has been related to increased virulence in mammals (Mettier et al., 2021), while PA-X truncation has been linked to altered replication *in vitro* and virulence in mice, chickens, and ducks (Gao et al., 2015; Hu et al., 2015). However, we chose to maintain the 90% coverage threshold considering the absence of a significant and unequivocal impact of these truncated segments on mammals. Thus, any truncated protein may have been potentially excluded from the analysis.


Despite the identified biases, by considering the high number of zoonotic markers included and the screening of over 1.4 million sequences from public databases, this research unquestionably serves as a starting point to investigate the zoonotic risks associated with avian influenza viruses and to develop a comprehensive risk assessment to enhance preparedness and mitigation strategies.

## References

- Abdelwhab EM, Mettenleiter TC. Zoonotic Animal Influenza Virus and Potential Mixing Vessel Hosts. *Viruses*. 2023 Apr 16;15(4):980. doi: 10.3390/v15040980. PMID: 37112960; PMCID: PMC10145017.
- AbuBakar U, Amrani L, Kamarulzaman FA, Karsani SA, Hassandarvish P, Khairat JE. Avian Influenza Virus Tropism in Humans. *Viruses*. 2023 Mar 24;15(4):833. doi: 10.3390/v15040833. PMID: 37112812; PMCID: PMC10142937.
- Agüero, M., Monne, I., Sánchez, A., Zecchin, B., Fusaro, A., Ruano, M. J., Del Valle Arrojo, M., Fernández-Antonio, R., Souto, A. M., Tordable, P., Cañas, J., Bonfante, F., Giussani, E., Terregino, C., & Orejas, J. J. (2023). Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*, 28(3), 2300001. <https://doi.org/10.2807/1560-7917.ES.2023.28.3.2300001>
- Akin, L., & Gözel, M. G. (2020). Understanding Dynamics of Pandemics. *Turkish Journal of Medical Sciences*, 50(SI-1), 515–519. <https://doi.org/10.3906/sag-2004-133>

- Berhane Y, Joseph T, Lung O, Embury-Hyatt C, Xu W, Cottrell P, Raverty S. Isolation and Characterization of Novel Reassortant Influenza A(H10N7) Virus in a Harbor Seal, British Columbia, Canada. *Emerg Infect Dis.* 2022 Jul;28(7):1480-1484. doi: 10.3201/eid2807.212302. PMID: 35731188; PMCID: PMC9239883.
- Bi Y, Fu G, Chen J, Peng J, Sun Y, Wang J, Pu J, Zhang Y, Gao H, Ma G, Tian F, Brown IH, Liu J. Novel swine influenza virus reassortants in pigs, China. *Emerg Infect Dis.* 2010 Jul;16(7):1162-4. doi: 10.3201/eid1607.091881. PMID: 20587196; PMCID: PMC3321907.
- Bedford T, Moncla L (2024). Avian-flu-USDA-cattle. GitHub. Available from: <https://github.com/moncla-lab/avian-flu-USDA-cattle/tree/main>
- Boivin S, Cusack S, Ruigrok RW, Hart DJ. Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. *J Biol Chem.* 2010 Sep 10;285(37):28411-7. doi: 10.1074/jbc.R110.117531. Epub 2010 Jun 10. PMID: 20538599; PMCID: PMC2937865.
- Bonilla-Aldana, D. K., Calle-Hernández, D. M., Ulloque-Badaracco, J. R., Alarcón-Braga, E. A., Hernández-Bustamante, E. A., Cabrera-Guzmán, J. C., Quispe-Vasquez, S. M., Huayta-Cortez, M. A., Benites-Zapata, V. A., & Rodríguez-Morales, A. J. (2024). Highly pathogenic avian influenza A(H5N1) in animals: A systematic review and meta-analysis. *New microbes and new infections*, 60-61, 101439. <https://doi.org/10.1016/j.nmni.2024.101439>
- Cui H, Shi Y, Ruan T, Li X, Teng Q, Chen H, Yang J, Liu Q, Li Z. Phylogenetic analysis and pathogenicity of H3 subtype avian influenza viruses isolated from live poultry markets in China. *Sci Rep.* 2016 Jun 7;6:27360. doi: 10.1038/srep27360. PMID: 27270298; PMCID: PMC4895239.
- Du W, Guo H, Nijman VS, Doedt J, van der Vries E, van der Lee J, Li Z, Boons GJ, van Kuppeveld FJM, de Vries E, Matrosovich M, de Haan CAM. The 2nd sialic acid-binding site of influenza A virus neuraminidase is an important determinant of the hemagglutinin-neuraminidase-receptor balance. *PLoS Pathog.* 2019 Jun 10;15(6):e1007860. doi: 10.1371/journal.ppat.1007860. PMID: 31181126; PMCID: PMC6586374.
- ECDC (European Centre for Disease Prevention and Control). Risk assessment H5 clade 2.3.4.4b viruses. Available from: <https://www.ecdc.europa.eu/en/infectious-disease-topics/z-disease-list/avian-influenza/threats-and-outbreaks/risk-assessment-h5>
- EFSA (European Food Safety Authority). 2024, April 17. Avian influenza. Available from: <https://www.efsa.europa.eu/en/topics/topic/avian-influenza>
- EFSA (European Food Safety Authority), ECDC (European Centre for Disease Prevention and Control), EURL (European Union Reference Laboratory for Avian Influenza), Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Mirinavičiūtė G, Niqueux É, Ståhl K, Staubach C, Terregino C, Willgert K, Baldinelli F, Chuzhakina K, Delacourt R, Georganas A, Georgiev M and Kohnle L, 2023. Scientific report: Avian influenza overview September–December 2023. *EFSA Journal*, 21(12), 8539. doi: 10.2903/j.efsa.2023.8539
- EFSA (European Food Safety Authority), ECDC (European Centre for Disease Prevention and Control), EURL (European Union Reference Laboratory for Avian Influenza), Alexakis L, Fusaro A, Kuiken T, Mirinavičiūtė G, Ståhl K, Staubach C, Svartström O, Terregino C, Willgert K, Delacourt R, Goudjihounde SM, Tampach S and Kohnle L, 2024. Scientific report: Avian influenza overview March–June 2024. *EFSA Journal* 2024;22(7):8930, 72 pp. <https://doi.org/10.2903/j.efsa.2024.8930>

- Gao, H., Sun, H., Hu, J., Qi, L., Wang, J., Xiong, X., Wang, Y., He, Q., Lin, Y., Kong, W., Seng, L. G., Pu, J., Chang, K. C., Liu, X., Liu, J., & Sun, Y. (2015). Twenty amino acids at the C-terminus of PA-X are associated with increased influenza A virus replication and pathogenicity. *The Journal of general virology*, 96(8), 2036–2049. <https://doi.org/10.1099/vir.0.000143>
- Global initiative on sharing all influenza data (GISAID). GISAID EpiFlu™ Database and Data. Available from: (<https://platform.epicov.org/epi3/frontend#502094>)
- Hu, J., Mo, Y., Wang, X., Gu, M., Hu, Z., Zhong, L., Wu, Q., Hao, X., Hu, S., Liu, W., Liu, H., Liu, X., & Liu, X. (2015). PA-X decreases the pathogenicity of highly pathogenic H5N1 influenza A virus in avian species by inhibiting virus replication and host response. *Journal of virology*, 89(8), 4126–4142. <https://doi.org/10.1128/JVI.02132-14>
- Liang Y. (2023). Pathogenicity and virulence of influenza. *Virulence*, 14(1), 2223057. <https://doi.org/10.1080/21505594.2023.2223057>
- Leguia M, Garcia-Glaessner A, Muñoz-Saavedra B, Juarez D, Barrera P, Calvo-Mac C, Jara J, Silva W, Ploog K, Amaro L, Colchao-Claux P, Uhart MM, Nelson MI, Lescano J. Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. *bioRxiv*. 2023 Mar 3. doi: 10.1101/2023.03.03.531008.
- Mettier J, Marc D, Sedano L, Da Costa B, Chevalier C, Le Goffic R. Study of the host specificity of PB1-F2-associated virulence. *Virulence*. 2021 Dec;12(1):1647-1660. doi: 10.1080/21505594.2021.1933848. PMID: 34125653; PMCID: PMC8205076
- Monjane, I. V. A., Djedje, H., Tamele, E., Nhabomba, V., Tivane, A. R., Massicame, Z. E., Arone, D. M., Pastori, A., Bortolami, A., Monne, I., Woma, T., Lamien, C. E., & Dundon, W. G. (2024). H7N6 highly pathogenic avian influenza in Mozambique, 2023. *Emerging microbes & infections*, 13(1), 2321993. <https://doi.org/10.1080/22221751.2024.2321993>
- OIE/FAO Network of Expertise on Animal Influenza (OFFLU). 2022. Influenza A Cleavage Sites. Available from: <https://www.offlu.org/wp-content/uploads/2022/01/Influenza-A-Cleavage-Sites-Final-04-01-2022.pdf>
- Puryear, W., Sawatzki, K., Hill, N., Foss, A., Stone, J. J., Doughty, L., Walk, D., Gilbert, K., Murray, M., Cox, E., Patel, P., Mertz, Z., Ellis, S., Taylor, J., Fauquier, D., Smith, A., DiGiovanni, R. A., Jr, van de Guchte, A., Gonzalez-Reiche, A. S., Khalil, Z., ... Runstadler, J. (2023). Highly Pathogenic Avian Influenza A(H5N1) Virus Outbreak in New England Seals, United States. *Emerging infectious diseases*, 29(4), 786–791. <https://doi.org/10.3201/eid2904.221538>
- Shortridge KF. Pandemic influenza: a zoonosis? *Semin Respir Infect*. 1992 Mar;7(1):11-25. PMID: 1609163.
- Suttie, A., Deng, Y. M., Greenhill, A. R., Dussart, P., Horwood, P. F., & Karlsson, E. A. (2019). Inventory of molecular markers affecting biological characteristics of avian influenza A viruses. *Virus genes*, 55(6), 739–768. <https://doi.org/10.1007/s11262-019-01700-z>
- Szaluś-Jordanow, O., Golke, A., Dzieciatkowski, T., Chrobak-Chmiel, D., Rzewuska, M., Czopowicz, M., Sapiernyński, R., Kardas, M., Biernacka, K., Mickiewicz, M., Moroz-Fik, A., Łobaczewski, A., Stefańska, I., Kwiecień, E., Markowska-Daniel, I., & Frymus, T. (2023). A Fatal A/H5N1 Avian Influenza Virus Infection in a Cat in Poland. *Microorganisms*, 11(9), 2263. <https://doi.org/10.3390/microorganisms11092263>

- 
- Tan S, Sjaugi MF, Fong SC, Chong LC, Abd Raman HS, Nik Mohamed NE, August JT, Khan AM. Avian Influenza H7N9 Virus Adaptation to Human Hosts. *Viruses*. 2021 May 10;13(5):871. doi: 10.3390/v13050871. PMID: 34068495; PMCID: PMC8150935.
- Tong S, Li Y, Rivaller P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA, Turmelle AS, Moran D, Rogers S, Shi M, Tao Y, Weil MR, Tang K, Rowe LA, Sammons S, Xu X, Frace M, Lindblade KA, Cox NJ, Anderson LJ, Rupprecht CE, Donis RO. A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci U S A*. 2012 Mar 13;109(11):4269-74. doi: 10.1073/pnas.1116200109. Epub 2012 Feb 27. PMID: 22371588; PMCID: PMC3306675.
- Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, Chen LM, Johnson A, Tao Y, Dreyfus C, Yu W, McBride R, Carney PJ, Gilbert AT, Chang J, Guo Z, Davis CT, Paulson JC, Stevens J, Rupprecht CE, Holmes EC, Wilson IA, Donis RO. New world bats harbor diverse influenza A viruses. *PLoS Pathog*. 2013;9(10):e1003657. doi: 10.1371/journal.ppat.1003657. Epub 2013 Oct 10. PMID: 24130481; PMCID: PMC3794996.
- U.S. Centers for Disease Control and Prevention (CDC). H5N1 Bird Flu: Current Situation Summary. CDC; July 5, 2024. Available from: <https://www.cdc.gov/flu/avianflu/avian-flu-summary.htm>
- U.S. Centers for Disease Control and Prevention (CDC). How CDC is monitoring influenza data among people to better understand the current avian influenza A (H5N1) situation. Available from: <https://www.cdc.gov/bird-flu/h5-monitoring/index.html>
- Zhang, Z., & Lei, Z. (2024). The Alarming Situation of Highly Pathogenic Avian Influenza Viruses in 2019-2023. *Global medical genetics*, 11(3), 200–213. <https://doi.org/10.1055/s-0044-1788039>