Check for updates

Comparative analysis of PB2 residue 627E/K/V in H5 subtypes of avian influenza viruses isolated from birds and mammals

Kelsey Briggs and Darrell R. Kapczynski*

Exotic and Emerging Avian Diseases Research Unit, Southeast Poultry Research Laboratory, U.S. National Poultry Research Center, Agricultural Research Service, Athens, GA, United States

Avian influenza viruses (AIVs) are naturally found in wild birds, primarily in migratory waterfowl. Although species barriers exist, many AIVs have demonstrated the ability to jump from bird species to mammalian species. A key contributor to this jump is the adaption of the viral RNA polymerase complex to a new host for efficient replication of its RNA genome. The AIV PB2 gene appears to be essential in this conversion, as key residues have been discovered at amino acid position 627 that interact with the host cellular protein, acidic nuclear phosphoprotein 32 family member A (ANP32A). In particular, the conversion of glutamic acid (E) to lysine (K) is frequently observed at this position following isolation in mammals. The focus of this report was to compare the distribution of PB2 627 residues from different lineages and origins of H5 AIV, determine the prevalence between historical and contemporary sequences, and investigate the ratio of amino acids in avian vs. mammalian AIV sequences. Results demonstrate a low prevalence of E627K in H5 non-Goose/Guangdong/1996-lineage (Gs/GD) AIV samples, with a low number of mammalian sequences in general. In contrast, the H5-Gs/GD lineage sequences had an increased prevalence of the E627K mutation and contained more mammalian sequences. An approximate 40% conversion of E to K was observed in human sequences of H5 AIV, suggesting a non-exclusive requirement. Taken together, these results expand our understanding of the distribution of these residues within different subtypes of AIV and aid in our knowledge of PB2 mutations in different species.

KEYWORDS

avian influenza virus, PB2, ANP32A, virus replication, wild birds, poultry

Introduction

Highly pathogenic (HP) avian influenza virus (AIV) outbreaks in domesticated poultry were rare prior to the 1990's (1, 2). However, in 1996, an HPAIV H5N1 was detected in a domesticated goose (A/Goose/Guangdong/1/1996) that crossed species barriers and was detected in humans in 1997 (3, 4). This AIV lineage created what we now know as the H5-Gs/GD lineage and is responsible for mortality in wild birds, poultry, mammals, and humans. The H5-Gs/GD lineage viruses have become adapted and distributed across the world via migratory waterfowl. This lineage has evolved into 10 genetically distinct clades (0–9) (1). Clade 2 versions of this lineage have become the most successful in terms of viral fitness and global distribution (1, 5, 6). The subclade 2.3.4 was first detected in 2008 in China and has continued to evolve into the current 2.3.4.4a-h viruses (1). The United States (U.S.) saw its first incursion of H5-Gs/GD lineage clade 2.3.4.4c viruses in 2014, which resulted in the deaths of over 47 million poultry, resulting in an estimated loss of 3.3 billion dollars

(1, 7). Many countries, including the U.S., are experiencing large-scale outbreaks from clade 2.3.4.4b viruses, which appear very adapted to migratory waterfowl based on their global spread over the past few years (8, 9).

One interest of this subclade is the adaption to other species (10–14). A well-described avian-to-mammalian genetic adaptation is an amino acid change in the PB2 protein at residue 627. Traditionally, avian sequences contain glutamate (E) at this position, while mammalian/human sequences contain a lysine (K) (15). It was originally thought that this residue played a role in the ability of PB2 to replicate at lower temperatures and provided an explanation for the inefficient replication of AIV in non-avian hosts (16). While the temperature may still play a role, more recent structural studies have determined that PB2 interacts with host protein acidic nuclear phosphoprotein 32 family member A (ANP32A) at residue 627, and this interaction is the driving force behind the E627K mutation (17, 18). The interaction is critical for the stabilization of the AIV polymerase complex (vPol), and the amino acid composition of ANP32A that surrounds residue 627 plays a major role in supporting AIV replication (18). However, other PB2 mutations have been found to support AIV replication in the absence of E627K, namely at positions 271, 590, and 591 (19–21).

Most avian ANP32A proteins encode an additional 33 amino acids (ANP32A$_{33}$) in between the two domains that are critical for AIV polymerase activity (22). The first four amino acids of the insert are a SUMO interacting motif (SIM) site that has been shown to increase the binding efficiency of ANP32A$_{33}$ with the vPol and is located directly above the PB2 627 residue (18, 23). The SIM site contains a mixture of acidic and basic residues, which provides the stabilization of the complex, and likely allows for replication with either a PB2 627E or K. The other 27 additional amino acids duplicated from exon 4 (amino acid residues 149–175) are believed to strengthen the interaction between vPol and ANP32A. Humans, mammals, and ratite species lack the 33 amino acid insertion (ANP32A$_\Delta$), which results in a weaker interaction between ANP32A and the vPol leading to lower polymerase activity (17, 18, 23). The E627K mutation appears to compensate for the weaker interaction and helps to restore the vPol activity of AIV in hosts lacking ANP32A$_{33}$ (18, 23).

Avian ANP32A contains natural variations in splicing patterns that result in three major isoforms of the protein, whereas humans only carry the ANP32A$_\Delta$ isotype (22, 24). Chicken, turkey, and duck produce all three transcripts of ANP32A (ANP32A$_{33}$, ANP32A$_{29}$ (lacking SIM), and ANP32A$_\Delta$), but the ANP32A$_{33}$ isotype is the preferentially expressed isoform (approximately 65%). However, some species of wild waterfowl and other migratory birds express higher proportions of the human-like ANP32A (ANP32A$_\Delta$) and the partial insertion that lacks the SIM site (ANP32$_{29}$), indicating that ANP32A expression is strongly associated with host range (22–24).

In this study, we investigate the number of submitted sequences containing PB2 627E, K, or V from H5 AIV. We compare the prevalence of sequences with 627K between non-Gs/GD and Gs/GD lineages and examine the ratio of sequences within the Gs/GD lineage. Finally, we examine residue specificity within the broad host range of clade 2 AIV, including current global 2.3.4.4 viruses.

Methods

Sequence analysis

All sequences were obtained from the Global Initiative on Sharing All Influenza Data (GISAID EPIFLU$^{\text{TM}}$) database (25). The search parameters on GISAID always included type A, H5, and required a complete segment of PB2. Other parameters, such as host(s) and clade(s) were chosen as needed. Downloaded datasets were aligned in Geneious Prime (Boston, MA) using a MUSCLE alignment. Only complete and correct PB2 protein sequences with coverage over residue 627 were used for analysis. Sequences that did not meet this criterion were discarded. Tables were created based on the number of PB2 sequences that matched the criteria set, such as host species, total number, amino acid residue, and clade. Totals were calculated by adding up each group in the table, and in some cases, the total does not represent the entire dataset. Very few sequences had 627 residues that were not glutamate (E), lysine (K), or valine (V), so they were not included in the analysis total. The percentage of K residues was obtained by dividing the number of K residues by the total of that group and then multiplying by 100. When a clade or subclade was chosen in the "clade" search panel, all subclades were included in the analysis, unless otherwise noted. Tables including a species section were chicken/turkey only, all other avian species (in addition to chicken/turkey), mammals (not including humans), and humans. It is important to note that all timespans listed in this report are based on the dataset used for analysis. They do not represent the exact circulation of those virus clades. The findings of this study are based on data from 125,996 PB2 sequences available on GISAID as of March 2023.

Results

Prevalence of 627K in non-A/Gs/GD/96 lineage H5Nx viruses

Using the GISAID "clade" search panel, we examined sequences that were not classified as part of the Gs/Gd lineage (25). The American-non-Gs/GD/96 (Am_non-GsGD) lineage had 789 complete PB2 sequences with 627 coverages, and the earliest viruses in the dataset were from Wisconsin, USA, in 1975. The subtypes in 1975 included H5N2, H5N6, and H5N1. The species in this lineage consisted of chicken/turkey (220), all other avian (567), mammalian (2), and human (0). Of the 789 sequences examined, only two had PB2 627K, the animals were a rhea and an emu (both from Texas, USA, 1993, H5N2). The percent of PB2 627K sequences for the Am_non-GsGD lineage was 0.25%. Unexpectedly, there were 17 chicken sequences with a V at position 627, which accounted for 2.15% of the total. They originated from an H5N2 isolation in Mexico in 2019 (Table 1, top).

For the Eurasian-non-Gs/GD/96 (Ea_non-GsGD) lineage, 390 sequences were examined. The earliest sequences were from Scotland in 1959 (H5N1). In this lineage, there were 46 chicken/turkey, 342 other avian, 2 mammals, and 0 human sequences. Of the 390 sequences, four had PB2 627K. All four sequences came from ostriches in South Africa in 2011 and 2015. The percentage of PB2 627K sequences from the Ea_non-GsGD

TABLE 1 Prevalence of PB2 627 E/K/V in non-A/Gs/GD/96 H5 lineage viruses.

Lineage	Species	Total numbers	627E	627K	627V	Percent 627E	Percent 627K
American-non-Gs/GD/96 (1975-present)	Chicken/Turkey	220	203	0	17	92.3	0
	Other Avian	567	564	2[a]	0	99.5	0.4
	Mammalian	2	2	0	0	100.0	0
	Human	0	0	0	0	0	0
	Total	**789**	**769**	**2**	**17**	**97.3**	**0.25**
Eurasian-non-Gs/GD/96 (1959-present)	Chicken/Turkey	46	0	0	0	0	0
	Other Avian	342	338	4[b]	0	98.8	1.2
	Mammalian	2	2	0	0	100.0	0
	Human	0	0	0	0	0	0
	Total	**390**	**2**	**4**	**0**	**99.4**	**1.03**

[a] 1 rhea and 1 emu.
[b] 4 Four ostriches.

lineage was 1.03% (Table 1, bottom). Interestingly, there were only four mammalian sequences in the dataset within the two lineages. All four were H5N2 sequences from swine, two were from Mexico in 2014/2015, and two were from Korea in 2008 (Table 1). Based on the available data, the American and Eurasian lineages appear to have low mammalian/human spillover events and a low percentage of PB2 627K adaptations.

Prevalence of PB2 627K in A/Gs/GD/96 H5 lineage clades 0−9

Next, we examined the percentage of PB2 627K residues in clades 0–9 of the Gs/GD lineage, which began in 1996. PB2 sequence availability ranged from 17 to 10,734 sequences between the clades (Table 2). Clade 0 had 11.7% (7/60) PB2 sequences with 627K, which included an ostrich and six human samples. The dataset contained sequences from 1996 to 2008 (Table 2). Clade 1 had 395 sequences available, 39 of which had PB2 627K (9.9 %), which consisted of 14 avian species, 24 mammalian/human species, and 1 unknown, and were sequenced from 1996 to 2014 (Table 2). Clade 2 had the most sequences available (10,734); of those, 738 contained PB2 627K and 16 contained a 627V. Clade 2 had 6.9% of total sequences with 627K; of those, 608 were avian and 130 were mammalian/human. Available clade 2 sequences ranged from 1996 to the present (Table 2). Clade 3 sequences contained 2.9% PB2 627K residues (6/206), 1 was from the environment and 5 were human cases. The sequences were collected from 1997 to 2015 (Table 2). Clade 6 (2002–2022) had two human cases with PB2 627K out of 33 total, making the percentage 6.1% (Table 2). Clade 7 had 2.5% sequences with PB2 627K, there were 2 avian samples out of 80 from 2002 to 2015 (Table 2). Clade 9 sequences ranged from 1997 to 2006; only 2 avian sequences had PB2 627K out of 49 total (4.1%) (Table 2). Clades 4, 5, and 8 had no sequences containing PB2 627K (Table 2). The overall average of PB2 627K in the Gs/GD lineage was 4.9%, which is 13x more than the Am_non-GsGd and 4.8x more than the Ea_non-GsGd lineages.

Prevalence of PB2 627K in A/Gs/GD/96 H5 lineage clade 2

To examine clade 2 more thoroughly, we determined the percent PB2 627K in subclades 2.1–2.5, based on available GISAID sequences (25). Clade 2.1 had a low proportion of avian sequences (34.6%) available compared to mammalian/human sequences (65.4%). All 19 (8.3%) sequences with PB2 627K in clade 2.1 were mammalian/human in origin (Table 3, top). The sequences from clade 2.1 were obtained between 2003 and 2015. Clade 2.2 was previously shown to have a high incidence of PB2 627K residues (20). We found that 92.1% of available PB2 sequences classified as clade 2.2 had 627K residue. Unexpectedly, 578 sequences with PB2 627K were from avian species, and only 36 were mammalian/human. Clade 2.2 sequences ranged from 1997 to 2017 in the dataset (Table 3, top). Clade 2.2 contributed to 83% (614/738) of the total clade 2 PB2 627K population (Table 2). Clade 2.3 (2003–present) has the largest total number of sequences available (9,797), although only 105 of them had PB2 627K. Of the sequences with 627K, 75 were mammalian/human and 30 were of avian origin. The total percentage of PB2 627K sequences for clade 2.3 was 1.1% (Table 3, top). Of note, only clade 2.3 contained PB2 627V sequences. The species containing PB2 627V were mixed, eight chickens, two other avians, three red foxes, one tiger, and one human. Clades 2.4 and 2.5 had low sample sizes, and none of the sequences contained PB2 627K residue. Additionally, the sequences available were only from 2003 to 2004 to 2003–2006, respectively (Table 3, top). The data indicate that the majority of PB2 627K sequences are from clade 2.2, not 2.3.

The most current and prevalent Gs/GD lineage viruses belong to clade 2.3.4.4 (1). We wanted to examine which species in subclades 2.3.4.4a-h had PB2 627K residue. Because the subclades 2.3.4.4d and 2.3.4.4f did not have "clade" options in the GISAID database, they were not included in the analysis (25). Clade 2.3.4.4a contained 1.2% (6/502) sequences with PB2 627K, including two avians and four humans among the species. Sequences in this clade were collected between 2012 and 2018 (Table 3, bottom). The 2.3.4.4b viruses are responsible for the current outbreak in the U.S. and have been reported in many countries (12, 25).

TABLE 2 Prevalence of PB2 627 E/K/V in A/Gs/GD/96 H5 lineage clades 0−9.

Clade	Total numbers	627E	627K	627V	Percent 627E	Percent 627K
0	60	53	7[a]	0	88.3	11.7
1	395	356	39[b]	0	90.1	9.9
2	10,734	9,996	738[c]	15	93.1	6.9
3	206	200	6[d]	0	97.1	2.9
4	18	18	0	0	100.0	0.0
5	22	22	0	0	100.0	0.0
6	33	31	2[e]	0	93.9	6.1
7	80	78	2[f]	0	97.5	2.5
8	17	17	0	0	100.0	0.0
9	49	47	2[g]	0	95.9	4.1

[a]Clade 0−1 avian (ostrich) and 6 humans.
[b]Clade 1−14 avians (1 ostrich), 24 mammalians/humans, 1 unknown.
[c]Clade 2−608 avian, 130 mammalians/humans.
[d]Clade 3−1 environment, 5 humans.
[e]Clade 6−2 humans.
[f]Clade 7−2 avians.
[g]Clade 9−2 avians.

Of the 5,311 sequences analyzed from 2003 to the present, 53 had PB2 627K making the percentage only 1.0%. However, 48 of the 53 sequences were from 2021 to the present. Bird sequences containing 627K accounted for 23 of the 53 sequences and included chickens/turkeys, ducks, ratites, and common terns. There was a higher diversity of mammalian species containing PB2 627K in the 2.3.4.4b viruses, compared to all other groups analyzed. Most of the mammalian/human sequences in clades 2.1 and 2.2 were human, whereas the 2.3.4.4b sequences included 1 bear, 2 skunks, 3 cats, 2 ferrets, 1 raccoon, 14 foxes, 4 seals, 1 otter, and 2 humans (Tables 2, 3, bottom). Of note, the 2.3.4.4c sequences, which were responsible for the first outbreak of a Gs/GD lineage H5Nx virus in the U.S., did not have any sequences with PB2 627K (8). The dataset contained 1,056 samples and included sequences from 2014 to 2020 (Table 3, bottom). Three human samples (3/551) from 2.3.4.4e had PB2 627K from 2014 to 2018 (Table 3, bottom). Clade 2.3.4.4g (2014–2020) had the smallest sample size (84), and none of the sequences contained PB2 627K (Table 3, bottom). Finally, 2.3.4.4h had the highest percentage of PB2 627K sequences at 4.02% (15/373), spanning from 2014 to 2021. The sequences were from two chickens/turkeys, four minks, and nine humans (Table 3, bottom). Taken together, the data indicate that the current 2.3.4.4b viruses have a low overall percentage of PB2 627K sequences, but an increased propensity to infect multiple species.

Prevalence of PB2 627E/K/V in human sequences from different hemagglutinin subtypes

Finally, we examined the prevalence of PB2 627K in human AIV sequences with more common HA subtypes (26). First, the percent of PB2 627K in human sequences in subtypes (H1–H3) was examined. Interestingly, 95% of the H1 sequences sequenced contained the avian-like PB2 627E. We only observed 4.7% of all H1 sequences demonstrating the PB2 627K residue. However, the majority of the 39,418 sequences examined were from the pandemic H1N1 (pdm09) lineage, which began in 2009 and was known to contain a PB2 segment from an avian North American virus (Table 4) (27). As expected, both the H2 and H3 sequences contained higher percentages of PB2 627K residues, with prevalence rates of 91.9% and 99.4%, respectively (Table 4). Next, we examined the proportion of PB2 627K in avian-adapted AIV subtypes (H5, H7, and H9). H5 viruses obtained from human samples contained 39.1% PB2 627K sequences (Table 4). Of the 1,207 human sequences classified as subtype H7, 851, or 70.4% contained a K at position 627 (Table 4). Finally, the H9 subtype contained only two human sequences with PB2 627K (Table 4). Interestingly, the H7 and H9 viruses contained a higher proportion of 627V residues, 2.7% and 36%, respectively. Apart from the pdm09 H1N1 viruses, the data illustrate that human-adapted viruses typically contain PB2 627K.

Discussion

This research aimed to compare the distribution of PB2 627 residues between avian and mammalian sequences in H5 from non-Gs/GD and Gs/GD lineages and determine the prevalence between previous subclades and current ones. The current clade 2.3.4.4b H5Nx AIVs have a global distribution and raise concerns about mammalian adaptations as recent isolations have occurred in domestic and peridomestic mammals (12, 28). In this study, we focused on one well-known avian-to-mammalian adaption residue, E627K/V, to determine if the 2.3.4.4 viruses have an increased propensity to mutate in that direction.

All sequences were obtained from GISAID, so the number of sequences was limited to what was available in the database. We chose GISAID because it contained the most sequences of

TABLE 3 Comparative prevalence of PB2 E/K/V in Gs/GD/96 H5 lineage clades 2.1−2.5 (Top) and distribution within various species in subclade 2.3.4.4 (Bottom).

Clade	Total numbers	Percent Avian sample	627E	627K	627V	Percent 627E	Percent 627K
2.1[a]	228	34.6	209	19[d]	0	91.7	8.3
2.2[b]	667	94.2	53[e]	614[f]	0	7.9	92.1
2.3[c]	9,797	92.4	9,672	105[g]	15	98.7	1.1
2.4	16	100.0	16	0	0	100.0	0.0
2.5	26	100.0	26	0	0	100.0	0.0

Species	2.3.4.4a (502)[l]	2.3.4.4b (5,311)	2.3.4.4c (1,056)	2.3.4.4e (551)	2.3.4.4g (84)	2.3.4.4h (373)
Chicken/Turkey	0	16	0	0	0	2
Other Avian	2[h]	7[i]	0	0	0	0
Mammalian	0	28[j]	0	0	0	4[k]
Human	4	2	0	3	0	9
Percent 627E	98.8	99.0	100.0	99.5	100.0	96.0
Total with 627K	6	53	0	3	0	15
Percent 627K	1.2	1.0	0.0	0.5	0.0	4.0

[a] 2.1 includes 2.1.1, 2.1.2, 2.1.3, 2.1.3.1, 2.1.3.2, 2.1.3.2a, 2.1.3.2b, 2.1.3.3.
[b] 2.2 includes 2.2, 2.2.1, 2.2.1.1, 2.2.1.1a, 2.2.1.2, 2.2.2, 2.2.2.1.
[c] 2.3 includes 2.3.1, 2.3.2, 2.3.2.1, 2.3.2.1a, 2.3.2.1b, 2.3.2.1c, 2.3.3, 2.3.4, 2.3.4.1, 2.3.4.2, 2.3.4.3, 2.3.4.4, 2.3.4.4b, 2.3.4.4c, 2.3.4.4e, 2.3.4.4g, 2.3.4.4h.
[d] 19 mammalians/humans. Of mammalians/humans (149), 12.8% have 627K.
[e] 3 mammalians and 50 avian.
[f] 36 mammalians/humans and 578 avians (4 ostriches).
[g] 75 mammalians/humans and 30 avians (1 emu, 2 rhea).
[h] 1 cormorant and 1 pheasant.
[i] 2 rhea, 1 emu, 1 duck, 2 common terns, and 1 unknown avian.
[j] 1 bear, 2 skunks, 3 cats, 2 ferrets, 1 raccoon, 14 foxes, 4 seals, and 1 otter.
[k] 4 minks.
[l] Total number of samples per 2.3.4.4 subclade.

TABLE 4 Prevalence of PB2 627 E/K/V from human sequences among different influenza hemagglutinin subtypes.

HxNx	Total numbers	627E	627K	627V	Percent 627E	Percent 627K
H1	39,418	37,559	1,855	4	95.3	4.7[a]
H2	123	10	113	0	8.1	91.9
H3	71,994	419	71,560	15	0.6	99.4
H5	402	244	157	1	60.7	39.1
H7	1,207	323	851	33	26.8	70.4[b]
H9	59	36	2	21	61.0	3.4

[a] Pandemic H1N1 (pdm09) contained PB2 from an avian North American virus.
[b] Recent H7N9 cases predominate sequences in this subtype.

current (2021–present) AIV, but it is possible that older strains were missed in the analysis because they were not added retroactively (prior to 2008) (25). AIV surveillance and reporting are not a standard practice in all countries; consequently, samples are limited to countries that do report, and this may contribute to the low dataset numbers and sample bias in some clades. Obtaining samples from only dead birds may also result in sampling bias, but it is not possible to tell whether samples were taken during active or passive surveillance.

In the non-Gs/GD lineage, the majority of sequences contained the 627E residue. We observed only six avian sequences that contained PB2 627K and they belonged to the Ratite family (Ostrich, Rhea, Emu, Cassowary, and Kiwi) of birds (Table 1). Ratites were the only species of birds that contained PB2 627K residues in this group. Phylogenetic analysis of the ANP32A gene demonstrates that the Ratite family lacks the 33 amino acid insertion in exon 5 that most other avian species have, which presumably makes it more mammalian-like, and may explain why viruses isolated from these species select for 627K during replication (14, 22). Of further note is that both rhea and emu sequences from the Am_non-GsGD lineage contained multiple polybasic residues at the HA cleavage site, suggesting that they were HPAI viruses. In the EA_non-GsGD lineage, the two ostriches from 2011 contained an HPAI cleavage

sequence, whereas the two ostriches from 2015 contained an LPAI sequence.

Clade 2 is the most evolutionarily successful clade from the Gs/GD lineage of AIV based on the sample size and the number of subclades (Tables 2, 3) (1, 29). Within subclade 2, the proportion of avian and mammalian samples in clade 2.1 was unexpected. The largest portion of samples was human sequences from Indonesia, ranging from 2003 to 2015 (Table 3). Of note is the observation that most of the sequences in clade 2.2 containing 627K were of avian origin (Table 3). Clade 2.2 viruses were shown to be shed via the respiratory route in waterfowl rather than the typical cloacal route; therefore, it was proposed that clade 2.2 viruses maintained the 627K to compensate for the cooler temperature of the respiratory tract (20, 30, 31). Recently, a study using RNA-seq showed that some waterfowl, land fowl, and pelicans preferentially express a human-like ANP32A, which could also have contributed to a larger proportion of avian species containing 627K in clade 2.2, as there was no selective pressure to maintain 627E (22). Additionally, Long et al. found that the PB2 627K mutation did not affect pathogenesis or transmission in ducks, suggesting that mammalian adaption could be maintained in an avian species, which also supports the notion that there is little selection pressure to go from K627E (20). The clade 2.3 dataset was comprised of 17 subclades, including the current 2.3.4.4b viruses. Interestingly, 98.7% (9,672) of the 2.3 sequences contained a 627E residue despite only 92.4% (9,052) being of avian origin, demonstrating that non-avian species also maintained the 627E residue. This observation went both ways in that of the 105 clade 2.3 sequences demonstrating 627K, only 71.4% (75) were mammalian (Table 3).

Within clade 2.3.4, more than half of the PB2 627K sequences were in the clade 2.3.4.4b subclade (Tables 2, 3). There were 16 chicken/turkey species and 7 other avians that contained PB2 627K in clade 2.3.4.4b, this may indicate that the virus is spilling over from mammals into avians and maintaining the residue at the time of isolation. Several accounts of mammalian spillover events have occurred since the 2.3.4.4b viruses have become predominant and the diversity of species being infected is unprecedented (Table 3) (9, 12, 13, 32). Nevertheless, based on our data, the number of 2.3.4.4b mammalian sequences with PB2 627K is still lower than the sequences with a 627E. Interestingly, we found that five sea mammals (four seals and one otter) contained PB2 627K, whereas a recent study examined dolphin and sea lion sequences from Peru and found they contained a 627E. Leguia et al. also proposed that transmission between sea lions off the coast of Peru could be occurring rather than independent avian spillover events because of the massive die-offs being observed (28). More investigation is required to determine if mammal-to-mammal and mammal-to-avian transmission are occurring, and which residues are allowing for efficient replication of the virus.

Despite the fact that residue PB2 627 is almost exclusively a K or E in all influenza A viruses, a small portion contains a valine (V). A study by Chin et al. found that inducing random mutations at position 627 allowed for the 627V mutation when purified in a mammalian system but not in an avian system (33). This may indicate that the 627V mutations observed in all avian species were transmitted from a mammalian host (Tables 1–4). The study observed a slight reduction in replication compared to the 627K

mutation using the culture-adapted A/Puerto Rico/8/34 (H1N1) virus (33). However, a more recent study used the polymerase genes from an avian H5N1 (A/Muscovy duck/Vietnam/TY93/2007; clade 2.3.4) to rescue a virus containing the PB2 627V mutation. Taft et al. found that the 627V mutation had significantly increased viral replication in mammalian cell culture and virulence in mice that was comparable to the 627K virus (34). Additionally, Luk et al. found that an H7N9 virus containing PB2 627V was extremely fit and transmissible between chickens and mammals (35). The number of V residues in H5 sequences remains low; however, the H7 and H9 human cases had a considerable number of sequences containing a 627V, many of which are from recent years (Table 4).

The dependency of viral polymerases on host ANP32A is one factor in crossing host barriers exemplified by the propensity for AIV sequences to mutate to 627K in mammalian sequences (Tables 1–4) (23). The Gs/GD lineage viruses are commonly found in wild waterfowl and appear to transcribe higher rates of the human-like ANP32A, which may account for shifts in 627 residue specificity (23, 32). While PB2 627K is an established marker for mammalian adaption, it is not solely responsible for it (15, 27, 33). The pdm09 H1N1 viruses contained a PB2 627E, due to the avian origin of the segment but had other compensatory mutations that allowed for efficient replication in mammals and humans (19, 20, 36). As more mammalian species become infected by 2.3.4.4b AIV additional residues may yet be identified. It is known that efficient replication in a certain host requires adaption to the machinery within, and that adaption from one host species to another requires mutation and selection. In this study, we performed a differential analysis of PB2 residue 627 and demonstrated a non-exclusive requirement for conversion.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

KB and DK: conceptualization, methodology, formal analysis, writing—original draft and preparation, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

Conflict of interest

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

Publisher's note

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

References

1. Lee DH, Bertran K, Kwon JH, Swayne DE. Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 23 44. *J Vet Sci.* (2017) 18:269–80. doi: 10.4142/jvs.2017.18.S1.269

2. Joseph U, Su YC, Vijaykrishna D, Smith GJ. The ecology and adaptive evolution of influenza A interspecies transmission. *Influenza Other Respi Viruses.* (2017) 11:74–84. doi: 10.1111/irv.12412

3. Guo Y, Xu X, Wan X. Genetic characterization of an avian influenza A (H5N1) virus isolated from a sick goose in China. *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi.* (1998) 12:322–5.

4. Wan X. Lessons from emergence of A/goose/Guangdong/1996-like H5N1 highly pathogenic avian influenza viruses and recent influenza surveillance efforts in southern China. *Zoonoses Public Health.* (2012) 59:32–42. doi: 10.1111/j.1863-2378.2012.01497.x

5. Claas EC, de Jong JC, van Beek R, Rimmelzwaan GF, Osterhaus AD. Human influenza virus A/HongKong/156/97 (H5N1) infection. *Vaccine.* (1998) 16:977–8. doi: 10.1016/S0264-410X(98)00005-X

6. Alkie TN, Lopes S, Hisanaga T, Xu W, Suderman M, Koziuk J, et al. A threat from both sides: Multiple introductions of genetically distinct H5 HPAI viruses into Canada via both East Asia-Australasia/Pacific and Atlantic flyways. *Virus Evolut.* (2022) 8:veac077. doi: 10.1093/ve/veac077

7. Kapczynski DR, Sylte MJ, Killian ML, Torchetti MK, Chrzastek K, Suarez DL. Protection of commercial turkeys following inactivated or recombinant H5 vaccine application against the 2015 US H5N2 clade 2.3. 44 highly pathogenic avian influenza virus. *Vet Immunol Immunopathol.* (2017) 191:74–9. doi: 10.1016/j.vetimm.2017.08.001

8. *Final Report for the 2014–2015 Outbreak of Highly Pathogenic Avian Influenza (HPAI) in the United States*, in *USDA APHIS HPAI Response.* Riverdale, MD: United States Department of Agriculture. (2016). p. 59.

9. *2022–2023 Confirmations of Highly Pathogenic Avian Influenza in Commercial and Backyard Flocks.* Avian Influenza. (2023). Available online at: https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/avian/avian-influenza/hpai-2022/2022-hpai-commercial-backyard-flocks (accessed April 11, 2023).

10. Kaplan BS, Webby RJ. The avian and mammalian host range of highly pathogenic avian H5N1 influenza. *Virus Res.* (2013) 178:3–11. doi: 10.1016/j.virusres.2013.09.004

11. Claas EC, Osterhaus AD, Van Beek R, De Jong JC, Rimmelzwaan GF, Senne DA, et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet.* (1998) 351:472–7. doi: 10.1016/S0140-6736(97)11212-0

12. Elsmo EJ, Wunschmann A, Beckmen KB, Broughton-Neiswanger LB, Buckles EL, Ellis J, et al. Pathology of natural infection with highly pathogenic avian influenza virus (H5N1) clade 2.3. 4.4 b in wild terrestrial mammals in the United States in 2022. *bioRxiv.* (2023) 2023:03. doi: 10.1101/2023.03.10.532068

13. Klopfleisch R, Wolf PU, Wolf C, Harder T, Starick E, Niebuhr M, et al. Encephalitis in a stone marten (Martes foina) after natural infection with highly pathogenic avian influenza virus subtype H5N1. *J Comp Pathol.* (2007) 137:155–9. doi: 10.1016/j.jcpa.2007.06.001

14. Shinya K, Makino A, Ozawa M, Kim JH, Sakai-Tagawa Y, Ito M, et al. Ostrich involvement in the selection of H5N1 influenza virus possessing mammalian-type amino acids in the PB2 protein. *J Virol.* (2009) 83:13015–8. doi: 10.1128/JVI.01714-09

15. Subbarao EK, London W, Murphy BR, A. single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J Virol.* (1993) 67:1761–4. doi: 10.1128/jvi.67.4.1761-1764.1993

16. Massin P, van der Werf S, Naffakh N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J Virol.* (2001) 75:5398–404. doi: 10.1128/JVI.75.11.5398-5404.2001

17. Long JS, Giotis ES, Moncorgé O, Frise R, Mistry B, James J, et al. Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature.* (2016) 529:101–4. doi: 10.1038/nature16474

18. Carrique L, Fan H, Walker AP, Keown JR, Sharps J, Staller E, et al. Host ANP32A mediates the assembly of the influenza virus replicase. *Nature.* (2020) 587:638–43. doi: 10.1038/s41586-020-2927-z

19. Mehle A, Doudna JA. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc Nat Acad Sci.* (2009) 106:21312–6. doi: 10.1073/pnas.0911915106

20. Long JS, Howard WA, Núñez A, Moncorgé O, Lycett S, Banks J, et al. The effect of the PB2 mutation 627K on highly pathogenic H5N1 avian influenza virus is dependent on the virus lineage. *J Virol.* (2013) 87:9983–96. doi: 10.1128/JVI.01399-13

21. Mok CK, Yen HL Yu MY, Yuen KM, Sia SF, Chan MC, et al. Amino acid residues 253 and 591 of the PB2 protein of avian influenza virus A H9N2 contribute to mammalian pathogenesis. *J Virol.* (2011) 85:9641–5. doi: 10.1128/JVI.00702-11

22. Baker SF, Ledwith MP, Mehle A. Differential splicing of ANP32A in birds alters its ability to stimulate RNA synthesis by restricted influenza polymerase. *Cell Rep.* (2018) 24:2581–8. doi: 10.1016/j.celrep.2018.08.012

23. Domingues P, Hale BG. Functional insights into ANP32A-dependent influenza A virus polymerase host restriction. *Cell Reports.* (2017) 20:2538–46. doi: 10.1016/j.celrep.2017.08.061

24. Domingues P, Eletto D, Magnus C, Turkington HL, Schmutz S, Zagordi O, et al. Profiling host ANP32A splicing landscapes to predict influenza A virus polymerase adaptation. *Nat Commun.* (2019) 10:3396. doi: 10.1038/s41467-019-11388-2

25. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges.* (2017) 1:33–46. doi: 10.1002/gch2.1018

26. Mänz B, Schwemmle M, Brunotte L. Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. *J Virol.* (2013) 87:7200–9. doi: 10.1128/JVI.00980-13

27. Long JS, Mistry B, Haslam SM, Barclay WS. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol.* (2019) 17:67–81. doi: 10.1038/s41579-018-0115-z

28. Leguia M, Garcia-Glaessner A, Munoz-Saavedra B, Juarez D, Barrera P, Calvo-Mac C, et al. Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. *bioRxiv.* (2023) 4:2023–03. doi: 10.1101/2023.03.03.531008

29. Li YT, Su YC, Smith GJ. H5Nx viruses emerged during the suppression of H5N1 virus populations in poultry. *Microbiology Spectrum.* (2021) 9:e01309–21. doi: 10.1128/Spectrum.01309-21

30. Löndt BZ, Nunez A, Banks J, Nili H, Johnson LK, Alexander DJ. Pathogenesis of highly pathogenic avian influenza A/turkey/Turkey/1/2005 H5N1 in Pekin ducks (Anas platyrhynchos) infected experimentally. *Avian Pathology.* (2008) 37:619–27. doi: 10.1080/03079450802499126

31. Wasilenko JL, Arafa AM, Selim AA, Hassan MK, Aly MM, Ali A, et al. Pathogenicity of two Egyptian H5N1 highly pathogenic avian influenza viruses in domestic ducks. *Arch Virol.* (2011) 156:37–51. doi: 10.1007/s00705-010-0813-y

32. Rijks JM, Hesselink H, Lollinga P, Wesselman R, Prins P, Weesendorp E, et al. Highly pathogenic avian influenza A (H5N1) virus in wild red foxes, the Netherlands, 2021. *Emerg Infect Dis.* (2021) 27:2960. doi: 10.3201/eid2711.211281

33. Chin AW Li OT, Mok CK, Ng MK, Peiris M, Poon LL. Influenza A viruses with different amino acid residues at PB2-627 display distinct replication properties in vitro and in vivo: revealing the sequence plasticity of PB2-627 position. *Virology.* (2014) 468:545–55. doi: 10.1016/j.virol.2014.09.008

34. Taft AS, Ozawa M, Fitch A, Depasse JV, Halfmann PJ, Hill-Batorski L, et al. Identification of mammalian-adapting mutations in the polymerase complex of an avian H5N1 influenza virus. *Nat Commun.* (2015) 6:7491. doi: 10.1038/ncomms8491

35. Luk GS, Leung CY, Sia SF, Choy KT, Zhou J, Ho CC, et al. Transmission of H7N9 influenza viruses with a polymorphism at PB2 residue 627 in chickens and ferrets. *J Virol.* (2015) 89:9939–51. doi: 10.1128/JVI.01444-15

36. Bussey KA, Bousse TL, Desmet EA, Kim B, Takimoto T. PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. *J Virol.* (2010) 84:4395–406. doi: 10.1128/JVI.02642-09