1

Evolutionary features of a prolific subtype of avian influenza A virus in European waterfowl

3

4

Running Title: Long term evolutionary patters of avian influenza

6

Michelle Wille[1, *, #], Conny Tolf[1], Neus Latorre-Margalef[1], Ron A. M. Fouchier[2], Rebecca A. Halpin[3], David E. Wentworth[3], Jayna Ragwani[4], Oliver G. Pybus [4,5], Björn Olsen[6], and Jonas Waldenström[1, #]

1. Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, SE - 391 82 Kalmar, Sweden.

2. Department of Virology, Erasmus Medical Centre, Rotterdam, The Netherlands

3. J. Craig Venter Institute, Rockville, Maryland, USA

4. Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom

5. Department of Pathobiology and Population Sciences, The Royal Veterinary College, AL9 7TA London, United Kingdom

6. Section of Infectious Diseases, Department of Medical Sciences, Uppsala University, SE751 85, Uppsala, Sweden

* Present address: Sydney Institute for Infectious Diseases, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, New South Wales, Australia

Corresponding authors: Michelle Wille michelle.wille@syndey.edu.au; Jonas Waldenström Jonas.waldenstrom@lnu.se

29 **Abstract**

30

31 Avian influenza A virus (AIV) is ubiquitous in waterfowl, and detected annually at high prevalence

32 in waterfowl during the Northern Hemipshere autumn. Some AIV subtypes are globally common in

33 waterfowl, such as H3N8, H4N6, and H6N2, and are detected in the same populations at high

34 frequency, annually. In order to investigate genetic features associated to the long-term

35 maintenance of common subtypes in migratory ducks, we sequenced 248 H4 viruses isolated across

36 8 years (2002-2009) from Mallards (*Anas platyrhynchos*) sampled in southeast Sweden.

37 Phylogenetic analyses showed that both H4 and N6 sequences fell into in three distinct lineages,

38 structured by year of isolation. Specifically, across the eight years of the study, we observed lineage

39 replacement, whereby a different HA lineage circulated in the population each year. Analysis of

40 deduced amino acid sequences of the HA lineages illustrated key differences in regions of the

41 globular head of hemagglutinin that overlap with established antigentic sites in homologous

42 hemagglutinin H3, suggesting the possibility of antigenic differences among these HA lineages.

43 Beyond HA, lineage replacement was common to all segments, such that novel genome

44 constellations were detected across years. A dominant genome constellation would rapidly amplify

45 in the duck population, followed by unlinking of gene segments as a result of reassortment within 2-

46 3 weeks following introduction. These data help reveal the evolutionary dynamics exhibited by AIV

47 on both annual and decadal scales in an important reservoir host.

48

49   **Introduction**

50

51   Influenza A viruses have a significant impact on human and animal health worldwide. These

52   viruses cause seasonal outbreaks and pandemics in humans, but also cause disease in domestic

53   animals such as horses, pigs and poultry, and wild animals including birds and seals (Anthony et al,

54   2012; Daly et al, 2011; Vincent et al, 2014; Webster et al, 1992; Zohari et al, 2014). Central to the

55   epidemiology of influenza A viruses are wild birds, particularly the *Anseriformes* (ducks, geese,

56   swans) and *Charadriiformes* (shorebirds, gulls) (Olsen et al, 2006; Webster et al, 1992).Within this

57   wild bird reservoir, influenza A viruses have been isolated from more than 105 species across 26

58   different families. Wild birds maintain a large diversity of avian influenza A viruses (AIV)

59   subtypes; 16 of the possible 18 haemagglutinin (HA) subtypes and 9 of the 11 neuraminidase (NA)

60   subtypes are found in wild birds (Olsen et al, 2006; Olson et al, 2014). Of importance in the context

61   of emerging and re-emerging disease is that this large diversity of AIV in waterfowl frequently

62   spillover into poultry (Verhagen et al, 2017), which may facilitate the introduction of AIVs into

63   future human pandemic viruses (Smith et al, 2009; Wille & Holmes, 2020).

64

65   In the Northern Hemipshere, waterfowl have high AIV prevalence in the autumn, linked to the

66   congregation of migrating birds including large numbers of immunologically-naïve juveniles

67   (Latorre-Margalef et al, 2014; van Dijk et al, 2014). A large diversity of HA subtypes are

68   maintained in waterfowl, however certain HA-NA subtype combinations, such as H3N8, H4N6,

69   H6N2, are over-represented at waterfowl surveillance sites and may comprise the majority of

70   viruses detected and isolated. Conversely, a number of HA-NA subtype combinations are

71   uncommon or absent from these study sites (Latorre-Margalef et al, 2014; Munster & Fouchier,

72   2009; Wilcox et al, 2011; Wille et al, 2018). The consistent detection of diverse subtypes, combined

73   with low rates of evolutionary change of amino acid sequences led early studies to postulate that

74   avian AIV was in evolutionary stasis (Hatchette et al, 2004; Sharp et al, 1997; Webster et al, 1992).

75   However, high rates of molecular evolution (~10^{-3} substitutions per site per year) and inference of

76   positive selection (dN/dS) in AIV genomes have firmly refuted this hypothesis (Chen & Holmes,

77   2006).

78

79   In the last decade, a number of interconnected hypotheses of AIV evolution have been put forward

80   based on observed AIV genetic structure. Central to these hypotheses are the evolutionary features

81   of antigenic drift and antigenic shift. Antigenic drift is the fixation, by natural selection, of

82   mutations in the HA and NA that enable the virus to evade the host immune response. These

83   mutations arise through an error prone RNA-dependant RNA polymerase (Chen & Holmes, 2006;

84   Gething et al, 1980). Antigenic shift, or process of reassortment following coinfection allows for the

85   generation of novel genome constellations with altered antigenic properties (Lowen, 2017; Steel &

86   Lowen, 2014). Together, these processes result in the capacity for rapid genetic and antigenic

87   change, and while the same subtypes may appear annually in a population, these subtypes consist of

88   multiple genetic lineages, which accumulate mutations, resulting in high levels of genetic diversity

89   in the viral population over time.

90

91   In this study, we aimed to reveal the genetic and evolutionary features that allow for the long-term

92   maintenance of AIV subtypes in waterfowl popuations. We collected and sequenced 248 H4

93   isolates across 8 years from migratory Mallards (*Anas platyrhynchos*) at a stopover site in SE

94   Sweden. H4 is not only the most common AIV in European ducks (Latorre-Margalef et al, 2014;

95   Munster et al, 2007), but also in Asian wild and domestic ducks (Cheng et al, 2010; Kang et al,

96   2013; Wisedchanwet et al, 2011; Zhang et al, 2012). We used Mallards as a model, as they are

97   considered to be an important reservoir for AIV diversity and have formed the basis of the current

98   understanding of spatial and temporal avian AIV dynamics in nature (Latorre-Margalef et al, 2014;

99   Olsen et al, 2006). Specifically, from these data, we aimed to reveal the role of different genetic

100 lineages, examine how this genetic variation may relate to antigenic differences, and address the

101 role of reassortment in both within-year and among-year maintenance of this virus subtype. We

102 suggest that a complex combination of natural selection, genetic hitchhiking, and reassortment is

103 responsible for the evolutionary patterns of AIV observed in wild birds.

104

105 **Methods**

106 *Ethics Statement*

107 All trapping and handling of mallards was done in accordance with regulations provided by the

108 Swedish Board of Agriculture under permits from the Linköping Animal Research Ethics Board

109 (permit numbers 8-06, 34-06, 80-07, 111-11, and 112-11).

110

111 *Study Site and Virus Collection*

112 Wild Mallards were captured as part of an ongoing long-term AIV surveillance scheme (Latorre-

113 Margalef et al, 2014) at Ottenby Bird Observatory, Sweden (56° 12'N, 16° 24'E). Details on

114 trapping, AIV surveillance and virus isolation have been published previously (Latorre-Margalef et

115 al, 2016; Latorre-Margalef et al, 2014). Briefly, for each captured duck, either a cloacal swab or a

116 faecal sample was collected and placed in virus transport medium (VTM). All samples were stored

117 in a -80 °C freezer within 2-6 hours of collection. Viral RNA was extracted from the VTM samples,

118 and assayed for a short fragment of the matrix gene using reverse-transcription real-time PCR (rRT-

4

119   PCR) (Spackman et al, 2002). Samples positive for AIV were inoculated into 10-12 day old specific

120   pathogen free (SPF) embroynated chicken eggs by the allantoic route. HA subtypes of positive

121   samples were determined using hemagglutination inhibition, and NA subtypes were determined by

122   sequencing (Latorre-Margalef et al, 2016; Latorre-Margalef et al, 2014).

123

124   *Sequence Dataset*

125   Full genomes were sequenced as part of the Influenza Genome Project

126   (http://gcid.jcvi.org/projects/gsc/influenza/index.php), an initiative by the National Institute of

127   Allergies and Infectious Diseases (NIAID) as previously described (Wille et al, 2018). Briefly, AIV

128   RNA was extracted and the entire genome was amplified using a multi-segment RT-PCR strategy

129   (Zhou et al, 2009; Zhou & Wentworth, 2012). The amplicons were sequenced using the Ion Torrent

130   PGM (Thermo Fisher Scientific, Waltham, MA, USA) and/or the Illumina MiSeq v2 (Illumina,

131   Inc., San Diego, CA, USA) instruments. When sequencing data from both platforms was available,

132   the data were merged and assembled together; the resulting consensus sequences were supported by

133   reads from both technologies.

134   We attemped sequencing 289 viruses, and 248 were successfully sequenced. Viruses that were not

135   successfully sequenced did not pass QC at the sequencing facility, and we did not attempt

136   resubmission of samples or resequencing.

137

138   All sequences generated in this study have been deposited in GenBank (accession numbers

139   CY164135-CY166103).

140

141   *Phylogenetic Analysis*

142   Sequences generated in this study, in addition to Eurasian reference sequences mined from the

143   Influenza Research Database (http://www.fludb.org/) were aligned using MAFFT (Katoh et al,

144   2009). Maximum likelihood trees were used to explore the temporal signal and clock-like behaviour

145   of each data set by performing linear regressions of root-to- tip distances against year of sampling,

146   using TempEst (Rambaut et al, 2016). Using BEAST v10.4 or v1.8, time-stamped data were

147   analysed under the uncorrelated lognormal relaxed molecular clock (Li & Drummond, 2012), and

148   the SRD06 codon-structured nucleotide substitution model (Shapiro et al, 2006) and the Bayesian

149   skyline coalescent tree prior. Three independent analyses of 100 million states were performed,

150   which were then combined in LogCombiner v1.8 following the removal of a burnin of 10 per cent.

151   Convergence was assessed using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum

152   credibility phylogenies were generated using TreeAnnotator v1.8 and visualized in FigTree v1.4.

153

154 The N2 phylogeny was estimated using Mr Bayes 3.2.1 (Ronquist et al, 2012). All trees were run

155 until convergance, which was assessed using Tracer v1.4. The tree was visualised using FigTree

156 v1.4.

157

158 Phylogenetic lineages were identified through a combination of pairwise identity plots (summarized

159 in Fig S1) and phylogenetic tree shape. From these analyses, 95% sequencing identity was

160 established as a cut-off for genetic lineage delineation, similar to previous assessments (Huang et al,

161 2014; Reeves et al, 2011; Wille et al, 2013). Lineage proportion plots were plotted using *ggplot2*

162 library in R v. 3.5.1 integrated into RStudio v. 1.0.143 .

163

164 Discrete trait analysis was performed using the asymmetric trait evolution model, and social

165 networks were inferred with Bayesian Stochastic Search Variable Selection (BSSVS) (Lemey et al,

166 2009). Connectivity among locations was determined using Bayes Factor analysis as implemented

167 in SpreaD3 (Bielejec et al, 2016). We considered Bayes Factors of greater 10 to be strong support,

168 and greater than 100 to be decisive support (Hill et al, 2021; Jeffreys, 1961).

169

170 Initial computations were performed on resources provided by SNIC through Uppsala

171 Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project

172 b2013122.

173

174 *Comparison of H4 HA1 peptide sequences of Swedish H4N6 virus*

175 In order to evaluate possible antigenic properties governing between different lineages of H4, HA

176 amino acid sequences of each lineage were aligned using MAFFT, and consensus sequence for each

177 lineage determined using Geneious Prime 2020.2.4. Consensus sequences from different years were

178 aligned to identify substitutions that had been introduced between years. In these alignments,

179 sequence regions corresponding to established antigenic sites (A to E) in the closely related and

180 well studied H3 HA1 was indicated as a reference (Broecker et al, 2018). These substitutions were

181 also visualized in the HA monomer derived from the crystal structure of an AIV H4 HA (PDB ID:

182 5XL1), using Geneious Prime 2020.2.4.

183

184 **Results**

185 *H4 appears annually, at high frequency, in migratory Mallards*

186 As part of a long-term AIV surveillance scheme, 18,643 cloacal/fecal samples from migratory

187 Mallards were collected between 2002–2009 at Ottenby Bird Observatory in southeast Sweden, as

6

188   outlined in (Latorre-Margalef et al, 2014). From these samples, 289 H4 viruses were successfully

189   isolated, representing 27% of all viruses isolated from Mallards in the study period (Fig 1).

190

191   For 2005–2009, the years for which we have the most data, we find "outbreak-like" patterns of H4

192   in the population, wherein during October or November, this subtype comprised > 50% of all

193   isolated viruses (Fig 1). In 2002 and 2003, years in which we have less data, we find a bimodal H4

194   prevalence peak, with high prevalence in September and November, respectively. No prevalence

195   pattern could be observed in 2004 due to limited number of isolates (Fig 1). Moreover, for H4

196   viruses isolated from 2002 to 2009, all NA subtypes except N4 and N7 were detected. The majority

197   (79%), however, were H4N6 viruses (Fig S2A).
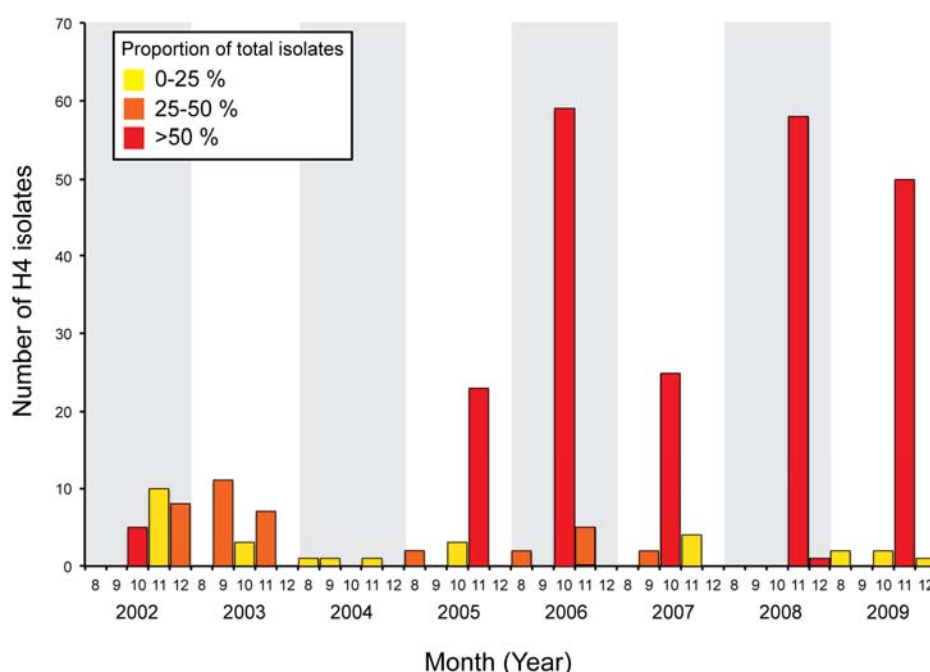
198

199

200   **Figure 1.** H4 AIVs are isolated at high frequency and constitute a substantial proportion of viruses

201   from Mallards sampled during the autumn migration at Ottenby Bird Observatory.

202

203   *Long term phylodynamics of H4N6*

204   A total of 246 H4 viruses isolated from 2002–2009, were sucessfully sequenced. Reconstruction of

205   the H4 phylogenly of Eurasian sequences revealed two main lineages: one lineage comprising

206   sequences from Asia and Australia (1980–present) (Fig 2A, collapsed node) and a second lineage

207   dominated by European sequences, most of which were generated in this study, in addition to a

208   smaller lineage of Asian sequences (Fig 2A). The estimated time of the most recent common

209   ancestor (tMRCA) of all Eurasian sequences, including both the Asian/Australian and European

210   lineages, is 1926–1957 (95% higest posterior density [HPD], mean 1946).

211   Given that two-thirds of H4 sequences from Eurasia between 2000–2009 are sequences generated in

212   this study, the European lineage of the H4 phylogenetic tree is largely defined by the sequences

213   from our study site. In addition to these, there are a smaller number of sequences from the

214   Netherlands and Czech Republic in GenBank. All H4 sequences from Sweden are within 95%

215   sequence identity (Fig S1) and the European lineages comprising the sequences generated in this

216   study diverged at 1997.7 (tMRCA, 95% HPD 1994–1999) (Fig 2A). We find the sequences

217   generated in this study to fall into 3 lineages with the exception of four sequences (including two

218   sequences from 2010) that group closely with Asian sequences nested within this European

219   lineage. We refer to those lineages detected in Swedish Mallards as Lineage 1–3 for clarity only

220   (Fig 2A).

221   There are three general observations regarding the distribution of sequences across the three

222   lineages. First, sequences from 2002 and 2003 fall into the same lineage (Lineage 2). Second,

223   unlike all other years in this dataset, sequences from 2005 fall into two disctinct lineages with equal

224   proportion: the lineage comprising sequences from 2006 (Lineage 3) and that of 2007 (Lineage 1).

225   And finally, starting in 2006 (for which we have the most data), sequences from within each year

226   fall into a single lineage with ~99% nucleotide similarity, and that these lineages are distinct from

227   the dominating lineage in the prior and subsequent year. Indeed, the lineage containing sequences in

228   2006 and 2009 (Lineage 3), do not contain more than 3 sequences from intermediate years. Taken

229   together, we find that lineage replacement in H4 is high between years (Fig 2B).

230

231   To attempt to better elucidate the source of annual viral introduction into Sweden, we undertook a

232   phylogeographic analysis. Based on the available sequences from 1980-2010, there is strong

233   evidence for virus movement between Sweden and other European locations (Bayes factor 88067.9,

234   posterior probability 1), between Sweden and western Asia (here western Russia [Moscow] and

235   Republic of Georgia) (Bayes factor 264, posterior probability 0.987), and between Sweden and

236   eastern parts of Asia (Bayes factor 24.7, posterior probability 0.88). In this analysis, virus

237   movement outside Sweden was poorly supported (Bayes factor <10 and posterior probability <0.5),

238   most likely because of the low number of sequences from other regions (Fig 2C). Overall, due to

239   poor sequence coverage from Eurasian locations outside of Sweden it is challenging to infer the role

240   of different geographic locations in the introduction of viruses into Swedish-caught Mallards.
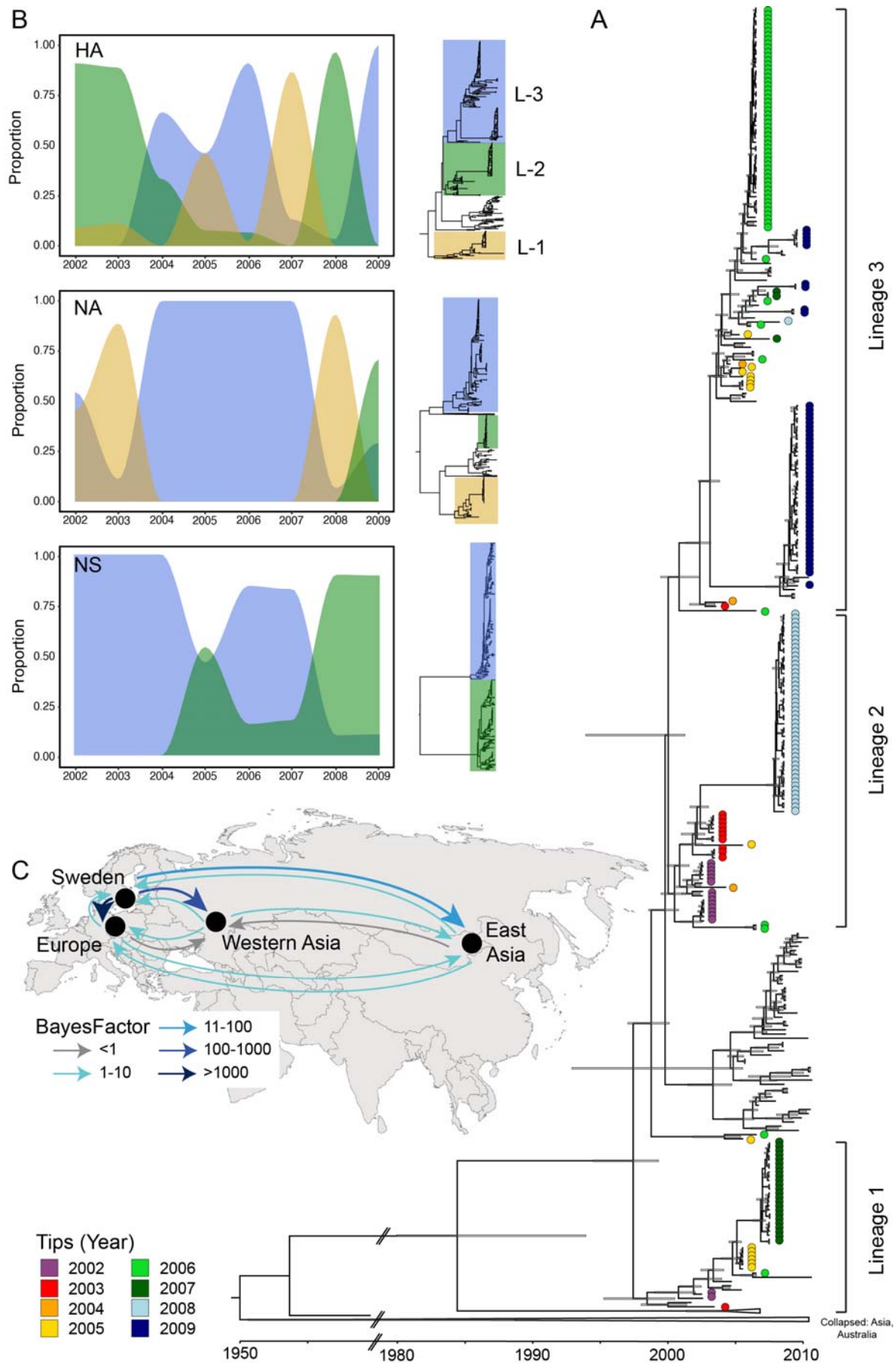
241

242

243

244

8

245

9

Figure 2: Features of long-term evolution of H4. (A) Time-scaled phylogeny of the Eurasian lineage of H4. Phylogeny includes all H4 sequences in GenBank from 1950-2010. Sequences generated in this study are denoted by a filled circle, with different colours referring the year of sample collection. Branch nodes correspond to the 95% HPD of node height. (B) Proportion of lineages of each H4, N6, and NS detected in Sweden between 2002 and 2010. For each segment the phylogeny with highlighted lineages is shown. For NS, the lineages correspond to the A and B alleles. In both (A) and (B), lineages circulating in Sweden are refered to as Lineage (L) 1-3 for clarity only. (C) Viral diffusion between geographic locations. Sequences included in Europe include Netherlands, Czech Republic, and England; Western Asia includes western Russia and the Republic of Georgia; Eastern Asia includes China, Japan, Korea, Mongolia, Russian Siberia. Arrow colour refers to strength of support for inclusion in a phylogeograhic model (*i.e.* Bayes Factors) where values greater than 10 represent strong support, whereas values greater than 100 are considered decisive support.

Similar to the H4 tree, sequences from this study play an important role in defining the shape of the Eurasian N6 phylogenetic tree (Fig S2). The N6 lineages are more distantly related, as compared to H4 lineages, where the three distinct N6 lineages detected in Ottenby are ~93% identical and postulated to have evolved from their MRCA ranging from 1977–1987 (95% HPD). The second most common NA subtype identified was N2, and as with N6, Swedish N2 sequences were similar to those circulating in Europe, most often Sweden, during the same time period. N2 sequences from 2005 and 2008 formed discrete lineages sharing 99% sequence identity, whereas sequence variation was greater in 2002 and 2007 (Fig S3). In addition to N6 and N2, the dataset included N1 (n=1), N3 (n=3), N5 (n=1), and N9 (n=3). In all instances these sequences were most closely related to other European AIV sequences, and the top 10 Blast matches often included other sequences from Sweden (Table S1).

Globally, non-structural protein (NS) sequences fall into two major lineages, termed allele A (including viruses with broad host range) and allele B (viruses specific to birds). NS sequences of viruses sequenced in this study fell into both allele A and B, and were most similar to NS sequences of AIV from Eurasian wild birds (Fig S4). Startgin in 2015, both alleles were found in the population each year, although with differing proportions. The shapes of the phylogenetic trees for the "internal" segments (PB2, PB1, PA, NP, M) were similar to those previously characterized (Fig S5). Within the "internal" segments, we found a pattern similar to that of HA and NA, wherein sequences from 2006-2009 fell into discrete year-specific lineages.

10

281

282   *Structural variation in HA protein sequences of Swedish H4N6 virus*

283   To evaluate putative antigenic differences in amino acid consensus sequences of these lineages,

284   genetic changes were evaluated by aligning H4 HA sequences across years and by mapping

285   identified substitutions onto the crystal structure of H4 (Fig 3). In line with the nucleotide

286   phylogeny, the phylogeny of HA1 amino acid consensus sequences from 2006–2009 exhibit

287   temporally discrete lineages. Across years, amino acid sequence variation predominately occured in

288   the HA1, and more specifically in, or in close proximity to, regions corresponding to antigenic sites

289   identified in the H3 HA1 (Broecker et al, 2018). While antigenic sites for H4 have not been

290   mapped, H3 and H4 belong to the same lineage within group 2 HA subtypes (Latorre-Margalef et

291   al, 2013), and we therefore postulate that these sites should be rather conserved across these two

292   subtypes.

293

294   When assessing the amino acid changes between consecutive years, we identified between 6–9

295   amino acid substitutions in the HA1 region. Specifically, between 2006–2007 (Lineage 3-> Lineage

296   1, Fig 3A) we found seven substitutions, 2007–2008 (Lineage 1-> Lineage 2, Fig 3B) nine

297   substitutions and 2008–2009 (Lineage 2-> Lineage 3, Fig 3C) six substitutions. When comparing

298   the sequences from 2006 compared to 2007, amino acid substitutions were located around the top of

299   of the globular head of HA1, close to secondary structures forming the receptor binding site,

300   including the 130 loop, the 190 helix and the 220 loop, which partly overlap with H3 antigenic site

301   A, B and D, respectively (Fig 3A). The majority of these substitutions are functionally favourable

302   (Table S2), with the exception of K186T and T208I. In the transition from 2007 to 2008, four out of

303   nine substitutions are found in loop structures closer to the stem region of HA1, including the 260

304   loop. These include Asn88Ser, Ser259Asn in a region corresponding to antigenic site E in H3 as

305   well as I270T, K283R and I295V adjacent to site C (Fig 3B). All substitutions except I208T and

306   I270T, are functionally favourable. It is noteworthy that four substitutions in position 124, 204, 208,

307   259 and 270 are reversions to amino acid residues expressed in the 2006 consensus sequence.

308   Finally, in the transition between 2008 and 2009, the only unique substitution is Q42H located in

309   what is consistent with antigenic site C in H3 HA1. This substitution is neutral from a functional

310   standpoint. In addition, five of six substitutions found in the consensus sequence from 2009

311   compared to that from 2008 represent reversions to amio acids expressed in the 2007 sequence.

312   These include amino acid at position 88, 166, 259, 283 and 295 (Fig 3C). For clarity, antigenic sites

313   A-E correspond to those in the HA1 peptide of H3 as there is no characterisation of H4 available
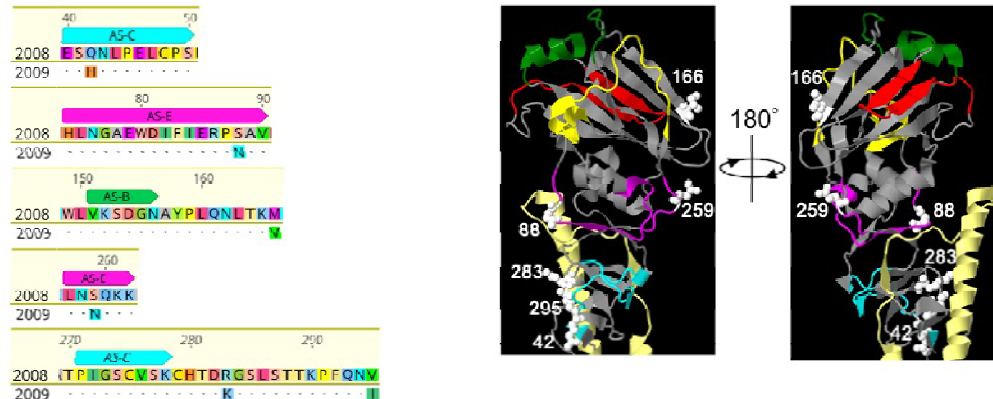
314   (Broecker et al, 2018).

315

11
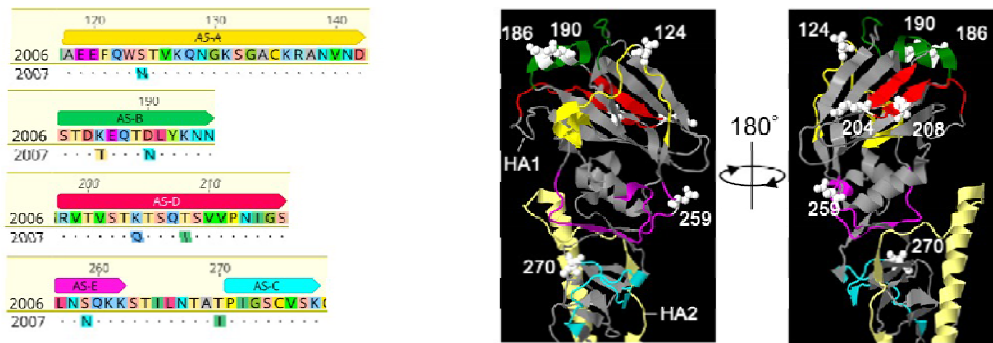
Figure 3: Amio acid substitutions across years between distinct H4 HA1 lineages of Swedish H4N6 strains. Specfically, amino acid substitions between sequences from (A) 2006-2007, (B) 2007-2008, and (C) 2008-2009. Amino acid sequence alignments showing substitutions between years indicated in, or in close proximity to, regions corresponding to established antigenic site A-E (AS-A – AS-E) in the HA1 peptide of H3 (Broecker et al 2018). In sequence comparisons, consensus sequences from 2006, 2007 and 2008 serve as reference and substitutions in sequences from the following year is indicated by single letter abbreviation of amino acids whereas identical residues are represented by dots. In B, the dotted vertical line indicates a truncation for clarity only. Substitutions visualized by white colour in the crystal structure of H4 HA1. Antigenic site regions

326   are indicated and colour coordinated with corresponding regions in alignments. HA1 (grey) and

327   HA2 (light yellow) peptide monomers are indicated the left structure of the 2006-2007 panel

328

329

330   *Lineage replacement and no perpetuation of genome constellations across years*

331   Analysis of temporal changes of H4N6 genome constellations revealed frequent lineage

332   replacement across years (Fig 2B). Plots of lineage proportions of the glycoproteins and the NS

333   alleles illustrate that, while lineage replacement is frequent for HA, concomitant replacement

334   among multiple segments occurred less frequently (Fig 2B). For example, between 2003–2004,

335   and 2008–2009, there was a replacement of both HA and NA lineages, and between 2007–2008

336   there was a simultaneous replacement of HA, NA and NS (Fig 2B). The replacement between 2007

337   and 2008 was not limited to only segments with established antigenic properites (HA and NA), but

338   rather new lineages were introduced for all segments in addition to the replacement from NS allele

339   A to allele B (Fig 2B, Fig S2, S4-S5). This global replacement also resulted in a reintroduction of

340   lineages in 2008 that were circulating among sampled birds six years earlier, in 2003.

341

342   Bringing all lineage data together, we may assess whether genome constellations remain in the

343   population across multiple years (Fig 4, Fig S6). Importantly, despite not detecting annual global

344   replacements, there was sufficient lineage replacement across the different segments for novel

345   constellations to proliferate each year. That is, genome constellations identified in one year, were

346   not found again in the population the following year (Fig 4, Fig S6). Using 2005–2006 as an

347   example: no constellations identified in 2005 were detected again in 2006, despite the fact that

348   some H4 sequences from 2005 were found in the same phylogenetic lineage as sequences from

349   2006. In this case, the lineages of four gene segments were different between the dominant genome

350   constellation in 2005 compared to the dominant constellation from 2006 (Fig 4). This example is

351   not an outlier, rather is consistent with findings between all years of this dataset (Fig S6)

352

353   *Rapid unlinking of genome constellations within year*

354   Within each year, particularly between 2006–2009, we identified a number of different genome

355   constellations. Despite the diversity, more than 50% of viruses characterized within each year

356   comprised a single, dominant constellation (Table S3). Data from 2006–2009 further indicate that

357   the first genome constellations detected each year were not the most frequently-observed

358   constellation overall that year. Following the introduction of the dominant constellation into the

359   population, it was consistently detected and isolated from Mallard samples for 1-2 weeks, followed

360   by a period (1–2 weeks) wherein this constellation rapidly unlinked and a diversity of new genome

361   constellations were detected in the population. These novel and diverse constellations most likely

362   contained gene lineages originating from other cocirculating viruses with different HA-NA subtype

363   combinations and formed through the process of reassortment (Fig 4, Fig S6-S7).

364

365   Using 2006 as an example, the first virus constellations were detected in week 23 followed by a

366   detection in 32. These constellations were not detected again, and were replaced by a frequently-

367   detected constellation that accounted for 73% of isolates (Fig 4A) in week 41. This constellation

368   initially appeared at low numbers in week 41 (n=2 sequenced isolates), however it was the only H4

369   genome constellation detected in week 42 (n=18 sequenced H4 isolates) and remained at high

370   proportion relative to other constellations in week 43 (n=13 sequenced H4 isolates). Seven other

371   genome constellations were also identified in week 43, and one additional novel constellation was

372   detected in week 44 (Fig 4). Data from 2005 revealed a genetic variaton that was different

373   compared to other years, wherein isolates comprised high levels of constellation diversity,

374   presumably due to the presence of two different HA lineages and NS alleles (Table 1, Fig S6). In

375   2006, we detected 15 "mixed" viruses – these are viruses containing more than one sequence for

376   any one segment. These mixed infections are likely reflective of co-infection in the Mallard host,

377   but as we sequenced viruses propagated in eggs we cannot make strong inferences about these

378   viruses (Fig 4, Table S4). These mixed viruses spanned the periods of highestprevalence and

379   HA/NA diversity during the autumn.The year 2006 accounted for a large proportion of mixed

380   viruses in our dataset, and in this year there was substantial cocirculation of other viral subtypes in

381   the Ottenby duck population (*e.g.* H1N1, H3N6, H6N2; Latorre-Margalef *et al* 2014) (Fig 4, Table
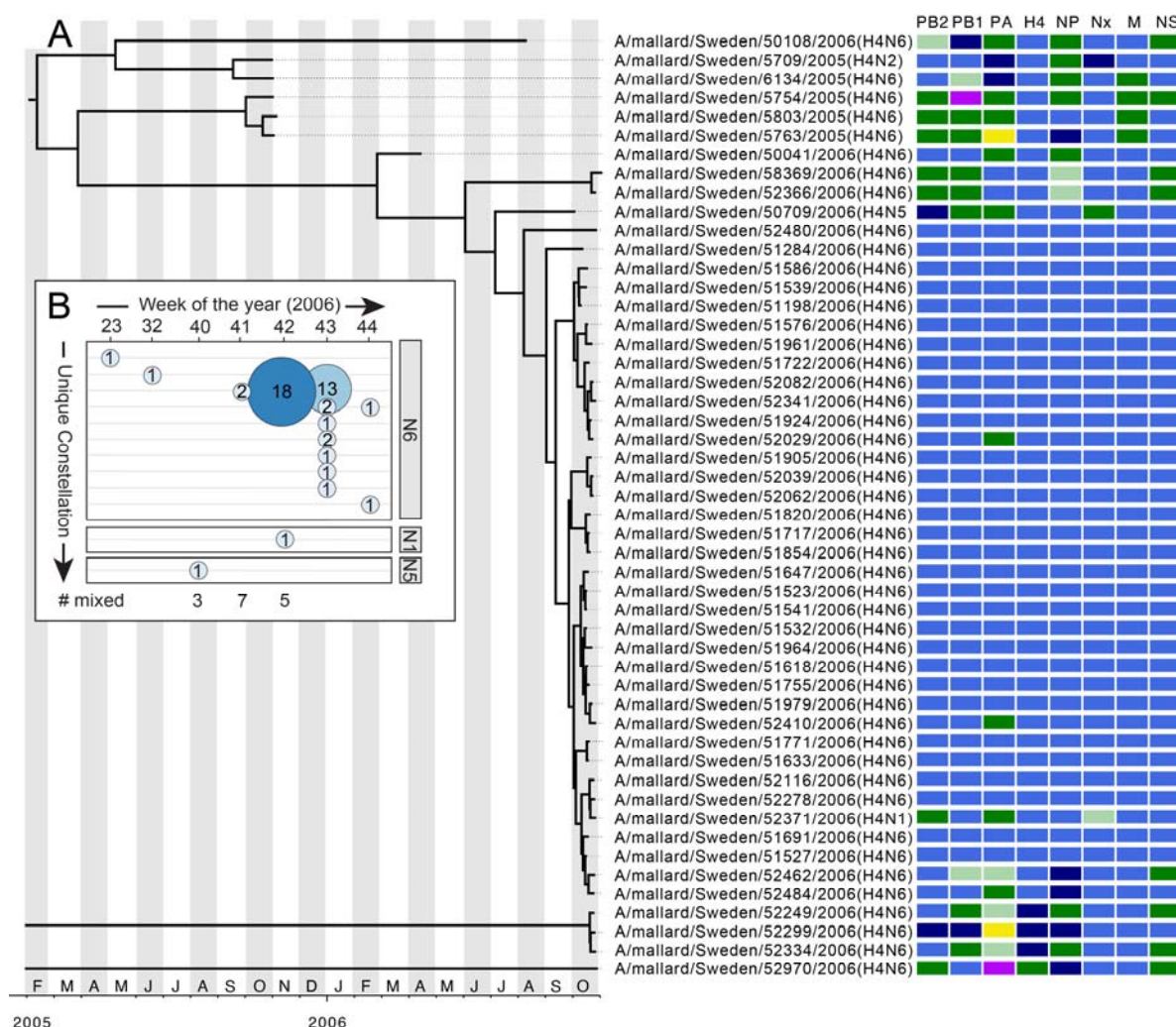
382   S4)

383

384

385

14

Figure 4: Within-year reassortment. (A) Time-structured phylogeny of H4 sequences from 2005 and 2006 comprising Lineage 1. Scale bar indicates time, with year and month indicated. Four H4 sequences from viruses detected in 2006 did not fall into this lineage, and therefore tMRCA is not shown in order to preserve resolution. Each tile column represents a segment, ordered by size: PB2, PB1, PA, H4, NP, Nx, M and NS, and each row comprises the genome constellation of a virus. Different colours pertain to different viral lineages. Genome constellations for all viruses is presented in Fig S6. (B) Genome constellation for each week of 2006. Week number is presented on the X axis and each unique constellation is on the Y axis. Bubble size and colour refer to the number of times each genome constellation was detected. Different NA subtypes are in separate panels. At the bottom the number of mixed H4 infections detected in each week is indicated. Change of genome constellations over time for all years is presented in Fig S7.

15

399 **Discussion**

400 *How are subtypes maintained in waterfowl on a decadal scale ?*

401 Low pathogenic AIV H4N6 is one of the most abundant virus subtypes isolated from waterfowl in

402 Europe. Indeed, these viruses have been circulating at high frequency in migratory Mallards at our

403 study site, in southeast Sweden, since the start of AIV surveillance in 2002, and continue to

404 circulate at high frequency today (Latorre-Margalef et al, 2014; Munster et al, 2007; Wallensten et

405 al, 2007). An important question, however, is how this subtype is maintained in a Mallard

406 population across many years, given that high rates of infection should confer immunity in the

407 population. Through the analysis of a dataset comprising viruses from the same species and

408 location, but across multiple years, we were able to reveal key genetic features likely playing a key

409 role in this phenomenon.

410

411 Unlike human influenza A virus, AIV HA segments do not exhibit a classic "ladder-shaped"

412 phylogeny (Fitch et al, 1997). These ladder-shaped phylogenies are the result of strong positive

413 selection on the HA, leading to antigenic drift, or the fixation of mutations in the HA (and NA) that

414 enables the virus to evade the immune system (Koel et al, 2013; Vijaykrishna et al, 2015; Wille &

415 Holmes, 2020). Rather, multiple lineages co-circulate in wild bird populations (Chen & Holmes,

416 2006, 2010). Herein, we detect 3 lineages of H4 in the mallard population. These lineages are

417 genetically similar (within 95% nucleotide similarty), although none are present in all years. We

418 find a phenomenon whereby there is a different HA lineage circulating each year, and within years

419 2006-2009, one lineage comprises ~90% of all isolates sequenced. Indeed, this is reflective of

420 antagonistic co-evolution in a sympatric model of circulating AIV, wherein host immune pressure

421 will drive phylogenetic branching of viruses into discrete antigenic lineages, and eventually,

422 subtypes, with limited overlap in antigenic space (Recker et al, 2007). Alternatively, or in concert,

423 these patterns could rather be driven by migration mediated metapopulation dynamics, which

424 frequently re-introduces lineages to the site.

425

426 Our data is consistent with the hypothesis proposed by Chen & Holmes (2010), whereby genetic

427 structure of AIV is shaped by a combination of occasional selective sweeps in the HA and NA

428 segments, coupled with transient genetic linkage to the internal gene segments. Indeed, we

429 identified annual replacement of HA, although simultaneous replacement of HA, NA and NS were

430 less frequent. While NS is less studied as a target for cellular and humoral immune responses, there

431 are several reports on the inhibitory effect of NS1 on the innate immunity and fitness differences of

432 the different alleles (Adams et al, 2013; Gack et al, 2009; Nogales et al, 2018; Rajsbaum et al,

433 2012; Yodsheewan et al, 2013). In 2008, we detected lineage replacement in all segments, which

16

434   coincided with the shift of NS A to B allele in the virus population. This event does support the

435   transient linkage hypothesis, where a change from NS A to NS B may have conferred some type of

436   fitness advantage, promoting the genome constellation with the new NS allele to become fixed and

437   resulting in a local sweep. Despite lineage replacement, previously circulating lineages do not go

438   extinct, but rather have a marked decrease in frequency in the system which may be explained by

439   negative frequency dependent selection (Gandon & Michalakis, 2002). That is, the fitness of a

440   phenotype or genotype decreases as it becomes more common, likely due to increased immunity

441   against these lineages in the population. In addition to competitive interactions between lineages,

442   the migratory behaviour of avian hosts and their distinct geographic distribution may contribute to

443   lineage maintenance in different host populations and to the patterns of replacement between

444   lineages in one site. Given these lineages do not go extinct, it is likely they circulate in other

445   dabbling duck populations, including Mallard populations, in other geographic locations, including

446   Asia, or in other less characterized host species, such as diving ducks, geese, or even shorebirds,

447   which are not as intensively monitored in surveillance programs.

448

449   Unlike in humans, reassortment of AIV (both inter and intra-subtypic reassortment) is prolific in

450   avian systems (Dugan et al, 2008; Macken et al, 2006; Steel & Lowen, 2014; Wille et al, 2013), and

451   is likely critical in both the long-term (decadale scale) and short-term (within year) arms race of

452   AIV against Mallard immunity. Specifically, it is through reassortment that novel HA or NA

453   lineages are introduced and also the mechanism by which entire genome constellations are formed

454   within the population between years. Indeed, it is the generation of novel genome constellations that

455   drives the emergence of pandemic influenza viruses in human populations (Smith et al, 2009; Wille

456   & Holmes, 2020). In our study, the role of reassortment is most evident when assessing H4N6

457   genome constellations within a year. Specifically, during periods of high infection burden of H4

458   viruses in the mallard population, we may expect that the dominant genome constellation has high

459   fitness allowing for rapid proliferation in the mallard population. However, within ~2 weeks of

460   entering/detection in the mallard population the initially dominant H4 constellation rapidly unlinks

461   through reassortment, and is replaced by a variety of auxillary genome constellations. We

462   hypothesize that this is driven by increasing immunity in the host population, resulting in a decrease

463   in fitness for this common constellation. While most antibodies are directed at the HA, and to a

464   lesser extent the NA, during infection all influenza proteins are expressed in infected cells and can

465   potentially induce an antibody response (Krammer, 2019). However, some proteins are more

466   accessible than others, including the NP and M segments (Haaheim, 1977; Sukeno et al, 1979).

467   There is some evidence from monoclonal antibody isolation and from antigenic fingerprinting that

468   natural infection also induces antibodies against internal segment proteins, although the magnitude

17

469   and quality of these responses is not well defined (Krammer, 2019; Krejnusova et al, 2009;

470   Thathaisong et al, 2008; Yodsheewan et al, 2013). As such, it is likely that through the generation

471   of novel genome constellations, the H4 subtype may be maintained in the population beyond the

472   point at which there is an increase HA-directed immunity and a subsequent decrease in viral fitness.

473   Novel segments incorporated into novel genome constellations originate from other viruses co-

474   circulating in the mallard population (Wille et al, 2013). Indeed, in years with the highest number of

475   mixed infections, a number of other subtypes were co-circulating at relatively high frequency:

476   H1N1 in 2006 and H11N2 in 2008.

477

478   *Do genetic patterns reflect antigenic patterns?*

479   Both the phenomenon of lineage replacement and reassortment are important drivers in this system,

480   and would only be selected for if they provided a fitness advantage. One such advantage would be

481   escape from acquired population immunity. Following infection, mallards develop a immune

482   response that is homosubtypic, with potential for heterosubtypic effects (Latorre-Margalef et al,

483   2017; Latorre-Margalef et al, 2013). Although strength and duration of the immune response is

484   largely unknown in ducks, it is hypothesized that immune memory may be weak and short-lived

485   (Magor, 2011). However, repeated sampling of individual sentinel mallards illustrate maintainance

486   of anti-NP antibodies for months following AIV infection. Furthermore, these same mallards were

487   unlikely to be reinfected with the same subtypes in their second year (Tolf et al, 2013). Indeed, it is

488   hypothesized that this pattern of both homo- and heterosubtypic immunity drives the temporal

489   sucession of different HA lineages across the sampling season (Latorre-Margalef et al, 2014).

490

491   Based on the phenomenon of lineage replacement or succession of H4 between years, we

492   hypothesize that the different HA lineages observed must confer homosubtypic immune escape.

493   Molecular and structural comparisons of sequences in each of the three HA lineages found in the

494   mallard population had substitutions in regions of HA1, corresponding to antigentic sites in the

495   homologous H3 peptide (Broecker et al, 2018). Unfortunately, dedicated work on identifying

496   domains has not been completed for H4, however H3 and H4 comprise the H3 Clade of Group 2

497   viurses (Latorre-Margalef et al, 2013), such that we predict the antigenic sites of H3 should

498   correspond to H4 as well. We found that the 6 to 9 amino acid differences delineating the H4

499   lineages described in this study partly map to different regions in the crystal structure of H4 HA1,

500   where substitutions in the 2007 sequence were located at the top of the globular head and

501   substitutions in 2008 and 2009 sequences located closer to the stem region of HA1. This may

502   suggest that at a few changes in structural regions in putative antigenic sites is suffcent to evade

503   immunity. Indeed, from work on human influenza A virus H3, it is clear that only a few

18

504   substitutions are required for immune escape in humans, given they are located in or adjacent to the

505   receptor binding domain (Koel et al, 2013; Smith et al, 2004). In contrast to the HA of human

506   influenza A (*e.g.* (Koel et al, 2013), antigenic diversity of avian HAs has not been well explored,

507   with a few exceptions (*e.g.* (Hill et al, 2016; Koel et al, 2014; Verhagen et al, 2020). However, a

508   recent study investigating infection probability given previous exposure illustrated H3 vaccine

509   escape despite >95% sequence differences in ducks (Wille et al, 2016) In this study, ducks were

510   vaccinated with an inactivated H3 virus, isolated in 2010, and were challenged with naturally

511   circulating AIVs in 2013. Despite a raised immune response due to the vaccination, ducks were not

512   protected against H3 viruses and it was hypotheised that the lack of immunity was associated to

513   substitutions in immunogenic epitopes of more recent H3 strains (e.g. vaccine escape) (Wille et al,

514   2016) . Despite the parallels, without performing a characterization of H4 antigenic sites or through

515   antigenic cartography, we are unable to confirm whether the different lineages (L1-L3) of H4

516   confer antigenic differences, and/pr whether reassortment confers partial escape from population

517   immunity.

518

519   Understanding of the evolutionary processes of AIV is critical. In humans, the detailed interrogation

520   of seasonal influenza viruses are key to vaccine selection and detection of antiviral resistance (Koel

521   et al, 2013; Van Poelvoorde et al, 2020; Vijaykrishna et al, 2015; Wille & Holmes, 2020). On a

522   larger scale, it is through the process of reassortment that novel pandemic viruses are generated

523   (Gething et al, 1980; Smith et al, 2009). Within the avian reservoir, there is a key focus on HA

524   subtypes that cause significant morbidity and mortality in food production birds (eg H5, H7) or

525   have zoonotic potential (eg H5, H7, H9) (Chang et al, 2020; Chang et al, 2018; Cui et al, 2020;

526   Mahmoud et al, 2019). We generally have a limited understanding of the evolution of LPAIV in the

527   wild bird reservoir (Chen & Holmes, 2006, 2010; Dugan et al, 2008; Wille et al, 2013). But genetic

528   diversity is shaped by virus-related factors including antigenic drift and reassortment, as well as

529   immunological and ecological features of the hosts such as migratory behaviour (van Dijk et al,

530   2018). Low pathogenicity viruses circulating in wild birds play an important role by providing

531   genetic variation to all influenza A viruses. For example, all human pandemic influenza A viruses

532   have had least one gene of avian origin (Kawaoka et al, 1989; Lindstrom et al, 2004; Rabadan et al,

533   2006; Runstadler et al, 2013; Scholtissek et al, 1978; Taubenberger et al, 2005), and HPAIV emerge

534   following rapid evolution of LPAIV in poultry (Seekings et al, 2018). Thus, in order to understand

535   evolutionary factors governing AIV dynamics, including host ecology affects, long-term evolution,

536   and genetic structure, it is pivotal to characterize viruses circulating in the wild bird reservoir.

537

538

19

Data Availability

All sequences generated in this study have been deposited in GenBank (accession numbers CY164135-CY166103).

20

References

Adams, S, Xing, Z, Li, J, Mendoza, K, Perez, D, Reed, K, & Cardona, C. (2013). The effect of avian influenza virus NS1 allele on virus replication and innate gene expression in avian cells. *Mol Immunol, 56*(4), 358-368.

Anthony, SJ, St Leger, JA, Pugliares, K, Ip, HS, Chan, JM, Carpenter, ZW, . . . Lipkin, WI. (2012). Emergence of fatal avian influenza in New England harbor seals. *MBio, 3*(4), e00166-00112.

Bielejec, F, Baele, G, Vrancken, B, Suchard, MA, Rambaut, A, & Lemey, P. (2016). SpreaD3: Interactive visualization of spatiotemporal hstory and trait evolutionary processes. *Mol Biol Evol, 33*(8), 2167-2169.

Broecker, F, Liu, STH, Sun, W, Krammer, F, Simon, V, & Palese, P. (2018). Immunodominance of antigenic site B in the hemagglutinin of the current H3N2 Influenza virus in humans and mice. *J Virol, 92*(20), e01100-01118. doi: 01110.01128/JVI.01100-01118.

Chang, P, Sealy, JE, Sadeyen, JR, Bhat, S, Lukosaityte, D, Sun, Y, & Iqbal, M. (2020). Immune escape adaptive mutations in the H7N9 avian influenza hemagglutinin protein increase virus replication fitness and decrease pandemic potential. *J Virol, 94*(19), e00216-00220. doi: 00210.01128/JVI.00216-00220.

Chang, P, Sealy, JE, Sadeyen, JR, & Iqbal, M. (2018). Amino acid residue 217 in the hemagglutinin glycoprotein is a key mediator of avian influenza H7N9 virus antigenicity. *J Virol, 93*, e01627-01618. doi: 01610.01128/JVI.01627-01618.

Chen, R, & Holmes, EC. (2006). Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol, 23*, 2336-2341.

Chen, R, & Holmes, EC. (2010). Hitchhiking and the population genetic structure of avian influenza virus. *J Mol Evol, 70*, 98-105.

Cheng, MC, Lee, MS, Ho, YH, Chyi, WL, & Wang, CH. (2010). Avian influenza monitoring in migrating birds in Taiwan during 1998-2007. *Avian Dis, 54*(1), 109-114.

Cui, Y, Li, Y, Li, M, Zhao, L, Wang, D, Tian, J, . . . Chen, H. (2020). Evolution and extensive reassortment of H5 influenza viruses isolated from wild birds in China over the past decade. *Emerg Microbes Infect, 9*(1), 1793-1803.

Daly, JM, MacRae, S, Newton, JR, Wattrang, E, & Elton, DM. (2011). Equine influenza: a review of an unpredictable virus. *Vet J, 189*(1), 7-14.

Dugan, VG, Chen, R, Spiro, DJ, Sengamalay, N, Zaborsky, J, Ghedin, E, . . . Taubenberger, JK. (2008). The evolutionary genetics and emergence of avian influenza A viruses in wild birds. *PLoS Pathog, 4*, e1000076. doi: 1000010.1001371/journal/ppat/1000076.

Fitch, WM, Bush, RM, Bender, CA, & Cox, NJ. (1997). Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc Natl Acad Sci USA, 94*(15), 7712-7718.

Gack, MU, Albrecht, RA, Urano, T, Inn, KS, Huang, IC, Carnero, E, . . . Garcia-Sastre, A. (2009). Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe, 5*(5), 439-449.

Gandon, S, & Michalakis, Y. (2002). Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. *J Evol Biol, 15*, 451-462.

Gething, MJ, Bye, J, Skehel, JJ, & Wakefield, M. (1980). Cloning and DNA sequence of double-stranded copies of hemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. *Nature, 287*, 310-306.

Haaheim, R. (1977). Single-radial-complement-fixation: a new immunodiffusion technique. 2. Assay of the antibody response to the internal antigens (MP and NP) of influenza A virus in human sera after vaccination and infection. *Dev Biol Stand, 39*, 481-484.

614 Hatchette, TF, Walker, D, Johnson, C, Baker, A, Pryor, SP, & Webster, RG. (2004). Influenza A
615   viruses in feral Canadian ducks: extensive reassortment in nature. *J Gen Virol, 85*, 2327-
616   2337.
617 Hill, NJ, Smith, LM, Muzaffar, SB, Nagel, JL, Prosser, DJ, Sullivan, JD, . . . Takekawa, JY. (2021).
618   Co-habiting in a disease hotspot: Overlap between wild and domestic birds in Egypt impacts
619   transmission of highly pathogenic H5N1. *Virus Evol, veaa093*, doi:
620   https://doi.org/10.1093/ve/veaa1093.
621 Hill, SC, Manvell, RJ, Schulenburg, B, Shell, W, Wikramaratna, PS, Perrins, C, . . . Pybus, OG.
622   (2016). Antibody responses to avian influenza viruses in wild birds broaden with age. *Proc
623   Royal Soc B, 283*(1845), doi: 10.1098/rspb.2016.2159.
624 Huang, Y, Wille, M, Dobbin, A, Walzthoni, NM, Robertson, GJ, Ojkic, D, . . . Lang, AS. (2014).
625   Genetic structure of avian influenza viruses from ducks of the Atlantic flyway of North
626   America. *PLoS One, 9*(1), e86999. doi: 86910.81371/journal.pone.0086999.
627 Jeffreys, H. (1961). *The theory of probability*. Oxford, U. K.: Oxford University Press.
628 Kang, HM, Choi, JG, Kim, KI, Park, HY, Park, CK, & Lee, YJ. (2013). Genetic and antigenic
629   characteristics of H4 subtype avian influenza viruses in Korea and their pathogenicity in
630   quails, domestic ducks and mice. *J Gen Virol, 94*(Pt 1), 30-39.
631 Katoh, K, Asimenos, G, & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT.
632   *Methods Mol Biol, 537*, 39-64.
633 Kawaoka, Y, Krauss, S, & Webster, RG. (1989). Avian-to-human transmission of the PB1 gene of
634   influenza A viruses in the 1957 and 1968 pandemics. *J Virol, 63*(11), 4603-4608.
635 Koel, BF, Burke, DF, Bestebroer, TM, van der Vliet, S, Zondag, GC, Vervaet, G, . . . Smith, DJ.
636   (2013). Substitutions near the receptor binding site determine major antigenic change during
637   influenza virus evolution. *Science, 342*(6161), 976-979.
638 Koel, BF, van der Vliet, S, Burke, DF, Bestebroer, T, Bharoto, EE, Yasa, IWW, . . . Fouchier,
639   RAM. (2014). Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino
640   acid substitutions immediately adjacent to the receptor binding site. *mBio, 5*, doi:
641   https://doi.org/10.1128/mBio.01070-01014.
642 Krammer, F. (2019). The human antibody response to influenza A virus infection and vaccination.
643   *Nat Rev Immunol, 19*(6), 383-397.
644 Krejnusova, I, Gocnikova, H, Bystricka, M, Blaskovicova, H, Polakova, K, Yewdell, J, . . . Russ, G.
645   (2009). Antibodies to PB1-F2 protein are induced in response to influenza A virus infection.
646   *Arch Virol, 154*(10), 1599-1604.
647 Latorre-Margalef, N, Avril, A, Tolf, C, Olsen, B, & Waldenstrom, J. (2016). How does sampling
648   methodology influence molecular detection and isolation success in Influenza A virus field
649   studies? *Appl Environ Microbiol, 82*(4), 1147-1153.
650 Latorre-Margalef, N, Brown, JD, Fojtik, A, Poulson, RL, Carter, D, Franca, M, & Stallknecht, DE.
651   (2017). Competition between influenza A virus subtypes through heterosubtypic immunity
652   modulates re-infection and antibody dynamics in the mallard duck. *PLoS Pathog, 13*(6), doi:
653   10.1371/journal.ppat.1006419.
654 Latorre-Margalef, N, Grosbois, V, Wahlgren, J, Munster, VJ, Tolf, C, Fouchier, RAM, . . . Olsen,
655   B. (2013). Heterosubtypic immunity to influenza A virus infections in Mallards may explain
656   existence of multpile virus subtypes. *PLoS Pathog, 9(6)*, e1003443. doi:
657   1003410.1001371/journal.ppat.1003443.
658 Latorre-Margalef, N, Tolf, C, Grosbois, V, Avril, A, Bengtsson, D, Wille, M, . . . Waldenström, J.
659   (2014). Long-term variation in influenza A virus prevalence and subtype diversity in a
660   migratory Mallards in Northern Europe. *Proc Royal Soc B, 281*, doi:
661   10.1098/rspb.2014.0098
662 Lemey, P, Rambaut, A, Drummond, AJ, & Suchard, MA. (2009). Bayesian phylogeography finds
663   its roots. *PLoS Comput Biol, 5*(9), e1000520. doi: 1000510.1001371/journal.pcbi.1000520.
664 Li, WL, & Drummond, AJ. (2012). Model averaging and Bayes factor calculation of relaxed
665   molecular clocks in Bayesian phylogenetics. *Mol Biol Evol, 29*(2), 751-761.

22

666 Lindstrom, SE, Cox, NJ, & Klimov, AI. (2004). Genetic analysis of human H2N2 and early H3N2
667   influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment
668   events. *Virology, 328*, 101-119.
669 Lowen, AC. (2017). Constraints, drivers, and implications of Influenza A virus reassortment.
670   *Annual Review of Virology, Vol 4, 4*, 105-121.
671 Macken, CA, Webby, RJ, & Bruno, WJ. (2006). Genotype turnover by reassortment of replication
672   complex genes from avian influenza A virus. *J Gen Virol, 87*(Pt 10), 2803-2815.
673 Magor, KE. (2011). Immunoglobulin genetics and antibody responses to influenza in ducks. *Dev
674   Comp Immunol, 35*, 1008-1016.
675 Mahmoud, SH, Mostafa, A, El-Shesheny, R, Seddik, MZ, Khalafalla, G, Shehata, M, . . . Ali, MA.
676   (2019). Evolution of H5-type avian Influenza A virus towards mammalian tropism in Egypt,
677   2014 to 2015. *Pathogens, 8*(4), 224. doi: 210.3390/pathogens8040224.
678 Munster, V, & Fouchier, RAM. (2009). Avian influenza virus: of virus and bird ecology. *Vaccine,
679   27*, 6340-6344.
680 Munster, VJ, Baas, C, Lexmond, P, Waldenström, J, Wallensten, A, Fransson, T, . . . Fouchier,
681   RAM. (2007). Spatial, temporal, and species variation in prevalence of influenza A viruses
682   in wild migratory birds. *PLoS Pathog, 3*, e61. doi: 10.1371/journal.ppat.0030061.
683 Nogales, A, Martinez-Sobrido, L, Topham, DJ, & DeDiego, ML. (2018). Modulation of innate
684   immune responses by the Influenza A NS1 and PA-X proteins. *Viruses, 10*(12), 708. doi:
685   710.3390/v10120708.
686 Olsen, B, Munster, VJ, Wallensten, A, Waldenström, J, Osterhaus, ADME, & Fouchier, RAM.
687   (2006). Global patterns of influenza A virus in wild birds. *Science, 312*, 384-388.
688 Olson, SH, Parmley, J, Soos, C, Gilbert, M, Latorre-Margalef, N, Hall, JS, . . . Joly, D. (2014).
689   Sampling strategies and biodiversity of influenza A subtypes in wild birds. *PLoS ONE, 9*(3),
690   e90826. doi: 90810.91371/journal.pone.0090826.
691 Rabadan, R, Levine, AJ, & Robins, H. (2006). Comparison of avian and human influenza A viruses
692   reveals a mutational bias on the viral genomes. *J Virol, 80*, 11887-11891.
693 Rajsbaum, R, Albrecht, RA, Wang, MK, Maharaj, NP, Versteeg, GA, Nistal-Villan, E, . . . Gack,
694   MU. (2012). Species-specific inhibition of RIG-I ubiquitination and IFN induction by the
695   influenza A virus NS1 protein. *PLoS Pathog, 8*(11), e1003059.
696 Rambaut, A, Lam, TT, Carvalho, LM, & Pybus, OG. (2016). Exploring the temporal structure of
697   heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol, 2*, doi:
698   10.1093/ve/vew1007.
699 Recker, M, Pybus, OG, Nee, S, & Gupta, S. (2007). The generation of influenza outbreaks by a
700   network of host immune responses against a limited set of antigenic types. *Proc Natl Acad
701   Sci U S A, 104*(18), 7711-7716.
702 Reeves, AB, Pearce, JM, Ramey, AM, Meixell, BW, & Runstadler, JA. (2011). Interspecies
703   transmission and limited persistence of low pathogenic avian influenza genomes among
704   Alaska dabbling ducks. *Infect Genet Evol, 11*, 2004-2010.
705 Ronquist, F, Teslenko, M, van der Mark, P, Ayres, DL, Darling, A, Hohna, S, . . . Huelsenbeck, JP.
706   (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a
707   large model space. *Syst Biol, 61*(3), 539-542.
708 Runstadler, J, Hill, N, Hussein, IT, Puryear, W, & Keogh, M. (2013). Connecting the study of wild
709   influenza with the potential for pandemic disease. *Infect Genet Evol, 17*, 162-187.
710 Scholtissek, C, Rohde, W, Von Hoyningen, C, & Rott, R. (1978). On the origin of human influenza
711   virus subtypes H2N2 and H3N2. *Virology, 87*, 13-20.
712 Seekings, AH, Slomka, MJ, Russell, C, Howard, WA, Choudhury, B, Nunez, A, . . . Brown, IH.
713   (2018). Direct evidence of H7N7 avian influenza virus mutation from low to high virulence
714   on a single poultry premises during an outbreak in free range chickens in the UK, 2008.
715   *Infect Genet Evol, 64*, 13-31.
716 Shapiro, B, Rambaut, A, & Drummond, AJ. (2006). Choosing appropriate substitution models for
717   the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol, 23*(1), 7-9.

Sharp, GB, Kawaoka, Y, Jones, DJ, Bean, WJ, Pryor, SP, Hinshaw, V, & Webster, RG. (1997). Coinfection of wild ducks by influenza A viruses: distribution patterns and biological significance. *J Virol, 71*(8), 6128-6135.

Smith, DJ, Lapedes, AS, de Jong, JC, Bestebroer, TM, Rimmelzwaan, GF, Osterhaus, AD, & Fouchier, RA. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science, 305*(5682), 371-376.

Smith, GJ, Vijaykrishna, D, Bahl, J, Lycett, SJ, Worobey, M, Pybus, OG, . . . Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature, 459*(7250), 1122-1125.

Spackman, E, Senne, DA, Myers, TJ, Bulaga, LL, Garber, LP, Perdue, ML, . . . Suarez, DL. (2002). Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. *J Clin. Microbiol., 40*(9), 3256-3260.

Steel, J, & Lowen, AC. (2014). Influenza a virus reassortment. *Curr Top Microbiol Immunol, 385*, 377-401.

Sukeno, N, Otsuki, Y, Konno, J, Yamane, N, Odagiri, T, Arikawa, J, & Ishida, N. (1979). Anti-nucleoprotein antibody response in influenza A infection. *Tohoku J Exp Med, 128*(3), 241-249.

Taubenberger, JK, Reid, AH, Lourens, RM, Wang, R, Jin, G, & Fanning, TG. (2005). Characterization of the 1918 influenza virus polymerase genes. *Nature, 437*, 889-893.

Thathaisong, U, Maneewatch, S, Kulkeaw, K, Thueng-In, K, Poungpair, O, Srimanote, P, . . . Chaicumpa, W. (2008). Human monoclonal single chain antibodies (HuScFv) that bind to the polymerase proteins of influenza A virus. *Asian Pac J Allergy Immunol, 26*, 23-35.

Tolf, C, Latorre-Margalef, N, Wille, M, Bengtsson, D, Gunnarsson, G, Grosbois, V, . . . Waldenström1, J. (2013). Individual variation in influenza A virus infection histories and long-term immune responses in mallards. *PLoS ONE, 8*(4), e61201. doi: 61210.61371/journal.pone.0061201.

van Dijk, JGB, Hoye, BJ, Verhagen, JH, Nolet, BA, Fouchier, RAM, & Klaassen, M. (2014). Juveniles and migrants as drivers for seasonal epizootics of avian influenza virus. *J Anim Ecol, 83*(1), 266-275.

van Dijk, JGB, Verhagen, JH, Wille, M, & Waldenström, J. (2018). Host and virus ecology as determinants of influenza A virus transmission in wild birds. *Curr Opin Virol, 28*, 26-36.

Van Poelvoorde, LAE, Saelens, X, Thomas, I, & Roosens, NH. (2020). Next-generation sequencing: an eye-opener for the surveillance of antiviral resistance in influenza. *Trends Biotechnol, 38*(4), 360-367.

Verhagen, JH, Lexmond, P, Vuong, O, Schutten, M, Guldemeester, J, Osterhaus, ADME, . . . Fouchier, RAM. (2017). Discordant detection of avian influenza virus subtypes in time and space between poultry and wild birds; Towards improvement of surveillance programs. *PLoS ONE, 12*(3), doi: 10.1371/journal.pone.0173470.

Verhagen, JH, Poen, M, Stallknecht, DE, van der Vliet, S, Lexmond, P, Sreevatsan, S, . . . Lebarbenchon, C. (2020). Phylogeography and antigenic diversity of low-pathogenic avian influenza H13 and H16 viruses. *J Virol, 94*(13), e00537-00520. doi: 00510.01128/JVI.00537-00520.

Vijaykrishna, D, Holmes, EC, Joseph, U, Fourment, M, Su, YC, Halpin, R, . . . Barr, IG. (2015). The contrasting phylodynamics of human influenza B viruses. *Elife, 4*, e05055. doi: 05010.07554/eLife.05055.

Vincent, A, Awada, L, Brown, I, Chen, H, Claes, F, Dauphin, G, . . . Ciacci-Zanella, J. (2014). Review of influenza A virus in swine worldwide: a call for increased surveillance and research. *Zoonoses Public Health, 61*(1), 4-17.

Wallensten, A, Munster, VJ, Latorre-Margalef, N, Brytting, M, Elmberg, J, Fouchier, RAM, . . . Olsen, B. (2007). Surveillance of influenza A virus in migratory waterfowl in Northern Europe. *Emerg Infect Dis, 13*, 404-411.

24

769 Webster, RG, Bean, WJ, Gorman, OT, Chambers, TM, & Kawaoka, Y. (1992). Evolution and
770   ecology of influenza A viruses. *Microbiol Rev, 56*(1), 152-179.
771 Wilcox, BR, Knutsen, GA, Berdeen, J, Goekjian, VH, Poulson, R, Goyal, S, . . . Stallknecht, DE.
772   (2011). Influenza A viruses in ducks in Northwestern Minnesota: fine scale spatial and
773   temporal variation in prevalence and subtype diversity. *PLoS ONE, 6*, e24010. doi:
774   24010.21371/journal.pone.0024010.
775 Wille, M, & Holmes, EC. (2020). The ecology and evolution of influenza viruses. *Cold Spring
776   Harb Perspect Med, 10*(7), doi: 10.1101/cshperspect.a038489.
777 Wille, M, Latorre-Margalef, N, Tolf, C, Halpin, R, Wentworth, DE, Olsen, B, & Waldenström, J.
778   (2018). Where do all the subtypes go? Temporal dynamics of H8–H12 influenza A viruses
779   in waterfowl. *Virus Evol, 4*(2), doi: 10.1093/ve/vey1025.
780 Wille, M, Latorre-Margalef, N, Tolf, C, Stallknecht, DE, & Waldenström, J. (2016). No evidence
781   for homo- or heterosubtypic immunity of influenza H3 in Mallards following vaccination in
782   a natural-experimental system. *Mol Ecol, 26*(5), 1420-1431.
783 Wille, M, Tolf, C, Avril, A, Latorre-Margalef, N, Wallerstrom, S, Olsen, B, & Waldenstrom, J.
784   (2013). Frequency and patterns of reassortment in natural influenza A virus infection in a
785   reservoir host. *Virology, 443*(1), 150-160.
786 Wisedchanwet, T, Wongphatcharachai, M, Boonyapisitsopa, S, Bunpapong, N, Kitikoon, P, &
787   Amonsin, A. (2011). Genetic characterization of avian influenza subtype H4N6 and H4N9
788   from live bird market, Thailand. *Virol J, 8*, 131. doi: 110.1186/1743-1422X-1188-1131.
789 Yodsheewan, R, Maneewatch, S, Srimanote, P, Thueng-In, K, Songserm, T, Dong-Din-On, F, . . .
790   Chaicumpa, W. (2013). Human monoclonal ScFv specific to NS1 protein inhibits
791   replication of influenza viruses across types and subtypes. *Antiviral Res, 100*(1), 226-237.
792 Zhang, H, Chen, Q, & Chen, Z. (2012). Characterization of an H4N2 avian influenza virus isolated
793   from domestic duck in Dongting Lake wetland in 2009. *Virus Genes, 44*(1), 24-31.
794 Zhou, B, Donnelly, ME, Scholes, DT, George, KS, Hatta, M, Kawaoka, Y, & Wentworth, DE.
795   (2009). Single-reaction genomic amplification accelerates sequencing and vaccine
796   production for classical and swine origin human influenza A viruses. *J Virol, 83*(19), 10309-
797   10313.
798 Zhou, B, & Wentworth, DE. (2012). Influenza A virus molecular virology techniques. *Methods Mol
799   Biol, 865*, 175-192.
800 Zohari, S, Neimanis, A, Harkonen, T, Moraeus, C, & Valarcher, JF. (2014). Avian influenza
801   A(H10N7) virus involvement in mass mortality of harbour seals (*Phoca vitulina*) in Sweden,
802   March through October 2014. *Euro Surveill, 19*(46), 20967. doi: 20910.22807/21560-
803   27917.es22014.20919.20946.20967.
804
805