

1 ***Avian influenza viruses in wild birds: virus evolution in a multi-host***
2 ***ecosystem***

3 Divya Venkatesh¹, Marjolein J. Poen², Theo M. Bestebroer², Rachel D. Scheuer²,
4 Oanh Vuong², Mzia Chkhaidze³, Anna Machabishvili³, Jimsher Mamuchadze⁴, Levan
5 Ninua⁴, Nadia B. Fedorova⁵, Rebecca A. Halpin⁵, Xudong Lin⁵, Amy Ransier⁵, Timothy
6 B Stockwell⁵, David E. Wentworth^{5*}, Divya Kriti⁶, Jayeeta Dutta⁶, Harm van Bakel⁶,
7 Anita Puranik⁷, Marek J Slomka⁷, Steve Essen⁷, Ian H. Brown⁷, Ron A.M.
8 Fouchier², **Nicola S. Lewis^{1,7#}**

9 ¹Department of Zoology, University of Cambridge, Downing Street, Cambridge
10 CB2 3EJ, United Kingdom

11 ²Department of Viroscience, Erasmus MC, P.O. Box 2040, 3000CA Rotterdam,
12 Netherlands

13 ³National Centre for Disease Control, Tbilisi, Georgia

14 ⁴Institute of Ecology, Ilia State University, 3/5 Cholokashvili, Tbilisi, Georgia.

15 ⁵J. Craig Venter Institute, Rockville, Maryland, United States of America

16 ⁶Icahn School of Medicine at Mount Sinai, New York, United States of America

17 ⁷Animal and Plant Health Agency-Weybridge, United Kingdom

18 ***Running title: Evolution of avian influenza viruses in wild birds***

19 #Address correspondence to **Nicola S. Lewis**: nsl25@cam.ac.uk

20 **Abstract**

21 Wild ducks and gulls are the major reservoirs for avian influenza A viruses (AIVs). The
22 mechanisms that drive AIV evolution are complex at sites where various duck and gull
23 species from multiple flyways breed, winter or stage. The Republic of Georgia is
24 located at the intersection of three migratory flyways: Central Asian Flyway, East
25 Asian/East African Flyway and Black Sea/Mediterranean Flyway. For six consecutive
26 years (2010-2016), we collected AIV samples from various duck and gull species that
27 breed, migrate and overwinter in Georgia. We found substantial subtype diversity of
28 viruses that varied in prevalence from year to year. Low pathogenic (LP)AIV subtypes
29 included H1N1, H2N3, H2N5, H2N7, H3N8, H4N2, H6N2, H7N3, H7N7, H9N1, H9N3,
30 H10N4, H10N7, H11N1, H13N2, H13N6, H13N8, H16N3, plus two H5N5 and H5N8
31 highly pathogenic (HP)AIVs belonging to clade 2.3.4.4. Whole genome phylogenetic
32 trees showed significant host species lineage restriction for nearly all gene segments
33 and significant differences for LPAIVs among different host species in observed
34 reassortment rates, as defined by quantification of phylogenetic incongruence, and in
35 nucleotide diversity. Hemagglutinin clade 2.3.4.4 H5N8 viruses, circulated in Eurasia
36 during 2014-2015 did not reassort, but analysis after its subsequent dissemination
37 during 2016-2017 revealed reassortment in all gene segments except NP and NS.
38 Some virus lineages appeared to be unrelated to AIVs in wild bird populations in other
39 regions with maintenance of local AIV viruses in Georgia, whereas other lineages
40 showed considerable genetic inter-relationship with viruses circulating in other parts
41 of Eurasia and Africa, despite relative under-sampling in the area.

42

43

44 **Importance**

45 Waterbirds (e.g., gulls/ducks) are natural reservoirs of avian influenza viruses (AIVs)
46 and have been shown to mediate dispersal of AIV at inter-continental scales during
47 seasonal migration. The segmented genome of influenza viruses enables viral RNA
48 from different lineages to mix or re-assort when two viruses infect the same host. Such
49 reassortant viruses have been identified in most major human influenza pandemics
50 and several poultry outbreaks. Despite their importance, we have only recently begun
51 to understand AIV evolution and reassortment in their natural host reservoirs. This
52 comprehensive study illustrates of AIV evolutionary dynamics within a multi-host
53 ecosystem at a stop-over site where three major migratory flyways intersect. Our
54 analysis of this ecosystem over a six-year period provides a snapshot of how these
55 viruses are linked to global AIV populations. Understanding the evolution of AIVs in
56 the natural host is imperative to both mitigating the risk of incursion into domestic
57 poultry and potential risk to mammalian hosts including humans.

58 Introduction

59 Avian influenza viruses (AIVs) have been identified in a wide diversity of wild and
60 domestic bird species but wild waterbirds of the Orders *Anseriformes* and
61 *Charadriiformes*, such as ducks, geese, swans and shorebirds (1, 2) form their natural
62 reservoir. These birds maintain diverse group of low pathogenic avian influenza A
63 viruses (LPAIVs), which cause limited morbidity in these host species in experimental
64 settings (3). The effect of AIV infection in wild birds in non-experimental settings is
65 more contradictory. Body mass was significantly lower in infected mallards (*Anas*
66 *platyrhynchos*) and the amount of virus shed by infected juveniles was negatively
67 correlated with body mass. However, there was no general effect of infection on
68 staging time (duration of stopover for migratory birds), except for juveniles in
69 September and LPAIV infection did not affect speed or distance of subsequent
70 migration (4). Conversely, a recent mallard study demonstrated no obvious detriment
71 to the bird as movement patterns did not differ between LPAIV infected and uninfected
72 birds. Hence, LPAIV infection probably does not affect mallard movements during
73 stopover, consequently resulting in the potential for virus spread along the migration
74 route (5). The precise role of migrants and resident birds in amplifying and dispersing
75 AIVs however, remains unclear. In another study the migrant arrivals played a role in
76 virus amplification rather than seeding a novel variant into a resident population (6). It
77 has also been suggested that switching transmission dynamics might be a critical
78 strategy for pathogens such as influenza A viruses associated with mobile hosts such
79 as wild waterbirds, and that both intra and inter-species transmission are important to
80 maintaining gene flow across seasons (7).

81

82 AIVs continue to cause both morbidity and mortality in poultry worldwide. Increased
83 mortality is strongly related to infection with highly pathogenic influenza A viruses

84 (HPAIVs), characterised by mortality in gallinaceous poultry (8). Periodically, human
85 infections associated with HPAIV of both the H5 and H7 subtypes have been detected.
86 In particular, parts of Asia and Africa have been significantly affected by the Eurasian
87 (goose/Guangdong/1996) lineage H5 HPAIV epizootic for two decades, becoming
88 enzootic in some areas and multiple waves of influenza with evolving viruses in others
89 (9). More recently, H5Nx reassortants of the Eurasian lineage HPAIVs from clade
90 2.3.4.4 have been introduced into wild birds from poultry and spread to new
91 geographic regions (10).

92 The Caucasus, at the border of Europe and Asia, is important for migration and over-
93 wintering of wild waterbirds. Three flyways, the Central Asian, East Africa-West Asia,
94 and Mediterranean/Black Sea flyways, converge in this region (11, 12). Understanding
95 the ecology and evolution of AIVs in wild birds is complex, particularly at sites where
96 multiple species co-habit and in those ecosystems which support different annual life-
97 cycle stages and where multiple migratory flyways intersect.

98 At a population level, Eurasian dabbling ducks were found to be more frequently
99 infected than other ducks and Anseriformes (13) with most AIV subtypes detected in
100 ducks, except H13 and H16 subtypes which were detected primarily in gulls (13, 14).
101 Temporal and spatial variation in influenza virus prevalence in wild birds was
102 observed, with AIV prevalence varying by sampling location. In this study site in the
103 Republic of Georgia, we observed peak prevalence in large gulls during the autumn
104 migration (5.3-9.8%), but peak prevalence in Black-headed Gulls (*Chroicocephalus*
105 *ridibundus*) in spring (4.2-13%)(15). In ducks, we observed increased AIV prevalence
106 during the autumn post-moult aggregations and migration stop-over period (6.3%) but
107 at lower levels to those observed in other more northerly post-moult areas in Eurasia.

108

109 In North America, studies have primarily focused on Anseriformes species with
110 sampling during late summer and autumn southern migration (16-18), rather than
111 longitudinally throughout the annual lifecycle of the host or within an ecosystem. The
112 southwestern Lake Erie Basin is an important stopover site for waterfowl during
113 migration periods, and over the past 28 years, 8.72% of waterfowl sampled in this
114 geographic location have been positive for AIV recovery during summer and autumn
115 (June – December) (19). More recent studies which targeted overwintering and
116 returning migratory birds during February – April showed the presence of diverse AIV
117 subtypes in waterbirds at northern latitudes in the United States (19).

118

119 Previous genetic studies of the viruses isolated from wild birds have focused on gene
120 flow at an intra- or intercontinental level involving multiple hosts, rather than on virus
121 gene flow among species within an ecosystem (18, 20-22). Indeed, the conclusions of
122 such studies have been somewhat limited at times by statistical power owing to
123 insufficient sequence data from enough hosts relevant to virus dynamics across the
124 geographic study area. (23). In Eurasia, frequent reassortment and co-circulating
125 lineages were observed for all eight genomic RNA segments over time. Although,
126 there was no apparent species-specific effect on the diversity of the AIVs, there was
127 a spatial and temporal relationship between the Eurasian sequences and significant
128 viral migration of AIVs from West Eurasia towards Central Eurasia (24).

129

130 This study presents novel findings concerning the ecology and evolution of both
131 LPAIVs and HPAIVs circulating in wild birds in a key active surveillance site in Eurasia.

132 We investigated the diffusion of AIV gene segments within different wild bird hosts
133 occupying the same ecosystem. There was substantial diversity in surface
134 glycoprotein HA (heamagglutinin) and NA (neuraminidase) subtypes, which varied
135 year to year and with the host species. M, NS, NP, PB1, PB2 and PA (henceforth
136 referred to as “internal” gene segments) also showed host restriction to various
137 degrees. There were differences in genetic diversity, reassortment rates, and inter-
138 species transmission rates in the internal gene segments associated with different
139 host species and HA subtypes. We also examined how closely related the Georgian
140 AIV gene segments were to AIV globally. We found evidence for genetic inter-
141 relationship of Georgian AIV with AIV in mainly Africa and Eurasia but several lineages
142 appear to be maintained locally.

143

144

145 **Methods**

146 **Surveillance**

147 Active surveillance for influenza A viruses was carried out from 2010-2016 in the
148 Republic of Georgia as described previously (15). The study area and sample
149 collection methods remain predominantly the same. In this analysis, the study area is
150 divided into three groups based on bird annual lifecycle and geography: the wetlands
151 in Ajara, Guria and Samegrelo constitute the Black Sea coast region; Samtskhe-
152 Javakheti form the Georgian uplands sampling area; and finally, Tbilisi and Kakheti
153 are grouped as Eastern Georgia. Sampling was targeted towards Anatidae (ducks)
154 and Charadriiformes (gulls) and other birds commonly found in the wetland
155 ecosystems. Details of the host species considered can be found in (15). We used
156 several methods to catch birds depending on the species and location, including mist
157 nets, spring traps and manual capture using hand-held nets, lamping and sampling
158 hunted birds. We took paired oropharyngeal and cloacal swabs, serum and in some
159 cases, feather samples from all live-caught birds.

160 To sample live-caught or hunted birds, a sterile plain cotton swab was inserted into
161 the trachea or oropharynx (in smaller bird species), or approximately 5 mm into the
162 cloaca of the bird and then gently turned to moisten the swab. All swabs were then
163 inserted into viral transport storage media (Hanks balanced salt solution containing
164 10% glycerol, 200 U/ml penicillin, 200 mg/ml streptomycin, 100 U/ml polymixin B
165 sulfate and 250 mg/ml gentamycin) and the shaft of the swab broken just above the
166 cotton tip. abs were stored at -70°C no more than 6 hours after collection and were
167 chilled at $1-4^{\circ}\text{C}$ on ice or in a portable refrigerator in the interim. Surveillance was
168 carried out throughout the year, but there was seasonal fluctuation in bird density. In

169 addition to previously described methods, we built a duck trap in the Javakheti uplands
170 close to the gull colony sampling site in 2015.

171 ***Dataset and genomic sequencing***

172 Over a period of six years, 30,911 samples from 105 different bird species were
173 analysed for the presence of AIVs. Positive isolates were obtained by standard
174 approaches (25), and where possible, subtyped and sequence generated from
175 extracted RNA as described below.

176 For virus samples from 2010-2012, codon complete genomes of IAV were
177 sequenced as part of the Influenza Genome Project
178 (<http://gcid.jcvi.org/projects/gsc/influenza/index.php>), an initiative by the National
179 Institute of Allergies and Infectious Diseases (NIAID). IAV viral RNA (vRNA) was
180 isolated from the samples/specimens, and the entire genome was amplified from 3 ul
181 of RNA template using a multi-segment RT-PCR strategy (M-RTPCR) (26, 27). The
182 amplicons were sequenced using the Ion Torrent PGM (Thermo Fisher Scientific,
183 Waltham, Massachusetts, USA) and/or the Illumina MiSeq v2 (Illumina, Inc., San
184 Diego, California, USA) instruments. When sequencing data from both platforms was
185 available, the data were merged and assembled together; the resulting consensus
186 sequences were supported by reads from both technologies. Sequence data for
187 Georgia was downloaded from the NIAID Influenza Research Database (IRD) (28)
188 through the web site at <http://www.fludb.org> on 11/5/2016. To this dataset, we added
189 sequence data for isolates from 2013 and 2016 which were sequenced at either
190 Erasmus MC, Animal and Plant Health Agency (APHA) or the Icahn School of
191 Medicine at Mount Sinai (ISMMS). At Erasmus MC sequencing was performed as
192 described previously by V. J. Munster et al. (29), with modifications. Primer sequences
193 are available upon request.

194 At APHA, viral RNA was extracted using the QIAquick Viral RNA extraction kit
195 (Qiagen, UK) without the addition of carrier. Double stranded cDNA (cDNA synthesis
196 system, Roche, UK) was generated from RNA according to the manufacturer's
197 instructions. This was quantified using the fluorescent PicoGreen reagent and 1ng was
198 used as a template for the preparation of the sequencing library (NexteraXT, Illumina,
199 Cambridge, UK). Sequencing libraries were run on a MiSeq instrument (Illumina,
200 Cambridge, UK) with 2x75 base paired end reads. Data handling of raw sequence
201 reads and extraction of consensus sequences were performed at APHA.

202 For the Icahn School of medicine at Mount Sinai, RNA was extracted using the
203 QIAamp Viral RNA Mini Kit (52904, Qiagen, UK). MS-RT-PCR amplification was
204 performed with the Superscript III high-fidelity RT-PCR kit (12574-023, Invitrogen)
205 according to manufacturer's instructions using the Opti1 primer set: Opti1-F1 5'
206 GTTACGCGCCAGCAAAAGCAGG, Opti1-F2 5'GTTACGCGCCAGCGAAAGCAGG
207 and Opti1-R15'GTTACGCGCCAGTAGAAACAAGG. DNA amplicons were purified
208 using Agencourt AMPure XP 5ml Kit (A63880, Beckman Coulter). At the Icahn School
209 of Medicine, sequencing libraries were prepared and sequencing was performed on a
210 MiSeq instrument (Illumina, Cambridge, UK) with 2x150 base paired end reads. Data
211 handling of raw sequence reads and extraction of consensus sequences were
212 performed at ISMMS, as described previously (30).

213 ***Genetic analyses***

214 ***Sequence alignment preparation***

215 Whole genome sequences from 81 Georgian strains isolated between 2010
216 and 2016 are used in this analysis. We aligned sequences from each gene segment
217 separately using MAFFT v7.305b (31) and trimmed to starting ATG and STOP codon
218 in Aliview v1.18. Hemagglutinin (HA) sequences were further trimmed to exclude the

219 initial signal sequence (32, 33). Sequences were then aligned using “muscle-codon”
220 option with default settings in MEGA7 (34).

221 The NS gene has two alleles A and B, with significant difference in sequence
222 composition, which could skew analyses of sequence diversity. The NS gene
223 sequences were therefore considered both as a complete dataset (NS) and
224 subdivided into NS-A and NS-B datasets where required. As only six out of 81
225 sequenced strains had the NS-A allele, only NS and NS-B datasets were used in the
226 analyses.

227 We then subdivided the complete datasets of each gene according to viral
228 traits, namely:

- 229 • host group (gull and duck)
- 230 • host type
 - 231 ○ **BMG**: Black-headed Gulls (*Chroicocephalus ridibundus*) and
 - 232 Mediterranean Gulls (*Ichthyiaetus melanocephalus*).
 - 233 ○ **YAG**: Yellow-legged Gulls (*Larus michahellis*) and Armenian Gulls
 - 234 (*Larus armenicus*).
 - 235 ○ **MD**: Mallards (*Anas platyrhynchos*).
 - 236 ○ **OD**: Other ducks. This includes the common teal (*Anas crecca*),
 - 237 domestic duck (*Anas platyrhynchos domesticus*), garganey (*Anas*
 - 238 *querquedula*), northern shoveler (*Anas clypeata*), common coot (*Fulica*
 - 239 *atra*), and tufted duck (*Aythya fuligula*).
- 240 • HA subtype. Dataset was reduced to include subtypes H1, 2, 3,4, 5, 6, 7, 9,10,
- 241 11, 13 where greater than three sequences were available for statistical
- 242 analyses.

243 ***Visualisation of phylogenetic incongruence***

244 We inferred Maximum Likelihood (ML) phylogenetic trees for each gene
245 segment using IQ-TREE, 1.5.5 (35) and ModelFinder (36) and obtained branch
246 supports with SH-like approximate Likelihood Ratio Test (aLRT) and standard non-
247 parametric bootstrap. All trees were rooted using the “best-fitting-root” function in
248 Tempest v1.5 (37) and visualised in FigTree v1.4.2, with increasing node-order. To
249 visualise incongruence, we traced the phylogenetic position of each sequence,
250 coloured according to host, across unrooted ML trees for all internal gene segments.
251 Figures were generated by modifying scripts from a similar analysis (38).

252 ***Quantification of nucleotide diversity***

253 Complete alignments of each internal gene, as well as alignment subsets by host
254 group, host type and HA subtype were used in “PopGenome” package in R v3.2 (39)
255 to estimate nucleotide diversity. Per-site diversity was calculated by dividing the
256 nucleotide diversity output by number of sites present in each alignment. As each
257 subset contained different numbers of sequences, this value was normalised by
258 dividing by the number of sequences in each respective dataset. Heat maps from this
259 data were plotted in R v3.2.

260 ***Correlating traits with phylogeny (BaTS)***

261 Null hypothesis of no association between phylogenetic ancestry and traits (host
262 group, host type and HA subtype) was tested using Bayesian Tip-association
263 Significance Testing (BaTS) beta build 2 (40) for all internal gene segments. Bayesian
264 posterior sets of trees were inferred using MrBayes v3.2.6 (41) using the same
265 segment-wise alignments generated for ML tree estimation. Ratio of clustering by
266 each trait on the gene segment trees that is expected by chance alone (Null mean),
267 with the association that is observed in the data (Observed mean) was calculated.

268 These expected/observed ratios were summarized in a heat-map with the y-axis
269 ordered by the amount of reassortment observed. Data manipulation and figure
270 preparation was done in R v3.2.

271 ***Quantification of diversity and between host transmission***

272 Alignments generated for ML trees were also used in Bayesian phylodynamic
273 analyses using BEAST v1.8.4 (42). We employed a strict molecular clock, a
274 coalescent constant tree prior and the SRD06 site model with two partitions for codon
275 positions (1st+2nd positions, 3rd position), with base frequencies unlinked across all
276 codon positions. The MCMC chain was run twice for 100 million iterations, with sub-
277 sampling every 10,000 iterations. All parameters reached convergence, as assessed
278 visually using Tracer (v.1.6.0). Log combiner (v1.8.4) was used to remove initial 10%
279 of the chain as burn-in, and merge log and trees files output from the two MCMC runs.
280 Maximum clade credibility (MCC) trees were summarized using TreeAnnotator
281 (v.1.8.4). After removal of burn-in, the trees were analysed using PACT (Posterior
282 analysis of coalescent trees) (<https://github.com/trvrb/PACT.git>) to determine
283 measures of diversity, and migration rates between hosts over time.

284 ***Geographical context for 'Georgian origin' internal protein coding gene*** 285 ***segments***

286 Internal gene sequences from, avian hosts, sampled across the world between 2005
287 and 2017 were obtained from gisaid.org (downloaded November 2017). Sequences
288 (each segment separately) were divided into regions namely Asia (including Oceania),
289 Europe, Africa, North America and South America. The program cd-hit-est (43, 44)
290 was used to down-sample each regional dataset to 0.9 similarity cut-off level. These
291 down-sampled sequences were then merged with the Georgian dataset. Discrete trait
292 ancestral reconstruction with symmetric and asymmetric models were implemented in

293 BEAST v1.8.4 (42) together with marginal likelihood estimation using path-
294 sampling/stepping-stone analysis. The symmetric model was chosen over the
295 asymmetric (log Bayes factor =14). The MCMC chain was run twice for 100 million
296 iterations, with sub-sampling every 10,000 iterations. All parameters reached
297 convergence, as assessed visually using Tracer (v.1.6.0). Log combiner (v1.8.4) was
298 used to remove initial 10% of the chain as burn-in, and merge log and trees files output
299 from the two MCMC runs. Maximum clade credibility (MCC) trees were summarized
300 using TreeAnnotator (v.1.8.4). PACT was used to extract overall migration rates
301 between trait locations.

302 **Results**

303 ***Prevalence, subtype diversity and host-specificity of AIVs***

304 Over the six-year period between 2010 and 2016, 30,911 samples from 105 different
305 bird species were analysed for the presence of AIVs. Approximately 3000-5000
306 samples were collected every year. The total number of samples collected, and the
307 total number of positives, for each host group each year are shown in the Figure 1.
308 The prevalence of AIV varied year to year, and between the two major host groups
309 (gulls and ducks). Between 2010-12, the prevalence of AIV between gull and ducks
310 was comparable (Figure 2A). The fall in prevalence in gulls from 2013 onwards could
311 be partially explained by reproductive failure in consecutive years two of the gull
312 species (Yellow-legged and Armenian gulls). The data also show strong seasonality
313 with most positives sampled during the autumn migration season (Figure 2B). When
314 we consider the three different regions of sampling sites (Figure 2D), we see that most
315 of the gull and duck positives from 2010-12 were sampled from the Black Sea coast
316 region. After the installation of a duck trap in 2015 in the Javakheti uplands, there is

317 an increase in prevalence in ducks (and “other” birds) from 2015 onward in the
318 uplands, during the migratory season.

319 24 HA/NA subtypes of influenza A virus, including 12 different HA subtypes (H1, 2, 3,
320 4, 5, 6, 7, 9, 10, 11, 13, and 16) were isolated (Figure 2C). The diversity of subtypes
321 varied from year to year, and associated with the level of prevalence in duck versus
322 gull hosts. Moreover, only a proportion of those samples that tested positive yielded
323 virus isolates which could be typed and sequenced. Within our sampling in Georgia,
324 H9 and H13 subtypes are found exclusively in gulls, while H1, H5, and H7 were
325 detected exclusively in mallards. H3, H4, H6, and H10 were found in mallards and
326 various other ducks. Positive evidence for multiple-species infection (ducks and gulls)
327 was found only for H2 and H11 viruses in this dataset even though globally, many
328 other subtypes are found in multiple hosts.

329 Between 2010-12, up to seven different HA subtypes were found every year,
330 consistent with the relatively high prevalence in both host groups in these years.
331 Subtypes included H1, 2, 3, 4, 6, 10, 11, 13, and 16. H13, which was found in the
332 greatest proportion of sequenced samples in 2011 and 2012 and was the sole HA
333 subtype sequenced in 2013. In 2014, again only a single subtype was found (H10).
334 The absence of more subtypes in these years could be explained by the comparatively
335 low prevalence of AIV in these years, in both gulls and ducks in 2014 and especially
336 ducks in 2013 (Figure 2A). In 2015, the prevalence was nearly zero in gulls, but in
337 ducks, we saw HPAI H5 type viruses detected along with an H6. H4, which was
338 previously isolated only in 2011, was the predominant type in 2016, followed by H5
339 and H7.

340 ***Genetic structure of AIV detected in Georgia in 2010-16***

341 For all gene segments except PA, there were two major subdivisions in tree topology
342 – one clade containing sequences predominantly from ducks and one clade entirely
343 derived from gull sequences (Figure 3,4). The internal protein coding gene segments
344 from certain subtypes formed sub-clades that were defined by year of circulation
345 suggesting single-variant epidemic-like transmission within the population. This was
346 seen in H13N8 in gulls and H4N6 and H5N8 in ducks. There were several examples
347 of gull-derived viruses, which had several internal gene segments (other than NP)
348 located in the ‘duck’ clade, mostly derived from Black-headed and Mediterranean
349 Gulls (BMG). Only the PA gene phylogeny had an occurrence of a small sub-clade of
350 Yellow-legged and Armenian Gull-derived (YAG) viruses clustered within the duck-
351 derived viruses. For M gene segment, there were two major clades entirely defined by
352 host species (except for 2 BMG viruses), and an outlier sub-clade consisting of H2
353 and H9 gull lineage viruses from BMGs. In PB1, PB2 and PA, these outlier- sub-clade
354 viruses were found in various configurations in the tree. For NS, the tree topology
355 divided into two alleles as reported previously (45). However, there were only six
356 viruses from Allele *A* isolated from four mallards (MD), a garganey (OD) and a
357 common teal (OD). Allele *B* splits into two sub-clades again defined by whether the
358 viruses were isolated from gulls or ducks. The ‘duck’ sub-clade includes the outlier
359 BMG viruses identified above for M. The long branch length to the gull sub-clade from
360 the duck sub-clade in Allele *B* would suggest that there might be host-specificity in NS
361 evolution, perhaps in response to differences between avian host innate immune
362 responses.

363 ***Variation in nucleotide diversity***

364 We used the PopGenome package in R to calculate the per-site nucleotide
365 diversity for all internal gene segments (Figure 5A-C). Nucleotide diversity of the

366 internal gene segments in one surveillance site may be an indication of the breadth of
367 sources where the viruses have been derived from. We found greater diversity in both
368 gulls and ducks in gene segment NS (possibly because of the presence of both A and
369 B alleles of this gene in the dataset) and PB2 (Figure 5A). When further sub-divided
370 into “host types” as described in the methods, we found that the group of Black-headed
371 and Mediterranean Gulls (BMG) had the highest per-site diversity. In comparison, the
372 mallards (MD), the Yellow-legged and Armenian Gulls (YAG) and other ducks (OD)
373 had relatively lower values across all internal gene segments, despite the OD
374 comprising of a variety of ducks. Only the PA gene had greater diversity in Yellow-
375 legged and Armenian Gulls than in Black-headed and Mediterranean Gulls (Figure
376 3B). When subset by HA subtype (Figure 5C), the internal gene segments associated
377 with H4 and H13, the most abundant types found in our dataset, had the lowest
378 diversity – possibly because several of the isolates were detected at the same time.
379 Those less commonly isolated, such as H11 was detected in different years (2011,
380 2014) which may explain the high diversity of its NS, M, NP, PA, PB1, and PB2 gene
381 segments. However, H3, which also has relatively high diversity were both detected
382 at the same time (September 2011). Both NS and NS-B datasets were used in the
383 analysis and as expected, the exclusion of sequences of NS-A (found exclusively in
384 viruses from duck hosts), lowers the overall diversity within the ducks even when the
385 values are normalised for the number of sequences found in each subset.

386 We tested the root-to-tip regression for ML trees for each of the six internal protein
387 coding gene segments using Tempest v1.5 (37) to look for temporal signatures. All
388 except NS gene showed positive correlation of distance with time, despite the short
389 window of six years (Figure 6). NS root to tip regression shows a negative slope, and
390 it is likely confounded by the presence of two alleles A and B. Therefore, only NS-B

391 allele, which forms a dominant portion of the NS gene segments in the data-set (75
392 out of 81), and shows clock-likeness (Figure 6) were used for further analysis using
393 BEAST v1.8.4. PACT analysis showed that the overall and yearly host-related
394 diversity measures (Figure 7A and B) show similar trends as seen in Figure 5.

395 ***Correlation of traits with phylogeny***

396 We tested the null hypothesis that there is no association between phylogenetic
397 ancestry and traits (host group, host type and HA subtype) using Bayesian Tip-
398 association Significance Testing (BaTS). Ratio of clustering by each trait on the gene
399 segment trees that is expected by chance alone (Null mean), with the association that
400 is observed in the data (Observed mean) are presented in Figure 8A-C. The higher
401 the value of null/observed, the lower is the support for phylogenetic clustering of the
402 given trait. Therefore, a higher value indicates a different ancestry. Hence, when we
403 consider the HA subtype trait as “lineage”, it provides a measure of reassortment as
404 described (46). Again, NS-B dataset was considered along with the complete NS
405 dataset but no significant differences in trends were found. Panel A shows that gull
406 viruses are more likely to cluster together in a phylogenetic tree than duck viruses in
407 general. When viruses of gulls and ducks were further subdivided, panel B shows that
408 OD viruses are less likely to cluster together in the tree, which is expected given that
409 we have grouped together several duck species under this category. Among the rest,
410 again it is the duck species (MD) that exhibit dynamic phylogenetic placing compared
411 to both the gull types. The only exception is with the PB2 gene segment, for which the
412 BMG show a lower level of phylogenetic clustering by species indicating putative
413 reassortment events. When we consider the HA subtype (lineage) of the viruses, we
414 find that H4 and H13, which showed the lowest nucleotide diversity, also show very
415 low levels of reassortment, as does H5. There was not enough statistical power to

416 interpret events in H1, 3, 6, 7, 9 or 11 viruses. Where statistically significant values
417 were found, lower levels of clustering were observed.

418 ***Directionality of viral gene segment transfer***

419 Figure 9 shows ancestral reconstruction of the host state along time-scaled
420 phylogenies for five of six internal gene segments. The results are summarised in
421 Figure 10A showing the mean number of host jump events from duck to gull and vice-
422 versa. For all gene segments, most of the host spillover events are in the direction
423 from ducks to gulls. In figure 10B we see that at a finer level, most of the host jump
424 events happen within the duck (mallards (MD) to other ducks (OD)) and gull (Black-
425 headed and Mediterranean Gulls (BMG) to Yellow-legged and Armenian Gulls (YAG)
426 and *vice versa*) species. In transmissions from ducks to gulls it is largely noticeable
427 only from MD to BMG. This likely explains the higher levels of nucleotide diversity and
428 reassortment rates in the BMG viruses relative to YAG seen above.

429 ***Geographical context for GE NS, M, NP, PA, PB1, PB2 segments***

430 To determine the origin and destination of the internal protein coding gene segments
431 found in viruses isolated in Georgia, we analysed our sequence dataset together with
432 avian influenza sequences from a broader timeframe (2005-2016) and regional
433 sampling. Figure 11 shows the genealogy for the NP gene for whose tips we know the
434 location of sampling and whose internal nodes are estimated using discrete-state
435 ancestral reconstruction in BEAST. Clades in which Georgian sequences occur are
436 highlighted. Figure 12 summarises the genealogy in a circularised graph in which the
437 arrowheads indicate the direction of transfer and the width of the arrow indicate the
438 rate of transfer to different locations. The analyses reveal viruses from the Atlantic and
439 Afro-Eurasian locations form largely separate clades, which is consistent with previous
440 studies (47, 48). However, we do find instances of transmission across this divide,

441 most notably to and from Asia and Europe. Many NP genes from Georgia cluster with
442 other Georgian NP genes, in some cases forming the terminal branches spanning
443 years indicating restriction to local spread. However, our dataset contains the latest
444 Georgian sequences, and sequences from this timeframe were not available from the
445 rest of Eurasia. Hence, we can expect to have missed identifying onward transmission.
446 From the transmission we do identify, it appears that there is considerable migration
447 into Africa and Europe and to a lesser extent to Southern/Eastern Asia. Most of the
448 sequences transmitted into Georgia come from Asia and Europe, along with a single
449 identified instance of direct transfer from North America.

450 Discussion

451 Wild birds have been shown to harbor substantial genetic diversity of avian
452 influenza viruses. This study showed the diversity not only varied by year but was
453 associated with the level of overall prevalence in different wild bird host species, perhaps
454 influencing the observed rates and diversity if prevalence were low. We observed
455 ecological fluctuations during the study period which might have influenced the results. In
456 2015, there was nearly complete reproductive failure on the breeding colony of Armenian gulls
457 which might have resulted in few susceptible juveniles and therefore altered influenza
458 prevalence. In 2013, the nest sites on the Chorokhi River Delta were flooded consecutively
459 again perhaps influencing disease dynamics. While the installation of the duck trap in the
460 Javakheti uplands improved the longitudinal window of duck sampling to include both
461 over-wintering and migratory populations, this initiative might have introduced
462 prevalence and subtype biases in the data by sampling a previously un-sampled
463 subpopulation. However, even allowing for these biases, the results from this study
464 show that there is little evidence that one species group maintains all influenza A virus
465 diversity, there appears to be relative host-restriction in many subtypes (except for H2

466 and H11 viruses) and there are differences in prevalence dynamics depending on
467 host. Therefore, one host is not representative of influenza A virus prevalence,
468 dynamics and diversity across the wild bird reservoir. Within both ducks and gulls
469 however, peak prevalence was consistently observed in hatch-year birds and with a
470 more restricted subtype diversity, suggesting that there is an initial influenza A virus
471 epidemic wave as naïve birds aggregate in their first year. Subsequently in the over-
472 wintering period, a wider subtype diversity was observed in both host groups and
473 adults were more frequently infected. This suggests that disease dynamics are
474 complex and influenced by multiple host factors including age and annual life cycle
475 stage.

476 It has previously been observed that some subtypes are routinely and nearly
477 exclusively isolated from certain host families/genus, the most notable example being
478 H13 and H16 viruses from gulls. However, mixed infections are relatively common but
479 might be masked if subtype characterization requires virus isolation, therefore putting
480 the clinical specimen through a culture bottleneck. Advances in sequencing direct from
481 clinical material would more accurately (remove possible culture selection bias)
482 establish the prevalence, subtype diversity and genetic diversity within wild birds.

483 In general, for all gene segments except PA, we identify strong patterns of clade
484 topology defined by host. This suggests that there is segregated gene flow through
485 these host populations with little inter-host reassortment. Additionally, within our study
486 period there were large scale perturbations in ecology which might also influence our
487 prevalence and subtype diversity estimates. For example, in 2014 and 2015 there was
488 widespread reproductive failure in two gull host species due to nest flooding (Yellow-
489 legged Gulls) and few returning adults to the colony (Armenian Gulls), and therefore

490 few juveniles from which to detect the annual epidemic wave. The occurrence and
491 significance of such ecological fluctuations on disease dynamics are unclear. We also
492 increased the ability to sample migrant ducks in late summer and early autumn from
493 August 2015 by constructing a duck trap in the newly created National Park. Again,
494 this addition to sampling strategy likely increased the detection of influenza in these
495 anseriform hosts as they were previously under-sampled.

496 We tested whether certain hosts maintained higher levels of nucleotide diversity in
497 the non-immune related internal genes. PB2 and NS were the most genetically diverse
498 in both gulls and ducks. Within host-group, Black-headed and Mediterranean Gull-
499 derived viruses showed highest per-site diversity, Yellow-legged and Armenian Gulls
500 lower diversity, likely because some of the viruses of the former were associated with
501 reassortants probably derived from ducks (or another unsampled host group). While
502 despite high rates of reassortment and spillover between duck subgroups mallards
503 (MD) and other ducks (OD), the absence of any gull derived viruses in these ducks
504 keeps their diversity levels lower compared to gulls/BMG.

505 Where gene flow does occur between host groups, for all gene segments, host-
506 spillover events were in the direction of ducks to gulls and from other ducks to Black-
507 headed and Mediterranean Gulls, likely explaining the higher levels of nucleotide
508 diversity in these gulls observed above. Where HA and NA gene segments were
509 acquired by gulls from ducks, there was a pre-requisite for a gull-clade internal gene
510 cassette suggesting a host-restrictive effect for onward maintenance within the gull
511 population (13, 49). Interestingly, Black-headed and Mediterranean Gulls only occur
512 on the study site in the over-wintering period where there are also high densities of
513 over-wintering ducks from other geographic areas. Although there is a duck-gull

514 interface on the breeding grounds in summer, the duck densities are very much lower,
515 perhaps suggesting that there is a threshold level of bird density that allows gene flow
516 among hosts.

517 If we look at diversity by HA subtype, H4 and H13 were the least diverse and
518 showed the lowest rates of reassortment and were also associated with hatch-year
519 bird infections, suggesting a clonal expansion and epidemic gene flow through these
520 birds. The 2014-2015 HPAI H5 epizootic also showed no reassortment unlike the
521 2016-2017 HPAI H5 viruses, perhaps indicating that the first wave of 2.3.4.4 viruses
522 diffused through the wild bird population similarly to a 'naïve' infection, and subsequent
523 epizootics have resulted in altered pathogen evolution strategies to maintain gene
524 flow, similar to those previously observed in North America when considering the effect
525 of latitude on gene flow (7).

526 When we examine the internal gene segments of the Georgian AIV in a broader
527 geographical context, we find significant gene flow to and from Georgia with Europe
528 and the rest of Asia, although data for Africa is very limited. Crossover into the Atlantic
529 flyway appears to be mediated largely by gulls with some exceptions, notably the
530 H5N1-NP gene that was transmitted between ducks.

531 From this study, the diffusion of avian influenza viruses within a multi-host
532 ecosystem is heterogeneous. One host group cannot therefore be used as a surrogate
533 for others. It is likely that virus evolution in these natural eco-systems is a complex mix
534 of host-pathogen interface and ecological factors. Understanding such drivers is key
535 to investigating these emerging pathogens, interpreting the data from different sites
536 around the world and ultimately informing risk of incursion of emerging variants from
537 one geographic region to another.

538 Acknowledgements:

539 This study including field work and sequencing was funded by National Institute of
540 Allergy and Infectious Diseases, National Institutes of Health, Department of Health
541 and Human Services contract No.HHSN2722000900007C
542 and HHSN266200700010C "NIAID Centres of Excellence for Influenza Research and
543 Surveillance"
544 <http://www.niaid.nih.gov/LabsAndResources/resources/ceirs/Pages/crip.aspx>, and a
545 DTRA FRCWMD Broad Agency Announcement (HDTRA1-09-14-FRCWMD
546 GRANT11177182). The funders had no role in study design, data collection and
547 analysis, decision to publish, or preparation of the manuscript. The sequencing data
548 for this manuscript was generated while D. E. Wentworth was employed at the J. Craig
549 Venter Institute. The opinions expressed in this article are the author's own and do not
550 reflect the view of the Centers for Disease Control and Prevention, the Department of
551 Health and Human Services, or the United States government.

552 References

- 553 1. Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus AD, Fouchier RA.
554 2006. Global patterns of influenza a virus in wild birds. *Science* 312:384-8.
- 555 2. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and
556 ecology of influenza A viruses. *Microbiol Rev* 56:152-79.
- 557 3. Gunnarsson G, Jourdain E, Waldenstrom J, Helander B, Lindberg P, Elmberg J,
558 Latorre-Margalef N, Olsen B. 2010. Zero prevalence of influenza A virus in two raptor
559 species by standard screening. *Vector Borne Zoonotic Dis* 10:387-90.
- 560 4. Latorre-Margalef N, Gunnarsson G, Munster VJ, Fouchier RA, Osterhaus AD, Elmberg
561 J, Olsen B, Wallensten A, Haemig PD, Fransson T, Brudin L, Waldenstrom J. 2009.
562 Effects of influenza A virus infection on migrating mallard ducks. *Proc Biol Sci*
563 276:1029-36.
- 564 5. Bengtsson D, Safi K, Avril A, Fiedler W, Wikelski M, Gunnarsson G, Elmberg J, Tolf
565 C, Olsen B, Waldenstrom J. 2016. Does influenza A virus infection affect movement
566 behaviour during stopover in its wild reservoir host? *R Soc Open Sci* 3:150633.
- 567 6. Verhagen JH, van Dijk JG, Vuong O, Bestebroer T, Lexmond P, Klaassen M, Fouchier
568 RA. 2014. Migratory birds reinforce local circulation of avian influenza viruses. *PLoS*
569 *One* 9:e112366.

- 570 7. Hill NJ, Ma EJ, Meixell BW, Lindberg MS, Boyce WM, Runstadler JA. 2016.
571 Transmission of influenza reflects seasonality of wild birds across the annual cycle.
572 *Ecol Lett* doi:10.1111/ele.12629.
- 573 8. Alexander DJ. 2007. An overview of the epidemiology of avian influenza. *Vaccine*
574 25:5637-44.
- 575 9. Lee DH, Bertran K, Kwon JH, Swayne DE. 2017. Evolution, global spread, and
576 pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. *J Vet Sci*
577 18:269-280.
- 578 10. Verhagen JH, Herfst S, Fouchier RA. 2015. Infectious disease. How a virus travels the
579 world. *Science* 347:616-7.
- 580 11. Hoyo JD, Elliott A, Sargatal J, Christie D. 1996. Handbook of the birds of the world, vol
581 1 and 3. Lynx Edicions, Barcelona, Spain.
- 582 12. Van de Kam J, Ens B, Piersma T, Zwartz L. 2004. Shorebirds: an illustrated
583 behavioural ecology. Brill.
- 584 13. Munster VJ, Baas C, Lexmond P, Waldenstrom J, Wallensten A, Fransson T,
585 Rimmelzwaan GF, Beyer WE, Schutten M, Olsen B, Osterhaus AD, Fouchier RA.
586 2007. Spatial, temporal, and species variation in prevalence of influenza A viruses in
587 wild migratory birds. *PLoS Pathog* 3:e61.
- 588 14. Latorre-Margalef N, Tolf C, Grosbois V, Avril A, Bengtsson D, Wille M, Osterhaus AD,
589 Fouchier RA, Olsen B, Waldenstrom J. 2014. Long-term variation in influenza A virus
590 prevalence and subtype diversity in migratory mallards in northern Europe. *Proc Biol*
591 *Sci* 281:20140098.
- 592 15. Lewis NS, Javakhishvili Z, Russell CA, Machabishvili A, Lexmond P, Verhagen JH,
593 Vuong O, Onashvili T, Donduashvili M, Smith DJ, Fouchier RA. 2013. Avian influenza
594 virus surveillance in wild birds in Georgia: 2009-2011. *PLoS One* 8:e58534.
- 595 16. Dusek RJ, Hallgrimsson GT, Ip HS, Jonsson JE, Sreevatsan S, Nashold SW, TeSlaa
596 JL, Enomoto S, Halpin RA, Lin X, Fedorova N, Stockwell TB, Dugan VG, Wentworth
597 DE, Hall JS. 2014. North Atlantic migratory bird flyways provide routes for
598 intercontinental movement of avian influenza viruses. *PLoS One* 9:e92075.
- 599 17. Lindsay LL, Kelly TR, Plancarte M, Schobel S, Lin X, Dugan VG, Wentworth DE, Boyce
600 WM. 2013. Avian influenza: mixed infections and missing viruses. *Viruses* 5:1964-77.
- 601 18. Fries AC, Nolting JM, Bowman AS, Lin X, Halpin RA, Wester E, Fedorova N, Stockwell
602 TB, Das SR, Dugan VG, Wentworth DE, Gibbs HL, Slemons RD. 2015. Spread and
603 persistence of influenza A viruses in waterfowl hosts in the North American Mississippi
604 migratory flyway. *J Virol* 89:5371-81.
- 605 19. Nolting JM, Fries AC, Gates RJ, Bowman AS, Slemons RD. 2016. Influenza A Viruses
606 from Overwintering and Spring-Migrating Waterfowl in the Lake Erie Basin, United
607 States. *Avian Dis* 60:241-4.
- 608 20. Bahl J, Vijaykrishna D, Holmes EC, Smith GJ, Guan Y. 2009. Gene flow and
609 competitive exclusion of avian influenza A virus in natural reservoir hosts. *Virology*
610 390:289-97.
- 611 21. Fourment M, Darling AE, Holmes EC. 2017. The impact of migratory flyways on the
612 spread of avian influenza virus in North America. *BMC Evol Biol* 17:118.
- 613 22. Chen R, Holmes EC. 2009. Frequent inter-species transmission and geographic
614 subdivision in avian influenza viruses from wild birds. *Virology* 383:156-61.
- 615 23. Anderson TK, Campbell BA, Nelson MI, Lewis NS, Janas-Martindale A, Killian ML,
616 Vincent AL. 2015. Characterization of co-circulating swine influenza A viruses in North
617 America and the identification of a novel H1 genetic clade with antigenic significance.
618 *Virus Res* 201:24-31.
- 619 24. Lewis NS, Verhagen JH, Javakhishvili Z, Russell CA, Lexmond P, Westgeest KB,
620 Bestebroer TM, Halpin RA, Lin X, Ransier A, Fedorova NB, Stockwell TB, Latorre-
621 Margalef N, Olsen B, Smith G, Bahl J, Wentworth DE, Waldenstrom J, Fouchier RA,
622 de Graaf M. 2015. Influenza A virus evolution and spatio-temporal dynamics in
623 Eurasian wild birds: a phylogenetic and phylogeographical study of whole-genome
624 sequence data. *J Gen Virol* 96:2050-60.

- 625 25. OIE. 2015. Avian influenza (Infection with avian influenza viruses), Avian influenza
626 (infection with avian influenza viruses): Manual of Diagnostic Tests and Vaccines for
627 Terrestrial Animals. World Organisation for Animal Health (OIE), Paris, France.
- 628 26. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE.
629 2009. Single-reaction genomic amplification accelerates sequencing and vaccine
630 production for classical and Swine origin human influenza A viruses. *J Virol* 83:10309-
631 13.
- 632 27. Zhou B, Wentworth DE. 2012. Influenza A virus molecular virology techniques.
633 *Methods Mol Biol* 865:175-92.
- 634 28. Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D,
635 Pickett BE, Zhang Y, Larsen CN, Ramsey A, Zhou L, Zaremba S, Kumar S, Deitrich J,
636 Klem E, Scheuermann RH. 2012. Influenza research database: an integrated
637 bioinformatics resource for influenza research and surveillance. *Influenza Other Respir*
638 *Viruses* 6:404-16.
- 639 29. Munster VJ, Baas C, Lexmond P, Bestebroer TM, Guldemeester J, Beyer WEP, de
640 Wit E, Schutten M, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. 2009.
641 Practical considerations for high-throughput influenza A virus surveillance studies of
642 wild birds by use of molecular diagnostic tests. *Journal of Clinical Microbiology* 47:666-
643 673.
- 644 30. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez R, Lara-Puente
645 JH, Castro-Peralta F, Cunha LF, Trovao NS, Lozano-Dubernard B, Rambaut A, van
646 Bakel H, Garcia-Sastre A. 2016. Origins of the 2009 H1N1 influenza pandemic in swine
647 in Mexico. *Elife* 5.
- 648 31. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
649 7: improvements in performance and usability. *Mol Biol Evol* 30:772-80.
- 650 32. McCauley J, Bye J, Elder K, Gething MJ, Skehel JJ, Smith A, Waterfield MD. 1979.
651 Influenza virus haemagglutinin signal sequence. *FEBS Lett* 108:422-6.
- 652 33. Burke DF, Smith DJ. 2014. A recommended numbering scheme for influenza A HA
653 subtypes. *PLoS One* 9:e112302.
- 654 34. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics
655 Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870-4.
- 656 35. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and
657 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol*
658 *Evol* 32:268-74.
- 659 36. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017.
660 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*
661 14:587-589.
- 662 37. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal
663 structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus*
664 *Evol* 2:vew007.
- 665 38. Bell SM, Bedford T. 2017. Modern-day SIV viral diversity generated by extensive
666 recombination and cross-species transmission. *PLoS Pathog* 13:e1006466.
- 667 39. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an
668 efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* 31:1929-
669 36.
- 670 40. Parker J, Rambaut A, Pybus OG. 2008. Correlating viral phenotypes with phylogeny:
671 accounting for phylogenetic uncertainty. *Infect Genet Evol* 8:239-46.
- 672 41. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu
673 L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic
674 inference and model choice across a large model space. *Syst Biol* 61:539-42.
- 675 42. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
676 BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969-73.
- 677 43. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-
678 generation sequencing data. *Bioinformatics* 28:3150-2.

- 679 44. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets
680 of protein or nucleotide sequences. *Bioinformatics* 22:1658-9.
- 681 45. Kawaoka Y, Gorman OT, Ito T, Wells K, Donis RO, Castrucci MR, Donatelli I, Webster
682 RG. 1998. Influence of host species on the evolution of the nonstructural (NS) gene of
683 influenza A viruses. *Virus Res* 55:143-56.
- 684 46. Nelson MI, Detmer SE, Wentworth DE, Tan Y, Schwartzbard A, Halpin RA, Stockwell
685 TB, Lin X, Vincent AL, Gramer MR, Holmes EC. 2012. Genomic reassortment of
686 influenza A virus in North American swine, 1998-2011. *J Gen Virol* 93:2584-9.
- 687 47. Lu L, Lycett SJ, Leigh Brown AJ. 2014. Reassortment patterns of avian influenza virus
688 internal segments among different subtypes. *BMC Evol Biol* 14:16.
- 689 48. Ma HC, Chen JM, Chen JW, Sun YX, Li JM, Wang ZL. 2007. The panorama of the
690 diversity of H5 subtype influenza viruses. *Virus Genes* 34:283-7.
- 691 49. Tonnessen R, Hauge AG, Hansen EF, Rimstad E, Jonassen CM. 2013. Host
692 restrictions of avian influenza viruses: in silico analysis of H13 and H16 specific
693 signatures in the internal proteins. *PLoS One* 8:e63270.

694

695

696

697

698

699

700

701

702

703

704

705

706

Main text figure legends:

Figure 1. Bar chart showing total number of positive samples (top) and total number of samples (bottom) collected each year. X-axis shows the year and Y-axis shows the number of samples. Bars coloured according to host from which samples were isolated: Duck – red, Gull – blue and Other birds – green.

Figure 2. Yearly prevalence of viruses in Georgia during 2010-16: (A) Overall (B) Seasonal (C) HA subtype-wise and (D) region-wise. In panel A, the Y-axis marks the prevalence of virus \pm standard deviation and bars are colored according to host from which virus was isolated (duck in pink and gull in green), and the X axis marks the time of isolation. In panels B and D, the Y-axis marks the prevalence of virus and the upper and lower bounds of 95% confidence intervals, and the X axis marks the time of isolation. In heat map in panel C, the Y-axis shows the HA subtypes of viruses isolated and squares are colored according to the number of isolates of each type identified.

Figure 3. Maximum-likelihood trees for all internal genes – PB2, PB1, MP, NS, NP and PA, from equivalent strains connected across the trees. Tips and connecting lines are coloured according to host type: BMG – Black-headed and Mediterranean gulls (light blue), YAG – Yellow-legged and Armenian gulls (blue), MD – Mallard (red), and OD – Other ducks (orange).

Figure 4. Maximum-Likelihood trees for each gene segment of AIV isolated in Georgia 2010-16. Branch supports are indicated by the approximate Likelihood Ratio Test (aLRT) values. Tip labels are coloured according to the type of bird the strain was isolated from: BMG – Black-headed and Mediterranean gulls (red), YAG – Yellow-legged and Armenian gulls (purple), MD – Mallard (blue), and OD – Other ducks (green).

Figure 5. Overall per-site nucleotide diversity defined as average number of nucleotide differences per site between two sequences in all possible pairs in the sample population, normalised to the number of sequences in each population. Comparison between (A) gulls and ducks. (B) host-types: BMG – Black-headed and Mediterranean gulls, YAG – Yellow-legged and Armenian gulls, MD – Mallard, and OD – Other ducks, and (C) HA type are shown.

Figure 6. Root to tip regression for ML trees generated from each internal gene of viruses (MP, NP, NS-A and B, PA, PB1, PB2 as well as the NS-B allele only), isolated from Georgia 2010-16 using Tempest v1.5 and plotted in R v3.2.

Figure 7. Overall/summary (A) and over-time/skyline (B) mean diversity for each segment from gulls (green) and ducks (pink) as determined by posterior analysis of coalescent trees (PACT). Here, diversity is defined as the average time to coalescence for pairs of lineages belonging to each host. Panel (C) shows overall/summary mean diversity values for ducks divided in to MD – Mallard, OD – Other ducks (light and dark blue), and gulls divided into BMG – Black-headed and Mediterranean gulls and YAG – Yellow-legged and Armenian gulls (light and dark green).

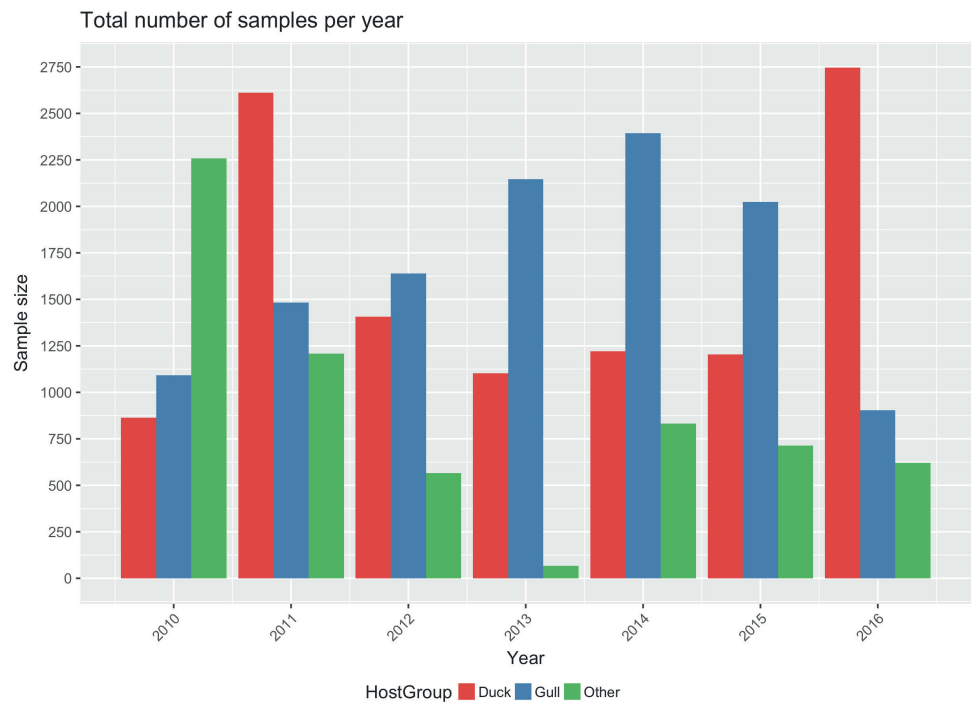
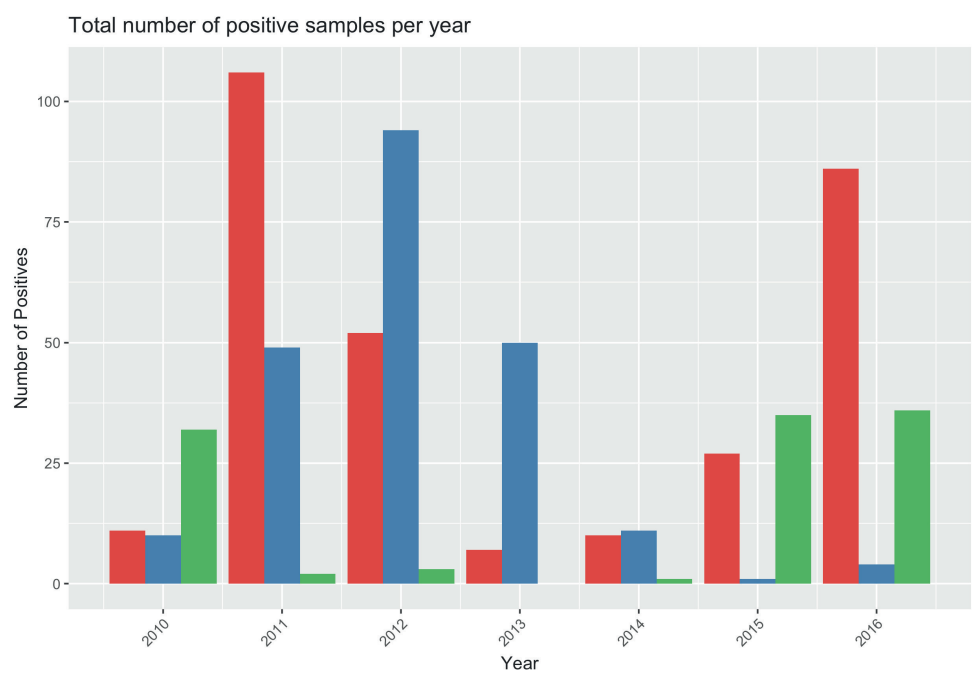
754 **Figure 8.** Summaries of expected/observed ratios from Bayesian Tip-association
755 Significance testing (BaTS) for all internal genes. Higher values indicate less
756 phylogenetic clustering by trait and hence higher rates of mixed ancestry. Comparison
757 between (A) gulls and ducks. (B) host-types (BMG – Black-headed and Mediterranean
758 gulls, YAG – Yellow-legged and Armenian gulls, MD – Mallard, and OD – Other ducks)
759 and (C) HA type are shown. Asterisks indicate p-values ($*** < 0.001$, $** < 0.01$, $* < 0.05$
760 and no asterisk > 0.05).

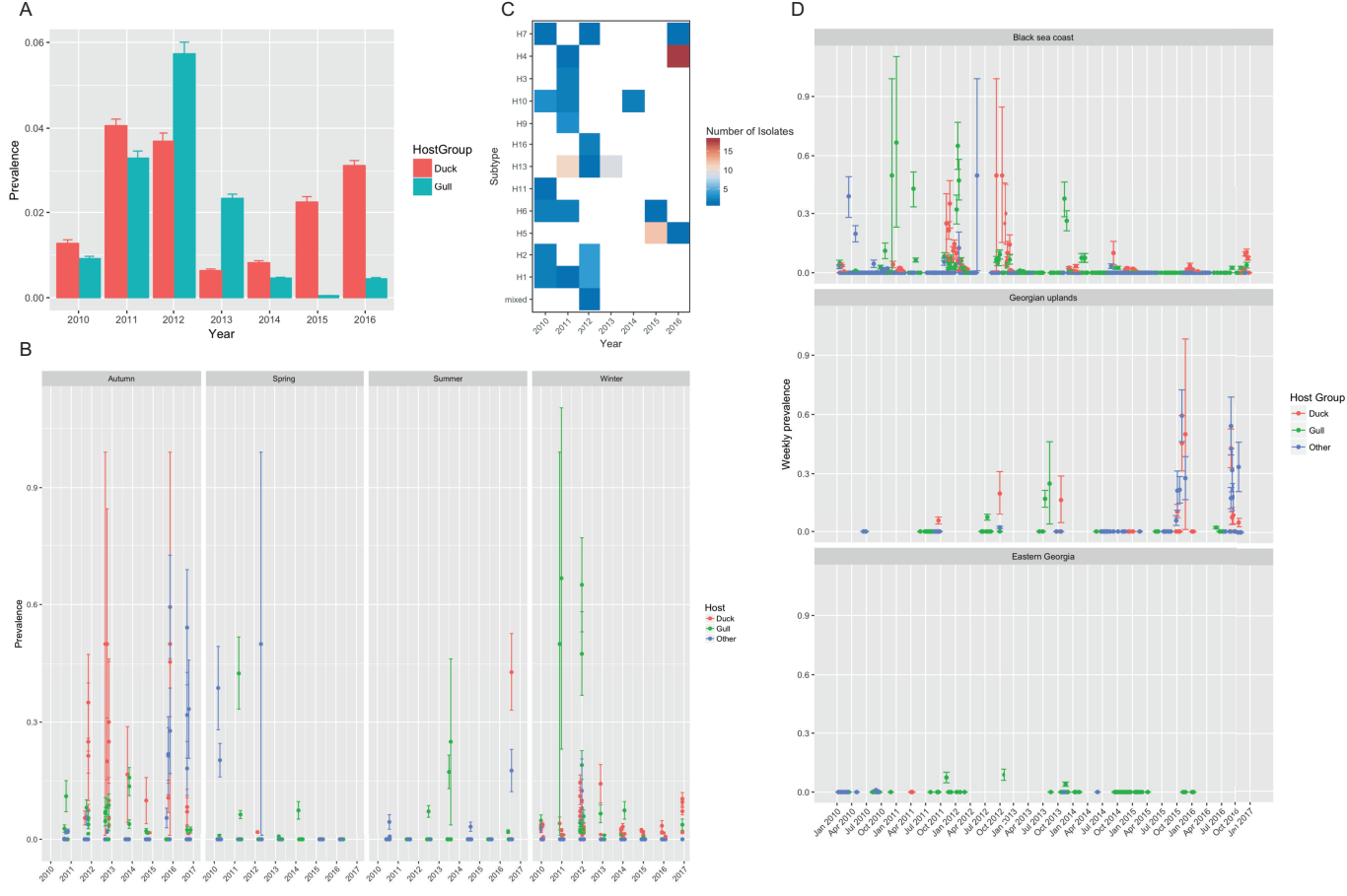
761
762 **Figure 9.** Maximum clade credibility (MCC) trees for five of six internal gene segments of
763 AIV isolated in Georgia 2010-16. Node icons are colored according to “host type” state inferred
764 by BEAST v1.8.4. BMG – Black-headed and Mediterranean gulls (red), YAG – Yellow-legged
765 and Armenian gulls (purple), MD – Mallard (blue), and OD – Other ducks (green).

766 **Figure 10.** Summary of mean migration events between hosts in the direction from
767 (A) duck to gull and gull to duck, and (B) between different host types (BMG – Black-
768 headed and Mediterranean gulls, YAG – Yellow-legged and Armenian gulls, MD –
769 Mallard, and OD – Other ducks) derived from the genealogy.

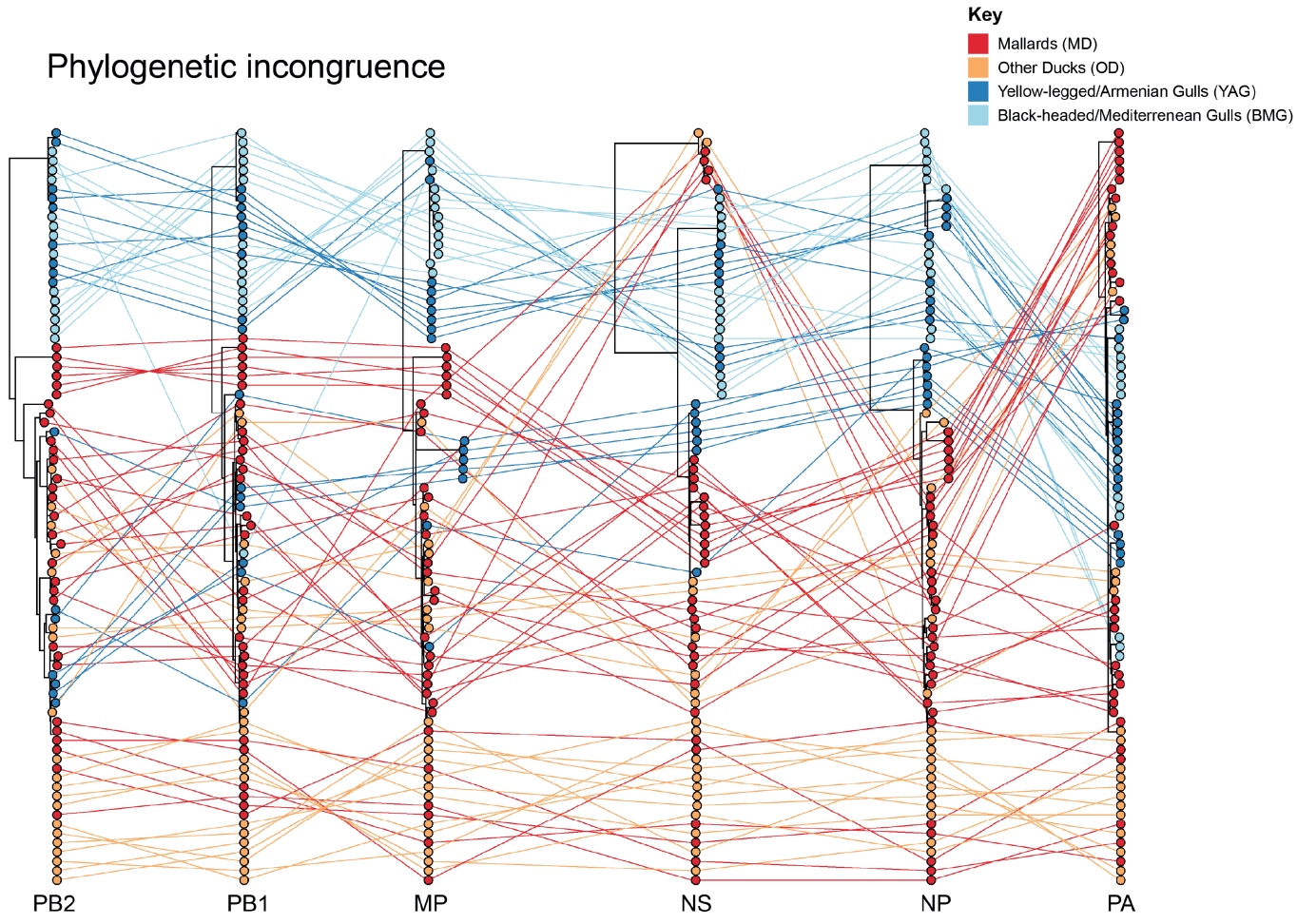
770
771 **Figure 11.** BEAST MCC (median-clade credibility) trees from viral sequences NP
772 gene sequences isolated world-wide from avian hosts between 2005 and 2016.
773 Branches are coloured according to location observed at the tips and estimated at
774 internal nodes by ancestral reconstruction of discrete trait. African strains in dark
775 green, Asian in orange, European in purple, Georgian in pink, North American in light
776 green, and South American in yellow. Nodes with posterior probability > 0.85 are
777 annotated with a diamond icon in the same colour as the branch.

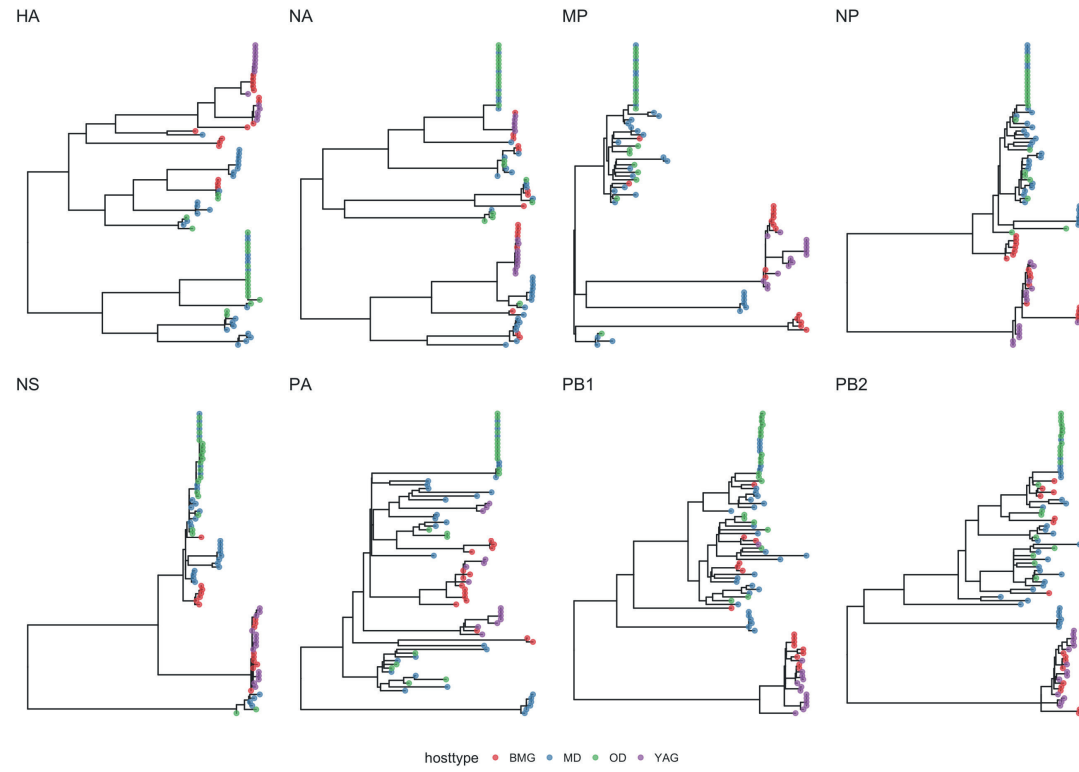
778
779 **Figure 12.** Circularised graph shows overall rates of migration, defined as the rate at
780 which labels (locations) change over the course of the genealogy, between Georgia
781 and other locations. Arrow heads indicate direction of migration; rates are measured
782 as migration events per lineage per year (indicated by the width of the arrow). Asia in
783 blood orange, Africa in orange, Georgia in yellow, Europe in green, South America in
784 teal and North America in blue.

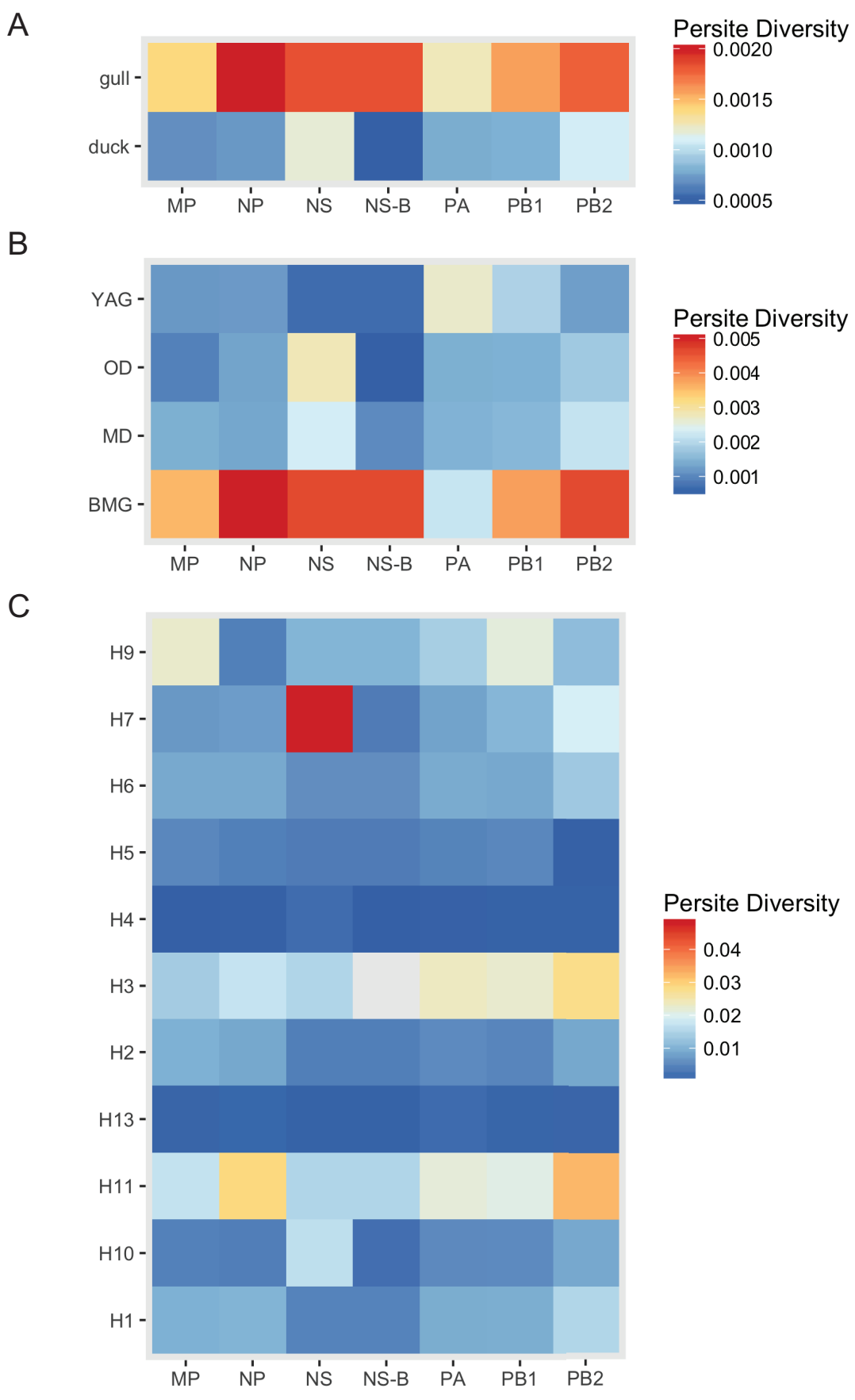


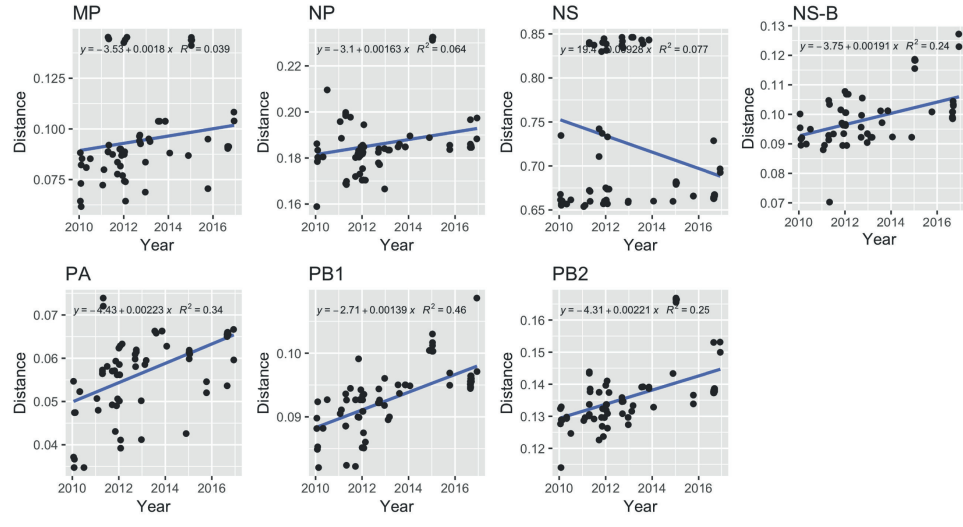


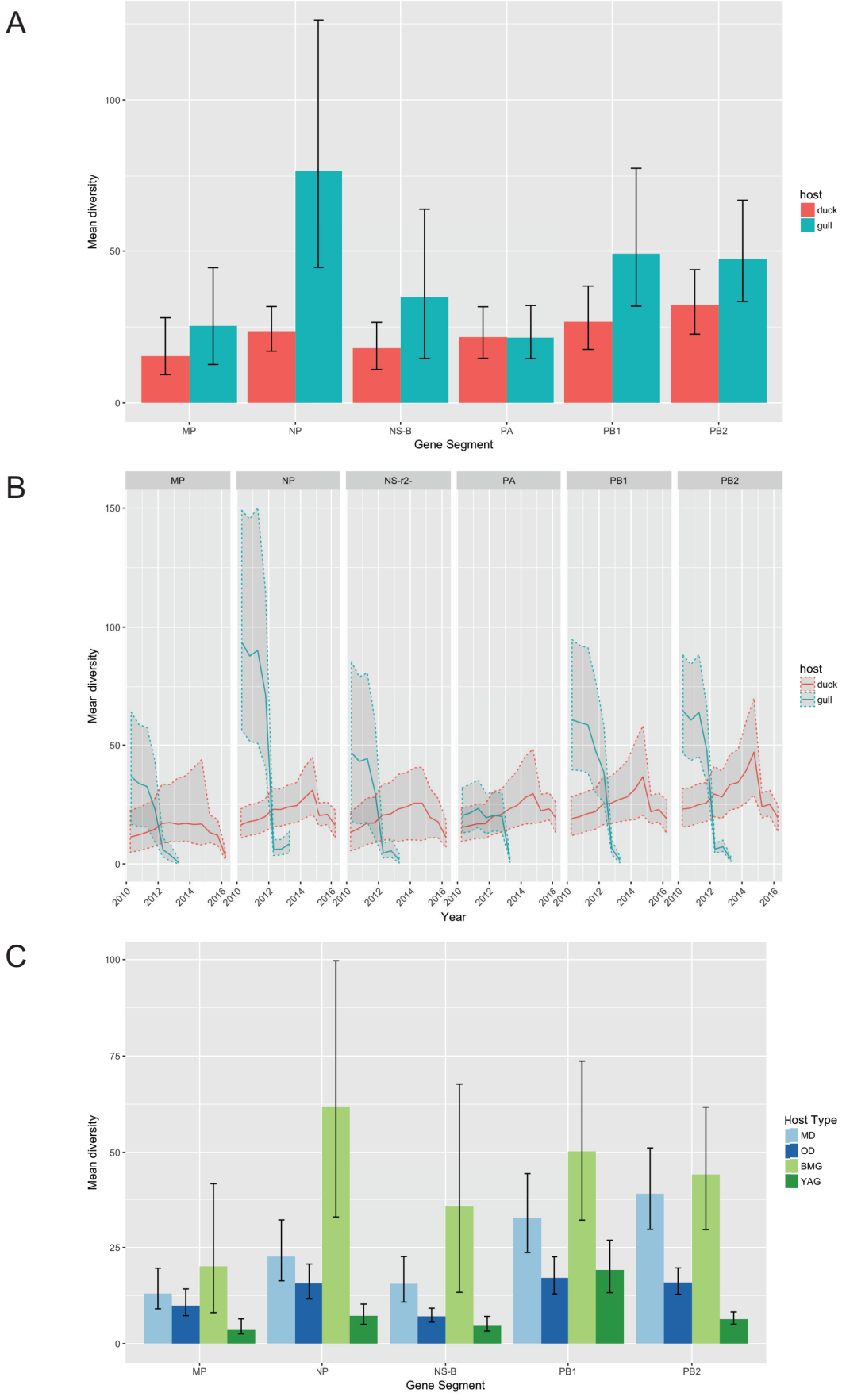
Phylogenetic incongruence

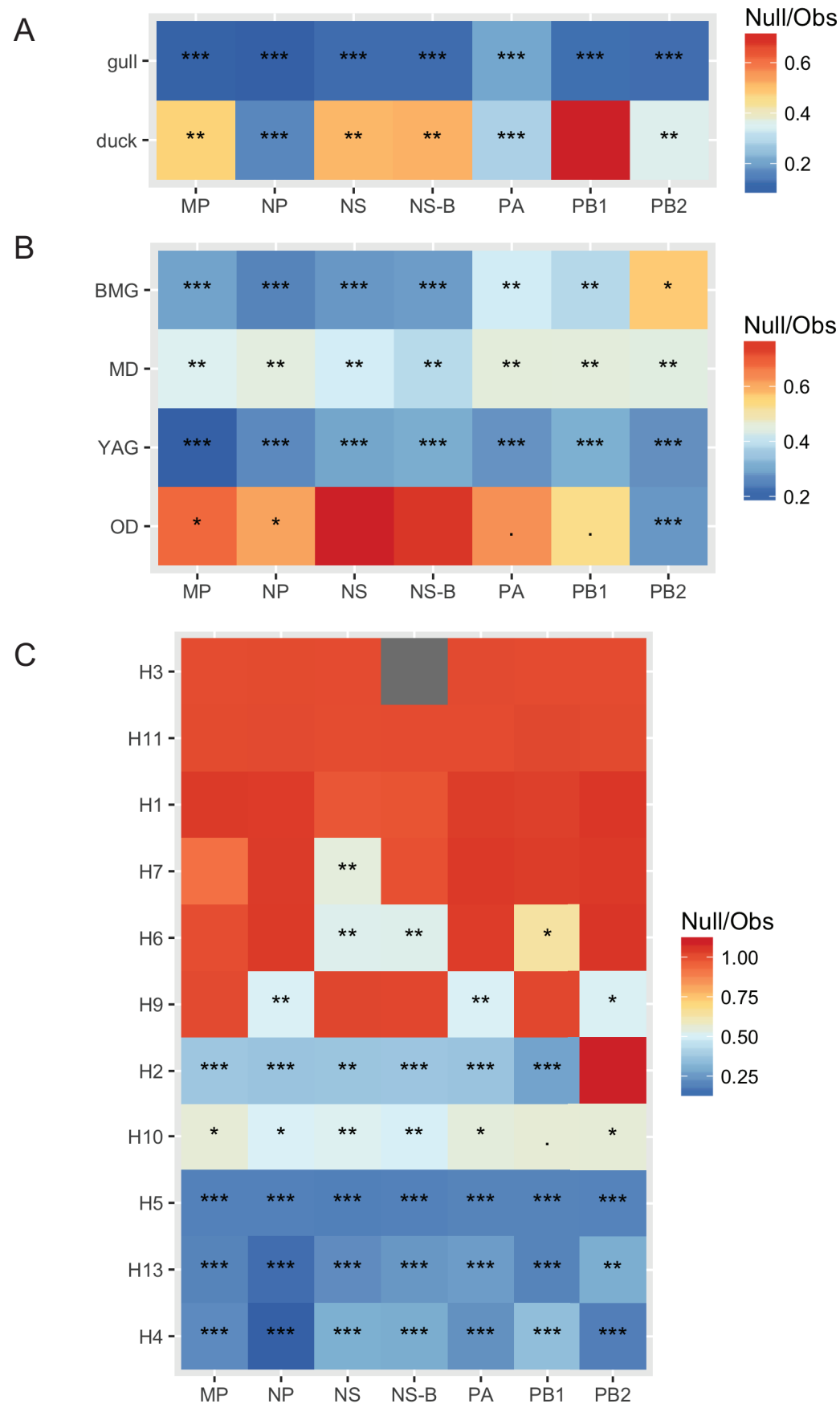


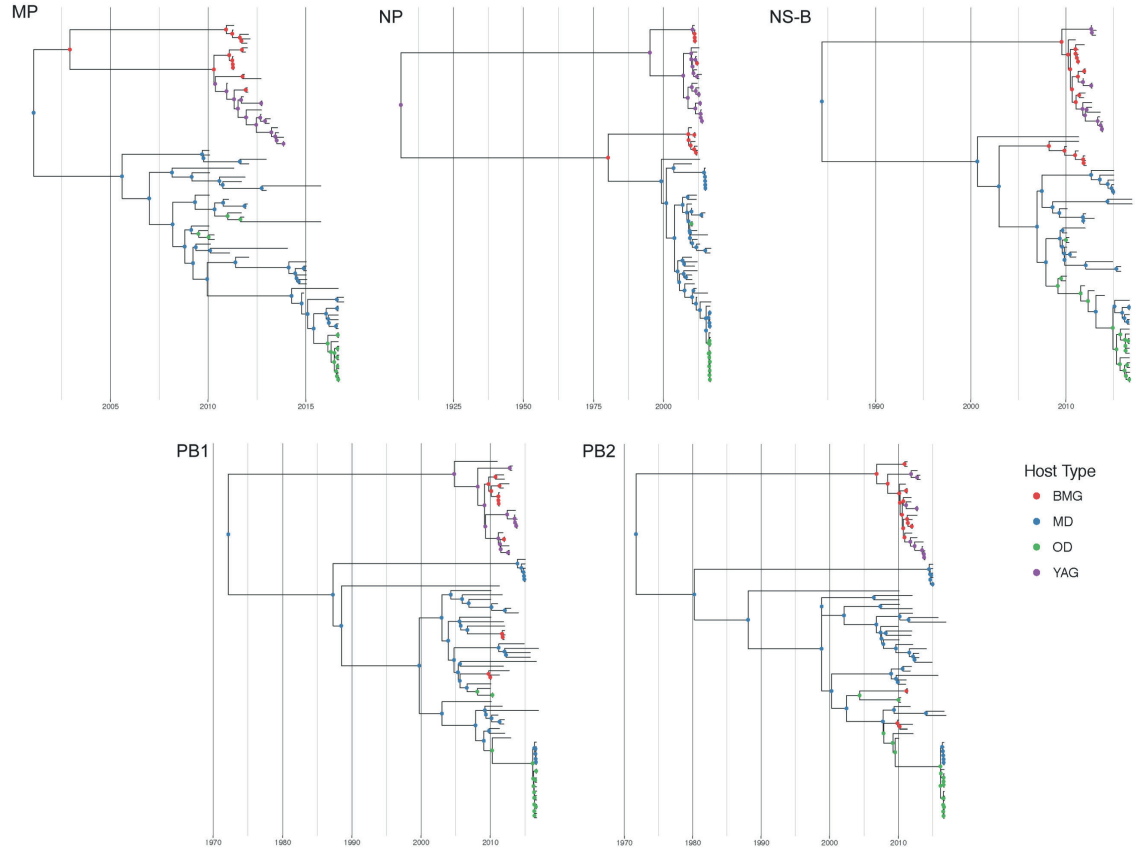




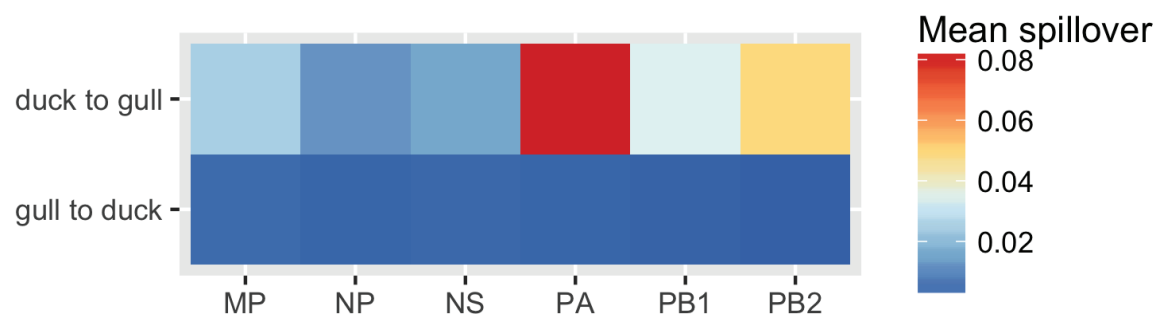








A



B

