Immune genes are hotspots of shared positive selection across birds and mammals

Allison J. Shultz (1,2,3), Timothy B. Sackton (1)

Affiliations:

1. Informatics Group, Harvard University, Cambridge, USA

2. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

3. Museum of Comparative Zoology, Harvard University, Cambridge, USA

Correspondance to AJS (allisonjshultz@gmail.com) or TBS (tsackton@fas.harvard.edu)

Consistent patterns of positive selection in functionally similar genes can suggest a common selective pressure across a group of species. We use alignments of orthologous protein-coding genes from 39 species of birds to estimate parameters related to positive selection for 11,000 genes conserved across birds. We show that functional pathways related to the immune system, recombination, lipid metabolism, and phototransduction are enriched for positively selected genes. By comparing our results with mammalian data, we find a significant enrichment for positively selected genes shared between taxa, and that these shared selected genes are enriched for viral immune pathways. Using pathogen-challenge transcriptome data, we show that genes up-regulated in response to pathogens are also enriched for positively selected genes. Together, our results suggest that pathogens, particularly viruses, consistently target the same genes across divergent clades, and that these genes are hotspots of host-pathogen conflict over deep evolutionary time.

INTRODUCTION

Central to the study of evolutionary biology is the desire to understand how natural selection operates across a diverse set of populations and species. While many selective pressures vary across taxa, some common selective pressures may result in consistent patterns of natural selection across a set of species. By taking an unbiased approach and scanning all orthologous genes across a set of species for signatures of positive selection, it may be possible to identify functional patterns that indicate shared selective pressures.

Early comparative genomic studies on primates, mammals, bees, ants, *Drosophila* and other organisms (Schlenke and Begun 2003; Sackton et al. 2007; Kosiol et al. 2008; Barreiro and Quintana-Murci 2009; Roux et al. 2014) that included unbiased selection scans identified immune system pathways as common targets of natural selection. This implies that pathogens, which elicit the immune response, may be strong and consistent selective forces across species. Furthermore, in several clades of invertebrates, receptor genes, or the genes interacting directly with pathogens are most often the target of positive selection (Sackton et al. 2007; Waterhouse et al. 2007; Ellis et al. 2012). Finally, recent studies of mammals show that proteins that interact with viruses experience about twice as many amino acid changes compared to proteins that do not (Enard et al. 2016) and proteins that interact with *Plasmodium* experience elevated rates of adaptation (Ebel et al. 2017). While this evidence clearly implicates pathogens as a major selective force shaping the evolution of genomes, the availability of many new genomes now allows detailed

41 comparisons between clades to test the degree to which the specific genes
42 represent shared hotspots of positive selection. Furthermore, linking the results
43 of positive selection scans to comparative functional data would provide greater
44 insight into the role of pathogens in driving shared selection across clades.

45 The number of bird (class Aves) genomes has increased dramatically in
46 recent years (e.g. Zhang, B. Li, et al. 2014), and provides an opportunity to study
47 genome-wide signatures of positive selection in this ecologically important group.
48 Birds are a radiation of approximately 10,000 species (Clements et al. 2016) that
49 possess diverse morphologies and behaviors (Gill 2007). They have a global
50 distribution and diverse range of habitats (Jetz et al. 2012), and many species
51 migrate thousands of miles annually (Gill 2007), making them excellent models
52 for studies of disease ecology. From a genomic perspective, they have small
53 genomes, generally stable chromosomes, little repeat content, and low rates of
54 gene loss and gain (Organ et al. 2007; Organ and Edwards 2011; Zhang and
55 Edwards 2012; Ellegren 2013; Zhang, C. Li, et al. 2014). Birds have the same
56 general blueprint of immune pathways as mammals, but with a slimmed down
57 gene repertoire and some small differences in the functions of specific genes
58 (Kaiser 2010; Chen et al. 2013; Juul-Madsen et al. 2014). Studies of the
59 evolutionary dynamics of avian immune genes have almost exclusively focused
60 on the major histocompatibility complex genes (MHC) or TLRs, with evidence of
61 positive selection across species in MHC class I genes (Alcaide et al. 2013),
62 MHC class II genes (Edwards et al. 1995; Edwards et al. 2000; Hess and
63 Edwards 2002; Burri et al. 2008; Burri et al. 2010) and TLRs (Alcaide and

Edwards 2011; Grueber et al. 2014; Velová et al. 2018). From a broader perspective, the conclusions drawn from more general studies of positive selection across birds have been limited by including only a few species (e.g. Nam et al. 2010), or using low-power analysis methods (e.g. comparing overall dN/dS values across GO-terms (Zhang, C. Li, et al. 2014)).

We use comparative genomics in birds to study genome-wide signatures of positive selection without any *a priori* assumptions of gene functions. We find that the strongest signatures of selection are concentrated in four general categories: immune system genes, genes involved in recombination and replication, genes involved in lipid metabolism, and phototransduction genes. By comparing avian and mammalian datasets, we show that genes under positive selection in birds are likely to be under positive selection in mammals, and that this signal is the strongest in viral defense immune pathways. Finally, we show that genes up-regulated following a pathogen challenge are more likely to be under positive selection in birds, that there is also an overlap in birds and mammals in genes up-regulated in response to pathogens, particularly viruses, and that some of the classic genes studied as targets of host-pathogen co-evolution (PKR, MX1), are under selection and differentially expressed in both clades. Together, all of our results support the hypothesis that pathogens consistently target the same genes across deep evolutionary timescales.

87 **RESULTS**

88 **Strong signatures of positive selection throughout the avian genome**

89 We used PAML and HyPhy site models to test genes for evidence of positive
90 selection across birds. We ran all models for 11,231 genes using the gene tree
91 as the input tree and for 8,669 genes using the species tree as the input tree. To
92 test for positive selection for each gene, we conducted likelihood ratio tests
93 between models that include an extra ω parameter for some proportion of sites
94 and models that do not include the extra ω parameter. An FDR-corrected p-value
95 from that likelihood ratio test less than 0.05 is considered evidence of positive
96 selection for that gene. For all model comparisons (PAML models described in
97 Table 1), we found that between 17% and 73% of genes are under positive

Table 1: PAML Model descriptions.

model	model description	parameters
M0	one ratio	ω
M1a	neutral	p0 (p1 = 1 - p0) ω0 < 1, ω1 = 1
M2a_fixed	neutral	p0,p1 (p1 = 1 - p0 - p1) ω0 < 1, ω1 = 1, ω2 = 1
M2a	selection	p0,p1 (p1 = 1 - p0 - p1) ω0 < 1, ω1 = 1, ω2 > 1
M7	neutral (beta distribution)	p, q
M8a	neutral (beta distribution)	p0 (p1 = 1 - p0), p, q, ωs = 1
M8	selection (beta distribution	p0 (p1 = 1 - p0), p, q, ωs > 1

98 selection (Table 2). About 20% of genes are positively selected with the more
99 conservative M1a vs. M2a tests or M2a vs M2a_fixed tests, with large overlaps
100 among the genes identified. The less conservative M7 vs. M8 tests show much
101 greater proportions of positively selected genes (~70%), although this is reduced

Table 2. Counts (above) and proportions (*below*) for all tests of individual, and combined tests of selection for gene trees and species trees.

dataset	n genes	m1a vs m2a	m2a vs m2a_fixed	m7 vs m8	m8 vs m8a	all PAML	BUSTED	all PAML + BUSTED
gene trees	11231	1925 *0.17*	2197 *0.20*	7504 *0.67*	3679 *0.33*	1901 *0.17*	6244 *0.56*	1562 *0.14*
species trees	8669	1783 *0.21*	2026 *0.23*	6293 *0.73*	3395 *0.39*	1752 *0.20*	3870 *0.45*	1203 *0.14*

to about 35% with the M8 vs. M8a test, indicating that the M8 model may often improve fit by adding a class of sites with ω very close to 1. HyPhy's BUSTED identified ~50% of genes as positively selected (FDR-corrected p-value less than 0.05). Fewer than half of these genes are also identified as being positively selected by all PAML tests - 1,562 genes with the genes tree as input and 1,203 genes with species tree as input. In total, 14% of analyzed genes are found to be under positive selection in all tests (Table 2, Supplemental Table 1 (raw gene tree results), Supplemental Table 2 (raw species tree results)). We consider these 1,562 genes to be a high-confidence positive selection gene for downstream functional analyses. Compared to all other genes, the high-confidence positive selection gene set has overall higher distributions of M0 model ω values, which assumes a single ω for all sites in a gene (Supplemental Figure 1; Mann-Whitney U-test: gene trees: not-significant median ω = 0.084, significant median ω = 0.344 , $p < 0.0001$; species trees: not-significant ω = 0.084, significant ω = 0.332, $p < 0.0001$).

Gene trees and species trees also had similar distributions overall ω values with the M0 model (K-S test: D=0.006, p=1, Supplemental Figure 2). The mean ω value is 0.15, the median ω is 0.10 and standard deviation is 0.14 using

either the gene or species tree as the input tree. Because of the similarity between gene tree and species tree results, and to minimize issues associated with hemiplasy in species trees (Hahn and Nakhleh 2015; Mendes and Hahn 2016), for all bird-specific analyses below, we use gene tree results to maximize the number of genes tested. However, all results are qualitatively similar with the species tree results as input.

Immune, recombination, lipid metabolism, and phototransduction pathways are enriched for positive selection in birds

For an unbiased perspective on whether or not positively selected genes are concentrated in particular functional pathways, we performed a pathway enrichment test of positively selected genes against a background of all genes tested. With chicken as the reference organism, 351 genes of the high-confidence positive selection gene set and 3,347 of all genes tested could be mapped to KEGG pathways for use as the test set and gene universe respectively. Out of the 166 KEGG pathways available for chicken, or any other bird species, 119 had at least one gene with evidence of positive selection (Supplemental Table 3). We found 18 KEGG pathways that were significantly enriched with positively selected genes (q-value less than 0.1; Figure 1A; Supplemental Table 3). These 18 pathways belong to seven KEGG functional categories: infectious disease, immune system, signaling molecules and interaction, replication and repair, lipid metabolism, and sensory system. Some

143

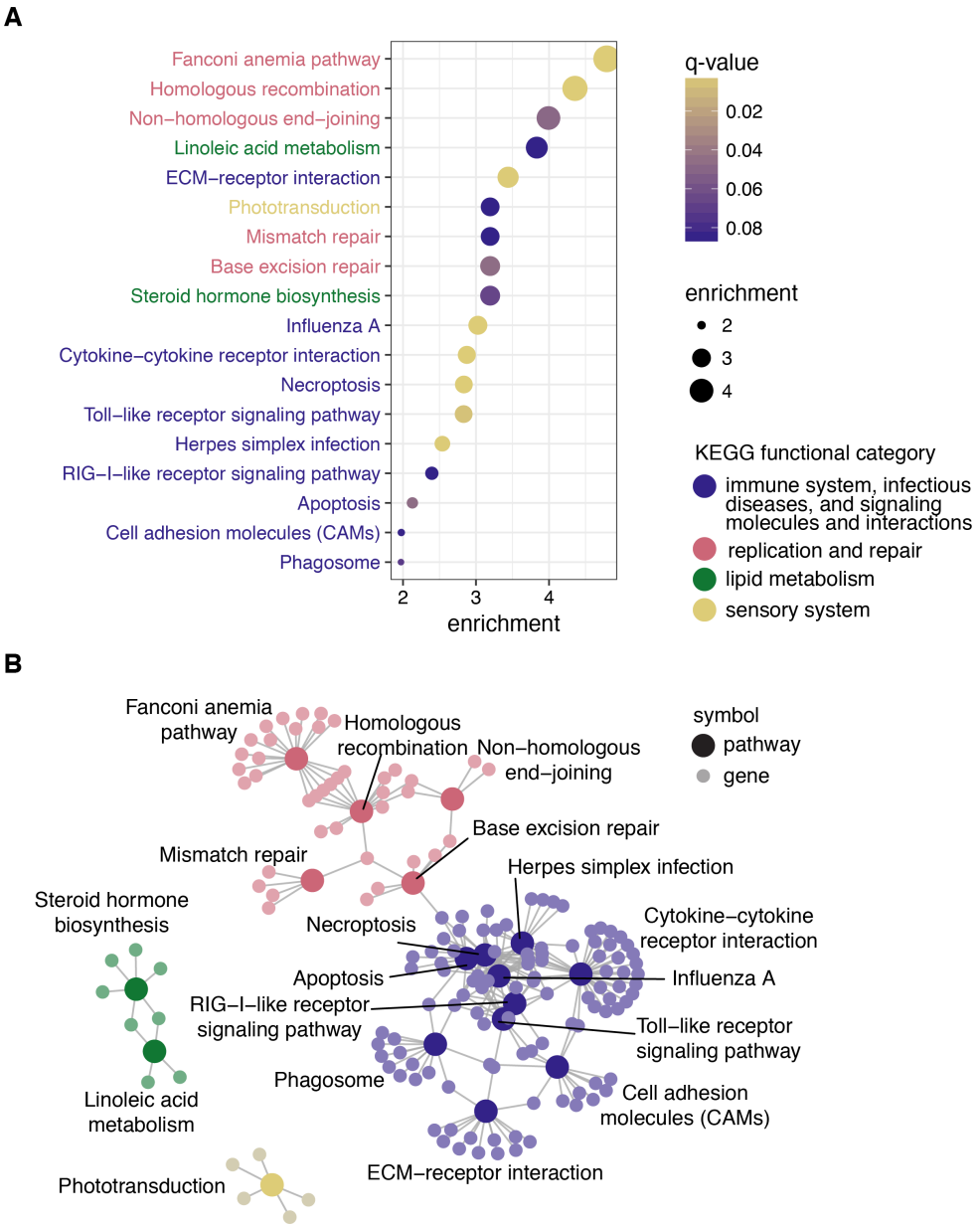


Figure 1. Pathway enrichment results to determine whether positively selected genes are functionally similar with chicken as the reference organism. A. The 18 pathways significant at q-value < 0.1 ordered by enrichment values, calculated as the proportion of genes under selection in the pathway over the proportion of genes significant in all KEGG pathways. Points are filled by q-value, and each pathway is colored by the broader KEGG functional category. B. Map depicting the relationships of all significant genes among pathways. Each gene (small, light circle) is connected by a line to each pathway it belongs to (large, dark circle). Each point is shaded according to the broader KEGG functional category.

144 immune or recombination-related functions. However, many genes are uniquely
145 enriched in a single pathway as well (Figure 1B), suggesting that these
146 enrichment results are not driven by a few core genes present in many pathways.
147 To test whether our pathway enrichment results were robust to reference
148 organism, we also conducted pathway enrichment tests using zebra finch and
149 human as the reference organism. Pathway enrichment results using zebra finch
150 showed similar results as those presented above using chicken, particularly for
151 immune-related pathways and recombination and repair pathways (Supplemental
152 Table 4). Pathway enrichment results using human, which has additional
153 annotated pathways, resulted 37 pathways significantly enriched with positively
154 selected genes (q-value less than 0.1; Supplementary Table 5) out of 269
155 pathways with at least one homologous gene in our dataset. Compared to the
156 enrichment results using chicken, the human pathways primarily added many
157 disease or immune pathways not available for birds, suggesting that the overall
158 functional results are robust to reference organism.
159
160
161 **Lineages clustered by genes under selection in birds are most strongly**
162 **related to body size and lifespan**
163 Codon-based site models typically can only detect positive selection when the
164 same sites in the protein are under selection in numerous lineages. In order to
165 detect selection limited to particular lineages, we relaxed this assumption, and
166 used aBS-REL to estimate of the probability of selection independently at each

167 branch of the phylogeny. To test consistency with our site-model results, we
168 calculated the number of lineages with evidence for positive selection from the
169 branch-site (aBS-REL) for each gene. We found that genes identified by
170 BUSTED as having sites with evidence of positive selection across avian
171 lineages also had significantly more lineages under selection using aBS-REL
172 (Mann-Whitney U-test: median proportion significant lineages under selection
173 given significant BUSTED result: 0.16, median proportion significant lineages
174 under selection given non-significant BUSTED result: 0.05, $p<10^{-16}$,
175 Supplemental Figure 3). While branch-site tests can be subject to false
176 inferences of positive selection due to multinucleotide mutations (Venkat et al.
177 2018), the similar patterns between our site-model results and the aBS-REL
178 results suggests that the overall patterns we observe hold true with this
179 alternative analysis.

180 To identify additional functional classes of genes that may be selected in
181 only a subset of lineages, we used a principle components analysis (PCA) to
182 summarize the variance of log-transformed p-values across genes for each
183 species, and then used phylogenetic comparative methods to identify species
184 traits associated with PC loadings for each gene. We find three PCs that together
185 explain 16.8% of the variance in aBS-REL p-values across species, while the
186 remaining PCs explain about 3% of the variance (Supplemental Figure 4). PC1
187 (Figure 2) identifies a diverse set of species from different sharing similar PC
188 scores, while PC2 and PC3 appear to identify clade-specific selection in

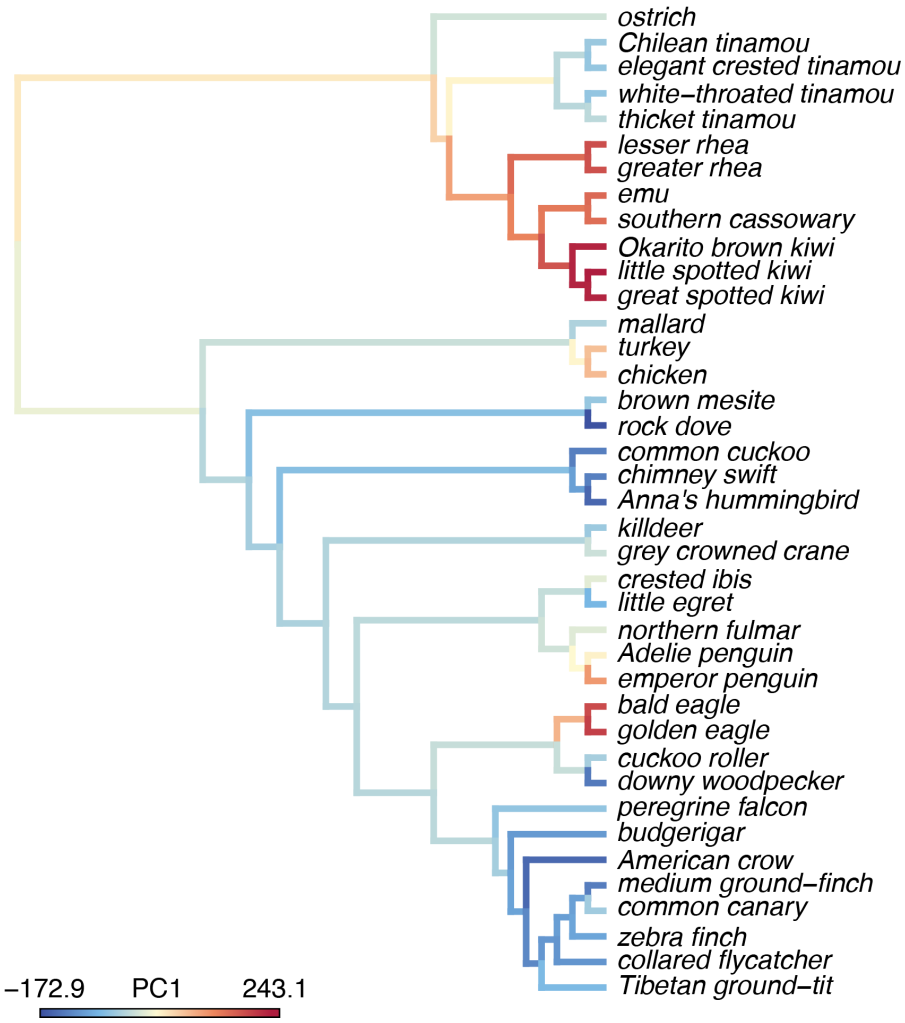


Figure 2. A visualization of PC1 scores on the phylogeny and the maximum likelihood reconstruction of the PC1 values for internal branches. The PC1 scores indicates species in different clades that have similar log-transformed p-values from aBS-REL tests for positive selection and explain 7.7% of the variance across all genes tested.

189 palaeognaths (Supplemental Figure 5) and passerines and allies (Supplemental
190 Figure 6) respectively. We focused on PC1, the only principle component to
191 cluster species in different lineages, in order to test whether it might be
192 associated with life history. We found a correlation between log-transformed
193 body mass, a proxy for many life history characteristics, and PC1 scores for each

194 species using phylogenetic generalized least squares (Figure 3, β = -24.97, SE =

195 6.46, t-value = -3.9, p-value = 0.0004).

196 To understand any functional signal in the genes most strongly correlated

197 with body size, we calculated a Spearman's rank correlation for each gene using

198 log-transformed aBS-REL p-values for each lineage compared to log-

199 transformed body mass. We performed KEGG pathway gene set enrichment

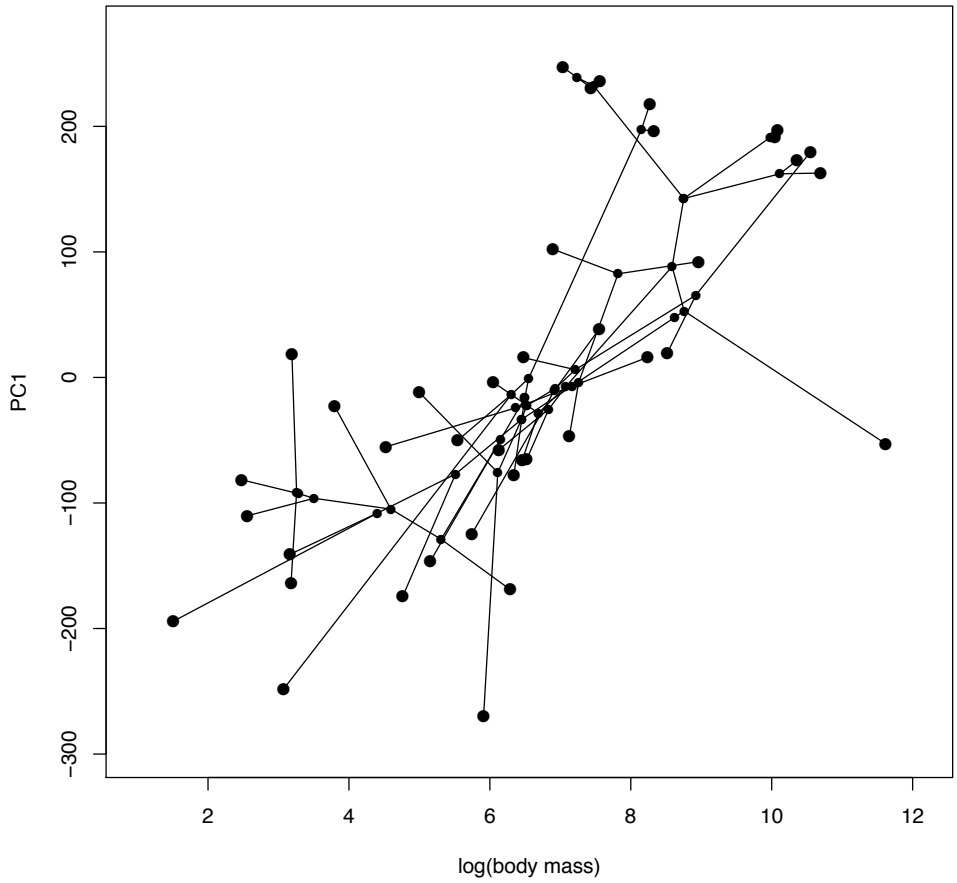


Figure 3. A phylomorphospace plot showing the association between log-transformed body mass on the x-axis and PC1 on the y-axis. Species values and reconstructed node values are connected by phylogeny. A PGLS analysis of these two traits showed a significant correlation (p = 0.0004).

200 using the ρ value for each gene. Only one pathway had a q-value less than one
201 (Supplemental Table 6). The cellular senescence pathway had a q-value of 0.26
202 and a normalized enrichment score of 1.62.
203
204
205 **Convergent signatures of selection in birds and mammals are enriched for**
206 **viral-interacting pathways**
207 We investigated whether we could detect signatures of pathogen-mediated
208 selection at deeper time scales by testing whether the same genes are
209 repeatedly under selection in both birds and mammals, and whether those genes
210 are clustered in functional pathways. We combined our results with those from
211 Enard et al. (2016), a study that used HyPhy's BUSTED program to test for
212 positive selection in 9,681 orthologous genes from 24 mammal genomes. To
213 best match the experimental procedures used by Enard et al. (2016), for bird-
214 mammal comparisons we only used our BUSTED results with the species tree as
215 the input phylogeny. The combined dataset consisted of 4,931 orthologous
216 genes with results in both clades.
217 We first tested for significant overlap in positively selected genes in both
218 clades with a Fisher's exact test. To understand whether these results were
219 driven by genes with different levels of evidence for positive selection, we used
220 four different FDR-corrected p-value cutoffs for significance, 0.1, 0.01, 0.001, and
221 0.0001. We found evidence for a significant overall overlap in positively-selected

genes at all four different FDR-corrected p-value cutoffs, with stronger signal at smaller FDR-corrected p-values (Figure 4A, Supplemental Table 7).

We tested for functional enrichment in shared selected genes compared to all genes under selection in birds. KEGG enrichment with a test set of genes under selection in both clades and gene universe of genes under selection in birds, showed that pathways with immune function, particularly viral-interacting pathways, are significantly enriched for convergent signatures of selection. These results are particularly significant at the lowest FDR-corrected p-value significance cutoffs (Figure 4B). As pathways enriched for positively selected genes in birds have higher enrichment values than other pathways (Figure 4B), we also conducted 1,000 randomized enrichment tests to make sure pathways with more genes under selection in birds are not more likely to show more positively selected genes in both lineages by chance. We calculated multiple test corrected p-values for the empirical enrichment scores compared to the randomly generated null distribution within each of the four FDR-corrected p-value cutoffs for significance. These results corroborate those of KEGG enrichment tests, with Influenza A and Herpes simplex infection pathways showing significantly higher enrichment values, particularly at lower FDR-corrected p-value cutoffs for significance (Figure 4C).

241

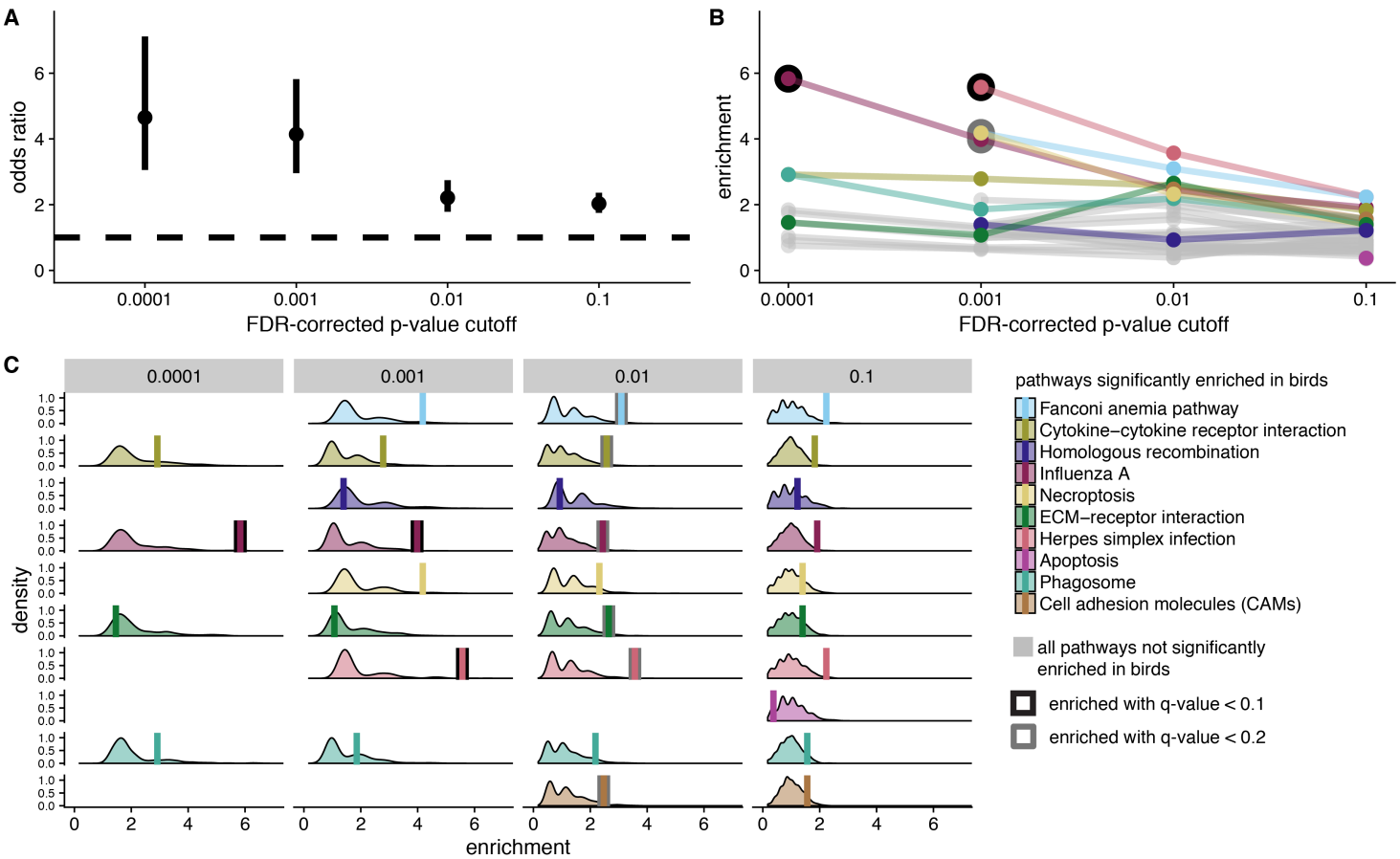


Figure 4. Signatures of convergent positive selection in birds and mammals. For all analyses, we considered four different FDR-corrected p-value cutoffs for significance (to identify genes under positive selection) A. Odds ratio of overlap in genes under selection in both bird and mammal datasets. B. Pathway enrichment scores from KEGG pathway enrichment tests with genes under selection in both birds and mammals as the test set, and genes under selection in birds as the background set. Ten pathways significantly enriched in birds with at least one gene under selection in both birds and mammals are color-coded. All other pathways are shown in grey. Significant enrichment values are outlined in black (q-value < 0.1) or grey (q-value < 0.2). C. Null distribution of enrichment scores generated from 1,000 randomization tests compared to empirical enrichment scores (vertical bars). Null distributions were generated by randomly selecting gene sets from the background set of genes (bird significant genes) for use as the test set. The randomized test set contained the same number of genes as empirical test set for each FDR-corrected p-value cutoff for significance. Empirical enrichment scores are depicted by a vertical bar, and with significant q-value scores outlined in black (q-value < 0.1) or grey (q-value < 0.2).

242 **Genes up regulated in response to pathogens are more likely to be under**
243 **positive selection in birds**
244 We used gene expression data to independently test whether pathogens are
245 likely driving immune-related patterns of positive selection. First, for birds, we
246 tested whether genes that were significantly differentially expressed following a
247 pathogen challenge are more likely to be under positive selection. We used avian
248 transcriptome data from 12 different studies representing seven different types of
249 pathogens, including four viruses, two bacteria, and one species of protist
250 (Supplemental Table 8). We compared both the proportions of positively-selected
251 genes (using the high-confidence positive selection gene set) that were up-
252 regulated following a pathogen challenge compared to those not differentially
253 expressed, and the proportions of positively selected genes that were down-
254 regulated following a pathogen challenge compared to those not differentially
255 expressed. We found that for all pathogens, up-regulated genes are significantly
256 more likely to be positively selected (Figure 5; Supplemental Table 9). The
257 pattern was less consistent for down-regulated genes, with overall smaller
258 numbers of down-regulated genes, weak evidence for a greater proportion of
259 positively selected genes for West Nile Virus and *Plasmodium*, and weak
260 evidence for a smaller proportion of positively selected genes for *E. coli* (Figure
261 5; Supplemental Table 9).
262 We tested for shared pathogen-mediated selection in birds and mammals
263 by comparing bird and mammal gene expression patterns when challenged with
264 the same, or a closely-related pathogen. There were five pathogens with publicly

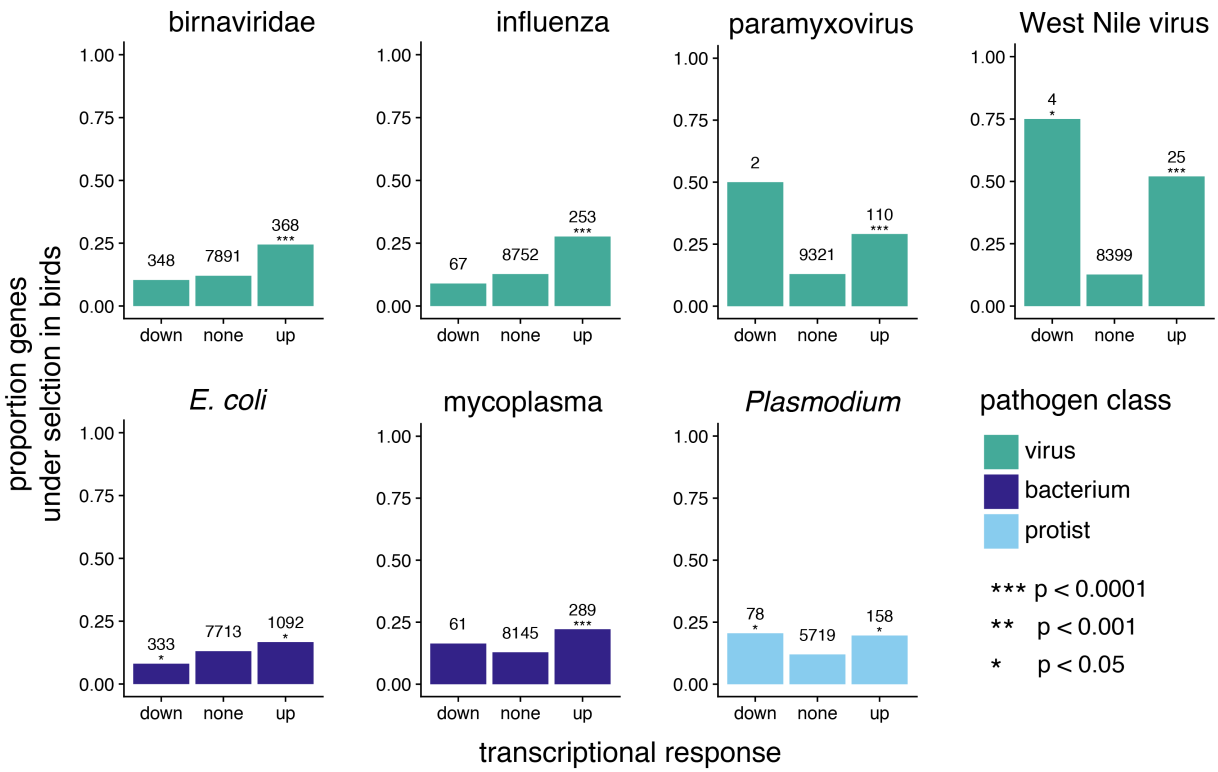


Figure 5. A comparison of genes under positive selection in birds and genes differentially expressed following pathogen challenge to test for patterns of pathogen-mediate selection. For different pathogens, we show the proportion of genes under positive selection in birds (defined as significant with FDR corrected p-value < 0.05 for all PAML and BUSTED model comparisons) for genes down significantly down regulated, significantly up regulated, or not significantly differentially regulated. The number above each bar indicates the number of genes in a given transcriptional response class. The significance of enrichment for positively-selected genes in up- or down-regulated expression classes, as calculated by logistic regression, is indicated by asterisks above the “down” and “up” bars.

265

266 available data for at least one bird and mammal species – two viruses: Influenza
267 A and West Nile virus, two bacteria: *E. coli* and mycoplasma, and one protist:
268 *Plasmodium* (Supplemental Table 8). All comparisons between bird and mammal
269 datasets for viral and bacterial pathogens showed that there was significant
270 overlap in up-regulated genes, but no significant overlap in down-regulated
271 genes (Table 3). Genes differentially expressed in response to *Plasmodium*
272 showed the opposite pattern, with significant overlap in down-regulated genes,
273 but no significant overlap in up-regulated genes (Table 3).

Table 3. Fisher's exact test results from bird and mammal transcriptome studies

pathogen	transcriptional response	n genes diff. expressed in both lineages	n genes expected by chance	p-value	odds ratio (95% conf. intervals)
Influenza	up	30	14.8	< 0.0001	2.48 (1.56,3.84)
Influenza	down	7	7.2	1.000	0.96 (0.35,2.29)
West Nile Virus	up	6	0.9	< 0.0001	25.06 (4.46,253.47)
West Nile Virus	down	0	0	1.000	0 (0, 884.62)
E. coli	up	84	52.4	< 0.0001	1.9 (1.45, 2.47)
E. coli	down	24	20.7	0.397	1.2 (0.73, 1.90)
Mycoplasma	up	16	8.4	0.010	2.1 (1.14, 3.62)
Mycoplasma	down	0	0.1	1.000	0 (0, 47.56)
Plasmodium	up	14	9.8	0.166	1.53 (0.79, 2.77)
Plasmodium	down	9	3.2	0.004	3.44 (1.42, 7.57)

274 Logistic regressions with genes under selection in birds as the response
275 variable and genes under selection in mammals, genes up or down-regulated in
276 birds, and their interaction as predictor variables showed that for all categories,
277 the selection status in mammals is the strongest predictor, followed by the
278 transcriptional response in birds for some pathogens, but no significant
279 interaction between the two (Table 4). Very few genes were under selection in
280 birds and mammals and also differentially expressed in both clades at all FDR-

Table 4. Logistic regression results testing whether genes under selection in birds could be predicted by selection status in mammals (sig_mammals), transcriptional regulation in birds, or their interaction

pathogen	transcriptional response	n genes	predictor variable	estimate	standard error	z score	p-value
influenza	down	4488	sig_mammals	0.76	0.09	8.45	< 0.0001
influenza	down	4488	down_reg_birds	-0.12	0.42	-0.29	0.771
influenza	down	4488	sig_mammals: down_reg_birds	-0.76	0.85	-0.89	0.372
influenza	up	4488	sig_mammals	0.77	0.09	8.52	< 0.0001
influenza	up	4488	up_reg_birds	0.79	0.22	3.69	0.0002
influenza	up	4488	sig_mammals: up_reg_birds	-0.88	0.5	-1.77	0.077
west nile virus	down	3774	sig_mammals	0.77	0.1	7.76	< 0.0001
west nile virus	down	3774	down_reg_birds	11.98	196.97	0.06	0.952
west nile virus	down	3774	sig_mammals: down_reg_birds	-	-	-	-
west nile virus	up	3774	sig_mammals	0.77	0.1	7.76	< 0.0001
west nile virus	up	3774	up_reg_birds	1.1	0.87	1.27	0.203
west nile virus	up	3774	sig_mammals: up_reg_birds	-1.46	1.66	-0.88	0.378
E. coli	down	4225	sig_mammals	0.74	0.09	7.95	< 0.0001
E. coli	down	4225	down_reg_birds	-0.16	0.19	-0.83	0.409
E. coli	down	4225	sig_mammals: down_reg_birds	0.18	0.5	0.36	0.717
E. coli	up	4225	sig_mammals	0.72	0.1	7.2	< 0.0001
E. coli	up	4225	up_reg_birds	0.24	0.1	2.39	0.017
E. coli	up	4225	sig_mammals: up_reg_birds	-0.03	0.26	-0.13	0.895
mycoplasma	down	4059	sig_mammals	0.73	0.09	7.74	< 0.0001
mycoplasma	down	4059	down_reg_birds	-0.59	0.53	-1.11	0.266
mycoplasma	down	4059	sig_mammals: down_reg_birds	-0.06	0.93	-0.06	0.948
mycoplasma	up	4059	sig_mammals	0.75	0.1	7.77	< 0.0001
mycoplasma	up	4059	up_reg_birds	0.51	0.21	2.42	0.016
mycoplasma	up	4059	sig_mammals: up_reg_birds	-0.71	0.41	-1.73	0.084
plasmodium	down	3222	sig_mammals	0.74	0.11	6.86	< 0.0001
plasmodium	down	3222	down_reg_birds	-0.02	0.37	-0.05	0.961
plasmodium	down	3222	sig_mammals: down_reg_birds	0.01	0.85	0.01	0.992
plasmodium	up	3222	sig_mammals	0.73	0.11	6.71	< 0.0001
plasmodium	up	3222	up_reg_birds	0.19	0.23	0.85	0.396
plasmodium	up	3222	sig_mammals: up_reg_birds	0.44	0.87	0.5	0.614

282 enrichment. However, a few genes with low FDR-corrected p-values selection
283 cutoffs in both datasets ($p < 0.0001$) were also up-regulated in response to
284 influenza (PKR, PARP9, and MX1), up-regulated in response to West Nile virus
285 (PKR), up-regulated in response to *E. coli* (F5), or down-regulated in response to
286 *E. coli* (RAD9A).

287 Due to the small number of genes under selection and differentially
288 expressed in both lineages, we also sought to test whether there was any
289 difference in differential expression effect size (β values) between genes under
290 selection in both lineages, genes under selection in birds, and genes not under
291 positive selection. A difference in overall differential expression effect size might
292 suggest the existence of general differences in gene expression patterns that
293 might not be strong enough to produce significant signal at individual genes. For
294 each gene, we calculated the harmonic mean of bird and mammal absolute,
295 standardized β values in response to infection with each pathogen and compared
296 the mean of each β distribution in the three selection categories with pairwise
297 Mann-Whitney U-tests. We found that genes under selection in both lineages
298 have larger β values than both other classes, particularly in response to viruses
299 (Figure 6, Supplemental Table 10). Genes under selection in birds also have
300 larger β values compared to genes not under selection in response to viruses,
301 but not other pathogens.

302

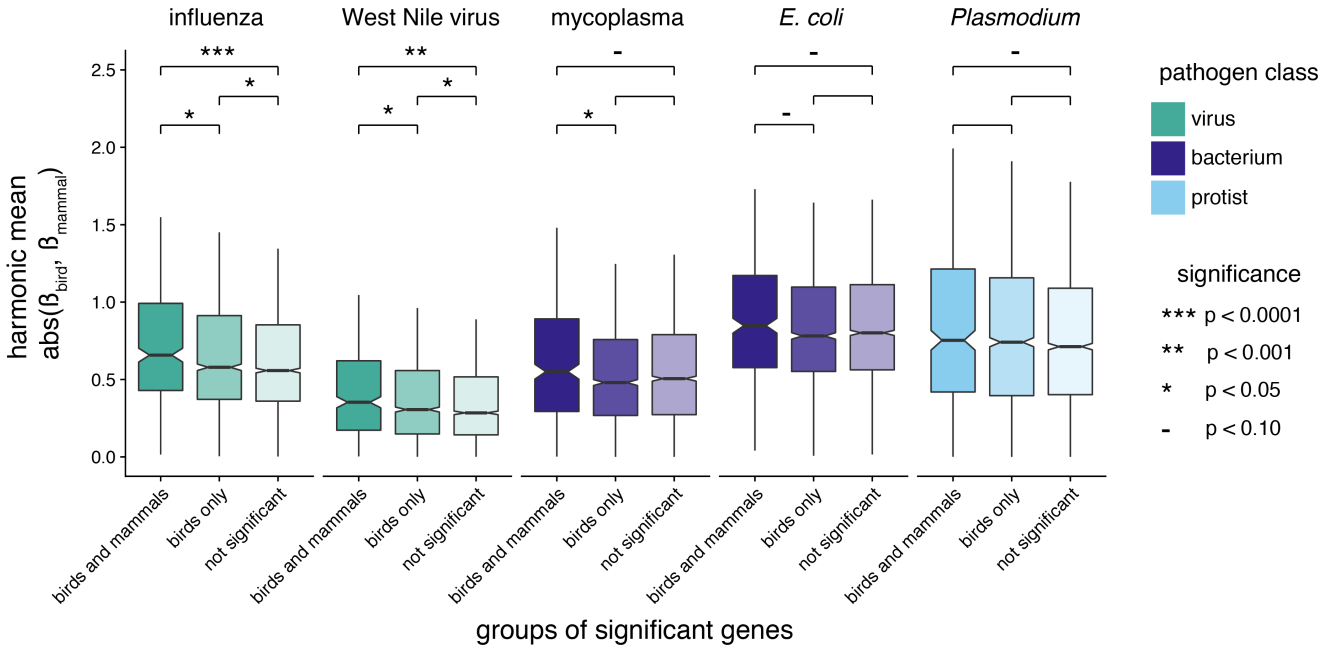


Figure 6. Comparison of differential expression effect values for groups of genes, across pathogens with transcriptome data available for both birds and mammals. Groups of genes are defined as being under positive selection in both birds and mammals, in birds only, or neither birds nor mammals (not significant). Differential expression effect values for each gene are calculated as the harmonic mean of the absolute β values of birds and mammals. We compared the mean of each category to that of the other two categories within each pathogen with Mann-Whitney U-tests, and the significance-level for each test is indicated by asterisks. Comparisons with $p > 0.10$ are left blank. Note that boxplot outliers are not depicted.

303

304 **DISCUSSION**

305 Here, we show that shared signatures of positive selection are consistent with
306 pathogen-mediated selection. First, across birds, genes involved in immune
307 system function, DNA replication and repair, lipid metabolism, and
308 phototransduction are targets of positive selection. Most of these pathways can
309 be directly or indirectly linked to immune response. Functional transcriptomic
310 data independently validates these results, showing that gene up-regulated in

311 response to pathogens contain a higher proportion of genes under positive
312 selection than those not regulated by infection. These results hold true at a
313 broader taxonomic scale. We not only show that genes under selection in birds
314 more likely to be under selection in mammals, but that these shared selected
315 genes are enriched for immune system processes, and in particular those related
316 to viral response. We find few genes differentially regulated and under positive
317 selection in both birds and mammals, but those we found are known to interact
318 directly with pathogens. We also find that genes under positive selection in both
319 lineages have significantly larger overall differential expression effect values
320 compared to those only under positive selection in birds or those not under
321 positive selection. Our results point to pathogens, and in particular viruses, being
322 the most consistent selective pressure across tetrapods.
323
324
325 ***Pathogens drive convergent signatures of selection across birds and***
326 ***mammals***
327 The strong overlap in genes under positive selection in mammals and birds
328 (Figure 4A), together with the functional enrichment and expression results,
329 support the hypothesis that pathogens consistently target the same genes across
330 deeper evolutionary timescales. Although there are some differences in the fine
331 details between avian and mammalian immune systems (e.g. different TLRs are
332 functionally similar in the types of pathogens they recognize (Kaiser 2010; Chen
333 et al. 2013), the overall immune responses are conserved between the clades

334 (Kaiser 2010). Schrom et al. (2017) theoretically demonstrated that there are a
335 limited number of network architecture configurations that are both inducible and
336 robust, and our results here further suggest that pathogens are constrained in
337 how they can interact with these networks to suppress an immune response.
338 Further work on signatures of selection in other tetrapod clades would help to
339 distinguish whether the shared patterns of selection we observe are due to
340 convergence or ancient shared tetrapod selection.

341 We also show that the same genes are likely to be up-regulated in
342 response to pathogen infection (Table 3), despite significant differences in the
343 overall transcriptomic study designs (Supplemental Table 8). We found few
344 genes both differentially expressed and under positive selection in both lineages,
345 although we did find a significant, but small tendency for genes under selection in
346 both clades to be differentially expressed (Figure 6). Those genes we did find are
347 either classic examples of well-documented host-pathogen arms races (PKR
348 (Samuel et al. 2006; Rothenburg et al. 2008; Enard et al. 2016) and MX1 (Ferris
349 et al. 2013)), or genes known to interact directly with pathogens or the immune
350 response (PARP9 (Zhang et al. 2015), RAD9 (An et al. 2010), F5 (Brunder et al.
351 1997)), which are new candidates for genes that may be involved in host-
352 pathogen arms races across tetrapods.

353

354 ***Viruses produce the strongest signatures of pathogen-mediated selection***

355 Our results highlight that shared signatures of selection are enriched for
356 pathways annotated to interact with viruses compared to those that interact with

357 other pathogens. We show similar findings with our differential expression results
358 - the differences in shared levels of differential expression in birds and mammals
359 are strongly significant for the viral infectious agents, and only marginally
360 significant for the other infectious agents (Figure 6). There was also a stronger
361 overlap in expressed genes for the viruses compared to the other two classes of
362 pathogens (Table 3). Finally, the Influenza A and Herpes simplex pathways were
363 significantly enriched for shared genes under selection (Figure 4B,C). There is a
364 near universal tendency to switch hosts in viruses (Geoghegan et al. 2017; Shi et
365 al. 2018) and retroviruses (Henzy et al. 2014), although there is some variation in
366 the prevalence of host switching in different viral families (Geoghegan et al.
367 2017). Examples of host-switching will only increase as more viruses are
368 sequenced. Across populations of *Drosophila melanogaster*, a recent study
369 observed higher rates of adaptation in viral genes only, not immune genes in
370 general, bacterial genes, or fungal genes (Early et al. 2017), suggesting that
371 these results may be more general across broader organisms as well.
372
373 ***Pathogens are a strong selective pressure in birds***
374 Our pathway enrichment results and differential expression results imply that
375 pathogens are one of the strongest selective pressures on amino acid sequence
376 of protein coding genes in birds. Our site test results, which require the same
377 sites to be under selection across species, suggest that host-pathogen
378 interactions are constrained to target specific sites in the same genes in different
379 species. Without these site constraints, genes may be under selection in many

380 different lineages of birds, but there may also be much greater variation in
381 molecular pathways under selection at more recent evolutionary timescales. Our
382 PCA of the probability of positive selection for each gene for each species
383 supports this hypothesis. Following PC1, axes of variation either separate clades
384 of species (e.g. ratites, song birds), or specific lineages. Three previous studies
385 that performed genome-wide scans for positive selection in specific bird lineages
386 further support this hypothesis. First, Nam et al. (2010) compared positive
387 selection acting on three avian lineages, and while 1,751 genes were evolving
388 more rapidly than average in one of these three lineages, only 208 were common
389 to all. Backstrom et al. (2013) compared signatures of positive selection in two
390 species of galliformes and two species of passerines, and found that only the
391 passerine lineages showed GO enrichment with terms related to fat metabolism,
392 neurodevelopment and ion binding. Finally, Zhang, C. Li et al. (2014) found
393 evidence for positive selection in the three vocal-learning bird lineages enriched
394 for neural-related GO terms.

395 Previous studies of immune gene evolution in birds focus on receptor
396 genes known to be hotspots of host-pathogen co-evolution, the Toll-like
397 Receptors (TLRs) and Major Histocompatibility Complex (MHC). Five of the 10
398 avian TLRs are present in our dataset, with TLRs 1A, 1B, 2A, and 2B likely
399 filtered out due to their recent duplication and TLR21 likely filtered out due to
400 missing data caused by sequencing difficulty (Alcaide and Edwards 2011;
401 Grueber et al. 2014). For the five TLRs in our dataset we observe that the M0 ω
402 values are similar to those observed by a previous study (Alcaide and Edwards

403 2001). In addition, our results confirm those of Alcaide and Edwards (2011)
404 Grueber et al. (2014), and Velová et al. (2018), which showed TLR5 as having
405 the highest proportion of selected sites, and the endosomal TLRs (TLR3, TLR7)
406 as having the smallest proportion of selected sites. These similarities from
407 independent studies focusing on just a few genes give us additional confidence
408 in the results of our larger dataset. Unfortunately, the complexity of the MHC
409 genes mean that they were not included in our dataset. However, a recent survey
410 of selection across birds found selection for both classes of MHC loci (Minias et
411 al. 2018), indicating that our overall patterns of selection likely hold true.
412 Receptor genes clearly have signatures of pathogen-mediated selection
413 but signaling pathways (e.g. ECM-receptor interaction and cytokine-cytokine
414 receptor interaction) and downstream genes in immune pathway are also under
415 selection in our dataset. Pathogens have evolved many ways to avoid the host
416 immune response, sometimes at receptors, but sometimes at signaling
417 molecules or genes involved in other cellular processes (Finlay and McFadden
418 2006; Randall and Goodbourn 2008; Pichlmair et al. 2012; Quintana-Murci and
419 Clark 2013; Sironi et al. 2015). The strong signatures of positive selection we
420 observe in these alternative pathways and locations suggest that pathogens not
421 only consistently target the same sites in receptor genes, but also the same sites
422 within genes with other functions. The gene with the highest proportion of
423 significant lineages in birds as estimated by aBS-REL in the Influenza A pathway
424 is not for a receptor gene, but a signaling gene – the gene TRIF, also known as
425 TICAM1. TRIF is recruited by TLR3, a viral sensing TLR in birds, and activates a

426 set of molecules that culminates in the activate of IRF7 or NF-κB (Santhakumar
427 et al. 2017). This gene is under selection in 61% of avian lineages, one of the
428 highest proportions in our dataset, and highlights that genes beyond the
429 classically studied MHC loci and TLRs may be interesting candidates for future
430 studies on the ecology and evolution of host-pathogen co-evolution.
431
432
433 ***Non-immune functional pathways under positive selection in birds***
434 In addition to pathways related to immune function, pathways related to DNA
435 replication and repair were also significantly enriched for positively selected
436 genes (Figure 1). These pathways promote chromosomal stability and remove
437 damaged DNA bases. Birds are known for their compact genomes that show
438 greater than average chromosomal stability (Zhang, C. Li, et al. 2014), and a
439 surprising paucity of transposable elements (TEs) (Cui et al. 2014; Zhang, C. Li,
440 et al. 2014; Kapusta and Suh 2017). One effect on genome structure during the
441 insertion of transposable elements is genome rearrangement due to homologous
442 recombination (Kazazian 2004). Cui et al. (2014) hypothesized that homologous
443 recombination may be responsible for purging transposable elements from the
444 genome, and even observed a galliform hepadnavirus in the process of being
445 removed via homologous recombination. Current host-pathogen evolutionary
446 arms races between birds and TEs are also observed in woodpeckers and allies
447 (Piciformes). There is evidence of different CR1 families expanding at least three
448 different times within the order, and purifying selection for polymorphic TEs in

three closely related woodpecker species (Manthey et al. 2018). Finally, Kapusta and Suh's (2016) observation that the non-recombining W chromosome and regions near centromeres had the highest TE richness also suggest that homologous recombination may prevent the insertion of TEs. Our pathway enrichment results support this hypothesis, and the similar dynamics of positive selection at specific sites as those observed with immune gene pathways suggest that birds may experience a form of host-pathogen co-evolution with TEs.

The process of double-strand break repair, potentially associated with the excision of TEs, can lead to genome-size reductions if biased toward deletions (Schubert and Vu 2016). This model, combined with the evidence for deletions in the ancestral bird lineage (Zhang, C. Li, et al. 2014; Kapusta et al. 2017), the lack of TEs throughout the bird genome (Cui et al. 2014; Kapusta and Suh 2017), and our observation of strong positive selection for homologous recombination, provide support that one of the drivers for the evolution of small genomes in birds is host-pathogen co-evolution against TEs (Kapusta et al. 2017). Powered flight and metabolic stress have also long been hypothesized as the selective pressure driving this decrease in genome size (Zhang and Edwards 2012; Wright et al. 2014; Kapusta et al. 2017). The strong positive selection for base excision repair suggests that there may be continued selection in functional pathways that may help correct breakage or DNA damage that could be a result of metabolic damage (Kapusta and Suh 2017).

471 Two other groups of pathways were enriched for positively selected
472 genes. The first category, lipid metabolism, includes the steroid hormone
473 biosynthesis pathway and linoleic acid metabolism pathway. Steroid hormone
474 biosynthesis is known to be related to diverse life history strategies in birds (Hau
475 et al. 2010), and linoleic acid, more common in seed-rich diets, is related to
476 thermoregulation and can vary across habitats (Ben-Hamo et al. 2013;
477 Andersson et al. 2015). These pathways could be under selection in the diverse
478 set of species included in our dataset. However, these are also known to be
479 factors modulating the immune system (Koutsos and Klasing 2014), and
480 pathogens are known to target cellular processes beyond the immune system
481 (Pichlmair et al. 2012). Further study including additional life history
482 characteristics may help distinguish between these two selective forces.
483 The second category, phototransduction, likely relates to different avian
484 life histories and the visual needs associated with those life histories. The
485 genetics of the avian visual system has traditionally focused on the evolution of
486 the cone receptor genes, and specifically variation in the short wave sensitivity
487 type 1 pigment, which has shifted multiple times between ultraviolet and short
488 wavelengths throughout birds with a single amino acid change (Ödeen and
489 Håstad 2003; Ödeen and Håstad 2013). However, a comparison of retinal
490 transcriptomes from owls, falcons, and hawks, groups that have visual systems
491 adapted to low-light environments (e.g. nocturnal or crepuscular species) or with
492 visual systems tuned for high visual acuity, found evidence for positive selection
493 on phototransduction genes (Wu et al. 2016). The strong signal of positive

494 selection across the broad array of species chosen for our dataset suggests that
495 these genes may be broadly important across many species, and an in-depth
496 analysis of species associated with specific visual needs may uncover additional
497 information on the evolution of this important avian sensory system.

498 Finally, our PCA to identify groups of species that have the same genes
499 under selection identified one principle component that separates species across
500 the avian phylogeny (Figure 2), which is significantly correlated with body mass
501 (Figure 3). By correlating the log-transformed p-values for each gene with body
502 mass to identify the genes driving this correlation, we find that the cellular
503 senescence pathway was the only associated pathway using gene set
504 enrichment (Supplemental Table 6). Lifespan is one trait that is highly correlated
505 with body mass (Furness and Speakman 2008), and cellular senescence may be
506 linked to lifespan through telomere dynamics (Monaghan and Haussmann 2006;
507 de Magalhães and Passos 2018). Within a species, telomeres typically degrade
508 as an organism ages, but few interspecific studies have found a correlation
509 between telomere length and lifespan (Monaghan and Haussmann 2006).
510 However, a recent comparative study in birds showed that telomeres shortened
511 more slowly in species with longer lifespans, and that these results are
512 conserved within families (Tricola et al. 2018). A study of genes associated with
513 telomeres in mammals did not find any correlation between the strength of
514 positive selection at these loci and body mass (Morgan et al. 2013). Our results,
515 and the unique pattern of telomere lengthening observed in birds may be an ideal
516 system to study the evolution of telomere dynamics, and the molecular

517 underpinnings of these processes. Finally, telomere length mediates lifespan and
518 lifetime fitness, both of which are reduced due to chronic malaria infection
519 (Asghar et al. 2015), and suggests that this lineage-specific signature of selection
520 may also be related to pathogen-mediated selection.
521
522
523 ***Conclusions and implications***
524 Across birds, and more generally across tetrapods, there is a clear signal
525 of positive selection acting on immune genes, whether against pathogens or
526 transposable elements. Our results demonstrate that the same genes and
527 potentially even the same codons may be common targets of pathogens to
528 subvert the immune response. Genes with particularly strong evidence of
529 selection may be good candidates for further study from a functional and
530 ecological perspective, and could broaden perspectives on the ecology and
531 evolution of immunity beyond MHC loci and TLRs typically examined. From an
532 applied perspective, there is a great need to understand which proteins or genes
533 in immune gene networks are important in pathogen resistance to improve
534 breeding strategies in economically important species (e.g. poultry; (Kaiser
535 2010). Our work is a first step in this direction, and we provide a rich resource for
536 the examination of specific genes and pathways.
537 Here we have only considered positive selection at a broad scale.
538 Combining these results with those from populations or specific clades within
539 birds and mammals may provide new insights on similarities or differences in

540 long and short-term selection. Pathogen load is the strongest driver of local
541 adaptation in humans (Fumagalli et al. 2011) and viruses are important drivers of
542 population adaptation in flies (Early et al. 2017). From a network perspective,
543 functional gene pathways under strong selection in humans are directly or
544 indirectly involved in immunity (Daub et al. 2013). Given the signatures of host-
545 pathogen co-evolution we observe across birds, we expect that pathogens are an
546 important driver of recent adaptation in bird populations as well.
547
548
549 **METHODS**
550
551 **Identification, alignment and filtering of avian orthologs**
552 Avian orthologs were identified, aligned, and filtered by Sackton et al. (2018). We
553 provide a brief outline of the methodology here, but full details and computer
554 code can be found in Sackton et al. (2018). The program OMA v.1.0.0 (Roth et
555 al. 2008; Altenhoff et al. 2013) was used to infer patterns of homology among
556 protein-coding genes across 39 sequenced bird (Figure 2) and three non-avian
557 reptile (*Alligator mississippiensis, Anolis carolinensis,* and *Chrysemys picta*)
558 genomes. For each gene set, the longest transcript was selected to represent
559 that protein in the homology search.
560 Once OMA had completed, alignments were built for each OMA-defined
561 homologous group using MAFFT v.7.221 (Katoh and Standley 2013), and a
562 HMM was built each protein alignment using HMMER v. 3.1b hmmbuild (Johnson

563 et al. 2010). Each HMM was then used to search the full set of OMA input both to
564 verify that the same proteins are recovered as belonging to a homologous group,
565 and to assign unassigned proteins if possible. Finally, a graph-based algorithm
566 was used to add gene models not assigned to any OMA group to the best match
567 if possible. This produced a new set of homologous groups, which we use in the
568 following analyses.

569 These 45,367 hierarchical orthologous groups, or HOGs, were filtered to
570 retain 16,151 HOGs with sequences for at least four species. Protein sequences
571 were aligned with MAFFT v. 7.245 (Katoh and Standley 2013), and filtered in
572 three steps. First, entire columns were excluded if missing in more than 30% of
573 species, had sequence in fewer than 10 taxa, or was missing in two of the three
574 of the main taxonomic groups (paleognaths, neognaths, or non-avian outgroups).
575 Second, poorly aligned regions were masked according to Jarvis et al. (2014)
576 using a sliding-window similarity approach. Third, columns were removed using
577 the same criteria as the first round. Next, entire sequences were removed from
578 each alignment if they were over 50% shorter than their pre-filtered length or
579 contained excess gaps. Finally, entire HOGs were removed if they contained
580 more than three sequences for any species, did not have more than 1.5x
581 sequences for the given number of species present in the alignment, or were less
582 than 100 base pairs long. Nucleotide sequences for all remaining HOGs were
583 aligned with the codon model in Prank v. 150803 (Loytynoja and Goldman 2008).
584 In total, 11,248 HOGs remained after all alignment and filtering steps.

585 Guide trees for use in the tests of selection were constructed for each
586 alignment with RAxML v. 8.1.4 (Stamatakis 2014) under a GTR+GAMMA
587 substitution model, partitioned into codon positions 1+2 and 3, with 200 rapid
588 bootstrap replicates and a maximum likelihood tree search. In cases where
589 species had more than one sequence in the alignment, we included all copies to
590 produce a gene tree for that HOG.
591
592 **Tests of selection**
593 Once HOGs had been identified and filtered, we considered them as
594 representatives for genes, and so will refer to them as genes. To identify
595 positively selected genes, we compared models of nearly neutral evolution to
596 those that included signatures of positive selection at a proportion of sites across
597 lineages in the avian phylogeny. Sites under positive selection are defined as
598 those with elevated nonsynonymous/synonymous substitution ratios ($\omega = d_N/d_S$)
599 compared to the expectation under neutral evolution, $\omega = 1$. We used two
600 different programs to identify genes with evidence for elevated ω values at
601 specific sites across avian lineages. First, we used the site models (Nielsen and
602 Yang 1998; Yang et al. 2000) implemented in the program Phylogenetic Analysis
603 by Maximum Likelihood v4.8 (PAML; (Yang 2007) to calculate likelihood scores
604 and parameter estimates for seven models of evolution (Table 1). Because some
605 genes contained gene duplicates, we ran all analyses of selection on gene trees
606 from all 11,248 genes, and separately on the species tree for 8,699 genes with
607 no duplicate sequences. We used the species tree generated by OMA from

608 Sackton et al. (2018) as the phylogenetic hypothesis. First, we fit the M0 model,
609 which estimates a single ω for all sites in the alignment. We used the branch
610 lengths estimated with the M0 model as fixed branch lengths for subsequent
611 models to decrease computational time. To identify genes with evidence of
612 positive selection, we conducted likelihood ratio tests between neutral models
613 and selection models (models with $\omega > 1$). We compared likelihood scores from
614 the M1a vs. M2a, M2a vs. M2a_fixed, M7 vs. M8, and M8 vs. M8a models
615 (Supplemental Table 8.1) (Nielsen and Yang 1998; Yang et al. 2000; W.S.W.
616 Wong et al. 2004). We computed p-values according to a χ^2 distribution with two,
617 one, two, and one degree of freedom respectively.
618 In addition to the site tests implemented in PAML, we used BUSTED
619 (Murrell et al. 2015), a modeling framework implemented in the program HyPhy
620 (Pond et al. 2005), to identify genes with evidence of positive selection at a
621 fraction of sites. BUSTED uses a model that allows branch-to-branch variation
622 across the entire tree (Murrell et al. 2015). Similar to the PAML models, BUSTED
623 uses a likelihood ratio test to compare a model including selection ($\omega > 1$ at a
624 proportion of sites) with one that does not. We parsed all PAML and HyPhy
625 results with custom code (available https://github.com/ajshultz/avian-immunity/)
626 and ran all downstream analyses in R. For both sets of tests, we used the
627 Benjamini-Hochberg approach to correct for multiple testing (Benjamini and
628 Hochberg 1995) with the p.adjust function in the stats packages in R v.3.5 (R
629 Core Development Team 2008). We considered an FDR-corrected p-value less

630 than 0.05 as evidence for positive selection in that gene for a given model
631 comparison.
632 Finally, in addition to testing for selection at particular sites across bird
633 lineages, we used the aBS-REL method in HyPhy with default parameters
634 (Kosakovsky Pond et al. 2011) to detect which specific lineages showed
635 evidence of selection for each gene. For each lineage, including both tip species
636 and internal branches, aBS-REL estimates a p-value for the presence of positive
637 selection. We considered both the raw p-value as well as a p-value corrected for
638 multiple testing within each gene. Fewer lineages showed evidence of selection
639 with an FDR-corrected p-value, but all subsequent results were qualitatively
640 consistent with both sets of tests. For simplicity and because the stringent
641 correction may remove biologically-interesting lineages with weak to moderate
642 selection, we present the results using the number of lineages considered
643 nominally significant without multiple-test correction. We also used a custom
644 script to parse all aBS-REL results and run all downstream analyses in R
645 (https://github.com/ajshultz/avian-immunity/).
646 Previous work has found that alignment errors can result in substantial
647 false positives (Markova-Raina and Petrov 2011). However, our strict alignment
648 filtering strategy and use of the evolution-aware PRANK aligner minimizes the
649 possibility that our results are solely false positives (Markova-Raina and Petrov
650 2011). Recombination also can elevate ω estimates, but the M7 vs M8 model
651 has been shown to be robust to recombination (Anisimova et al. 2003), and these
652 results give us the highest proportions of positively selected genes we observe in

653 our dataset (Table 2). Finally, despite observing high proportions of selected
654 genes, the overall trend of gene-wide estimates of $\omega << 1$ are consistent with
655 patterns of purifying selection on coding regions of the genome (Supplemental
656 Figure 2). Furthermore the similarity in estimated ω values between this study
657 and previous studies in birds with different sets of genome sequences or the use
658 of pairwise estimates between chicken and zebra finch (Nam et al. 2010; Zhang,
659 B. Li, et al. 2014) give us confidence that our results are robust.
660
661
662 **Gene annotation**
663 We annotated genes for downstream enrichment analyses using chicken (*Gallus*
664 *gallus* assembly version 4.0; G.K.-S. Wong et al. 2004) and zebra finch
665 (*Taeniopygia guttata* assembly version 3.2.4; Warren et al. 2010) NCBI gene IDs
666 from sequences of those species included in the alignment of each gene. Of the
667 11,248 HOGs, 10,890 could be assigned to a chicken NCBI gene id, 10,365
668 could be assigned to a zebra finch NCBI gene id, 10,143 could be assigned to
669 both, and 136 could not be assigned to a chicken or zebra finch NCBI gene ID. In
670 order to test additional pathways available for mammalian species (see below),
671 we converted chicken and zebra finch NCBI gene IDs to human (*Homo sapien*;
672 GRCh38.p10) NCBI gene IDs using the R biomaRt package version 2.36.1
673 (Durinck et al. 2005; Durinck et al. 2009). For both avian species, we
674 downloaded the ENSEMBL gene IDs, NCBI gene IDs, and human homolog
675 ENSEMBL gene IDs for each gene using the ggallus_gene_ensembl (chicken

genes, Gallus-gallus-5.0) and tguttata_gene_ensembl (zebra finch genes, TaeGut3.2.4) datasets. For humans, we downloaded the ENSEMBL gene IDs and NCBI gene IDs from the human hsapien_gene_ensembl (human genes, GRCh38.p10) dataset. We assigned each gene by first identifying all human ENSEMBL gene IDs and NCBI gene IDs that were chicken orthologs, and filled in missing IDs with zebra finch annotations. In total, 9,461 out of 11,248 genes could be annotated with human NCBI gene IDs.

Functional gene pathway enrichment for lineages under positive selection in birds

We looked for patterns of positive selection among groups of genes with similar functions using KEGG pathway enrichment tests (Kanehisa and Goto 2000; Kanehisa et al. 2011). We used our most conservative set of genes as our test set – those with FDR-corrected p-values less than 0.05 for all site tests (N=1,521), including the m1a vs. m2a PAML model comparison, m2a vs. m2a_fixed PAML model comparison, m7 vs. m8a PAML model comparison, m8 vs. m8a PAML model comparison, and BUSTED analysis (see Table 1 for PAML model descriptions). Because of the similarity between the model results using gene trees and species trees (see Results), we use the gene tree results as input to maximize the number of genes that could be included in a functional analysis. Preliminary analyses using the species tree results are qualitatively similar to those presented here.

698 To conduct KEGG pathway enrichment analyses, we used the
699 'enrichKEGG' command from clusterProfiler v. 3.8.1 (Yu et al. 2012) from
700 Bioconductor v. 3.7 (Gentleman et al. 2004) with chicken as the reference
701 organism. We used the genes included in both PAML and HyPhy analyses with
702 NCBI gene IDs (N = 10,874) as the gene universe for enrichment tests. To
703 ensure genes not present in the chicken genome, but present in other bird
704 species were not biasing our results, we also performed the functional
705 enrichment test using zebra finch as the reference organism. Finally, we
706 performed a final enrichment test using human as the reference organism to test
707 whether the expanded KEGG pathways of humans could provide insights beyond
708 those available for chicken and zebra finch. We visualized the results using
709 modified versions of the 'dotplot' and 'cnetplot' commands in clusterProfiler.
710
711
712 **Clustering genes under selection among bird lineages**
713 We used aBS-REL results to understand how groups of species that experience
714 similar selective pressures might show evidence for positive selection for the
715 same genes. To do this, we created a matrix of the p-values for the probability of
716 positive selection at each gene for each species. We used this matrix to conduct
717 a principle components analysis (PCA) to cluster species by the log-transformed
718 p-value of each species for each gene. We replaced any missing values with the
719 mean p-value for that gene, log-transformed all p-values, and performed the PCA
720 with the prcomp function in R.

721 Only the first principle component grouped unrelated species (see
722 Results), so we tested whether PC1 might be related to body mass, a
723 measurement correlated with many life history characteristics (Pienaar et al.
724 2013). We extracted body mass measurements from each species using the
725 CRC Handbook of Avian Body Masses (Dunning 2009) and used phylogenetic
726 generalized least squares (PGLS) (Martins and Hansen 1997) to test for a
727 correlation between the PC1 scores and log-transformed body mass. To obtain
728 branch lengths for our species tree topology, we randomly selected one gene
729 with one sequence for all species and used the branch lengths as calculated by
730 the M0 model in PAML. Our results were robust to tests with alternative genes.
731 We ran the PGLS analysis in R with the gls function from the nlme package 3.1-
732 137 (Pinheiro et al. 2013), with a both a Brownian motion (Felsenstein 1985) and
733 an Ornstein-Uhlenbeck (Hansen and Martins 1996) model of evolution. A
734 Brownian motion fit better than the Ornstein-Uhlenbeck model (AIC >2), so we
735 report those results. However, the results are qualitatively similar. We visualized
736 the two traits and the phylogeny using the 'phylomorphospace' function and the
737 evolution of PC1 on the phylogeny using the a modified version of the
738 'plotBranchbyTrait' function from phytools v. 0.6-44 (Revell 2012).
739 To better understand which genes and molecular functions were
740 contributing to the correlation between PC1 and body mass (see Results), we
741 calculated a p-value for the association between log-transformed p-values and
742 log-transformed body mass for each gene separately. Due to the non-normal
743 distribution of log-transformed p-values, we used Spearman's rank correlation

744 with the cor.test function from the stats package in R (R Core Development
745 Team 2008). Although Spearman's rank correlation does not include
746 phylogenetic correction, the aBS-REL p-values are estimated independently for
747 each branch, and so should not be biased by phylogeny. We used the Benjamini-
748 Hochberg approach to correct for multiple testing (Benjamini and Hochberg
749 1995).
750 We tested whether there might be any functional signal in these genes
751 using gene set enrichment with the Spearman's rank correlation values (ρ) as the
752 input for each gene. To avoid biases in genes with only one or a few lineages
753 under selection, we only tested genes with at least five lineages under selection
754 (preliminary results with alternative cutoff suggest that results are robust to the
755 specific cutoff used). We tested for gene set enrichment with the chicken KEGG
756 pathways using the 'gseKEGG' command from clusterProfiler (Yu et al. 2012).
757

758 **Comparisons of avian and mammalian selection datasets**

759 In order to identify shared signatures of selection in both birds and mammals, we
760 compared our results to those of Enard et al (2016). We used our BUSTED
761 results as calculated using the species tree to ensure our results were
762 comparable to their BUSTED tests of positive selection. We combined our
763 datasets using the human ENSEMBL gene ID annotations (conversion methods
764 described above). In total, we could identify 4,931 orthologous genes with results
765 from both datasets. With the set of genes included in both studies, we re-
766 calculated FDR-corrected p-values, and compared the proportion of genes

767 significant in both birds and mammals with a p-value cutoff of 0.1, 0.01, 0.001
768 and 0.0001 to understand whether genes under weak or strong selection might
769 produce different signals. We calculated significance of an increased overlap in
770 genes under selection in both birds and mammals with a Fisher's exact test.
771 We tested whether pathogen-mediated selection might be an important
772 factor in driving the overlap of these genes using KEGG pathway enrichment. We
773 ran these tests as described above, with the genes under positive selection in
774 both birds and mammals as the test set of genes, and the set of genes under
775 selection in birds as the background set of genes. We used the four different
776 FDR-corrected p-value cutoffs ($p < 0.1$, 0.01, 0.001, or 0.0001) to identify genes
777 under selection in each clade. Finally, we used permutation tests to ensure that
778 our pathway enrichment results were not biased toward genes commonly under
779 selection in birds. We randomly created test sets the same size as those
780 empirically defined from the set of genes significant in birds and performed
781 KEGG pathway enrichment. We calculated the enrichment score (proportion of
782 selected genes in the pathway compared to the proportion of selected genes in
783 the dataset) for each pathway significant in our bird-only results (described in
784 above section) and included in our test of empirical data. That is, the pathway
785 had to be significant in birds, and contain at least one gene under selection in
786 both birds and mammals. We performed each permutation for each p-value
787 cutoff 1,000 times to generate a null distribution of enrichment values to compare
788 to our empirical results.
789

790 **Association of genes under positive selection to pathogen-mediated**
791 **transcriptional responses in birds**
792 We independently tested whether genes under positive selection throughout
793 birds were associated with pathogen-mediated immune responses using publicly-
794 available transcriptome data. We tested whether genes that were differentially
795 expressed in response to a pathogen challenge were more likely to be under
796 positive selection. We identified 12 studies of birds that compared the
797 transcriptomes of control individuals and individuals experimentally infected with
798 a virus, bacterium or protist (Supplemental Table 8; (Smith, Burt, et al. 2015;
799 Smith, Smith, et al. 2015; Sun, Liu, Nolan, and Lamont 2015a; Sun, Liu, Nolan,
800 and Lamont 2015b; Videvall et al. 2015; Sun et al. 2016; Beaudet et al. 2017;
801 Deist, Gallardo, Bunn, Dekkers, et al. 2017; Deist, Gallardo, Bunn, Kelly, et al.
802 2017; Newhouse et al. 2017; Zhang et al. 2018). We downloaded all available
803 SRA files for each bioproject and extracted the fastq files with fastq-dump from
804 SRA-Tools v. 2.8.2.1 (Leinonen et al. 2010). We used kallisto v. 0.43.1 (Bray et
805 al. 2016) to quantify transcript abundance with 100 bootstrap replicates. We used
806 paired-end or single-end mode (assuming average fragment lengths of 250 base
807 pairs with a standard deviation of 50 base pairs) as appropriate for each
808 bioproject, using the ENSEMBL transcriptome reference for each species
809 (Supplemental Table 8). One species did not have an ENSEMBL reference
810 available (*Spinus spinus*), so we mapped to the transcriptome reference of the
811 closest available reference, *Serinus canaria*, downloaded from NCBI.

812 We tested for differential expression between experimentally infected
813 individuals and control individuals with sleuth v0.30 (Pimentel et al. 2017). In
814 cases where individuals were available at different timepoints, had different
815 phenotypes (e.g. resistant or susceptible), used different pathogen strains, or
816 sequenced transcriptomes from different organs, we tested each condition
817 separately. We considered a gene to be significantly differentially expressed if it
818 had a q-value less than 0.05 and an effect size, quantified as the absolute value
819 of β, greater than 1. We then combined results for each condition of each
820 bioproject. We considered a gene to be differentially expressed for that study if it
821 was significant in half of conditions defined as different time points and
822 phenotypes for each organ (Supplemental Table 8). For three studies, only a
823 single condition had any appreciable signal, we use used a relaxed our cutoff to
824 count a gene as significant if it was significantly differentially expressed in any
825 condition (Supplemental Table 8).
826 To compare genes across species, we translated all ENSEMBL gene IDs
827 to homologous chicken ENSEMBL gene IDs from the R biomaRt package
828 version 2.36.1 (Durinck et al. 2005; Durinck et al. 2009), except for *S. canaria*,
829 which we translated to chicken gene IDs by mapping genes IDs to sequences in
830 the same gene alignments in our dataset. Some pathogens were represented by
831 more than one study in our dataset. To combine the results for each pathogen,
832 we considered each gene to be significant for that pathogen if it was significant in
833 at least one study. We used logistic regression to test whether genes that were
834 up-regulated (compared to no difference in transcription) were more likely to be

835 under selection in birds (defined in above section), and to test whether genes
836 that were down-regulated were more likely to be under selection in birds.
837
838 **Comparisons of gene expression patterns and positively selected genes in**
839 **birds and mammals**
840 Finally, we compared the pathogen-mediated transcriptional responses in birds
841 to those in mammals. We used 14 previously published studies that generated
842 transcriptomes for control individuals and pathogen-challenged individuals to
843 identify differentially expressed genes in response to pathogen infection for a
844 species of mammal (Supplemental Table 8; (Qian et al. 2013; Langley et al.
845 2014; Ogorevc et al. 2015; Rojas-Peña et al. 2015; DeBerg et al. 2016; Lee et al.
846 2016; Tran et al. 2016; Chopra-Dewasthaly et al. 2017; Jong et al. 2018). We
847 chose studies that used similar pathogens as those used in the avian
848 experiments to compare the expression profiles of the two clades as closely as
849 possible, while acknowledging that such matching will necessarily be somewhat
850 imprecise. We used the same preprocessing steps as described in the above
851 avian transcriptomic section. In two studies, seven and nine different timepoints
852 were used, with a large number of individuals giving increased power to detect
853 differentially expressed genes. For these two studies, we required genes to be
854 significant in half of all timepoints as well as overall (all infected individuals
855 compared to control individuals). We translated all non-human ENSEMBL gene
856 IDs to human ENSEMBL gene IDs using biomaRt to compare results across all
857 bird and mammal species. Finally, for birds and for mammals, we summarized

858 results for each gene for each infectious agent, considering a gene to be
859 differentially expressed if it was differentially expressed in any study. Despite the
860 smaller number of genes identified in the joint bird and mammal datasets, results
861 comparing the enrichment in bird-only studies as described above were robust
862 (results not shown), so we have confidence that our combined bird and mammal
863 dataset captured the signal observed with birds alone.

864 With our combined bird and mammal dataset, we first tested whether
865 genes up-regulated in infected birds were also likely to be up-regulated in
866 infected mammals, or whether genes down-regulated in infected birds were also
867 likely to be down-regulated in infected mammals. We used a Fisher's exact test
868 to test whether the proportions of up- or down- regulated genes in both clades
869 deviated from null expectations. Then, we combined the gene expression results
870 with the significance results across birds and mammals. We sought to test
871 whether genes that were under positive selection in birds were likely to be under
872 positive selection in mammals and differentially expressed (either up- or down-
873 regulated in both clades). To do this, for each pathogen, we used logistic
874 regression with genes under selection in birds as the response variable (under
875 selection or not), and the mammalian selection status (under selection or not),
876 the differential expression status in birds (up- or down-regulated), and their
877 interaction as predictor variables. Finally, due to the variety of experimental
878 setups and small number of genes up- or down-regulated in both birds and
879 mammals, we used a more sensitive test to test whether the absolute value of
880 mammal and bird β values were significantly higher in genes under selection in

both lineages, or genes under selection in birds, compared to genes not detected as being under selection with our BUSTED site tests. A larger absolute value of β implies larger magnitudes of differential expression, regardless of the direction of selection or q-value significance. To ensure the β values were as comparable as possible among studies, we first standardized the β values to have a mean of 0 and standard deviation of 1 for each study. Then, for pathogen replicates within birds and mammals, we used the maximum β value observed as the bird or mammal β value for that gene (results were robust if the mean β value was used instead). For each gene, we calculated the harmonic mean of bird and mammal β values, and used a Mann-Whitney U-test to test whether mean β values were significantly different between genes under selection (q-value < 0.05) in birds and mammals and genes under selection in birds only, between genes under selection in birds and mammals and genes not under selection in either lineage, and between genes under selection in birds only and genes not under selection in either lineage.

ACKNOWLEDGEMENTS

We thank Scott Edwards, Hopi Hoekstra, and John Wakeley for feedback on the project, as well as members of the Edwards Lab and Harvard Informatics Group. We thank Alison Cloutier with assistance with alignments and filtering, and Julia Yu for early discussion. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

905
906
907
908
909 **LITERATURE CITED:**
910
911

912 Alcaide M, Edwards SV. 2011. Molecular evolution of the Toll-Like Receptor
913 multigene family in birds. Molecular Biology and Evolution 28:1703–1715.

914 Alcaide M, Liu M, Edwards SV. 2013. Major histocompatibility complex class I
915 evolution in songbirds: universal primers, rapid evolution and base
916 compositional shifts in exon 3. PeerJ 1:e86.

917 Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. 2013. Inferring hierarchical
918 orthologous groups from orthologous gene pairs. PLoS ONE 8:e53786.

919 An L, Wang Y, Liu Y, Yang X, Liu C, Hu Z, He W, Song W, Hang H. 2010. Rad9
920 is required for B cell proliferation and immunoglobulin class switch
921 recombination. J. biol. Chem 285:35267–35273.

922 Andersson MN, Wang H-L, Nord A, Salmón P, Isaksson C. 2015. Composition of
923 physiologically important fatty acids in great tits differs between urban and
924 rural populations on a seasonal basis. Front. Ecol. Evol. 3:522–13.

925 Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy
926 of the likelihood method for detecting positive selection at amino acid sites.
927 Genetics 164:1229–1236.

928 Asghar M, Hasselquist D, Hansson B, Zehtindjiev P, Westerdahl H, Bensch S.
929 2015. Hidden costs of infection: Chronic malaria accelerates telomere
930 degradation and senescence in wild birds. Science 347:436–438.

931 Backström N, Zhang Q, Edwards SV. 2013. Evidence from a House Finch
932 (*Haemorhous mexicanus*) spleen transcriptome for adaptive evolution and
933 biased gene conversion in passerine birds. Molecular Biology and Evolution
934 30:1046–1050.

935 Barreiro LB, Quintana-Murci L. 2009. From evolutionary genetics to human
936 immunology: how selection shapes host defence genes. Nature Reviews
937 Genetics 11:17–30.

938 Beaudet J, Tulman ER, Pflaum K, Liao X, Kutish GF, Szczepanek SM, Silbart
939 LK, Geary SJ. 2017. Transcriptional Profiling of the Chicken Tracheal
940 Response to Virulent *Mycoplasma gallisepticum* Strain Rlow. Infect Immun
941 85:e00343–17–15.

942 Ben-Hamo M, McCue MD, Khozin-Goldberg I, McWilliams SR, Pinshow B. 2013.
943 Ambient temperature and nutritional stress influence fatty acid composition of
944 structural and fuel lipids in Japanese quail (*Coturnix japonica*) tissues. Comp
945 Biochem Physiol A 166:244–250.

946 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A Practical
947 and Powerful Approach to Multiple Testing. J. R. Statist. Soc. B 57:289–300.

948 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic
949 RNA-seq quantification. Nat Biotechnol 34:525–527.

950 Brunder W, Schmidt H, Karch H. 1997. EspP, a novel extracellular serine
951 protease of enterohaemorrhagic Escherichia coli O157:H7 cleaves human
952 coagulation factor V. Mol. Microbiol. 24:767–778.

953 Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L. 2008. Evolutionary patterns
954 of MHC Class II B in owls and their implications for the understanding of
955 avian MHC evolution. Mol Biol Evol 25:1180–1191.

956 Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L. 2010. Adaptive divergence
957 of ancient gene duplicates in the avian MHC Class II. Mol Biol Evol 27:2360–
958 2374.

959 Chen S, Cheng A, Wang M. 2013. Innate sensing of viruses by pattern
960 recognition receptors in birds. Vet. Res. 44:82.

961 Chopra-Dewasthaly R, Korb M, Brunthaler R, Ertl R. 2017. Comprehensive RNA-
962 Seq profiling to evaluate the sheep mammary gland transcriptome in
963 response to experimental *Mycoplasma agalactiae* infection. PLoS ONE
964 12:e0170015–e0170017.

965 Clements JF, Schulenberg TS, Iliff MJ, Fredericks TA, Sullivan BL, Wood CL.
966 2016. The Clements checklist of the birds of the world: v2016. Available from:
967 http:// www.birds.cornell.edu/clementschecklist/downloadable-clements-
968 checklist

969 Cui J, Zhao W, Huang Z, Jarvis ED, Gilbert MTP, Walker PJ, Holmes EC, Zhang
970 G. 2014. Low frequency of paleoviral infiltration across the avian phylogeny.
971 Genome Biol 15:539.

972 Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi
973 M, Excoffier L. 2013. Evidence for polygenic adaptation to pathogens in the
974 human genome. Mol Biol Evol 30:1544–1558.

975 de Magalhães JP, Passos JF. 2018. Stress, cell senescence and organismal
976 ageing. Mech Ageing Dev 170:2–9.

977 DeBerg HA, Zaidi MB, Khaenam P, Gersuk V, Linsley PS, Estrada-Garcia T.
978 2016. Blood transcriptional profiling of childhood diarrheal diseases identifies
979 gene signatures of Shigella and rotavirus infections. J Immunol 196:66.15–
980 66.15.

981 Deist MS, Gallardo RA, Bunn DA, Dekkers JCM, Zhou H, Lamont SJ. 2017.
982 Resistant and susceptible chicken lines show distinctive responses to
983 Newcastle disease virus infection in the lung transcriptome. BMC Genomics
984 18:1–15.

985 Deist MS, Gallardo RA, Bunn DA, Kelly TR, Dekkers JCM, Zhou H, Lamont SJ.
986 2017. Novel mechanisms revealed in the trachea transcriptome of resistant
987 and susceptible chicken lines following infection with Newcastle disease
988 virus. Clin. Vaccine Immunol. 24:e00027–17–17.

989 Dunning J. 2009. CRC Handbook of Avian Body Masses. Second Edition. Boca
990 Raton: CRC Press

991 Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W.
992 2005. BioMart and Bioconductor: a powerful link between biological
993 databases and microarray data analysis. Bioinformatics (Oxford, England)
994 21:3439–3440.

995 Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the
996 integration of genomic datasets with the R/Bioconductor package biomaRt.
997 Nat Protoc 4:1184–1191.

998 Early AM, Arguello JR, Cardoso-Moreira M, Gottipati S, Grenier JK, Clark AG.
999 2017. Survey of global genetic diversity within the *Drosophila* immune
1000 system. Genetics 205:353–366.

1001 Ebel ER, Telis N, Venkataram S, Petrov DA, Enard D. 2017. High rate of
1002 adaptation of mammalian proteins that interact with *Plasmodium* and related
1003 parasites. PLoS Genet 13:e1007023–e1007027.

1004 Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF. 2000. A 39-kb
1005 sequence around a blackbird Mhc class II gene: ghost of selection past and
1006 songbird genome architecture. Mol Biol Evol 17:1384–1395.

1007 Edwards SV, Wakeland EK, Potts WK. 1995. Contrasting histories of avian and
1008 mammalian Mhc genes revealed by class II B sequences from songbirds.
1009 PNAS 92:12200–12204.

1010 Ellegren H. 2013. The evolutionary genomics of birds. Annu. Rev. Ecol. Evol.
1011 Syst. 44:239–259.

1012 Ellis JS, Turner LM, Knight ME. 2012. Patterns of selection and polymorphism of
1013 innate immunity genes in bumblebees (Hymenoptera: Apidae). Genetica
1014 140:205–217.

1015 Enard D, Cai L, Gwennap C, Petrov DA. 2016. Viruses are a dominant driver of
1016 protein adaptation in mammals. Elife 5:e12469.

1017 Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat 125:1–
1018 15.

1019 Ferris MT, Aylor DL, Bottomly D, Whitmore AC, Aicher LD, Bell TA, Bradel-
1020 Tretheway B, Bryan JT, Buus RJ, Gralinski LE, et al. 2013. Modeling host
1021 genetic regulation of influenza pathogenesis in the collaborative cross. PLoS
1022 Pathog 9:e1003196.

1023 Finlay BB, McFadden G. 2006. Anti-immunology: Evasion of the host immune
1024 system by bacterial and viral pathogens. Cell 124:767–782.

1025 Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. 2011.
1026 Signatures of environmental genetic adaptation pinpoint pathogens as the
1027 main selective pressure through human evolution. PLoS Genet 7:e1002355.

1028 Furness LJ, Speakman JR. 2008. Energetics and longevity in birds. AGE 30:75–
1029 87.

1030 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B,
1031 Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software
1032 development for computational biology and bioinformatics. Genome Biol
1033 5:R80.

1034 Geoghegan JL, Duchêne S, Holmes EC. 2017. Comparative analysis estimates
1035 the relative frequencies of co-divergence and cross-species transmission
1036 within viral families. PLoS Pathog 13:e1006215–e1006217.

1037 Gill FB. 2007. Ornithology. W H Freeman & Company

1038 Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic positive selection in the
1039 evolution of avian Toll-Like Receptor innate immunity genes. PLoS ONE
1040 9:e89632.

1041 Hahn MW, Nakhleh L. 2015. Irrational exuberance for resolved species trees.
1042 Evolution 70:7–17.

1043 Hansen T, Martins E. 1996. Translating between microevolutionary process and
1044 macroevolutionary patterns: the correlation structure of interspecific data.
1045 Evolution 50:1404–1417.

1046 Hau M, Ricklefs RE, Wikelski M, Lee KA, Brawn JD. 2010. Corticosterone,
1047 testosterone and life-history strategies of birds. Proc Biol Sci 277:3203–3212.

1048 Henzy JE, Gifford RJ, Johnson WE, Coffin JM. 2014. A novel recombinant
1049 retrovirus in the genomes of modern birds combines features of avian and
1050 mammalian retroviruses. J Virol 88:2398–2405.

1051 Hess CM, Edwards SV. 2002. The evolution of the major histocompatibility
1052 complex in birds. BioScience 52:423–431.

1053 Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P. 2014. Whole-genome analyses
1054 resolve early branches in the tree of life of modern birds. Science 346:1320–
1055 1331.

1056 Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global
1057 diversity of birds in space and time. Nature 491:444–448.

1058 Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic
1059 and iterative HMM search procedure. BMC Bioinformatics 11:431.

1060 Jong E, Hancock DG, Hibbert J, Wells C, Richmond P, Simmer K, Burgner D,
1061 Strunk T, Currie AJ. 2018. Identification of generic and pathogen-specific
1062 cord blood monocyte transcriptomes reveals a largely conserved response in
1063 preterm and term newborn infants. J Mol Med 96:1–11.

1064 Juul-Madsen HR, Viertlböeck B, Härtle S, Schmidt AL, Göbel TW. 2014. Innate
1065 Immune Responses. In: Schat KA, Kaspars B, Kaiser P, editors. Avian
1066 Immunology. Elsevier, Ltd. pp. 121–147.

1067 Kaiser P. 2010. Advances in avian immunology—prospects for disease control: a
1068 review. Avian Pathol 39:309–324.

1069 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2011. KEGG for
1070 integration and interpretation of large-scale molecular data sets. Nucleic
1071 Acids Res 40:D109–D114.

1072 Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes.
1073 Nucleic Acids Res 28:27–30.

1074 Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in
1075 birds and mammals. PNAS 114:E1460–E1469.

1076 Kapusta A, Suh A. 2017. Evolution of bird genomes-a transposon's-eye view.
1077 Ann. N.Y. Acad. Sci. 1389:164–185.

1078 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software
1079 Version 7: Improvements in Performance and Usability. Mol Biol Evol
1080 30:772–780.

1081 Kazazian HH. 2004. Mobile elements: drivers of genome evolution. Science
1082 303:1626–1632.

1083 Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler
1084 K. 2011. A random effects branch-site model for detecting episodic
1085 diversifying selection. Molecular Biology and Evolution 28:3033–3043.

1086 Kosiol C, Vinar T, da Fonseca RR, HUBISZ MJ, Bustamante CD, Nielsen R,
1087 Siepel A. 2008. Patterns of positive selection in six mammalian genomes.
1088 PLoS Genet 4:e1000144.

1089 Koutsos EA, Klasing KC. 2014. Factors Modulating the Avian Immune System.
1090 In: Schat KA, Kaspers B, Kaiser P, editors. Avian Immunology. Elsevier Ltd.
1091 pp. 299–314.

1092 Langley RJ, Tipper JL, Bruse S, Baron RM, Tsalik EL, Huntley J, Rogers AJ,
1093 Jaramillo RJ, O'Donnell D, Mega WM, et al. 2014. Integrative “omic” analysis
1094 of experimental bacteremia identifies a metabolic signature that distinguishes
1095 human sepsis from systemic inflammatory response syndromes. Am J Respir
1096 Crit Care Med 190:445–455.

1097 Lee E-Y, Lee H-C, Kim H-K, Jang SY, Park S-J, Kim Y-H, Kim JH, Hwang J, Kim
1098 J-H, Kim T-H, et al. 2016. Infection-specific phosphorylation of glutamyl-prolyl
1099 tRNA synthetase induces antiviral immunity. Nat Immunol 17:1252–1262.

1100 Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide
1101 Sequence Database Collaboration. 2010. The Sequence Read Archive.
1102 Nucleic Acids Res 39:D19–D21.

1103 Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors
1104 in sequence alignment and evolutionary analysis. Science 320:1632–1635.

1105 Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and high rate of
1106 false positives in the estimates of positive selection in the 12 Drosophila
1107 genomes. Genome Res 21:863–874.

1108 Martins E, Hansen T. 1997. Phylogenies and the comparative method: a general
1109 approach to incorporating phylogenetic information into the analysis of
1110 interspecific data. Am Nat 149:646–667.

1111 Mendes FK, Hahn MW. 2016. Gene tree discordance causes apparent
1112 substitution rate variation. Syst Biol 65:711–721.

1113 Minias P, Pikus E, Whittingham LA, Dunn PO. 2018. A global analysis of
1114 selection at the avian MHC. Evolution 97:133–16.

1115 Monaghan P, Haussmann MF. 2006. Do telomere dynamics link lifestyle and
1116 lifespan? Trends Ecol Evol 21:47–53.

1117 Morgan CC, Mc Cartney AM, Donoghue MTA, Loughran NB, Spillane C, Teeling
1118 EC, O'Connell MJ. 2013. Molecular adaptation of telomere associated genes
1119 in mammals. BMC Evol Biol 13:251.

1120 Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K,
1121 Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of
1122 episodic selection. Mol Biol Evol 32:1365–1371.

1123 Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backström N, Kunstner A,
1124 Balakrishnan CN, Heger A, Ponting CP, et al. 2010. Molecular evolution of
1125 genes in avian genomes. Genome Biol 11:R68.

1126 Newhouse DJ, Hofmeister EK, Balakrishnan CN. 2017. Transcriptional response
1127 to West Nile virus infection in the zebra finch (*Taeniopygia guttata*). R. Soc.
1128 open sci. 4:170296.

1129 Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected
1130 amino acid sites and applications to the HIV-1 envelope gene. Genetics
1131 148:929–936.

1132 Ogorevc J, Mihevc SP, Hedegaard J, Bencina D, Dove P. 2015. Transcriptomic
1133 response of goat mammary epithelial cells to *Mycoplasma agalactiae*
1134 challenge–a preliminary study. Anim Sci Pap Rep 33:155–163.

1135 Organ CL, Edwards SV. 2011. Major Events in Avian Genome Evolution. In:
1136 Dyke G, Kaiser G, editors. Living Dinosaurs: the evolutionary history of
1137 modern birds. Hoboken, NJ: John Wiley & Sons, Ltd. pp. 325–337.

1138 Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. 2007. Origin of avian
1139 genome size and structure in non-avian dinosaurs. Nature 446:180–184.

1140 Ödeen A, Håstad O. 2003. Complex distribution of avian color vision systems
1141 revealed by sequencing the SWS1 opsin from total DNA. Molecular Biology
1142 and Evolution 20:855.

1143 Ödeen A, Håstad O. 2013. The phylogenetic distribution of ultraviolet sensitivity
1144 in birds. BMC Evol Biol 13:36.

1145 Pichlmair A, Kandasamy K, Alvisi G, Mulhern O, Sacco R, Habjan M, Binder M,
1146 Stefanovic A, Eberle C-A, Goncalves A, et al. 2012. Viral immune modulators
1147 perturb the human molecular network by common and unique strategies.
1148 Nature 487:486–490.

1149 Pienaar J, Ilany A, Geffen E, Yom-Tov Y. 2013. Macroevolution of life-history
1150 traits in passerine birds: adaptation and phylogenetic inertia. Ecol Lett
1151 16:571–576.

1152 Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis
1153 of RNA-seq incorporating quantification uncertainty. Nat Meth 14:687–690.

1154 Pinheiro J, Bates D, DebRoy S, Sarkar D. 2013. nlme: Linear and nonlinear
1155 mixed effects models. R package.

1156 Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using
1157 phylogenies. Bioinformatics 21:676–679.

1158 Qian F, Chung L, Zheng W, Bruno V, Alexander R, Wang Z, Wang X, Kurscheid
1159 S, Zhao H, Fikrig E, et al. 2013. Identification of genes critical for resistance
1160 to infection by West Nile virus using RNA-Seq analysis. Viruses 5:1664–
1161 1681.

1162 Quintana-Murci L, Clark AG. 2013. Population genetic tools for dissecting innate
1163 immunity in humans. Nat Rev Immunol 13:280–293.

1164 R Core Development Team. 2008. R: A Language and Environment for
1165 Statistical Computing.

1166 Randall RE, Goodbourn S. 2008. Interferons and viruses: an interplay between
1167 induction, signalling, antiviral responses and virus countermeasures. J Gen
1168 Virol 89:1–47.

1169 Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology
1170 (and other things). Methods Ecol Evol 3:217–223.

1171 Rojas-Peña ML, Vallejo A, Herrera S, Gibson G, Arévalo-Herrera M. 2015.
1172 Transcription profiling of malaria-naïve and semi-immune Colombian
1173 volunteers in a *Plasmodium vivax* sporozoite challenge. PLoS Negl Trop Dis
1174 9:e0003978–18.

1175 Roth AC, Gonnet GH, Dessimoz C. 2008. Algorithm of OMA for large-scale
1176 orthology inference. BMC Bioinformatics 9:518.

1177 Rothenburg S, Seo EJ, Gibbs JS, Dever TE, Dittmar K. 2008. Rapid evolution of
1178 protein kinase PKR alters sensitivity to viral inhibitors. Nat Struct Mol Biol
1179 16:63–70.

1180 Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. 2014.
1181 Patterns of positive selection in seven ant genomes. Mol Biol Evol 31:1661–
1182 1685.

1183 Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP,
1184 Clarke JA, Baker AJ, Clamp M, Edwards, SV. 2018. Convergent regulatory
1185 evolution and the origin of flightlessness in palaeognathous birds. bioRxiv:1–
1186 31.

1187 Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007.
1188 Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet
1189 39:1461–1468.

1190 Samuel MA, Whitby K, Keller BC, Marri A, Barchet W, Williams BRG, Silverman
1191 RH, Gale M, Diamond MS. 2006. PKR and RNase L contribute to protection
1192 against lethal West Nile Virus infection by controlling early viral spread in the
1193 periphery and replication in neurons. J Virol 80:7009–7019.

1194 Santhakumar D, Rubbenstroth D, Martinez-Sobrido L, Munir M. 2017. Avian
1195 interferons and their antiviral effectors. Front. Immunol. 8:49.

1196 Schlenke TA, Begun DJ. 2003. Natural selection drives *Drosophila* immune
1197 system evolution. Genetics 164:1471–1480.

1198 Schrom EC, Prada JM, Graham AL. 2017. Immune signaling networks: Sources
1199 of robustness and constrained evolvability during coevolution. Mol Biol Evol
1200 35:676–687.

1201 Schubert I, Vu GTH. 2016. Genome stability and evolution: Attempting a holistic
1202 view. Trends Plant Sci 21:749–757.

1203 Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, Wang W, Eden J-S, Shen J-J,
1204 Liu L, et al. 2018. The evolutionary history of vertebrate RNA viruses. Nature
1205 540:1–16.

1206 Sironi M, Cagliani R, Forni D, Clerici M. 2015. Evolutionary insights into host-
1207 pathogen interactions from mammalian sequence data. Nat Rev Genet
1208 16:224–236.

1209 Smith J, Burt DW, Bencina D. 2015. The Avian RNAseq Consortium: a
1210 community effort to annotate the chicken genome. Cytogenet Genome Res
1211 145:78–179.

1212 Smith J, Smith N, Le Yu, Paton IR, Gutowska MW, Forrest HL, Danner AF, Seiler
1213 JP, Digard P, Webster RG, et al. 2015. A comparative analysis of host
1214 responses to avian influenza infection in ducks and chickens highlights a role
1215 for the interferon-induced transmembrane proteins in viral resistance. BMC
1216 Genomics 16:574.

1217 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
1218 analysis of large phylogenies. Bioinformatics (Oxford, England) 30:1312–
1219 1313.

1220 Sun H, Liu P, Nolan LK, Lamont SJ. 2015a. Avian pathogenic *Escherichia coli*
1221 (APEC) infection alters bone marrow transcriptome in chickens. BMC
1222 Genomics 16:690.

1223 Sun H, Liu P, Nolan LK, Lamont SJ. 2015b. Novel pathways revealed in bursa of
1224 fabricius transcriptome in response to extraintestinal pathogenic *Escherichia*
1225 *coli* (ExPEC) infection. PLoS ONE 10:e0142570–17.

1226 Sun H, Liu P, Nolan LK, Lamont SJ. 2016. Thymus transcriptome reveals novel
1227 pathways in response to avian pathogenic *Escherichia coli* infection. Poultry
1228 Sci 95:2803–2814.

1229 Tran TM, Jones MB, Ongoiba A, Bijker EM, Schats R, Venepally P, Skinner J,
1230 Doumbo S, Quinten E, Visser LG, et al. 2016. Transcriptomic evidence for
1231 modulation of host inflammatory responses during febrile *Plasmodium*
1232 *falciparum* malaria. Sci Rep 6:31291.

1233 Tricola GM, Simons MJP, Atema E, Boughton RK, Brown JL, Dearborn DC,
1234 Divoky G, Eimes JA, Huntington CE, Kitaysky AS, et al. 2018. The rate of
1235 telomere loss is related to maximum lifespan in birds. Philos Trans R Soc
1236 Lond B Biol Sci 373:20160445.

1237 Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-like receptor
1238 evolution in birds: gene duplication, pseudogenisation and diversifying
1239 selection. Mol Biol Evol msy119, https://doi.org/10.1093/molbev/msy119

1240 Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false
1241 inferences of lineage-specific positive selection. Nat Ecol Evol 2:1280–1288.

1242 Videvall E, Cornwallis CK, Palinauskas V, Valkiūnas G, Hellgren O. 2015. The
1243 avian transcriptome response to malaria infection. Mol Biol Evol 32:1255-
1244 1267.

1245 Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle
1246 S, White S, Vilella AJ, Fairley S, et al. 2010. The genome of a songbird.
1247 Nature 464:757–762.

1248 Waterhouse RM, Kriventseva EV, Meister S, Xi Z. 2007. Evolutionary dynamics
1249 of immune-related genes and pathways in disease-vector mosquitoes.
1250 Science 316:1738-1743.

1251 Wong GK-S, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, Meng Q, Zhou J, Li D,
1252 Zhang J, et al. 2004. A genetic variation map for chicken with 2.8 million
1253 single-nucleotide polymorphisms. Nature 432:717–722.

1254 Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of
1255 statistical methods for detecting adaptive evolution in protein coding
1256 sequences and for identifying positively selected sites. Genetics 168:1041–
1257 1051.

1258 Wright NA, Gregory TR, Witt CC. 2014. Metabolic “engines” of flight drive
1259 genome size reduction in birds. P R Soc B 281:20132780–20132780.

Wu Y, Hadly EA, Teng W, Hao Y, Liang W, Liu Y, Wang H. 2016. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. Sci Rep 6:33578.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. molecular biology and evolution 24:1586–1591.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16:284–287.

Zhang G, Li B, Li C, Gilbert MTP, Jarvis ED, Wang J, Avian Genome Consortium. 2014. Comparative genomic data of the Avian Phylogenomics Project. GigaSci 3:26.

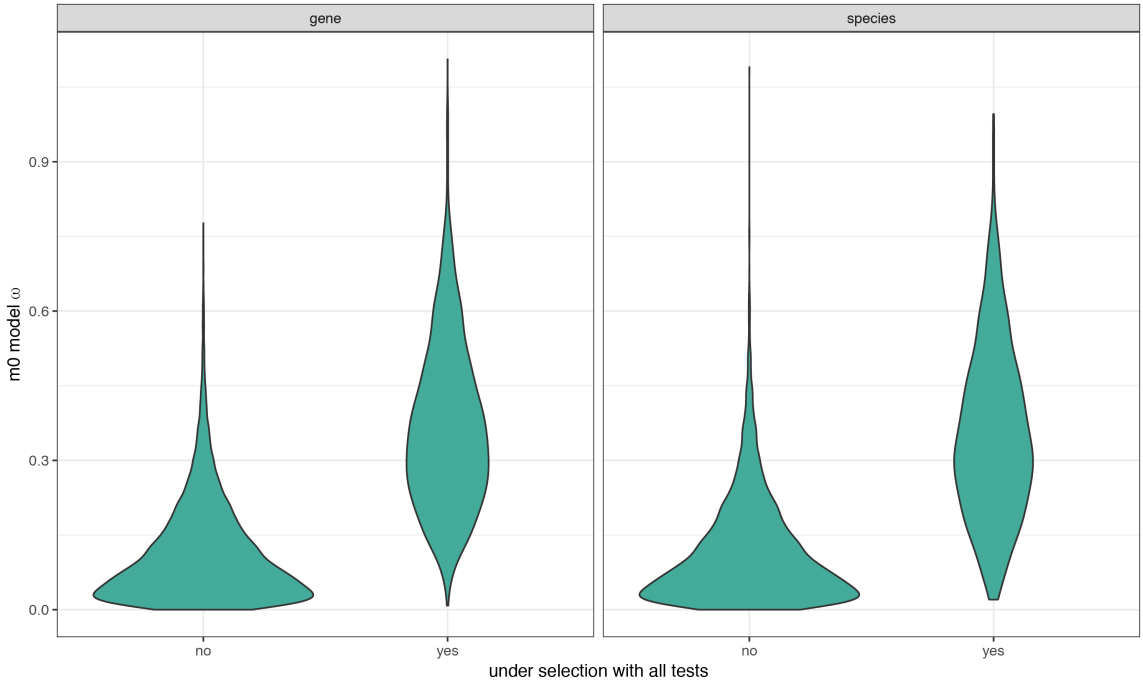
Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zhang J, Kaiser MG, Deist MS, Gallardo RA, Bunn DA, Kelly TR, Dekkers JCM, Zhou H, Lamont SJ. 2018. Transcriptome analysis in spleen reveals differential regulation of response to Newcastle disease virus in two chicken lines. Sci. Rep. 8:1278.

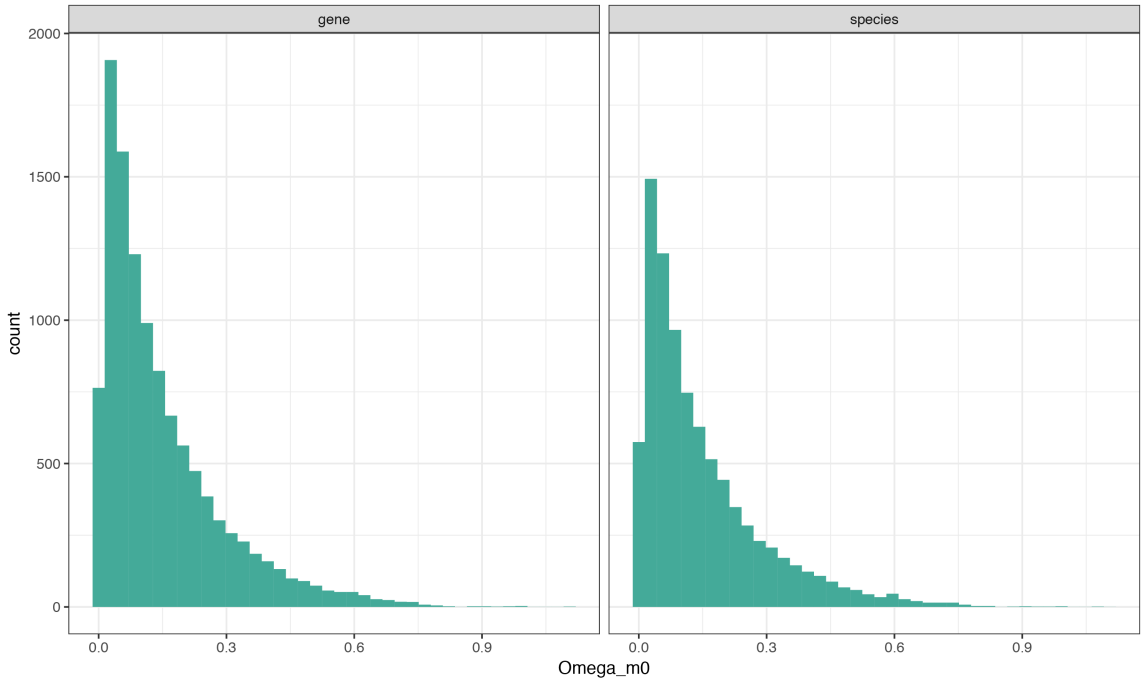
Zhang Q, Edwards SV. 2012. The evolution of intron size in amniotes: A role for powered flight? Gen Biol Evol 4:1033–1043.

Zhang Y, Mao D, Roswit WT, Jin X, Patel AC, Patel DA, Agapov E, Wang Z, Tidwell RM, Atkinson JJ, et al. 2015. PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to enhance interferon signaling and control viral infection. Nat Immunol 16:1215–1227.

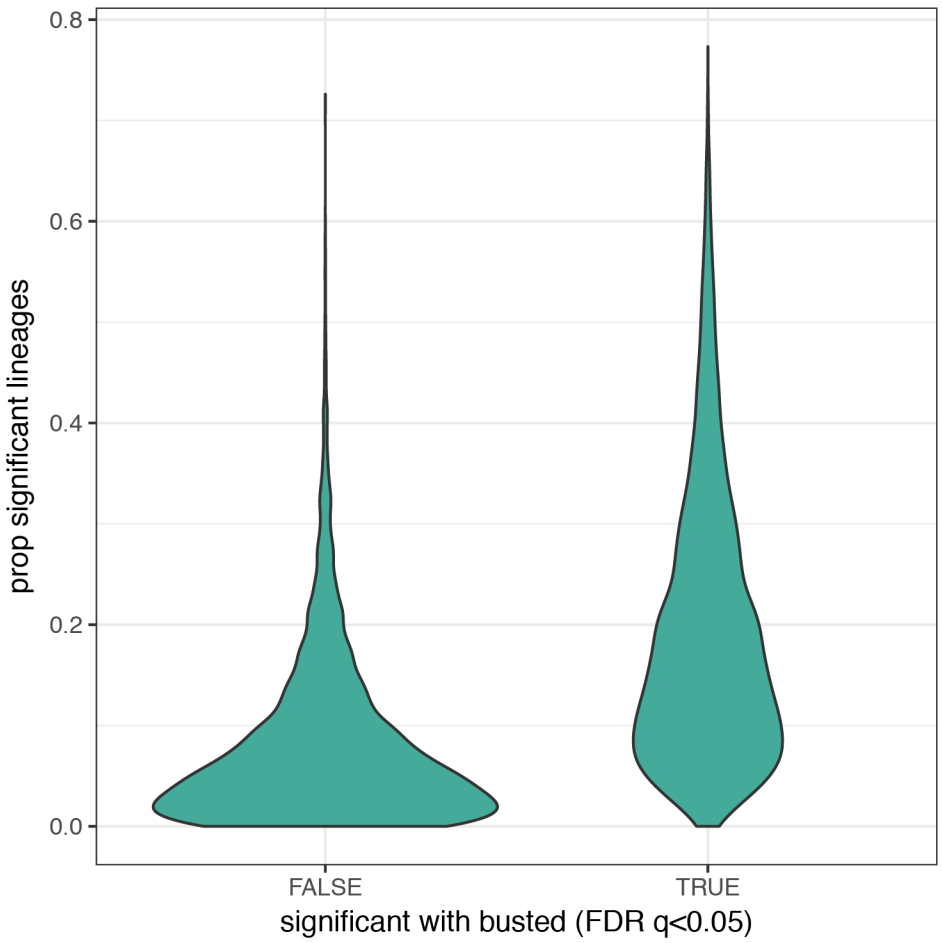
1288 **SUPPLEMENTAL FIGURES:**



1289
1290 **Supplemental Figure 1.** Comparison of M0 model ω values between genes under
1291 selection for all site tests, and those not under selection, for both gene trees and species
1292 trees. The mean ω values were significantly higher for genes under selection for both
1293 gene trees and species trees (Mann-Whitney U-test: gene trees: W = 1201387, $p < 0.0001$;
1294 species trees: W = 938934, $p < 0.0001$).
1295

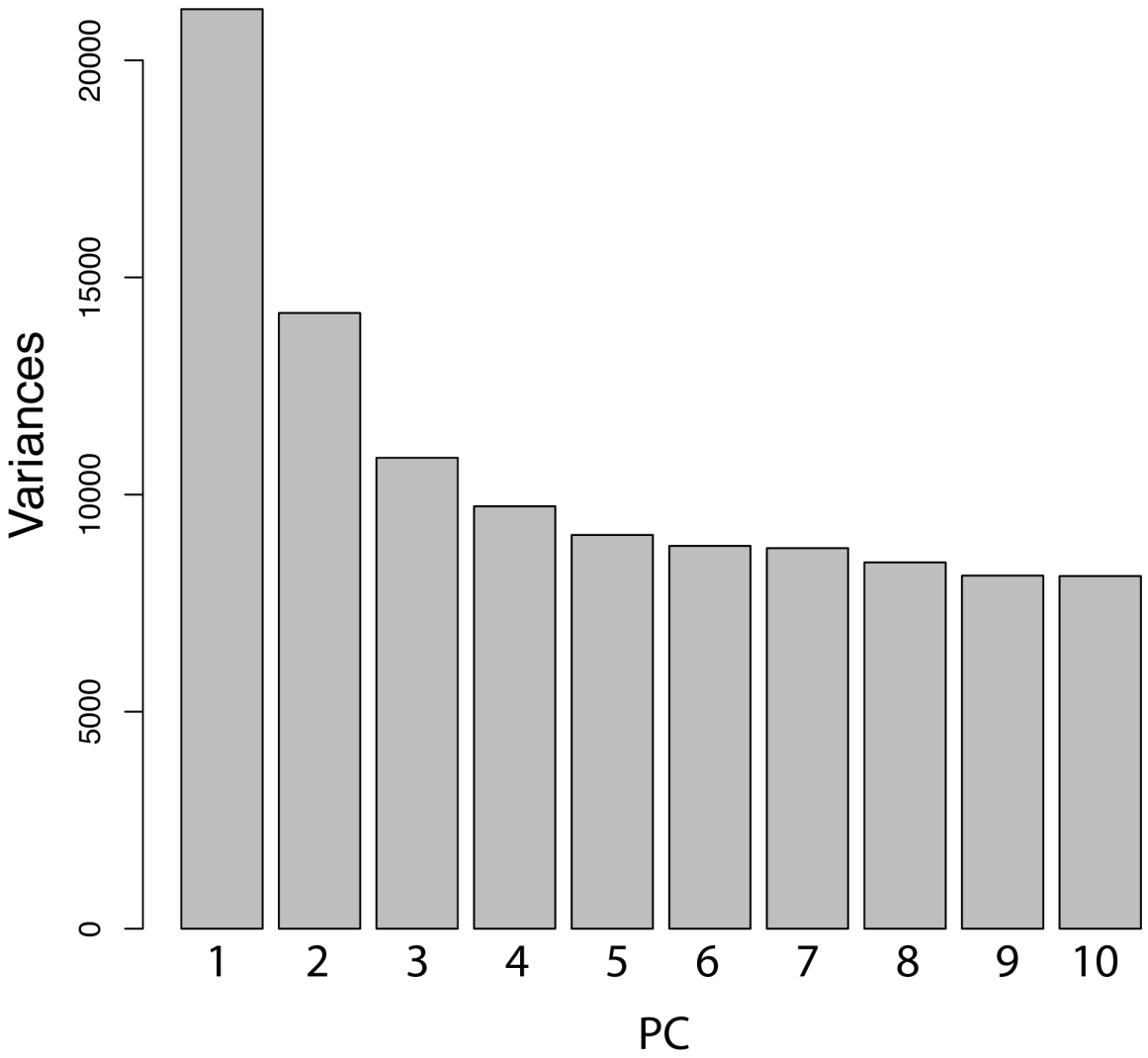


Supplemental Figure 2. Histogram of ω values from the PAML M0 model using either gene trees or species trees as the input phylogeny.

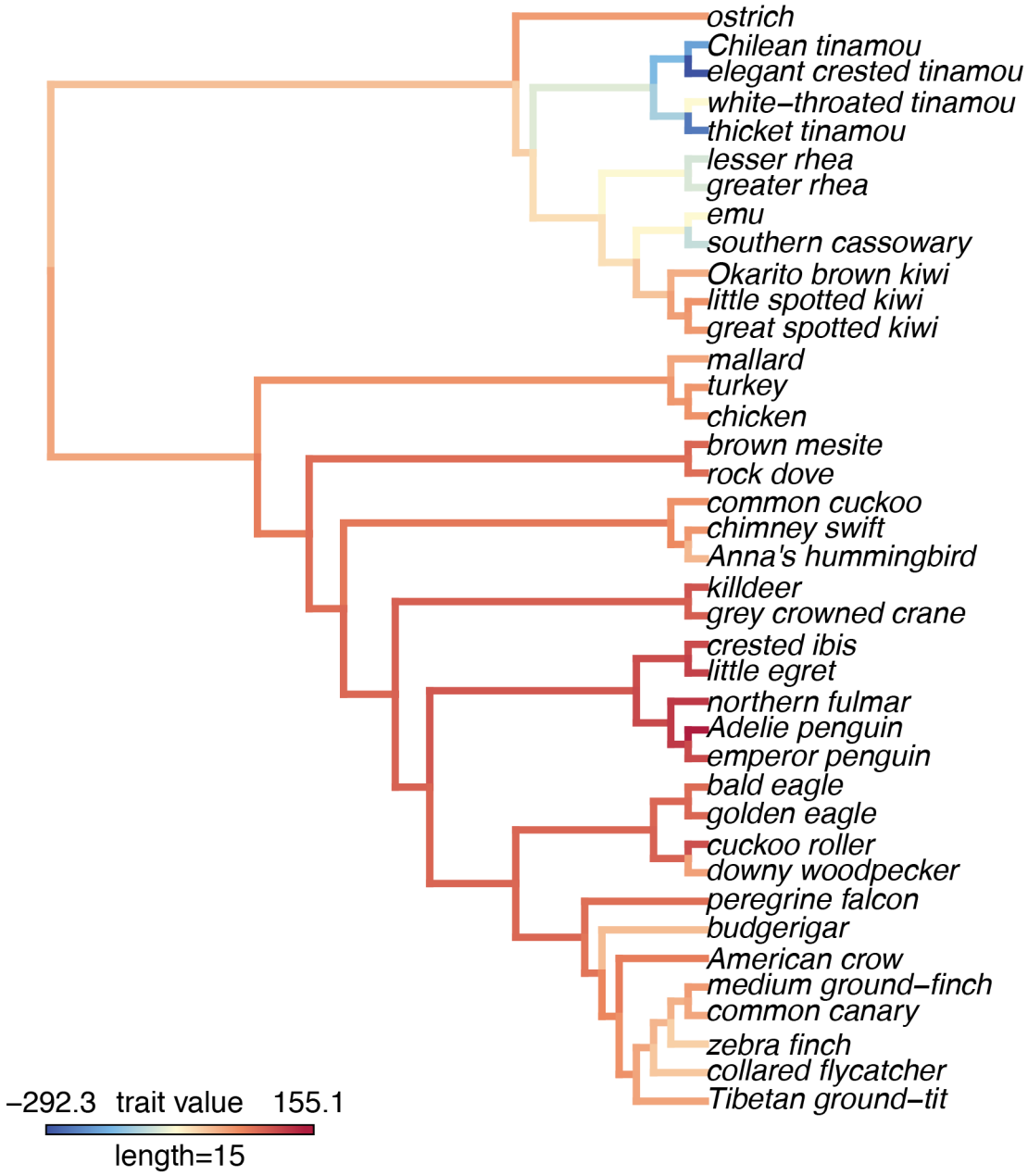


1299

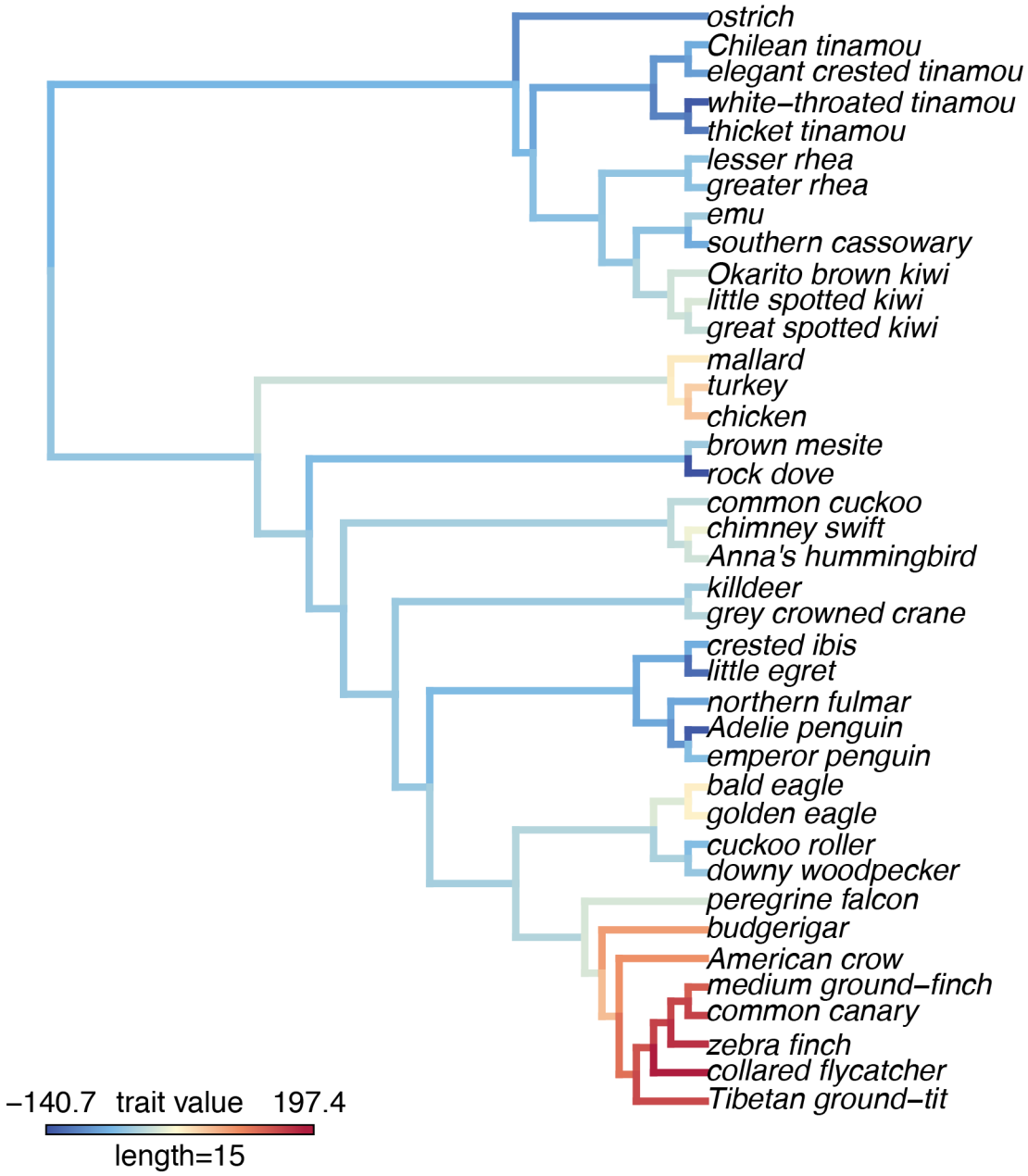
1300 **Supplemental Figure 3.** Distribution of the proportion of significant lineages for HOGs
1301 identified as not significant (FDR-corrected p-value >= 0.05), or significant (FDR-
1302 corrected p-values < 0.05) with BUSTED. The means of the two distributions are
1303 significantly different (Mann-Whitney U-test: W= 6205530, $p<10^{-16}$).



Supplemental Figure 4. Visualization of the variance explained by the first 10 PC axes (scree plot). PCA of the log-transformed p-values across genes for each species



Supplemental Figure 5. A visualization of PC2 scores on the phylogeny, the maximum likelihood reconstruction of the PC2 values for internal branches. PC2 explains 5.2% of the variance among log-transformed p-values across genes for each species.



Supplemental Figure 6. A visualization of PC3 scores on the phylogeny, the maximum likelihood reconstruction of the PC3 values for internal branches. PC3 explains 3.9% of the variance among log-transformed p-values across genes for each species.