Evolution of olfactory receptors in birds

By

Robert Driver

December, 2022

Director of Dissertation: Michael Brewer, PhD

Major Department: Biology

ABSTRACT

Olfaction is an evolutionary ancient sensation, and is the perception and interpretation of chemical stimuli from surrounding air or water. Olfaction is an essential sensory modality for nearly all animals, and is used to define territories, to identify kinship, to navigate to breeding sites, to select mates, and when selecting mates. Unlike vision, which detects different wavelengths of a single particle, the photon, olfaction must detect a wide range of odor molecules. Odor molecules can be simple or complex, be large or small, and have a wide range of elements and chemical structures. To detect these diverse compounds, animals employ olfactory receptors, which constitute the largest gene family in all vertebrates. The total number of olfactory receptors that a species possesses can be used as a measurement of that species' reliance upon smell in ecology and behavior. Despite the importance of smell and olfactory receptors in mammals, little is known about olfactory receptors in birds.

The lack of knowledge of olfactory receptors in birds stems from centuries old misconceptions about birds relying on vision over olfaction in their behavior, leading scientists to historically overlook the use of smell in birds. Recent behavioral work is gradually debunking the notion that birds cannot smell, showing that birds use smell in similar ways to mammals, in

foraging, individual recognition, and mate choice. However, research into olfactory receptors in birds continues to lag behind other vertebrate classes.

My dissertation shows that birds have much larger olfactory receptor repertoires than the scientific community previously appreciated. In chapter 1, I show the discovery of hundreds of new olfactory receptors in birds, overlooked in previous studies, and show that olfactory receptors in birds, particularly the bird-specific gamma-c OR subfamily, can only be properly counted using genome assemblies that employ long-read sequencing technology. Knowing the importance of long-read assemblies for obtaining accurate olfactory receptor counts, I then expand olfactory receptor counts to 70 bird species with publicly available long read genomes, showing large olfactory repertoires across the bird phylogeny. I also show the dynamic birth and death of olfactory receptors through bird evolution, with a particularly high rate of death in the early lineages of the Neoaves bird group. However, our genomic counts only tell us the number of olfactory receptor genes in the genome, and do not directly implicate the olfactory receptors in a role specific to smell. To do this, in chapter 3, I show that the vast majority of olfactory receptors detected in the genomes of birds are indeed expressed in the olfactory epithelium, the tissue located inside the bird's bill that is relevant to smell and the olfactory system. I further show that the gamma-c olfactory receptor subfamily is expressed in the olfactory epithelium, and that certain members of the family are expressed at high levels. These findings show that birds across the phylogeny likely use smell in their behavior and ecology, and that this sensory modality should not be overlooked in birds. My research paves the way for future studies to match bird olfactory receptors to the odors they respond to and to discover the odors that birds detect.

Evolution of olfactory receptors in birds

A Dissertation

Presented to the Faculty of the Department of Biology and the Interdisciplinary Doctoral Program in Biology, Biomedicine, and Chemistry

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy, Interdisciplinary Doctoral Program in Biology, Biomedicine, and Chemistry

By

Robert Driver

December, 2022

Director of Dissertation: Michael Brewer, PhD
Thesis Committee Members:
Christopher Balakrishnan, PhD
John Stiller, PhD
Maude Baldwin, PhD

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER 1: HIGHLY CONTIGUOUS GENOMES IMPROVE THE UNDERSTANDING OF AVIAN OLFACTORY RECEPTOR REPERTOIRES 1

Abstract 1

Introduction 1

Methods 4

Assembly selection 5

Olfactory receptor identification 8

Classification of final OR set 9

Results 10

OR totals 10

OR counts relative to previous studies 12

Assembly effects on OR subgenome 12

Physical location of ORs in avian genomes 13

Discussion 15

Long-read sequencing is critical for characterizing avian OR repertoire 15

Phylogenetic implications of updated OR counts 16

Towards a better understanding of the avian OR subgenome 17

Long read assembly methods better characterize complex gene families 19

Funding 20

Acknowledgements 20

References 21

CHAPTER 2: EVOLUTION OF OLFACTORY RECEPTOR REPERTOIRES ACROSS AVIAN PHYLOGENY 27

Abstract 27

Introduction 28

Methods 30
Assembly selection 30
OR identification and classification 31
Estimation of tree topology 32
Estimation of branch lengths 33
Trait analyses: data collection 33
Trait analyses: phylogenetic generalized least squares 35
Phylogenetic analyses of OR counts: ancestral state reconstructions 36
Phylogenetic analyses of OR counts: branch birth and death rates 37
Results 37
OR totals 37
Ancestral state reconstruction 38
Olfactory receptor birth and death rates 42
Comparisons of OR counts and traits 47
Discussion 51
Olfactory capabilities are potentially widespread in birds 51
OR counts declined in early diverging Neoaves 52
Dynamic birth and death of ORs across the bird tree 54
Olfactory bulb size correlates with OR repertoire counts 56
Diet and song learning are not correlated with OR counts 57
Nocturnality increases total OR counts 59
No evidence for influence of genome size on OR count 60
Conclusion 61
References 62
CHAPTER 3: FUNCTIONAL CHARACTERIZATION OF OLFACTORY RECEPTORS IN THE CONTEXT OF THEIR RADIATION IN BIRDS 69
Abstract 69
Introduction 69

Methods 73
Sample collection 73
RNA extractions and sequencing 74
Read mapping 75
Counting and differential expression 76
Results 77
ORs found in tissues 77
Differential expression 79
Discussion 80
Most genomic ORs are expressed in OE 80
High expression levels of ORs, including gamma-c ORs 85
Few differentially expressed ORs and high variance between samples 86
Conclusion 89
References 90

LIST OF TABLES

1. Table 1.1 List of assemblies used 6

2. Table 1.2 OR counts from short and long read assemblies..... 11

LIST OF FIGURES

1. Fig. 1.1. Maximum-likelihood phylogenetic reconstruction of OR repertoire from long-read assemblies 7
2. Fig. 1.2. Distribution of gamma-c ORs among chromosomes and scaffolds in chromosome-level assemblies 14
3. Fig. 1.3. Magnification of the largest OR cluster in the chicken genome on chromosome 33..... 19
4. Fig 2.1. Ancestral character states of bird ORs in two topologies..... 40
5. Fig 2.2. Ancestral state reconstruction of bird OR subfamily repertoires, generated by maximum likelihood using the fastAnc function in phytools in R 45
6. Fig 2.3. Highest OR birth and death rates across two topologies. 46
7. Fig. 2.4. Comparison of OR counts between diurnal and nocturnal bird species..... 49
8. Fig. 2.5. Significant correlations between OR counts and olfactory bulb size to telencephalon size ratio and whole brain ratio..... 50
9. Fig 3.1. OR gene expression in OE and pectoralis in log CPM..... 83
10. Fig 3.2. OR expression levels and the total number of ORs expressed varied substantially between OE samples within species. 84

I. HIGHLY CONTIGUOUS GENOMES IMPROVE THE UNDERSTANDING OF AVIAN OLFACTORY RECEPTOR REPERTOIRES

Abstract

Third generation (long read-based) sequencing technologies are reshaping our understanding of genome structure and function. One of the most persistent challenges in genome biology has been confidently reconstructing radiations of complex gene families. Olfactory receptors (ORs) represent just such a gene family with upwards of 1000s of receptors in some mammalian taxa. Whereas in birds olfaction was historically an overlooked sensory modality, new studies have revealed an important role for smell. Chromosome-level assemblies for birds allow a new opportunity to characterize patterns of OR diversity among major bird lineages. Previous studies of short read (second-generation) genome assemblies have associated OR gene family size with avian ecology, but such conclusions could be premature if new assembly methods reshape our understanding of avian OR evolution. Here we provide a fundamental characterization of OR repertoires in five recent genome assemblies, including the most recent assembly of golden-collared manakin (*Manacus vitellinus*). We find that short read-based assemblies systematically undercount the avian-specific gamma-c OR subfamily, a subfamily that comprises over 65 percent of avian OR diversity. Therefore, in contrast to previous studies we find a high diversity of gamma-c ORs across the avian tree of life. Building on these findings, ongoing sequencing efforts and improved genome assemblies will clarify the relationship between OR diversity and avian ecology.

Introduction

Our understanding of avian sensory biology has progressed substantially in recent years. Studies have discovered fantastic ways that birds experience the world, including the visual detection of non-spectral colors, the detection of sugar via a repurposed umami taste receptor in nectivorous species, and amphibious hearing in cormorants (Baldwin et al. 2014; Larsen et al. 2020; Stoddard et al. 2020). However, while studies investigating most senses, particularly bird vision, have received considerable attention, research into olfaction has lagged behind. Misconceptions about bird olfaction date back nearly 200 years, when John James Audubon falsely concluded that turkey vultures (*Cathartes aura*) could not smell carrion (Audubon 1826). Darwin also performed behavioral experiments on Andean condors (*Vultur gryphus*) to conclude that they could not smell meat (Darwin 1891). An examination of olfactory bulb size across a diversity of bird species concluded that birds could not have anything more than a rudimentary sense of smell (Hill 1905). In response to these conclusions, bird olfaction remained relatively unexplored until behavioral studies showed odor recognition in pigeons (Michelsen 1959). Following this study, there have been a wealth of morphological and behavioral studies testing for olfaction in both captive and wild birds (Bang and Cobb 1968; Hagelin 2007; Gwinner and Berger 2008; Nevitt et al. 2008; Krause et al. 2012; Van Huynh and Rice 2019).

To follow this appreciation for the behavioral and ecological roles of olfaction in birds, researchers have characterized bird olfactory receptors at a genomic level. Olfactory receptors (ORs) are seven transmembrane domain rhodopsin-like G protein-coupled receptors that detect odors when expressed in the olfactory sensory neurons of the nasal epithelium (Buck and Axel 1991; Mombaerts 2004). In the protruding cilia of the olfactory sensory neurons, ORs recognize specific volatile compounds in their transmembrane domain binding pocket, which creates a signaling cascade that depolarizes the cell membrane and sends an action potential to the

olfactory bulb glomeruli and later the brain (summarized in Breer 2003). Each OR may recognize one or multiple odorants, and each odorant may be detected by one or multiple ORs, and so in this way, species may perceive a wide array of odors (Saito et al. 2009).

ORs constitute one of the largest gene families in vertebrates. For example, the elephant genome contains about 2,000 intact ORs (Niimura et al. 2014). In birds, there are three major subfamilies of ORs, the alpha, gamma, and gamma-c (Niimura and Nei 2005). The alpha and gamma subfamilies are shared between all amniotes (Niimura and Nei 2005; Steiger et al. 2009). The third subfamily, gamma-c, is unique to birds (Niimura and Nei 2005; Silva et al. 2020). The gamma-c subfamily is numerous in some bird genomes comprising over 65% of the OR repertoire (Steiger et al. 2009; Khan et al. 2015). Gamma-c ORs are similar in sequence and sequences cluster by species rather than by orthologs among species (Steiger et al. 2009). Gamma-c ORs likely evolve with a high level of birth and death rates and gene conversion to maintain the species-specific clustering (Niimura and Nei 2005; Steiger et al. 2009).

The first genomic investigations to determine bird OR repertoire counts provided further evidence for the potential of birds to recognize a wide variety of odors. A total of 214 intact ORs were reported in chicken (*Gallus gallus*) and 134 reported in zebra finch (*Taeniopygia guttata*) by Steiger et al. (2009). The finding of hundreds of ORs in the chicken and zebra finch genomes were replicated using multiple OR identifying pipelines (Wang et al. 2013; Khan et al. 2015; Vandewege et al. 2016). All of these studies identified similar OR counts as well as similar proportions of each OR subfamily, with the gamma-c family dominating the OR repertoire. The majority of ORs, however, were located on unmapped contigs, including over 90% of gamma-c ORs in chicken and zebra finch.

Second-generation (Illumina, short read-based) genomes greatly broadened genome sampling across the tree of life, including birds (Jarvis et al. 2014). However, in these assemblies, intact OR numbers were significantly lower than had been observed in the Sanger-based chicken and zebra finch assemblies (Steiger et al. 2009; Khan et al. 2015). Particularly absent from analyses was the distinct avian radiation of the gamma-c OR subfamily, with 45 of 46 species with short-read assemblies yielding fewer than 25 gamma-c ORs (Khan et al. 2015). Despite sequencing technology being a common thread in the 45 assemblies with lower OR counts, technical explanations were ruled out in favor of evolutionary explanations for the observed patterns of diversity (Khan et al. 2015).

Chromosome-level reference genomes, using long-read sequencing technology, should provide more reliable information about OR repertoire diversity in birds. The Vertebrate Genomes Project recently expanded chromosome scale-assembly methods from model organisms across the vertebrate tree of life (Rhie et al. 2021). Combining these and other new assemblies, we are now able to characterize OR diversity in five bird genomes in which long-read approaches have been deployed (Feng et al. 2020; Liu et al. 2021; Rhie et al. 2021). Included in our species analyses is the new assembly of golden-collared manakin (*Manacus vitellinus*) that was sequenced as part of a collaborative effort within the National Science Foundation supported Research Coordination Network for biologists studying manakins (Pipridae). We directly compare Sanger, Illumina, hybrid, and Pac-Bio based assemblies to examine the ways in which our understanding of bird OR family repertoire, and our comprehension of avian olfactory capabilities, are shaped by these higher-quality assemblies.

Methods

Assembly selection

We sought to compare OR discovery rates and assembly quality by using select bird species with multiple publicly available genome assemblies on GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Assemblies for each species varied in the sequencing technology employed and assembly software used (Table 1). In order to examine how variation in genome sequencing methods impacts OR discovery and description, we included one genome of each species with long-read sequencing technology (Pacific Biosciences (Pac Bio), RSII or Sequel) as well as one genome without long-read sequencing. We obtained two assemblies from five bird species: emu (*Dromaius novaehollandiae*), chicken (*Gallus gallus*), Anna's hummingbird (*Calypte anna*), golden-collared manakin (*Manacus vitellinus*), and zebra finch (*Taeniopygia guttata*) (Warren et al. 2010; Zhang et al. 2014; Feng et al. 2020; Liu et al. 2021; Rhie et al. 2021, Table 1). In addition to the availability of multiple genomes, these 5 species are representative across the three major groupings of extant birds, including the Paleognathae and two groups within the Neognathae, Galloanseres and Neoaves, represent diverse ecology, and include two important avian models, chicken and zebra finch (Fig. 1A).

Table 1.1. List of assemblies used.

Species	Abbreviation	Accession	contigN50 (Mb)	Data types	Assembler
M. vitellinus[1]	Mvit1	GCF_000692015.1	0.04	Illumina	SOAPdenovo
M. vitellinus[2]	Mvit3	GCF_001715985.3	0.29	PacBio/Illumina	MaSuRCA
G. gallus[3]	Ggal4	GCF_000002315.3	0.30	Sanger/454	Celera
G. gallus[3]	Ggal6	GCF_000002315.6	17.65	PacBio	Falcon
T. guttata[4]	Tgut1	GCF_000151805.1	0.038	Sanger	PCAP
T. guttata[5]	Tgut2	GCA_003957565.3	12.00	PacBio/10x/Bionano/HiC	Falcon etc.
D. novaehollandiae	Dnov1	GCA_013396795.1	0.86	Illumina	Allpaths-LG
D. novaehollandiae[6]	Dnov2	GCA_016128335.1	14.09	PacBio	Falcon
C. anna [1]	Cann1	GCF_000699085.1	0.03	Illumina	SOAPdenovo
C. anna[5]	Cann2	GCF_003957555.1	14.52	PacBio/10x/Bionano/HiC	Falcon etc.

1. Zhang et al. 2014
2. Feng et al. 2020
3. International Chicken Genome Consortium
4. Warren et al. 2010
5. Rhie et al. 2021
6. Liu et al. 2021

Note: Each species included has two representative assemblies. Within each species, one assembly was sequenced with either the Illumina or the Sanger sequencing platform, while the other assembly was sequenced at least in part with PacBio.

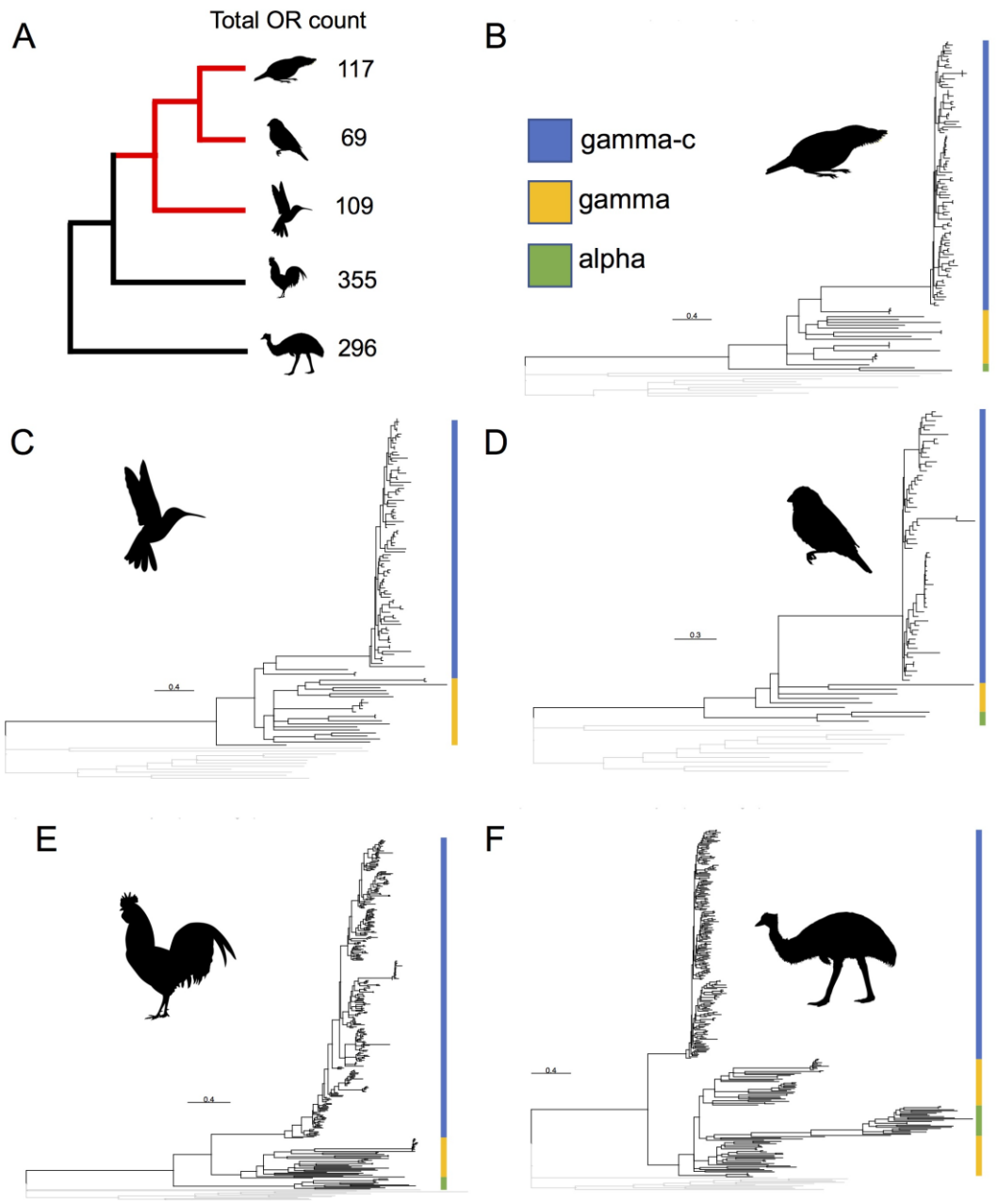

Fig. 1.1 Maximum-likelihood phylogenetic reconstruction of OR repertoire from long-read assemblies. Phylogenetic trees were assembled using IQ-TREE and only nodes with >50% support following a likelihood ratio test are shown. **(A)** Topological phylogeny of species used in this study is shown with red branches indicating Neoaves species. OR counts from long-read genomes for each species are given. The five species and assemblies shown are **(B)** *M. vitellinus* Mvit3, **(C)** *C. anna* Cann2, **(D)** *T. guttata* Tgut2, **(E)** *G. gallus* Ggut6, and **(F)** *D. novaehollandiae* Dnov2. The three OR subfamilies were assigned based on putative orthology to previously described bird ORs (Niimura and Nei 2005, Vandewege et al. 2016). Images not to scale. Image **(A)** is modified artwork by Kristen Orr.

Olfactory receptor identification

To detect putatively functional ORs in the selected genomes, we created a BLAST query with a set of 2,110 OR protein sequences from six mammals (*Ornithorhynchus anatinus*, *Didelphis virginiana*, *Bos taurus*, *Canis lupus*, *Rattus norvegicus*, *Macaca mulatta*), two birds (*Gallus gallus*, *Taeniopygia guttata*), and one crocodilian (*Gavialis gangeticus*). We obtained this query OR set by combining previously published OR subgenomes (Niimura and Nei 2007; Niimura 2009; Vandewege et al. 2016). Using this query file, we performed TBLASTN searches against all 11 bird genomes with a threshold of E <1e-20. To remove pseudogenized and truncated ORs, we filtered for hits >250 amino acids long. For any single location on the genome, we filtered out hits within 100bp of each other, and selected the lowest e-value associated with that location.

After obtaining unique BLAST hits, we extracted the associated nucleotide sequence from the genome as well as 300bp regions flanking the hit both upstream and downstream. We used a modified Perl script from Beichman et al. (2019) to detect open reading frames (ORFs) within each extracted region (Montague et al. 2014; Beichman et al. 2019). We then aligned these ORFs to each other as well as to the human Olfactory Receptor Family 2 Subfamily J Member 3 (OR2J3) sequence using the E-INS-I default parameters in MAFFT (Katoh and Standley 2013). Using the previously characterized transmembrane domains of OR2J3 as a guide, we removed any sequences that had five or more amino acid insertions or deletions within a transmembrane domain in the alignment (McRae et al. 2012; Beichman et al. 2019). This included ORFs with stop codons appearing prior to the end of the seventh transmembrane domain.

Using this alignment, we recorded the position of the first amino acid in the first transmembrane domain. To estimate the location of the ORF start codon, we used modified Perl scripts from Beichman et al. (2019) to find the most appropriate methionine upstream of this recorded transmembrane start position (Montague et al. 2014; Beichman et al. 2019). ORF sequences were then truncated at the 5' ends to begin with this methionine. This set of ORFs was then aligned using the E-INSI-I parameters in MAFFT to a set of *T. guttata* reference ORs as well as 11 non-OR rhodopsin-like G-protein coupled receptors (non-OR GPCRs) that functioned as an outgroup (Katoh and Standley 2013; Niimura 2013; Vandewege et al. 2016; Beichman et al. 2019). We then used clustalW to generate a neighbor-joining tree from this alignment with 1000 bootstraps, gaps removed, and Kimura's distance correction (Kimura 1980; Goujon et al. 2010). We then removed any ORFs that were phylogenetically more closely related to the non-OR GPCRs.

Classification of final OR set

We classified all remaining ORFs as functional ORs. Using this final set, we ran a maximum likelihood tree using IQ-TREE with automatic model selection and 1000 SH-like approximate likelihood ratio test replicates (Minh et al. 2020). Using ML support values, we collapsed all nodes <50% support into a polytomy using iTOL software, and rooted the tree using the ancestral branch leading to the 11 non-OR GPCRs (Letunic and Bork 2019). We classified bird ORs into subfamilies alpha, gamma, and gamma-c based on the subfamily of the query sequence used to identify the OR and the location of the OR in one of the three distinct avian OR clades (Steiger et al. 2009; Vandewege et al. 2016). We then counted the final number of OR sequences as well as the number of ORs from each subfamily.

Results

OR totals

We identified a total of 1496 ORs across all 10 bird assemblies from five species (Table 2). Of these ORs, the gamma-c subfamily constituted 77% (1158) of the total, while 18% (263) of the identified ORs were gamma, and 5% (74) were alpha subfamily ORs. For assemblies with long-read sequencing, we found 946 ORs, with 42 alpha (4%), 162 gamma (17%), and 741 gamma-c (78%). Within a single assembly, the chicken Ggal6 (see assembly abbreviations in Table 1) yielded the largest number of ORs, with 355 total, 303 (85%) of which were gamma-c ORs (Fig. 1E). Gamma-c represented 97% (179/184) of the ORs in zebra finch Tgut1, the highest percent gamma-c out of total OR repertoire for any assembly.

Table 1.2 OR counts from short and long read assemblies

Name	Total ORs	Alpha ORs	Gamma ORs	Gamma-c ORs	Proportion gamma-c	contigN50 (Mb)	OR total in literature
Mvit1	9	1	8	0	0.000	0.04	9[1]
Mvit3	117	2	18	97	0.829	0.29	
Ggal4	272	10	33	229	0.842	0.30	266[2]
Ggal6	355	11	41	303	0.854	17.65	
Tgut1	184	2	3	179	0.973	0.038	
Tgut2	69	3	6	60	0.870	12.00	190[2]
Dnov1	57	17	33	7	0.123	0.86	
Dnov2	296	26	75	195	0.659	14.09	
Cann1	27	2	23	2	0.071	0.03	21[1]
Cann2	109	0	23	86	0.761	14.52	

1. Khan et al. 2015
2. Vandewege et al. 2016

Emu, Dnov2, yielded both the highest gamma counts (75) and alpha counts (26), in addition to 195 gamma-c ORs (Fig. 1F). In the OR maximum-likelihood phylogenies we elected to present each species separately for clarity of visualization (Fig. 1B-E). We noted that in these analyses the gamma OR subfamily for *D. novaehollandiae* were not recovered as monophyletic (Fig. 1F). In other multi-species analyses we have done this is not the case and this is also not the case in our Dnov2 neighbor-joining tree (analyses not shown). This unusual pattern here seems to be driven by the long branch at the base of the Dnov2 alpha OR clade.

OR counts relative to previous studies

Overall, reanalysis of previously analyzed genomes were consistent with previous findings (Khan et al. 2015; Vandewege et al. 2016, Table 2). For Ggal4, we recovered 272 ORs, six more than previously recovered previously (Vandewege et al. 2016). Two previous searches of the Tgut1 assembly yielded 182 and 190 ORs, similar to our search of 184 ORs (Wang et al. 2013; Vandewege et al. 2016, respectively). We also found similar subfamily diversity of gamma and gamma-c ORs in zebra finch and chicken (Niimura 2009; Steiger et al. 2009; Wang et al. 2013; Khan et al. 2015; Vandewege et al. 2016). Similar patterns emerged for the other short read assemblies in our analysis (Table 2) giving us confidence in our methods of recovering ORs.

Assembly effects on OR subgenome

We found that assembly had substantial effects on the ability to reconstruct OR subgenomes. In 4 of the 5 surveyed species, inclusion of long-read sequencing (Pacific Biosciences) increased OR counts (Table 2). The most pronounced effect on OR repertoire was in the gamma-c family, which also constitutes the majority of known avian ORs. Between *D. novaehollandiae* assemblies, contigN50 improved from 0.86Mb in Dnov1 with Illumina sequencing to 14.09Mb in Dnov2. We detected an additional 239 ORs in Dnov2 of which 188 (78.6%) were from the gamma-c family. In the case of *M. vitellinus*, no gamma-c representatives were recovered from Mvit1 in our analysis or a previous analysis (Khan et al. 2015), yet our search of Mvit3 yielded 97 gamma-c ORs (Fig. 1B).

Improved assemblies also resulted in the identification of additional ORs in the alpha and gamma subfamilies as well, but these effects were less pronounced (Table 2). The gamma

subfamily of *D. novaehollandiae* more than doubled in count in both Dnov2 compared to Dnov1 (42 new ORs), and in *M. vitellinus* between Mvit3 and Mvit1 (10 new ORs). In other species, the relative increase in gamma was smaller (Table 2). Alpha OR counts were similar between within-species assemblies (Table 2). Unexpectedly, we identified no alpha ORs in *Calypte anna* Cann2 assembly though two were previously reported based on Cann1.

The Sanger sequencing-based *T. guttata* genome Tgut1 unexpectedly yielded a greater number of ORs than Tgut2, which was assembled with several technologies. This species was the only case in which a newer assembly based on long-read technology yielded fewer ORs than the assembly without long-read sequencing. Despite an overall higher OR count in Tgut1, we found more alpha ORs (3 versus 2) and more gamma ORs (6 versus 3) in Tgut2 than Tgut1. However, the overall count of ORs in Tgut2 was substantially lower as there were 119 fewer gamma-c ORs in Tgut2 than Tgut1. Therefore, the lower OR count in Tgut2 is entirely due to differences in gamma-c OR recovery.

Physical location of ORs in avian genomes

Although new, chromosome-scale assemblies have assigned the vast majority of genomic sequence data to chromosomes (Rhie et al. 2021), gamma-c OR regions remain primarily assigned to unmapped scaffolds (Fig. 2). The exception to this rule is the Ggal6 assembly in which 302 out of 303 identified ORs have been assigned to chromosomes. Most of these (274 ORs) map to a single microchromosome 33 (Lee et al. 2020, Fig 3). The remaining ORs are divided between two other microchromosomes, chromosome 16 (8 ORs) and chromosome 31 (16 ORs) with a single receptor on an unmapped scaffold. For *C. anna* Cann2 only 20% of gamma-c receptors were on scaffolds assigned to chromosomes. The main clusters of ORs were

a group of 10 assigned to the W chromosome and another 18 assigned to a single scaffold (RRCD01000065.1). For *D. novaehollandiae* Dnov2 only 8% of gamma-c ORs are assigned to chromosomes, with a large cluster of 108 on scaffold JABVCD010000554.1 (Fig. 2). The remaining gamma-c loci were distributed among 29 chromosomes and scaffolds. Finally, only 3% (2/60) of *T. guttata* Tgut2 gamma-c ORs were assigned to chromosomes with a cluster of 24 loci on scaffold VOHI02000029.1.

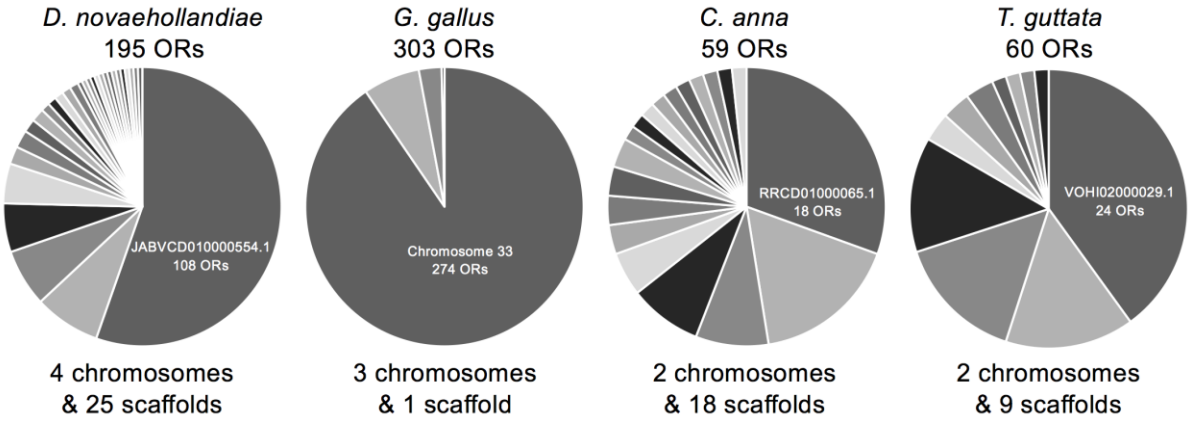

Fig. 1.2. Distribution of gamma-c ORs among chromosomes and scaffolds in chromosome-level assemblies. The largest gamma-c OR cluster in each long-read assembly is located on unmapped scaffolds except *G. gallus*, where the largest cluster is localized to chromosome 33, a microchromosome.

Discussion

Long-read sequencing is critical for characterizing avian OR repertoire

Our results show that Illumina short read-based approaches were not successful in accurately characterizing the gamma-c OR subfamily. The three Illumina-based assemblies we assessed undercounted gamma-c diversity, revealing fewer than 10 gamma-c ORs in each assembly. In all five of the assemblies sequenced with long-reads, we found that gamma-c ORs constituted at least 66% of the OR subgenome and on average there were 148 gamma-c loci per species. The hybrid assembly approach using MaSuRCA (Zimin et al. 2013) also substantially increased gamma-c recovery in Mvit3.

In the most phylogenetically comprehensive analysis of avian ORs to date, Khan et al. (2015) analyzed OR repertoire from 48 bird species. Forty-six of the assemblies surveyed were sequenced and assembled using short read-based methods. The two other species included were the Sanger-based chicken and zebra finch, and those had the highest OR diversity, which was attributed to ecological adaptations of these two particular species (Khan et al. 2015). Other reports in the literature also interpret a lack of gamma-c ORs as biologically meaningful without considering the shortcomings of short read-based assemblies (Zhan et al. 2013). Importantly, these issues extend beyond gamma-c ORs to other complex gene families such as the major histocompatibility complex (MHC). Long-read based studies are also improving our understanding of the avian MHC (He et al. 2021). Select MHC genes are linked to ORs in mammals, providing further evidence that this family repertoire may also be obscured in Illumina sequenced assemblies (Santos et al. 2010).

Prior to our analysis, chicken (Ggal3, Ggal4) and zebra finch (Tgut1) were the only Sanger-based assemblies analyzed (Niimura 2009; Steiger et al. 2009; Wang et al. 2013; Khan et al. 2015; Vandewege et al. 2016). Analyses of these assemblies, which involve longer read lengths and Bacterial Artificial Chromosome (BAC)-based scaffolding, provided the only previous evidence of substantial gamma-c OR diversity. Whereas the incorporation of long-read sequencing methods greatly increased the count of gamma-c ORs relative to Illumina-based assemblies, they also reduced the count of gamma-c in *T. guttata* compared to the previous Sanger-based assembly. We propose two potential reasons for this disparity. It is possible that the original Tgut1 assembly resolved alternative alleles as separate loci, artificially inflating the total gamma-c count with duplicate loci. Tgut2 and many third-generation assemblies are haplotype phased, mitigating this problem. Additionally, however, filtering steps at the end of the Tgut2 assembly curation process used for quality control may aggressively remove repetitive regions that harbor tandem gamma-c loci.

Phylogenetic implications of updated OR counts

Our finding of high OR diversity for *D. novaehollandiae* has potentially important implications for broad scale patterns of OR evolution in birds. High OR diversity in this paleognath genome contrasts with lower diversity found in previous analyses (Le Duc et al. 2015) and now suggests that the avian ancestor had high gamma-c OR diversity. Based on our assessment, all three OR subfamilies have lower diversity in the three neoavian lineages tested relative to *D. novaehollandiae* and *G. gallus.* Due to our limited taxonomic sampling, it remains unclear whether the differences reflect broader patterns of phylogenetic change or lineage-specific adaptations. For example, *D. novaehollandiae* and *G. gallus* are both omnivorous and

therefore may retain ORs as a variety of different odorants are relevant when foraging. The three Neoaves species we selected are not generalists, instead *C. anna* is mostly nectivorous, *M. vitellinus* is mostly frugivorous, and *T. guttata* is a granivore. Our understanding of phylogenetic patterns of OR diversity will continue to change with improved genome assembly and taxonomic sampling. For example by sequencing additional manakin species, we will be able to determine whether OR repertoire is elaborated within a frugivorous family or if counts vary substantially within individual lineages.

Towards a better understanding of the avian OR subgenome

Many key features of OR subgenomes and olfaction generally are well-characterized in mammals, but not in birds (Olender et al. 2008; Niimura et al. 2014). There are no previous reports of using assemblies with long-read sequencing to search for an OR repertoire in birds, and until now, no previous reports of an expansive gamma-c OR repertoire outside of *G. gallus* and *T. guttata* assemblies. To date, there is only one report of transcriptome sequencing an avian olfactory epithelium, a critical undertaking considering that ORs are expressed in non-olfactory tissues and some identified ORs may be nonfunctional despite having open reading frames (Pluznick et al. 2009; Sin et al. 2019). Although Sin et al. (2019) do show the expression of at least three gamma-c ORs in *Oceanodroma leucorhoa* olfactory epithelium, this still leaves the expression of a potentially large number of gamma-c genes uncharacterized. A current priority would be to sequence the olfactory epithelium transcriptome from species with high-quality genome assemblies to understand the extent to which a high number of gamma-c ORs recovered from the genome are functionally expressed in the OE.

Information about which ORs are expressed in the olfactory epithelium will also help with functional testing to identify the binding properties of avian ORs in a process known as "deorphanizing". To date, no avian ORs are deorphanized, and so it remains unclear what ligands are bound by any avian ORs. The diverse gamma-c ORs, unique to birds, are of particular interest. Avian olfaction is in many fundamental ways a frontier in the field of sensory biology.

There is also a great deal left to be learned about the molecular evolutionary mechanisms that have given rise to the diversity of gamma-c genes in birds. First, enhanced spatial information on the physical location of OR genes will be informative for understanding the genetic processes at play. With most loci still scattered among unmapped scaffolds it is somewhat unclear how clustered these loci are. That said, the *G. gallus* Ggal6 OR repertoire is highly spatially localized (Figures 2, 3), a pattern that is likely in the other species as well. Given spatial clustering, and extremely short branch-lengths, gamma-c species-specific clades could be a result of gene conversion among loci, but further study is needed. In the passerine bird MHC, endogenous retroviral elements may have played a role in generating gene family diversity in MHC Class II (Balakrishnan et al. 2010) and the same could be the case in the gamma-c radiation across birds. We note, however, that high repeat content is a general pattern across avian microchromosomes and is not restricted to OR and MHC regions (Fig. 3, Burt 2002).

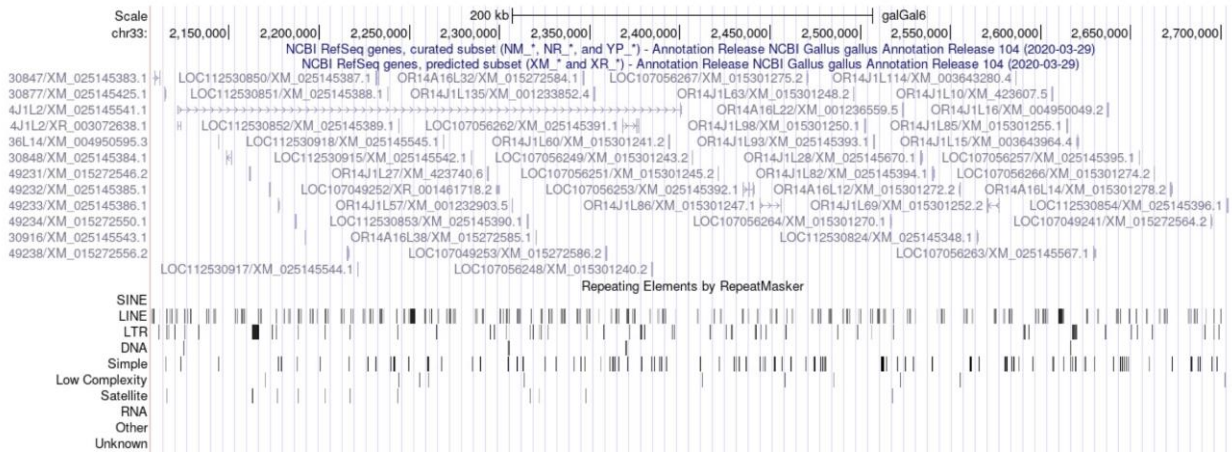

Fig. 1.3. Magnification of the largest OR cluster in the chicken genome on chromosome 33. Numerous ORs are present in the displayed region and are flanked by repetitive elements, potentially contributing to the difficulty of gamma-c OR subfamily discovery. Image from the UCSC Genome Browser (Lee et al. 2020).

Long read assembly methods better characterize complex gene families

Our increased OR counts in long-read assemblies contributes to the growing literature quantifying the advantages of third generation (long read) sequencing technology of second generation (short read) sequencing technology. Third generation sequencing has greatly improved the detection of tandem repeats generated by long terminal repeat retrotransposons, microsatellites, homonucleotide stretches, and repetitive regions (Mason et al. 2016; Kapusta and Suh 2017; Korlach et al. 2017). Large gene families found in clusters like the ORs described here may show the greatest improvement following the incorporation of long-read data. Long read sequencing has already led to better characterization of MHC loci in birds, and also greatly increased the resolution and count of vomeronasal receptors in mammals (Larsen et al. 2014; He et al. 2021).

Even with long read sequencing, the chromosomal location of all ORs remains largely unresolved. With the exception of the chicken, ORs in the long-read assemblies that we analyzed

largely mapped to unassigned scaffolds, including the largest OR cluster in each assembly (Fig. 2). The inability to assign these scaffolds to a chromosome is likely related to expansion of duplicate regions that contain OR loci and the high repeat element content found in microchromosomes (Fig. 3, Burt 2002). Indeed the assignment of an OR containing region to the hummingbird female-specific W chromosome is likely spurious and driven by the repetitive sequences in these regions. Bird chromosomes are highly syntenic, suggesting that the large OR cluster on the chicken microchromosome 33 likely match to homologous chromosomes (Nanda et al. 2011). Manual curation of these regions may be required to resolve remaining uncertainties. Other solutions to this complexity include physically mapping OR loci to chromosomes, and/or using approaches less dependent on assembly. Sin and colleagues (2019) incorporated assessment of depth of coverage in their OR quantification pipeline, an approach that should be informative in face of varying assembly quality.

Funding

This work was supported by National Science Foundation awards [grant numbers 1457541, IOS-1456612].

Acknowledgements

We would like to thank the editors two anonymous reviewers and Dr. Maude Baldwin for helpful discussion of the analyses presented here. In addition, we thank Drs. Michael Brewer and John Stiller are serving on RJDs dissertation committee. The *Dromaius novaehollandiae*, *Calypte anna*, and *Taeniopygia guttata* genomes were generated as part of the Vertebrate Genomes Project (https://vgp.github.io/genomeark/) and pre-publication public access to these now published data made these analyses possible. Likewise, we thank the Manakin Genomics

Research Coordination Network (https://www.manakinsrcn.org/) for generating public genomic resources for the manakin species analyzed here. This material is based upon work conducted while CNB was serving at the National Science Foundation.

References

Audubon J. J. (1826). Account of the habits of the turkey buzzard, *Vultur aura*, particularly with the view of exploding the opinion generally entertained of its extraordinary power of smelling. *Edinburgh New Philosophical Journal* 2:172-184.

Balakrishnan C. N., Ekblom R., Völker M., Westerdahl H., Godinez R., Kotkiewicz H., Burt D. W, Graves T., Griffin D. K., Warren W. C., Edwards S. V. (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology* 8:29.

Baldwin M. W., Toda Y., Nakagita T., O'Connell M. J., Klasing K. C., Misaka T., Edwards S. V., Liberles S. D. (2014). Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* 345:929–33.

Bang B. G., Cobb S. (1968). The size of the olfactory bulb in 108 species of birds. *The Auk* 85:55–61.

Beichman A. C., Koepfli K-P., Li G., Murphy W., Dobrynin P., Kliver S., Tinker M. T., Murray M. J., Johnson J., Lindblad-Toh K., Karlsson E. K., Lohmueller K. E., Wayne R. K. (2019). Aquatic adaptation and depleted diversity: a deep dive into the genomes of the sea otter and giant otter. *Molecular Biology and Evolution* 36:2631–55.

Breer H. (2003). Olfactory receptors: molecular basis for recognition and discrimination of odors. *Analytical and Bioanalytical Chemistry* 377:427–33.

Buck L., Axel R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 65:175–87.

Burt D. W. (2002). Origin and evolution of avian microchromosomes. *Cytogenetic and Genome Research* 96:97–112.

Darwin C. (1891). Journal of researches into the natural history and ecology of the countries visited during the voyage of H.M.S. Beagle round the world, Journal of researches into the natural history and ecology of the countries visited during the voyage of H.M.S. Beagle round the world London, England: Ward, Lock & Co.

Feng S., Stiller J., Deng Y., Armstrong J., Fang Q., Reeve A. H., Xie D., Chen G., Guo C., Faircloth B. C., Petersen B., Wang Z., Zhou Q., Diekhans M., Chen W., Andreu-Sánchez S., Margaryan A., Howard J. T., Parent C., Pacheco G., Sinding M-H. S., Puetz L., Cavill E., Ribeiro Â. M., Eckhart L., Fjeldså J., Hosner P. A., Brumfield R. T., Christidis L., Bertelsen M. F., Sicheritz-Ponten T., Tietze D. T., Robertson B. C., Song G., Borgia G., Claramunt S., Lovette I. J., Cowen S. J., Njoroge P., Dumbacher J. P.,

Ryder O. A., Fuchs J., Bunce M., Burt D. W., Cracraft J., Meng G., Hackett S. J., Ryan P. G., Jønsson K. A., Jamieson I. G., da Fonseca R. R., Braun E. L., Houde P., Mirarab S., Suh A., Hansson B., Ponnikas S., Sigeman H., Stervander M., Frandsen P. B., van der Zwan H., van der Sluis R., Visser C., Balakrishnan C. N., Clark A. G., Fitzpatrick J. W., Bowman R., Chen N., Cloutier A., Sackton T. B., Edwards S. V., Foote D. J., Shakya S. B., Sheldon F. H., Vignal A., Soares A. E. R., Shapiro B., González-Solís J., Ferrer-Obiol J., Rozas J., Riutort M., Tigano A., Friesen V., Dalén L., Urrutia A. O., Székely T., Liu Y., Campana M. G., Corvelo A., Fleischer R. C., Rutherford K. M., Gemmell N. J., Dussex N., Mouritsen H., Thiele N., Delmore K., Liedvogel M., Franke A., Hoeppner M. P., Krone O., Fudickar A. M., Milá B., Ketterson E. D., Fidler A. E., Friis G., Parody-Merino Á. M., Battley P. F., Cox M. P., Lima N. C. B., Prosdocimi F., Parchman T. L., Schlinger B. A., Loiselle B. A., Blake J. G., Lim H. C., Day L. B., Fuxjager M. J., Baldwin M. W., Braun M. J., Wirthlin M. , Dikow R. B. , Ryder T. B., Camenisch G., Keller L. F., DaCosta J. M., Hauber M. E., Louder M. I. M., Witt C. C., McGuire J. A., Mudge J., Megna L. C., Carling M. D., Wang B., Taylor S. A., Del-Rio G., Aleixo A., Vasconcelos A. T. R., Mello C. V., Weir J. T., Haussler D., Li Q., Yang H., Wang J., Lei F., Rahbek C., Gilbert M. T. P., Graves G. R., Jarvis E. D., Paten B., Zhang G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature* 587:252–57.

Goujon M., McWilliam H., Li W., Valentin F., Squizzato S., Paern J., Lopez R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research* 38:W695–99.

Gwinner H., Berger S. (2008). Starling males select green nest material by olfaction using experience-independent and experience-dependent cues. *Animal Behaviour* 75:971–76.

Hagelin J. C. (2007). The citrus-like scent of crested auklets: reviewing the evidence for an avian olfactory ornament. *Journal of Ornithology* 2:195–201.

He K., Minias P., Dunn P. O. (2021). Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution* 13:evaa270.

Hill A. (1905). Can birds smell? *Nature* 71:318–19.

Huynh A. V., Rice A. M. (2019). Conspecific olfactory preferences and interspecific divergence in odor cues in a chickadee hybrid zone. *Ecology and Evolution* 9:9671–83.

Jarvis E. D., Mirarab S., Aberer A. J., Li B., Houde P., Li C., Ho S. Y. W., Faircloth B. C., Nabholz B., Howard J. T., Suh A., Weber C. C., Fonseca R. R. da, Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M. S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W. C., Ray D., Green R. E., Bruford M. W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E. P., Bertelsen M. F., Sheldon F. H., Brumfield R. T., Mello C. V., Lovell P. V., Wirthlin M., Schneider M. P. C., Prosdocimi F., Samaniego J. A., Velazquez A. M. V., Alfaro-Núñez A., Campos P. F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert

D. M., Zhou Q., Perelman P., Driskell A. C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt FE., Braun M., Fjeldsa J., Orlando L., Barker F. K., Jønsson K. A., Johnson W., Koepfli K-P., O'Brien S., Haussler D., Ryder O. A., Rahbek C., Willerslev E., Graves G. R., Glenn T. C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S. V., Stamatakis A., Mindell D. P., Cracraft J., Braun E. L., Warnow T., Jun W., Gilbert M. T. P., Zhang G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–31.

Kapusta A., Suh A. (2017). Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences* 1:164–85.

Katoh K., Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–80.

Khan I., Yang Z., Maldonado E., Li C., Zhang G., Gilbert M. T. P., Jarvis E. D., O'Brien S. J., Johnson W. E., Antunes A. (2015). Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Molecular Biology and Evolution* 32:2832–43.

Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111–20.

Korlach J., Gedman G., Kingan S. B., Chin C-S., Howard J. T., Audet J-N., Cantin L., Jarvis E. D. (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience* 6.

Krause E. T., Krüger O., Kohlmeier P., Caspers B. A. (2012). Olfactory kin recognition in a songbird. *Biology Letters* 8:327–29.

Larsen O. N., Wahlberg M., Christensen-Dalsgaard J. (2020). Amphibious hearing in a diving bird, the great cormorant (*Phalacrocorax carbo sinensis*). *Journal of Experimental Biology* 223.

Larsen P. A., Heilman A. M, Yoder A. D. (2014). The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genomics* 15:720.

Le Duc D., Renaud G., Krishnan A., Almén M. S., Huynen L., Prohaska S. J., Ongyerth M., Bitarello B. D., Schiöth H. B., Hofreiter M., Stadler P. F., Prüfer K., Lambert D., Kelso J., Schöneberg T. (2015). Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology* 16:147.

Lee C. M., Barber G. P., Casper J., Clawson H., Diekhans M., Gonzalez J. N., Hinrichs A. S., Lee B. T., Nassar L. R., Powell C. C., Raney B. J., Rosenbloom K. R., Schmelter D., Speir M. L., Zweig A. S., Haussler D., Haeussler M., Kuhn R. M., Kent W. J. (2020). UCSC Genome Browser enters 20th year. *Nucleic Acids Research* 48:D756–61.

Letunic I., Bork P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:W256–59.

Liu J., Wang Z., Li J., Xu L., Liu J., Feng S., Guo C., Chen S., Ren Z., Rao J., Wei K., Chen Y., Jarvis E. D., Zhang G., Zhou Q. (2021). A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Research* 31:497–511.

Mason A. S., Fulton J. E., Hocking P. M., Burt D. W. (2016). A new look at the LTR retrotransposon content of the chicken genome. *BMC Genomics* 17:688.

McRae J. F., Mainland J. D., Jaeger S. R., Adipietro K. A., Matsunami H., Newcomb R. D. (2012). Genetic variation in the odorant receptor OR2J3 Is associated with the ability to detect the "grassy" smelling odor, cis-3-hexen-1-ol. *Chemical Senses* 37:585–93.

Michelsen W. J. (1959). Procedure for studying olfactory discrimination in pigeons. *Science* 130:630–31.

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., von Haeseler A., Lanfear R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–34.

Mombaerts P. (2004). Odorant receptor gene choice in olfactory sensory neurons: the one receptor–one neuron hypothesis revisited. *Current Opinion in Neurobiology* 14:31–36.

Montague M. J., Li G., Gandolfi B., Khan R., Aken B. L., Searle S. M. J., Minx P., Hillier L. W., Koboldt D. C., Davis B. W., Driscoll C. A., Barr C. S., Blackistone K., Quilez J., Lorente-Galdos B., Marques-Bonet T., Alkan C., Thomas G. W. C., Hahn M. W., Menotti-Raymond M., O'Brien S. J., Wilson R. K., Lyons L. A., Murphy W. J., Warren W. C. (2014). Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proceedings of the National Academy of Sciences* 111:17230–35.

Nanda I., Benisch P., Fetting D., Haaf T., Schmid M. (2011). Synteny conservation of chicken macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenetic and Genome Research* 132:165–81.

Nevitt G. A., Losekoot M., Weimerskirch H. (2008). Evidence for olfactory search in wandering albatross, Diomedea exulans. *Proceedings of the National Academy of Sciences* 105:4576–81.

Niimura Y. (2009). On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biology and Evolution* 1:34–44.

Niimura Y. (2013). Identification of Olfactory Receptor Genes from Mammalian Genome Sequences. In: Crasto CJ, editor. Olfactory Receptors: Methods and Protocols. Methods in Molecular Biology Totowa, NJ: Humana Press. p. 39–49.

Niimura Y., Matsui A., Touhara K. (2014). Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Research* 24:1485–96.

Niimura Y., Nei M. (2005). Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences* 102:6039–44.

Niimura Y., Nei M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2:e708.

Olender T., Lancet D., Nebert D. W. (2008). Update on the olfactory receptor (OR) gene superfamily. *Human Genomics* 3:87.

Pluznick J. L., Zou D-J., Zhang X., Yan Q., Rodriguez-Gil D. J., Eisner C., Wells E., Greer C. A., Wang T., Firestein S., Schnermann J., Caplan M. J. (2009). Functional expression of the olfactory signaling system in the kidney. *Proceedings of the National Academy of Sciences* 106:2059–2064.

Rhie A., McCarthy S. A., Fedrigo O., Damas J., Formenti G., Koren S., Uliano-Silva M., Chow W., Fungtammasan A., Gedman G. L., Cantin L. J., Thibaud-Nissen F., Haggerty L., Lee C., Ko B. J., Kim J., Bista I., Smith M., Haase B., Mountcastle J., Winkler S., Paez S., Howard J., Vernes S. C., Lama T. M., Grutzner F., Warren W. C., Balakrishnan C. N., Burt D., George J. M., Biegler M., Iorns D., Digby A., Eason D., Edwards T., Wilkinson M., Turner G., Meyer A., Kautt A. F., Franchini P., Detrich H. W., Svardal H., Wagner M., Naylor G. J. P., Pippel M., Malinsky M., Mooney M., Simbirsky M., Hannigan B. T., Pesout T., Houck M., Misuraca A., Kingan S. B., Hall R., Kronenberg Z., Korlach J., Sović I., Dunn C., Ning Z., Hastie A., Lee J., Selvaraj S., Green R. E., Putnam N. H., Ghurye J., Garrison E., Sims Y., Collins J., Pelan S., Torrance J., Tracey A., Wood J., Guan D., London S. E., Clayton D. F., Mello C. V., Friedrich S. R., Lovell P. V., Osipova E., Al-Ajli F. O., Secomandi S., Kim H., Theofanopoulou C., Zhou Y., Harris R. S., Makova K. D., Medvedev P., Hoffman J., Masterson P., Clark K., Martin F., Howe K., Flicek P., Walenz B. P., Kwak W., Clawson H., Diekhans M., Nassar L., Paten B., Kraus R. H. S., Lewin H., Crawford A. J., Gilbert M. T. P., Zhang G., Venkatesh B., Murphy R. W., Koepfli K-P., Shapiro B., Johnson W. E., Palma F. D., Margues-Bonet T., Teeling E. C., Warnow T., Graves J. M., Ryder O. A., Hausler D., O'Brien S. J., Howe K., Myers E. W., Durbin R., Phillippy A. M., Jarvis E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592:737–746.

Saito H., Chi Q., Zhuang H., Matsunami H., Mainland J. D. (2009). Odor coding by a mammalian receptor repertoire. *Science Signaling* 2:ra9–ra9.

Santos P. S. C., Kellermann T., Uchanska-Ziegler B., Ziegler A. (2010). Genomic architecture of MHC-linked odorant receptor gene repertoires among 16 vertebrate species. *Immunogenetics* 62:569–84.

Silva M. C., Chibucos M., Munro J. B., Daugherty S., Coelho M. M., Silva J. C. (2020). Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis for smell in procellariiform seabirds. *Scientific Reports* 10:543.

Sin S. Y. W., Cloutier A., Nevitt G., Edwards S. V. (2022). Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* 220:iyab210.

Steiger S. S., Kuryshev V. Y., Stensmyr M. C., Kempenaers B., Mueller J. C. (2009). A comparison of reptilian and avian olfactory receptor gene repertoires: Species-specific expansion of group γ genes in birds. *BMC Genomics* 10:446.

Stoddard M. C., Eyster H. N., Hogan B. G., Morris D. H., Soucy E. R., Inouye D. W. (2020). Wild hummingbirds discriminate nonspectral colors. *Proceedings of the National Academy of Sciences* 117:15112–22.

Vandewege M. W., Mangum S. F., Gabaldón T., Castoe T. A., Ray D. A., Hoffmann F. G. (2016). Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biology and Evolution* 8:470–80.

Wang Z., Pascual-Anaya J., Zadissa A., Li W., Niimura Y., Huang Z., Li C., White S., Xiong Z., Fang D., Wang B., Ming Y., Chen Y., Zheng Y., Kuraku S., Pignatelli M., Herrero J., Beal K., Nozawa M., Li Q., Wang J., Zhang H., Yu L., Shigenobu S., Wang J., Liu J., Flicek P., Searle S., Wang J., Kuratani S., Yin Y., Aken B., Zhang G., Irie N. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nature Genetics* 45:701–6.

Warren W. C., Clayton D. F., Ellegren H., Arnold A. P., Hillier L. W., Künstner A., Searle S., White S., Vilella A. J., Fairley S., Heger A., Kong L., Ponting C. P., Jarvis E. D., Mello C. V., Minx P., Lovell P., Velho T. A. F., Ferris M., Balakrishnan C. N., Sinha S., Blatti C., London S. E., Li Y., Lin Y-C., George J., Sweedler J., Southey B., Gunaratne P., Watson M., Nam K., Backström N., Smeds L., Nabholz B., Itoh Y., Whitney O., Pfenning A. R., Howard J., Völker M., Skinner B. M., Griffin D. K., Ye L., McLaren W. M., Flicek P., Quesada V., Velasco G., Lopez-Otin C., Puente X. S., Olender T., Lancet D., Smit A. F. A., Hubley R., Konkel M. K., Walker J. A., Batzer M. A., Gu W., Pollock D. D., Chen L., Cheng Z., Eichler E. E., Stapley J., Slate J., Ekblom R., Birkhead T., Burke T., Burt D., Scharff C., Adam I., Richard H., Sultan M., Soldatov A., Lehrach H., Edwards S. V., Yang S-P., Li X., Graves T., Fulton L., Nelson J., Chinwalla A., Hou S., Mardis E. R., Wilson R. K. (2010). The genome of a songbird. *Nature* 464:757–62.

Zhang G., Li B., Li C., Gilbert M. T. P., Jarvis E. D., Wang J., The Avian Genome Consortium. (2014). Comparative genomic data of the Avian Phylogenomics Project. *GigaScience* 3.

Zimin A. V., Marçais G., Puiu D., Roberts M., Salzberg S. L., Yorke J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29:2669–77.

II. EVOLUTION OF OLFACTORY RECEPTOR REPERTOIRES ACROSS AVIAN PHYLOGENY

Abstract

Olfaction is a critical sensory modality, allowing animals to process information from environmental chemicals. It plays a central role in recognizing food, mates, predators, territories, and kin. Olfactory receptors (ORs), a gene family largely expressed in the olfactory epithelium, are responsible for odor detection. To accommodate the incredible variety of odorants in nature, olfactory receptors constitute the largest gene family in vertebrates, with over 1,000 genes in some mammals and over 300 genes in some bird species. Birds are a highly diverse class of vertebrates, inhabiting nearly all land environments, with a broad range of social systems and foraging strategies. Yet, early 20th century studies dismissed the use of olfaction in birds, a misconception that at one time pervaded sensory biology. More recently, studies have shown that birds indeed rely on olfaction in behavior and ecology, such as locating food and nesting material, and in individual and species recognition. To contribute to the rapidly expanding knowledge of bird olfaction, in this study, we show that bird have many more OR genes that previously detected, and that the majority of bird ORs are in an OR subgroup unique to birds, called the gamma-c OR subfamily. Using a dataset of 70 long read bird genome assemblies, we show that the highest surveyed OR counts occur in rails (*Laterallus jamaicensis*) and with the lowest counts occurring in crows, specifically *Corvus monedula* and *Corvus corone*. We mapped ancestral OR repertoires and show that OR counts declined early in the Neoaves lineage 60-70 million years ago, but OR counts remained high through the Cretaceous-Paleogene extinction

event in Palaeognathae and Galloanserae. We show that nocturnality increases OR counts, and OR counts correlate with increased olfactory bulb size. Taken together, we show that the OR superfamily in birds experienced dynamic births and deaths throughout the bird tree, reflecting the ability of olfaction to adapt and support bird behavior and ecology.

Introduction

Olfaction is essential for survival and reproduction in many animals. It plays a central role in foraging, avoiding predation, kin recognition, and territorial behavior. In vertebrates, air or waterborne odor molecules are detected with olfactory receptors (ORs), a gene family of G protein-coupled receptors expressed in the olfactory sensory neurons (OSNs) of the olfactory epithelium (OE, Buck and Axel 1991, Strotmann et al. 1992). To accommodate the incredible variety of odorants in nature, ORs constitute the largest gene family in vertebrates, with over 1,000 genes in some mammals and over 300 genes in some birds (Niimura et al. 2014, Niimura and Nei 2005).

Birds are the most speciose class of terrestrial vertebrates, inhabiting nearly all land environments. Among birds there is high diversity of social structures and foraging strategies, yet birds were long thought to rely on visual rather than olfactory signals (Audubon 1826, Hill 1905). Recent behavioral work in birds has shown important roles for olfaction in foraging, locating nest sites, seed caching behavior, and species recognition, among other behaviors (Buitron and Nuechterlein 1985, Molina-Morales et al. 2020, Bonnadonna and Gagliardo 2021, Wikelski et al. 2021, Van Huynh and Rice 2021). Additionally, specific bird species rely on a

highly specialized olfactory system for foraging, including *Cathartes aura* (turkey vulture) and many seabirds (Procellariformes, Owre and Northington 1961, Grubb 1972).

In addition to the recent surge of interest in how olfaction influences bird behavior, we showed that birds have many more OR genes in their genomes that previously realized (Driver and Balakrishnan 2021, see Chapter 1). Genomic analysis divides bird species' OR repertoires into three phylogenetic subgroups: alpha, gamma, and gamma-c ORs (Niimura and Nei 2005, Steiger et al. 2009, Driver and Balakrishnan 2021). The alpha and gamma OR subgroups are shared across tetrapods: chicken alpha and gamma ORs form phylogenetic clades with alpha and gamma ORs from amphibians, reptiles, and mammals (Niimura and Nei 2005, Steiger et al. 2009, Vandewege et al. 2016). This illustrates a degree of sequence conservation in the OR repertoire of these subgroups despite at least 315 million years of divergence between mammal and bird lineages (Laurin and Reisz 1995). Contrastingly, the gamma-c OR subgroup is only present in birds (Niimura and Nei 2005, Steiger e al. 2009, Driver and Balakrishnan 2021). Previous studies show that the gamma-c OR subfamily was the most abundant OR clade in most species (Steiger et al. 2009, Khan et al. 2015). For example, the gamma-c subfamily constituted over 85% of all OR genes in the zebra finch (60 total gamma-c ORs) and chicken (303 total gamma-c ORs, Driver and Balakrishnan 2021). Phylogenetic analyses of OR repertoires containing multiple bird species reveal that gamma-c ORs cluster into species-specific clades as opposed to showing clear orthologous relationships among species (Zhan et al. 2013, Silva et al. 2020), suggesting possible species-specific roles for the gamma-c. Gamma-c ORs within a species also have shorter phylogenetic terminal branch lengths compared to alpha and gamma ORs, showing a high degree of sequence similarity between gamma-c genes (Steiger et al. 2009, Silva et al. 2020). Together, these patterns suggest that gamma-c ORs evolve through a dynamic

birth-and-death model of gene evolution, with ubiquitous duplication events occurring over short evolutionary time scales that post-date the divergence of many modern bird genera (Silva et al. 2020). However, without accurate counts of olfactory receptors across the bird phylogeny, we do not know the patterns of olfactory receptor turnover across the vast diversity of the bird phylogeny.

Only in the last five years have numerous long read bird assemblies become publicly available on NCBI's GenBank, making accurate comparisons of OR counts across the bird phylogeny possible, including across all three of the major bird lineages: Palaeognathae, Galloanserae, and Neoaves (Bravo et al. 2021). We therefore investigated OR gene family and subfamily counts across the bird phylogeny to detect any lineage-specific gains and losses in ORs. We tested for associations between OR counts and the diverse ecological niches and diets of the our bird species dataset. From these results, we hope to understand the evolutionary patterns of olfactory receptors, including the gamma-c, and gain a better understanding of the importance of smell in the life of birds.

Methods

Assembly selection

We investigated OR diversity in birds by selecting multiple publicly available genome assemblies on GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Assemblies for each species implemented some form of long-read sequencing technology, including Pacific Biosciences or Oxford Nanopore methods. Genomes varied in the assembly methods used and in the size and total number of contigs and scaffolds. We selected only assemblies using long read sequencing

due to the difficulty in recovering total OR counts in assemblies with shorter contigs (Driver and Balakrishnan 2021). In total, we analyzed 70 different bird assemblies, including species from the three main lineages of birds, the Palaeognathae, Galloanserae, and Neoaves. The set of 70 species represent diverse ecology, diets, and trophic levels.

OR identification and classification

To detect putatively functional ORs in the selected genomes, we created a BLAST query with a set of 2,110 OR protein sequences from 6 mammals (*Ornithorhynchus anatinus, Didelphis virginiana, Bos taurus, Canis lupus, Rattus norvegicus, Macaca mulatta*), 2 birds (*Gallus gallus, Taeniopygia guttata*), and 1 crocodilian (*Gavialis gangeticus*). We obtained this query OR set by combining previously published OR subgenomes (Niimura and Nei 2007; Niimura 2009; Vandewege et al. 2016). Using this query file, we performed TBLASTN searches against all 70 bird genomes with a threshold of $E < 1e{-}20$. The TBLASTN –num_alignments option was set to 200,000 to capture all genomic ORs similar to a single query sequence. To remove pseudogenized and truncated ORs, we filtered for hits > 250 amino acids long. For any single location on the genome, we filtered out hits within 100 bp of each other, and selected the lowest *E*-value associated with that location.

After obtaining unique BLAST hits, we extracted the associated nucleotide sequence from the genome as well as 300-bp regions flanking the hit both upstream and downstream. We used a modified Perl script from Beichman et al. (2019) to detect open reading frames (ORFs) within each extracted region (Montague et al. 2014; Beichman et al. 2019). We then aligned these ORFs to each other as well as to the human Olfactory Receptor Family 2 Subfamily J

Member 3 (OR2J3) sequence using the E-INS-I default parameters in MAFFT (Katoh and Standley 2013). Using the previously characterized transmembrane domains of OR2J3 as a guide, we removed any sequences that had five or more amino acid insertions or deletions within a transmembrane domain in the alignment (McRae et al. 2012; Beichman et al. 2019). This included ORFs with stop codons appearing prior to the end of the seventh transmembrane domain.

Using this alignment, we recorded the position of the first amino acid in the first transmembrane domain. To estimate the location of the ORF start codon, we used modified Perl scripts from Beichman et al. (2019) to find the most appropriate methionine upstream of this recorded transmembrane start position (Montague et al. 2014; Beichman et al. 2019). ORF sequences were then truncated at the 5' ends to begin with this methionine. This set of ORFs was then aligned using the E-INSI-I parameters in MAFFT to a set of *T. guttata* reference ORs as well as 11 non-OR rhodopsin-like G-protein coupled receptors (non-OR GPCRs) that functioned as an outgroup (Katoh and Standley 2013; Niimura 2013; Vandewege et al. 2016; Beichman et al. 2019). We then used clustalW to generate a neighbor-joining tree from this alignment with 1000 bootstraps, gaps removed, and Kimura's distance correction (Kimura 1980; Goujon et al. 2010). We then removed any ORFs that were phylogenetically more closely related to the non-OR GPCRs.

We classified all remaining ORFs as functional ORs. We classified bird ORs into subfamilies alpha, gamma, and gamma-c based on the subfamily of the query sequence used to identify the OR and the location of the OR in one of the three distinct avian OR clades (Steiger et al. 2009;

Vandewege et al. 2016). We then counted the final number of OR sequences as well as the number of ORs from each subfamily.

Estimation of tree topology

To analyze olfactory receptor counts in a phylogenetic context, we sought to create a phylogeny of the bird species set. The bird species used in this study are a unique set, with no preexisting published phylogenies containing all 70 species in a single tree. Therefore, we used topologies from seven existing phylogenies in the literature. We used previously published phylogenies to delineate relationships within the orders Accipitriformes and Passeriformes and within the families Falconidae, Corvidae, and Psittacidae (Wright et al. 2008, Haring et al. 2012, Mindell et al. 2018, Wink 2018, Oliveras et al. 2019). For topological relationships between orders, we referenced two established competing phylogenies in the literature (Jarvis et al. 2014, Prum et al. 2015). We created two separate topologies, both following the same topology for within-family level relationships, but one topology following the intra-order relationships in Jarvis et al. and one following Prum et al. This choice to include multiple competing topologies is due to the contentious nature of the phylogenetic relationships in birds following the Cretaceous-Paleogene extinction (Jarvis et al. 2014). Between 60 to 70 million years ago, the Neoaves lineage of birds underwent rapid diversification to form all modern day Neoaves orders, and the relative timing of various lineage divergences is disputed between different molecular datasets (Jarvis et al. 2014, Prum et al. 2015). Therefore, we created two topologies corresponding to each phylogeny (Jarvis et al. 2014, Prum et al. 2015).

Estimation of branch lengths

To determine the branch lengths for our literature-based topologies, we mined the 70 genomes for ultraconserved elements (UCEs). We used the UCE 5K probe set available in the phyluce pipeline to search for 5,472 UCEs from the 70 bird genomes (Faircloth et al. 2012). We recovered 5,044 UCEs from this search and, using custom shell scripts, extracted these UCEs from the bird assemblies. Using further shell scripts, we assigned the top hit in each bird assembly from each UCE query to a fasta file. In this way, we obtained 5,044 fasta files, each containing the top UCE hit from each bird assembly. We then aligned the individually-grouped UCEs using the E-INSI-I parameters in MAFFT (Katoh and Standley 2013). We then ran the FASconCAT perl script (Kuck and Meusemann 2010) to concatenate all UCEs from individual species. Together, this created one concatenated alignment of all UCEs for the 70 bird species.

Using the input topologies and the UCE concatenated alignment, we generated branch lengths using IQ-TREE (Minh et al. 2020). We used a partition file generated by FASconCAT (Kuk and Meusemann 2010) to partition the concatenated alignment into each individual input UCE. We set all partitions to the general time-reversible (GTR+FO) substitution mode, a partition rich substitution model that allows all substitution rates and base frequencies to occur at different rates (GTR), with base frequencies optimized by maximum likelihood (+FO) (Mihn et al. 2020). We then ran IQ-TREE twice, one for each input topology (Jarvis et al. 2014, Prum et al. 2015). We viewed output trees using iTOL software, and rooted the tree appropriately (Letunic and Bork 2019).

Trait analyses: data collection

For each bird species, we collected a variety of trait data for comparisons with olfactory receptor counts. As a positive control, previous research shows that olfactory receptor size positively correlates with olfactory bulb size (Steiger et al. 2008). We used a previously published dataset of olfactory bulb measurements, and recorded the ratio of log olfactory bulb volume to both telencephalon volume (the section of the brain where the olfactory bulb is located) and total brain volume (as recorded in Corfield et al. 2015). We omitted species in this analysis that were not represented in the Corfield et al. dataset. Using information from Birds of the World (Billerman et al. 2022), we recorded whether each species is nocturnal or diurnal, has a learned song or innate song, and whether the species is mostly terrestrial or aquatic. When selecting these traits, we considered the possible sensory trade-offs, such as decreased vision in nocturnal species and increased reliance on auditory cues in song learning species. We also recorded the trophic level and diet of each species from the EltonTratis 1.0 dataset (Wilman et al. 2014). To understand the potential relationship between transposable element proliferation and olfactory receptor counts, we also recorded the estimated overall genome size for each species, using the Animal Genome Size Database (Gregory 2022). Overall variation in genome size is driven in large part but the extent of repeat element proliferation (Kidwell 2002), and repeat element proliferation is associated with gene duplication events (Kidwell 2002). For species without a recorded genome size, sizes were averaged for recorded members of the same family. Species in families without any recorded genome sizes were not included in this analysis.

Trait analyses: phylogenetic generalized least squares

To control for the phylogenetic non-independence of our trait comparisons across bird species, we ran phylogenetic generalized least squares (PGLS) models. The phylogenetic trees with branch lengths generated from the UCE dataset were converted to a correlation structure in R using the *ape* package function corBrownian to estimate a Brownian motion (BM) model of trait evolution and corMartens to estimate an Ornstein-Uhlenbeck (OU) model (Paradis and Schliep 2019). The OU model may better replicate actual biological processes due to an additional parameter to the "random walk" of BM in that there is a greater attraction to an initial central value the further the trait is from this value. We then used the function gls in the R *nlme* package (Pinheiro and Bates 2022). This function fit a linear model to the traits of interest while considering either the BM or OU correlation structure as defined by one of the two phylogenetic trees. For each trait comparison, we compared the AIC values of each model to determine whether to select BM or the additional parameter in OU. These methods were repeated for both phylogenetic trees based on the two original topologies.

Phylogenetic analyses of olfactory receptor counts: ancestral state reconstructions

To estimate ancestral states across the bird phylogeny, we ran maximum likelihood estimates under a Brownian motion model using the function fastAnc in the R *phytools* package (Revell 2012). The character state input to these analyses was the log of the total intact OR count. We also obtained estimates of variance and 95% confidence intervals at each node. We used the *phytools* function contMap to set the ancestral state reconstructions on both of the phylogenetic trees, and used setMap and plot functions to generate the tree image (Revell 2012).

Phylogenetic analyses of OR counts: branch birth and death rates

To estimate rates of gene family birth and death across the bird phylogeny, we ran the program Badirate (Librado et al. 2012). Badirate uses either a gain and death or birth, death, and innovation stochastic population models in a phylogenetic context. Badirate has advantages over other gene family birth and death modeling tools such as being able to set separate birth and death rates, rather than the equal rates assumed by CAFE (Mendes et al. 2020). Badirate takes a phylogenetic tree and a gene family table as input. The gene family table (or "size file") can be divided into known subfamily groups, to reduce the amount birth and death rates mask each other. Here, the size file was divided into the total counts for the alpha, gamma, and gamma-c OR subfamilies for each bird species. A free rates branch model was selected, giving each branch its own birth and death rate. The birth and death estimation procedure used was a parsimony-based method. Here, birth, death, and innovation rates are determined from counting gain and loss events from the family members of internal nodes using the Wagner parsimony algorithm and two equations from Vieira et al. (2007). We ran a birth and death rates model along all tips and branch across both phylogenies. We recorded the birth and death rates at each branch with particular attention to branches with high birth and death rates.

Results

OR totals

Across all 70 bird species examined, we found a total of 8,880 ORs. This included 551 alpha (6.21% of total) and 2,427 gamma (27.33%) ORs. A total of 5,902 gamma-c ORs constituted nearly two-thirds (66.46%) of the total bird ORs found. Individual species OR repertoires ranged

from 7 in *Corvus corone* and 9 in *Corvus monedula* to 399 in *Laterallus jamaicensis* and 385 in *Gallus gallus*. Alpha OR counts in individual species ranged from 0-21, gamma counts range from 5-134 ORs, and gamma-c ranged from 0-351 ORs, revealing a wide range of individual OR subfamily counts across species. All ORs grouped into alpha, gamma, or gamma-c subfamilies. Although theta ORs were previously reported in *Gallus gallus* and *Taeniopygia guttata* (Steiger et al. 2009), we did not detect any ORs in the theta subfamily.

Ancestral state reconstruction

We generated ancestral state reconstructions of log-transformed total OR counts using maximum likelihood methods and the fastAnc function in phytools. We then visualized the ancestral state reconstructions across both topologies (Fig. 1a, b). Ancestral states were consistently highest in the deepest nodes of the tree, prior to the divergence of Galloanserae from Neoaves. Across the 69 nodes within the phylogeny, five nodes are not within the Neoaves clade. These five nodes in the top six highest ancestral character estimates in both topologies, with ancestral states in these branches ranging from 5.16–5.39 in the Prum topology (Jarvis topology is consistent with Prum topology unless stated otherwise). The only Neoavian branch within the top six highest ancestral OR counts within is the ancestor of the Rallidae. In the Jarvis topology this is the highest ancestral OR count (5.39, 95% CI 4.73–6.05), and in the Prum topology it is the second highest branch (5.37, 95% CI 4.72–6.03). OR counts were consistently the lowest in the Corvidae family, with all three nodes within Corvidae ranking lowest (ancestral state range within Corvide nodes from Prum topology 3.38–3.60). Other consistently low-ranking branches were in parrots (for example, node 117 in Prum 4.02, 95% CI 3.41–4.63) and in the node at the common

ancestor of all passerines (node 118 Prum 4.23, 95% CI 3.78–4.76).

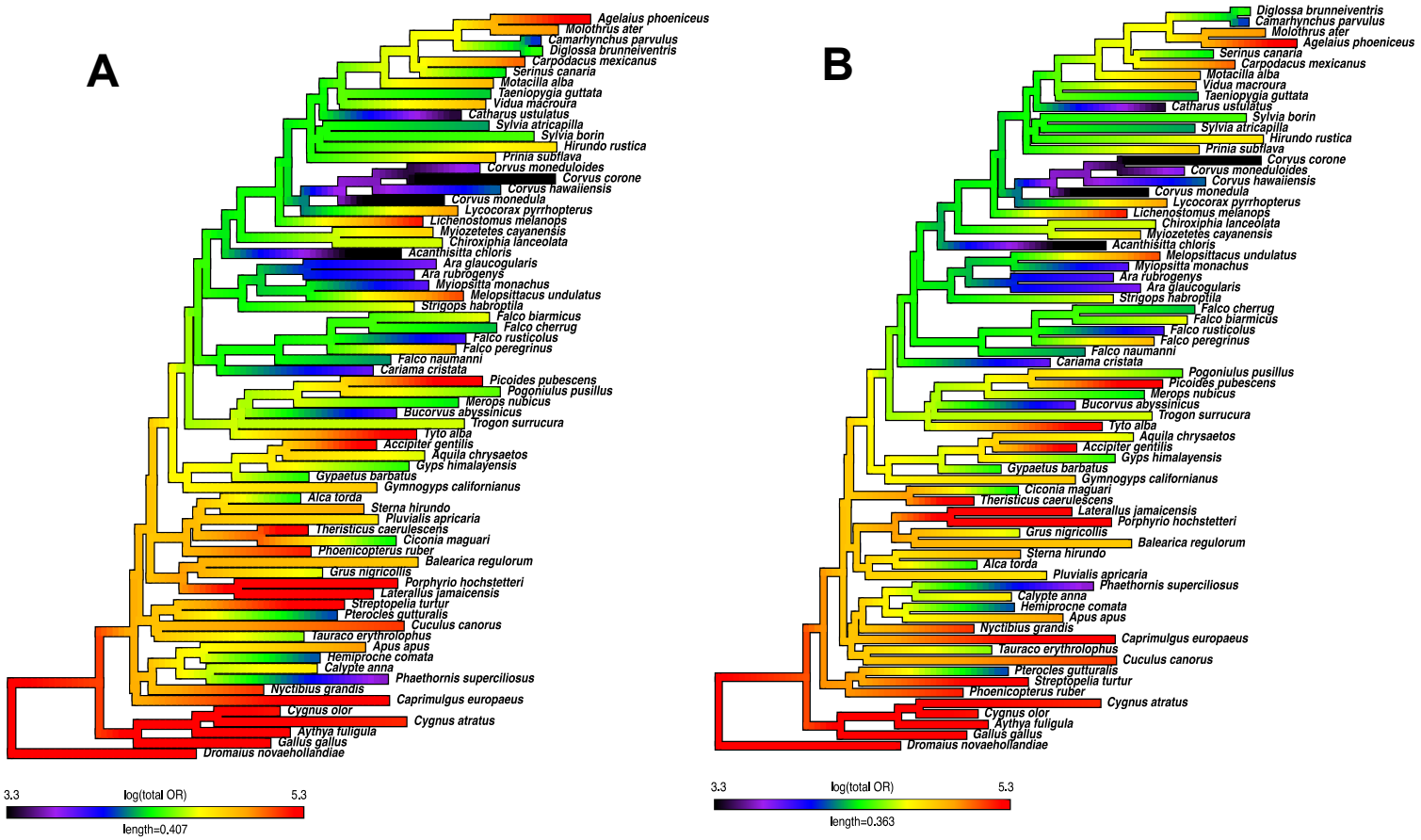

Fig 2.1. Ancestral character states of bird ORs in two topologies. (A) Ancestral character states of bird OR repertoires mapped onto the Prum et al. 2015 topology. (B) Ancestral character states of bird OR repertoires mapped onto the Jarvis et al. 2014 topology. We estimated ancestral character states using maximum likelihood methods an the phytools fastAnc function in R (Revell 2012). We mapped colors to the phylogeny using contMap.

Similar to overall OR counts, each of the three OR subfamilies showed a general pattern where some of the highest ancestral state reconstruction estimates occurred in the earliest diverging nodes of the tree, and show a decline following the divergence of Neoaves (Fig. 2a-c). In alpha ORs, the node with the highest ancestral character state is at the common ancestor of all modern birds (Prum, alpha = 2.61, 95% CI 1.95–3.26; Fig. 2a). Ancestral alpha OR counts also rebound in nodes leading to the carnivorous Accipitridae (for example node 97 Prum topology, 2.58, 95% CI 2.16–2.99). Alpha OR counts also increased at the Gruidae ancestral node (node 86 Prum topology, 2.41, 95% CI 1.98–2.85). Alpha ORs were lowest in Piciformes (node 103 Prum topology, 1.02, 95% CI 0.53–1.52). After the divergence of Psittaciformes and Passeriformes, alpha OR values decrease substantially, with the 11 nodes within this clade showing ancestral states below an average of 1.48.

The gamma OR subfamily also shows high ancestral values at the common ancestor of all modern birds (Prum, 4.00, 95% CI 3.42–4.59; Fig. 2b). Gamma ORs are high in different clades throughout the phylogeny, including Accipitridae (ie., Prum node 100, 4.00, 95% CI 3.64–4.37), and Psittaciformes (ie., Prum node 116, 3.91, 95% CI 3.50–4.33). Unlike alpha ORs, gamma ORs remain high in parrots, but decline in passerines, and do not recover. The twenty lowest ancestral state reconstructions for gamma ORs are the nodes within passerines (Prum topology, all below 3.02).

For gamma-c ORs, the three highest nodes are within Galloanserae (Prum nodes ie., 71-73 Prum node 72, 5.05, 95% CI 4.19–5.92; Prum Fig 2c), and is also high at the common ancestor of all birds (Prum 4.85, 95% CI 3.48–6.24). After a decrease in the Neoaves common ancestor, gamma-c counts increase in the ancestor of Gruidae (Prum 4.91, 95% CI 3.90–5.92).

Interestingly, despite an overall decrease in ORs in passerines, gamma-c ancestral states increased in one lineage of oscine passerines including Motacillidae, Fringillidae, Thraupidae, and Icteridae (ie., Prum node 135, 4.75, 95% CI 3.95–5.55). The smallest gamma-c values were the nodes within Psittaciformes (ie., Prum node 114, 2.33, 95% CI 1.37–3.29) and Corvidae (ie., Prum node 121, 2.61, 95% CI 1.75–3.48).

Olfactory receptor birth and death rates

To estimate the birth and death model of gene family evolution across our phylogeny, we ran Badirate (Librado et al. 2012). We input the three gene subfamilies in separate rows, allowing Badirate to estimate birth and death while simultaneously considering the three families independently. Across both topologies, the largest birth rate occurred on the branch leading to suboscines and oscines, following the divergence of Acanthisittidae (birth rate: Prum β = 58.29, Jarvis β = 57.71, but see Discussion). Following this branch, additional high birth rates occurred on various passerine lineages. Consistent with other results, a high birth rate occurred on the branch leading to Rallidae (Prum β = 17.88, Jarvis β = 19.48). Due to different birth and death rates among subfamilies, a small death rate was also found on the Rallidae ancestral branch (Prum δ = 0.10, Jarvis δ = 0.11). Other high birth rates occurred in Neoaves, including *Theristicus caerulescens* (Prum β = 20.86, Jarvis β = 21.69) and at the common ancestor of *Aquila chrysaetos* and *Accipiter gentilis* (Prum β = 11.98, Jarvis β = 13.85).

The highest gene death rates in both topologies occurred in the earliest diverging lineages of Neoaves (Fig. 3a,b). However, the relationships among modern Neoaves orders, occurring 60-70 million years ago, is highly debated, and is the main difference between the two topologies. Both

topologies showed an initial death rate in the ancestor of all Neoaves (Prum δ = 7.40, Jarvis δ = 14.10). However, this death rate was lower than subsequent death rates within different Neoaves lineages. In the Prum topology, two major OR declines occur on these branches, the first occurring following the first divergence within Neoaves, following the divergence of Strisores (Prum δ = 36.61). The Neoavian OR then undergo a subsequent second decline following the divergence of Gruidae, in the lineage leading to Aequorlitornithes, Accipitriformes, and all other Neoaves (Prum δ = 33.51). The Jarvis topology detects two losses as well, one following the divergence of Strisores (Jarvis δ = 60.47), and a second loss following the divergence of Cursorimorphae (Charadriiformes and Gruiformes) and leading to all other Neoaves (Jarvis δ = 29.36). While both topologies agree that Strisores diverged prior to a decline in OR receptor diversity, there is disagreement between the topologies on whether certain orders experienced any, some, or all of this OR loss. For example, Phoenicopteriformes (flamingos) diverge prior to either of these losses in the Jarvis topology, but diverge following both losses in the Prum topology.

Additional high OR death rates occurred within Coraciimorphae following the divergence of trogons (leading to barbets and woodpeckers, Prum δ = 20.21, Jarvis δ = 43.64; Fig. 3). Two passerine linages also experienced high death rates, Sylviidae (Prum δ = 30.46, Jarvis δ = 28.99), and *Camarhynchus parvulus* within Thraupidae (Prum δ = 19.48, Jarvis δ = 22.68). However, these lineages also experienced different rates of changes within subfamilies, as both experienced gene duplications as well (Prum Sylviidae 1.70, *Camarhynchus* β = 1.44; Jarvis Sylviidae β = 1.61, *Camarhynchus* β = 1.68).

In independent Badirate runs for the specific OR subfamilies, we detected the large death rates consistent in the subfamily-specific ancestral state reconstructions declines from fastAnc. A large reduction in gamma receptors occurred in the Australaves common ancestor (seriemas, falcons, parrots, passerines; δ = Prum 22.58, Jarvis δ = 19.76), and then again a substantial gamma decline occurred in the branch leading to all passerines (Prum δ = 17.57, Jarvis δ = 12.02). A large deline in alpha ORs occurred on a single branch leading to parrots and passerines (Prum δ = 41.56, Jarvis δ = 64.15), but subsequent increases occurred in specific lineages, such as Sylviidae (Prum β = 49.07, Jarvis β = 46.71). Gamma-c birth and death rates were similar to the three family analyses, given influence of the large gamma-c counts on this analysis.

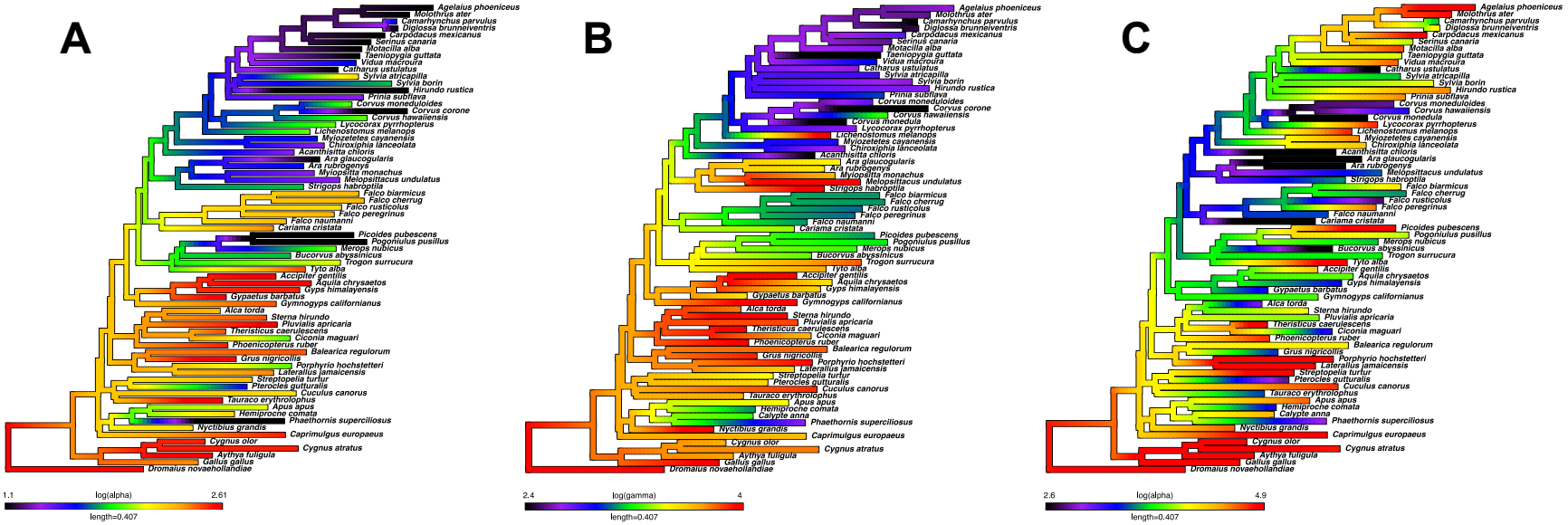

Fig 2.2. Ancestral state reconstruction of bird OR subfamily repertoires, generated by maximum likelihood using the fastAnc function in phytools in R (Revell 2012). Topology displayed is derived from Prum et al. 2015 topology. (A) Ancestral reconstruction of alpha OR subfamily, (B) gamma OR subfamily, and (C) gamma-c OR subfamily.

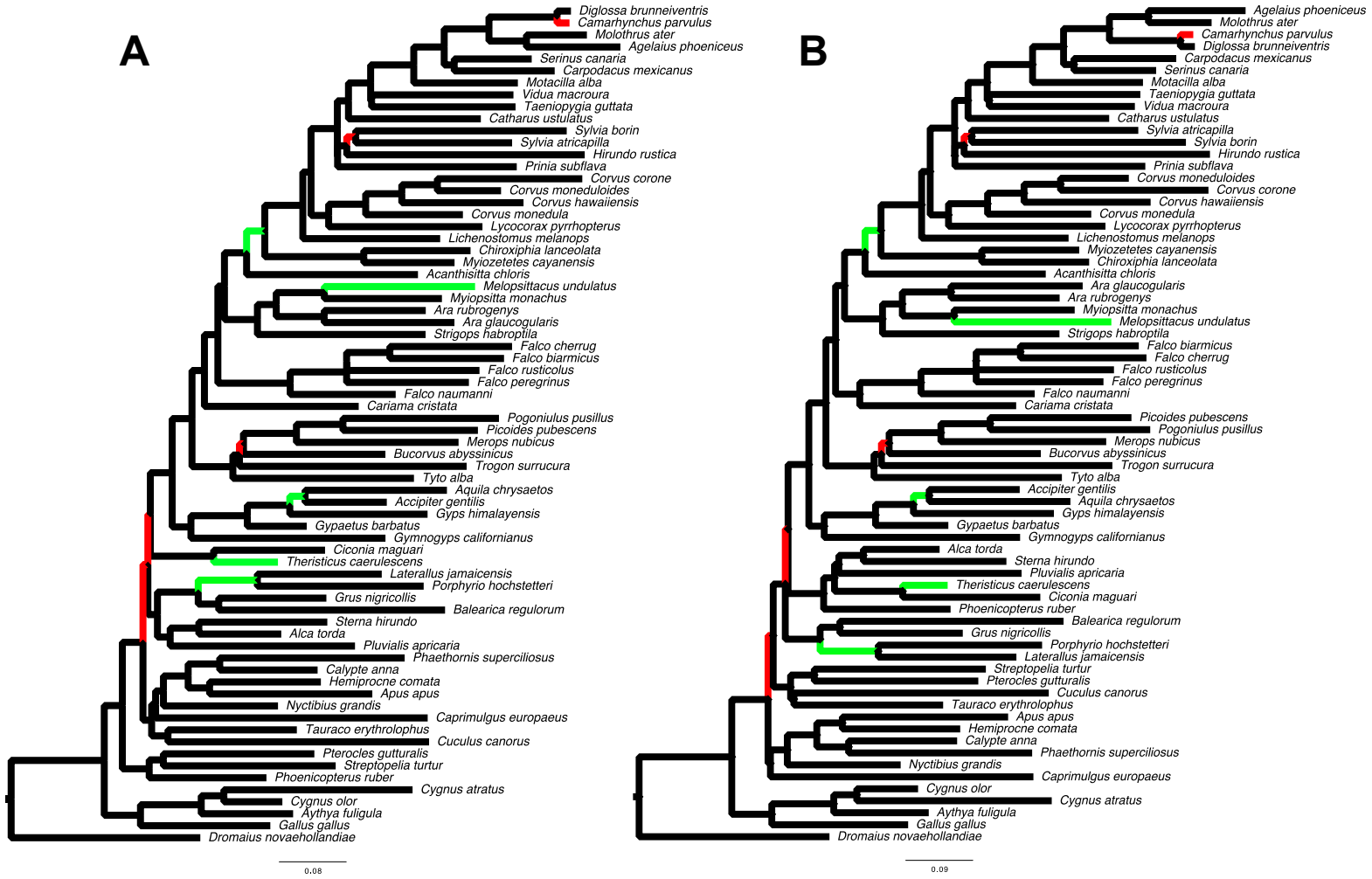

Fig 2.3. Highest OR birth and death rates across two topologies. (A) Prum et al. 2015 topology with the top five largest OR birth rate branches highlighted in green, and the top give largest OR death rate branches highlighted in red. (B) Jarvis et al. 2014 topology with the top five largest birth rate branches highlighted in green and the top five largest OR death rate branches highlighted in red.

Comparisons of OR counts and traits

Using phylogenetic least squares, we compared OR counts across all 70 species with behavioral and ecological phenotypes, including species diet, trophic level, environment type, song learning ability, and nocturnality. None of the eight measured diet types showed a correlation with OR counts. This lack of significant correlation included frugivore (t = 0.78, P = 0.44, BM model, Prum topology), granivore (t = -0.14, P = 0.89, OU model, Prum), aquatic herbivore (t = 0.21, P = 0.83, BM model, Prum), invertivore (t = 0.22, P = 0.82, OU model, Prum), nectarivore (t = 0.29, P = 0.29, OU model, Prum), omnivore (t = -0.57, P = 0.57, BM model, Prum), scavenger (t = -1.14, P = 0.26, BM model, Prum), and vertivore (t = 1.38, P = 0.17, BM model, Prum). Dividing the dataset into eight diet types may over partition the data and limit the number of independent gains of the trait across the phylogeny. Therefore, we also looked at trophic level, which more coarsely defines species as herbivores, carnivores, omnivores, and scavengers. Here too, however, we did not see any significant correlation with herbivory (t = -0.85, P = 0.40, BM model, Prum), carnivory (t = -1.08, P = 0.29, BM model, Prum), omnivory (t = -1.78, P = 0.08, BM model, Prum), or scavenging (t = -1.14, P = 0.26, BM model, Prum).

We also detected no significant correlation with OR total count when defining species as terrestrial or aquatic (t = 0.89, P = 0.38, BM model, Prum; t = 0.71, P = 0.48, BM model, Jarvis). To test for reliance on auditory cues, we saw no correlation between OR counts and song learning (t = -1.22, P = 0.22, BM model, Prum; t = -1.06, P = 0.30, BM model, Jarvis). Both topologies however showed a significant positive correlation between OR count and nocturnality (t = 2.83, P = 0.01, BM model, Prum; t = 3.00, P < 0.01, BM model, Jarvis; Prum Fig. 4).

In birds, olfactory bulb size is a long-standing measurement used to assess potential reliance on olfactory ability (Cobb 1959, Bang and Cobb 1968, Zelenitsky et al. 2011). Research has also shown a positive relationship between OR repertoire size and olfactory bulb size in birds (Steiger et al. 2008, Steiger et al. 2009, Khan et al. 2015). Using measurements from 24 species in Corfield et al. 2015, we found a significant positive correlation between the ratio of olfactory bulb size to telencephalon size and OR counts in both topologies (t = 2.19, P = 0.04, BM model, Prum, Fig. 5a). We saw the same significant correlation when measuring the ratio of olfactory bulb size to overall brain size and comparing to OR count (t = 2.16, P = 0.04, BM model, Prum, Fig. 5b).

We also compared counts of the three OR subfamilies, alpha, gamma, and gamma-c, with the set of traits. The majority of traits compared did not show a significant correlation with OR subfamily counts, however, several traits did show correlations with specific subfamilies. Alpha OR counts were negatively correlated with nectarivory, with low counts in all three nectivorous species, across two separate gains (Trochilidae, Thraupidae) (t = 2.59, P = 0.01, OU model, Prum). Alpha OR counts were also negatively correlated with song learning, with low alpha OR counts in passerines, parrots, and hummingbirds (t = -3.17, P < 0.01, BM model, Prum). Alpha OR counts increased in granivorous species (t = 2.47, P = 0.02, OU model, Prum). Gamma-c OR counts were also positively correlated with omnivorous species (t = -2.18, P = 0.03, BM model, Prum).

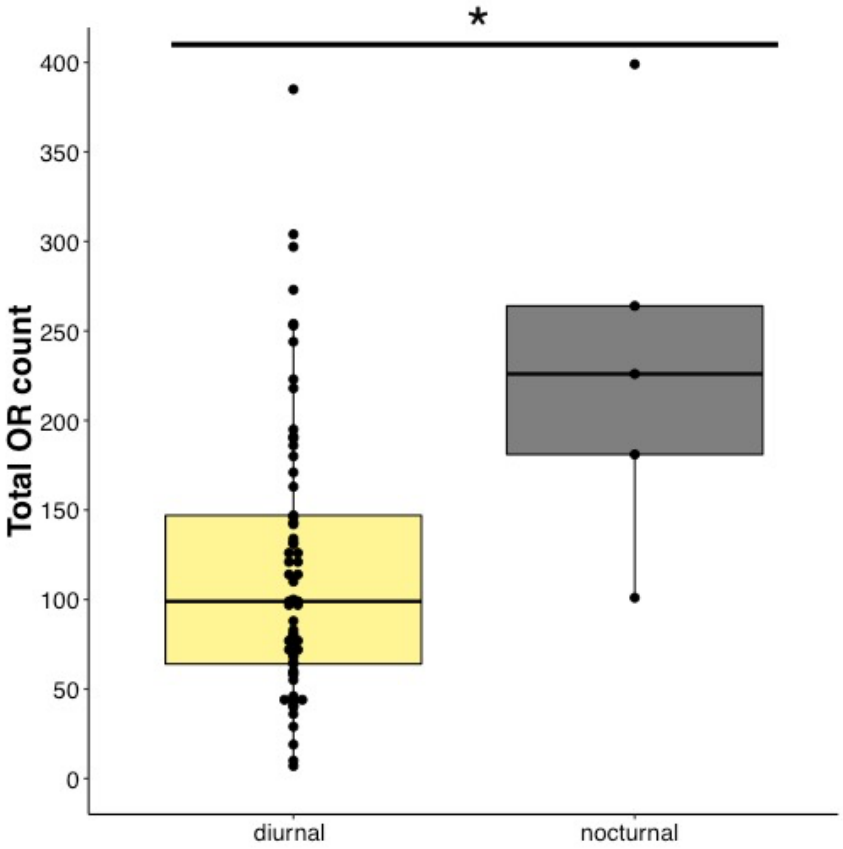

Fig. 2.4. Comparison of OR counts between diurnal and nocturnal bird species. Using phylogenetic generalized least squares methods, we detected a significant increase in nocturnal species (t = 2.83, P = 0.01, BM model, Prum topology).

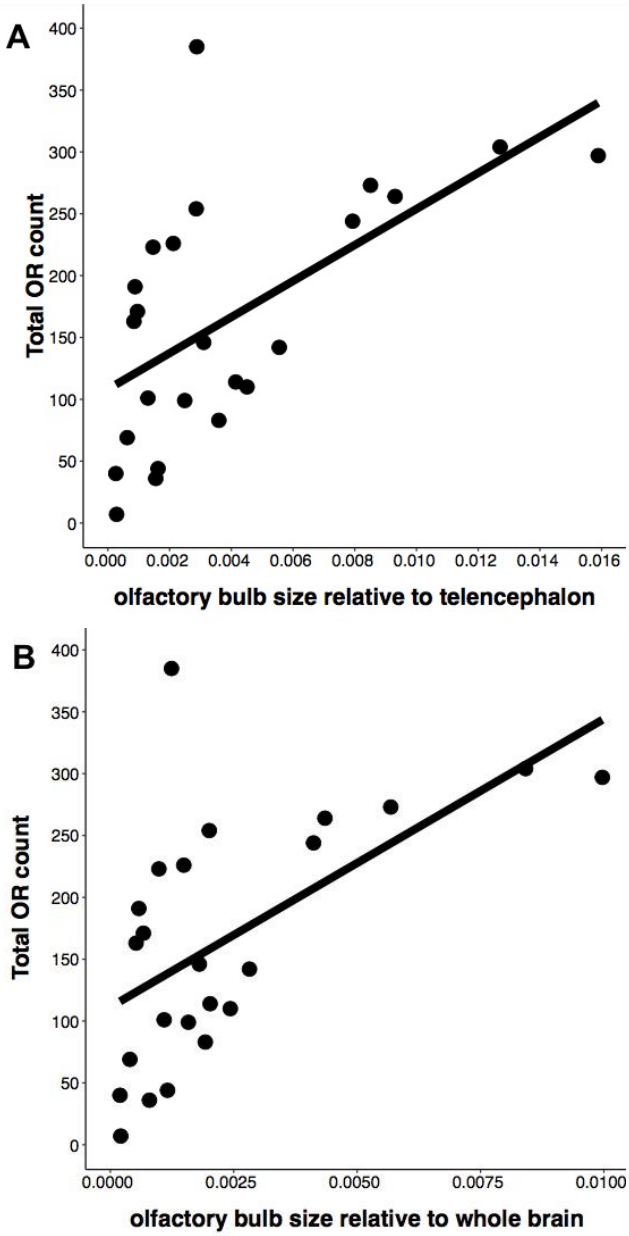

Fig. 2.5. Significant correlations between OR counts and olfactory bulb size to telencephalon size ratio and whole brain ratio. (A) Comparison of total OR counts and olfactory bulb size relative to telencephalon (Prum topology, $t = 2.19$, $P = 0.04$, BM model). (B) Comparison of total OR counts and olfactory bulb size relative to whole brain (Prum topology, $t = 2.16$, $P = 0.04$, BM model). Olfactory bulb, telencephalon, and whole brain measurements from Corfield et al. 2015.

Discussion

Olfactory capabilities are potentially widespread in birds

With the use of long read assemblies, we found that 47 of the 70 species analyzed had a repertoire size of at least 75 ORs, and these 47 species were present across diverse lineages of birds. This is in contrast to the previous study investigating olfactory receptor counts across the bird phylogeny, which found repertoire sizes >75 ORs in only three of 48 species (Khan et al. 2015). We therefore show the robust use of long read genomes for characterizing bird OR counts, as well as the potential importance of smell for birds across the phylogeny. These counts were largely supported by the bird-specific gamma-c OR subfamily, which had an average of 84 ORs per species, or 66.46% of the total ORs recovered. Due to the lack of functional studies of bird ORs, we know little about the gamma-c and their role in smell. While orthologs with characterized binding odors exist in mammals and reptiles for bird alpha and gamma ORs (Saito et al. 2009, Steiger et al. 2009, Vandewege et al. 2016), the gamma-c ORs do not have comparable orthologs in other vertebrate classes. In the first investigation of bird olfactory epithelium RNA expression, gamma-c OR expression was detected in the olfactory epithelium, suggesting a role in olfaction (Sin et al. 2022). The role of gamma-c in olfaction would suggest that gamma-c, and overall bird OR counts, are related to a species' behavioral or ecological reliance on smell, as is shown in other vertebrates, such as mammals (Niimura et al. 2014).

The highest OR count was in *allus jamaicensis* (black rail) at 399 ORs. *Laterallus jamaicensis* lives in dense marsh habitat, and is also nocturnal, and is one of the most challenging birds in North America to observe (Billerman 2022). The third highest OR count *Porphyrio hochstetteri* (takahe) was also in the Rallidae family. These two species show a

remarkable and consistent high repertoire size within Rallidae, which exceeds all other Neoaves species by at least 40 ORs. The second, fourth, and fifth highest OR counts (*Gallus gallus* 385, *Dromaius novaehollandiae* 297, *Aythya fuligula* 273 ORs) include one Palaeognathae and two Galloanserae species, illustrating the phylogenetic retention of high OR counts in these groups and the possible importance of smell in these species through evolutionary history to present day. Conversely, all four *Corvus* species in the dataset had relatively low counts, and in particular, the lowest counts among all birds were *Corvus monedula* (jackdaw, 10 ORs) and *Corvus corone* (carrion crow, 7 ORs). This could mean a decreased reliance on smell for these species, and perhaps a tradeoff with other senses, such as increased reliance on vision, or other energy investments in the brain, for example, increased cognitive performance (Cobb 1959).

OR counts declined in early diverging Neoaves

The most diverse ancestral nodes across the bird phylogeny were in Galloanserae, particularly Anatidae, in the ancestor of all modern birds, and in common ancestor of all Neognathae. Following the divergence of Neoaves, total OR counts decline, although the two major bird topologies (Jarvis et al. 2014, Prum et al. 2015) disagree on the placement of these declines. Both topologies agree that prior to the divergence of Strisores, ancestral OR counts remain high. Following the Neoaves radiation, OR total counts do not recover to their previous states, with one exception, in the Rallidae. In the Jarvis topology, the ancestral state of the common ancestor between *Laterallus jamaicensis* and *Porphyrio hochstetteri* exceeds ancestral states prior to the divergence of Neaoves. The ancestor of these two rail species possibly adapted to dense marsh habitat with limited visibility, and where potential prey items are beneath substrate, ecological

conditions that could promote olfactory abilities. This pattern across the phylogeny shows that in the birth and death model of gene family evolution, genes can decline substantially, but recover under specific circumstances, likely driven by ecological selection. Consistent with the OR counts in extant species, the lowest ancestral state OR counts occurred in crows, with the three ancestral nodes in this clade having the lowest states of across all birds. Due to our species sampling, it is unclear whether these low ancestral character states are unique to the genus *Corvus*, or if these low counts extend to other Corvids, such as jays or magpies. Mining additional Corvidae species for OR counts could help figure out where this decline took place on the phylogeny. Behavioral experiments indicate that magpies (*Pica hudsonia*) can more easily find cached food items scented with cod liver than unscented food, suggesting that perhaps the very low OR counts occurred within Corvidae, perhaps in an ancestral *Corvus* species (Buitron and Nuechterlein 1985).

Ancestral states of OR subfamily counts show that the history of each subfamily is unique, and that the composition of the total OR counts in birds has changed over evolutionary time. Three clades in particular show low alpha counts, the Trochilidae (hummingbird) clade, the hornbill, bee-eater, woodpecker clade within Coraciimorphae, and the Psittaciformes and Passeriformes clade. These two clades show consistently low levels of alpha ORs, despite including species with diverse diets and habitats. A similar result is present in the Passeriformes clade for gamma ORs, and in the Psittaciformes clade for gamma-c ORs. Although we do not know the specific reason why these clades saw declines in these OR subfamilies, is possible that the ORs no longer detect relevant odors in these clades, while other OR subfamilies retain relevance. Like crows, these clades, particularly woodpeckers, parrots, and passerines, are

considered to have highly developed cognitive abilities, a potential tradeoff with olfactory abilities (Cobb 1959).

Dynamic bird and death of ORs across the bird tree

Our analysis through Badirate detected non-zero birth and death rates for many branches across the tree, showing a dynamic birth and death model of gene family evolution for ORs across birds. For total OR counts, the largest expansions occurred on the branch separating the passerine Acanthisitti from suboscines and oscines. This result is partially due to a very low OR count in *Acanthisitta chloris*, which has the lowest OR count among all species (19 ORs), aside from crows. This could potentially be an issue with obtaining the original *Acanthisitta chloris* DNA sample, as this species is restricted to New Zealand and may be difficult to access. The contig N50 and total number of contigs for the assembly were consistent with other assemblies used, and the assembly was created with the standard Vertebrate Genomes Project pipeline. If this large birth rate is indeed accurate, then following the divergence of Acanthisittidae (New Zealand wrens), the ancestor of oscine and suboscine passerines experienced a birth rate nearly three times higher than at any other point in the bird phylogeny. This birth rate substantially impacted the gamma-c, as alpha and gamma ORs remain low across all passerines.

The ancestor of rails also had a high birth rate, and was one of the few branches on the tree that had a high birth rate and one of the highest ancestral state reconstructions. This suggests that an ancestral rail had one of the largest OR repertoires in birds, and that many of the ORs in this repertoire arose recently, following the divergence from Gruidae (cranes). The gene birth rate along this branch was also high in the gamma-c OR subfamily (β = 43.15). This paints the

possibility of an ancestral Gruiformes bird perhaps entering marsh habitat, experiencing gamma-c OR duplications, and retaining those genes to aid in olfaction. This is in contrast to other bird species with high OR counts and high ancestral state reconstructions, such as the chicken. The branch leading to the chicken does have a small birth rate ($\beta = 1.41$), but the chicken's large OR repertoire size is largely due to the maintained ancestral state throughout Galloanserae evolution.

The highest death rates in birds occurred along the early diverging branches in Neoaves. In the Prum topology, a high OR death rate occurs following the divergence of Strisores, then Columbaves (cuckoos and turacos and doves and sandgrouse) diverge, and then a second high death rate occurs in the rest of Neoaves. In the Jarvis topology, the loss is positioned following the divergence of Phoenicopteriformes (flamingos), Columbiformes (doves), Pterocliformes (sandgrouse), and Strisores. Following this death, the lineage leading to Charadriiformes and Gruiformes (including rails) diverges, and the branch leading to all other Neoaves experiences high rate of gene death. Between 60-70 million years ago, the Neoaves underwent a massive radiation, splitting into all of today's modern orders (Jarvis et al. 2014, Prum et al. 2015). This rapid radiation is difficult for phylogeneticists to resolve, and is unclear from even a variety of approaches, including using both coding and non-coding DNA regions to construct trees (Suh 2016). The time frame of this radiation includes the Cretaceous-Paleogene extinction, and following the extinction of non-avian dinosaurs, birds likely began to occupy into newly available niches. A previous study measuring the olfactory bulb size from fossilized Cretaceous bird species showed that bulb size increased in early Neornithine and Palaeognathae evolution, and perhaps olfaction aided these species during the Cretaceous-Paleogene extinction event (Zelenitsky et al. 2011). However, the authors detect only one olfactory bulb increase in early diverging Neoaves branches, in the branch leading to Gruiformes, Procellariformes, and other

mostly aquatic lineages, however, this comparison used a topology that we did not consider in the current study (Zelenitsky et al 2011). Many other early diverging Neoaves lineages experienced a decrease in olfactory bulb size (Zelenitsky et al. 2011). Therefore, although Palaeognathae and Galloansarae retained s large olfactory bulb that originated in the ancestor of all modern birds, this comparatively large olfactory bulb decreased in Neoaves. In agreement with our OR counts, we show that although olfaction may have aided Palaeognathae and Galloanserae through the Cretaceous extinction, we do not support the idea that during this same time smell played a major role in the Neoaves radiation, but rather, that reliance on smell decreased in most Neoaves lineages.

In OR subfamily birth and death rate analyses, Badirate similarly detected the decrease in alpha OR counts, occurring in the ancestor of parrots and passerines, and in gamma ORs, occurring in two ancestral branches, including the ancestor of passerines. Despite the low alpha and gamma counts in passerines, as mentioned earlier, the common passerine ancestor (excluding Acanthisittidae) experiences a radiation of gamma-c. This high gamma-c birth rate is furthered by marginal lineage-specific gamma-c gains, including the Icteridae ancestor (β = 2.84), and again in the Icterid *Agelaius phoeniceus* (β = 4.39). It is possible that over evolutionary time species shift reliance on different OR subfamilies to accommodate different ecologies.

Olfactory bulb size correlates with OR repertoire counts

Consistent with previous work, olfactory bulb to brain size ratios positively correlated with total OR repertoire counts. This further supports that both measurements can reliably be used as a

proxy for olfactory ability. One outlier species in comparison of bulb to brain size ratio with OR count was the chicken, which had a much larger OR count relative to its olfactory bulb size. While it is uncertain as to why chicken is such an outlier compared to other species (Fig. 5a,b), the DNA reference source for the chicken assembly used here for OR counts was from a domesticated bird. Previous studies show domesticated mammals, including rats, llamas, sheep, pigs, and dogs, have a lower volume of olfactory structures relative to wild "ancestral" species (Kruska 1980, Kruska 1988). Our measurements of olfactory bulb are slightly different, and consider the relative volume of the olfactory bulb to the telencephalon or overall brain (Corfield et al. 2015), however, decreased olfactory bulb volume could lead to the outlier position of the chicken that we observed. Although other species in both the OR count dataset and olfactory bulb size come from domesticated birds (for example, *Taeniopygia guttata*), people in the Indus Valley were estimated to domesticate the chicken about 4,500 years ago, far longer than any other bird species (Tixier-Boichard et al. 2011). Although there is no evidence of how OR repertoire size is impacted by domestication, it is possible OR repertoire does not change at the same rate as olfactory bulb size in response to domestication. In this case, the chicken may have a reduced olfactory bulb, while retaining much of its ancestral OR repertoire. Additional studies on different chicken breeds, as well as wild red junglefowl (*Gallus gallus*), could help reveal the impact of domestication on OR counts. It is also unclear if Corfield et al. obtained a wild red junglefowl in their morphological analysis or a domesticated chicken (Corfield et al. 2015).

Of the 24 species examined for olfactory bulb size, the two smallest olfactory bulbs were in the genus *Corvus*- *C. moneduloides* and *C. corone*, which matched perfectly with our extremely low counts of ORs in *Corvus*. The olfactory bulb of *Corvus macorhynchos* is very small relative to the cerebral hemisphere and in one study did not have distinct posterior conchae

present (Yokosuba et al. 2009, Kondoh et al. 2011). Although it can be challenging to define "intelligence" across many different bird species, it has been suggested that "intelligent" birds have smaller olfactory bulbs (Cobb 1959).

Diet and song learning are not correlated with OR counts

Across all observed diets and trophic niches, there were no correlations with total OR counts. This was true for a comparison that assigned species to one of eight potential diet niches and a comparison that broke species into four trophic levels. The lack of a relationship was surprising, because presumably different diet types attract specific foraging methods that vary in their reliance on olfaction. This negative result could be due to potentially diverse ways to arrive at a given diet. For example, the diet category 'invertivore' encompasses a variety of different foraging strategies. *Apus apus* (swift) is a diurnal aerial hunter, while *Cuculus canorus* (cuckoo) gleans arboreal insects, and *Picoides pubescens* (woodpecker) excavates insects from tree bark (Billerman et al. 2022). However, these diverse foraging behaviors are all considered 'invertivores' in EltonTraits (Wilman et al. 2014). Therefore, OR totals may better correlate with particular foraging strategies as opposed to diet. More species should be surveyed for OR counts to increase the number of species representing each foraging strategy.

Another possibility for the lack correlation between diet and OR counts is that dietary changes do not greatly impact the total number of ORs. It is possible that shifts in olfactory ability could occur due to change in sensitivity of existing ORs. Alternatively, only a small number of OR gains and losses could potentially confer great changes to olfactory abilities, but not be reflected when looking at the comparatively large number of ORs in the total repertoire.

For example, the three subfamilies could permit the detection of different types of odors, and so a change in diet would only impact a given group of ORs or subfamily. We show that alpha ORs significantly decreased in both nectivorous lineages (in Trochilidae and Thraupidae). Our PGLS analysis did not include zero values, and the hummingbird *Calypte anna* had zero alpha ORs. Therefore, our result, which only considers low counts in *Phaethornis* and *Diglossa* is further supported by *Calypte* counts. We also saw an increase in alpha ORs for granivorous species, and an increase in gamma-c ORs in omnivorous species, further suggesting that dietary shifts may fine tune subfamily repertoires, as opposed to drastically altering total counts.

Similar to diet, song learning did not correlate with overall OR counts, but did correlate with a decrease in alpha OR counts. This was due to exceptionally low alpha ORs in, hummingbirds, parrots, and oscine passerines. Woodpeckers also had very low alpha OR counts and are not song learners by standard measures, but forebrain nuclei used in territorial drum displays are the same as used in songbird vocal learning (Schuppe et al. 2022). Therefore, song learning may have a relationship with decreases in alpha ORs even moreso than detected in our traditional trait analysis.

Nocturnality increases total OR counts

Nocturnality was positively associated with higher OR counts. Our results agree with a previous morphological comparison that shows that nocturnality increases olfactory bulb size in birds, whereas other ecological variables, including diet, do not show a correlation (Healy and Guilford 1990). Across the phylogeny, there were five nocturnal species and four presumed gains of nocturnality– one in Strisores (*Camprimulgus europaeus* and *Nyctibius grandis*), one in Rallidae

(*Laterallus jamaicensis*), one in Strigiformes (*Tyto alba*), and one in Psittaciformes (*Strigops habroptila*). Increased OR repertoire in nocturnal species is significant despite the diverse behavior and ecology of the nocturnal species included. The Strisores species are aerial insectivores, *L. jamaicensis* is a skulking invertivore in dense marsh habitat, *T. alba* is a primarily mammalian predator, and *S. habroptila* is a giant herbivorous flightless parrot (Billerman et al. 2022). Despite disparate underlying ecology, a nocturnal lifestyle is a strong transition that greatly impacts the sensory biology of the organism, for example, owls lack a functional UV-sensing shortwave sensitive 1 opsin but have greater hearing abilities (Grothe 2018, Höglund et al. 2019). Therefore, although different diets may gives rise to a variety of foraging methods that may influence a species sensory biology in various ways, nocturnality has a consistent signal in birds, where olfactory receptors significantly increase in number.

No evidence for influence of genome size on OR count

While high OR counts may reflect a true reliance on olfaction, we wanted to know if the propensity of a genome to experience duplications, as measured by total genome size, was also responsible for OR count. The location of many ORs in the genome can be found in large clusters on unmapped contigs, flanked by repeat regions of DNA, and transposable elements (Glusman et al. 2000, Vandewege et al. 2016, Driver & Balakrishnan 2021). In humans, large OR clusters are interspersed with repetitive elements, particular LINEs (Glusman et al. 2000). LINES are a common source of reverse transcriptase and can retrotranspose intron-less paralogs into genomic DNA (Kidwell 2002). Additionally, transposable elements or DNA replication slippage could increase DNA content, and carry local ORs along in the duplication. However, we

did not see any relationship between OR counts and overall genome size. Although a weak correlation did appear, following phylogenetic correction we did not see a significant result. This was somewhat surprising, since we noticed that hummingbirds, particularly *Phaethornis superciliosus*, have low OR counts and hummingbirds have the smallest genome sizes of any bird family (Gregory et al. 2009). However, the correlation between OR counts and genome size is not significant when across the 70 species presented here.

Conclusion

We show a high level of dynamism in OR repertoire counts across the bird phylogeny. We show that some birds have large OR repertoires, such as in rails, where the OR total count of *Laterallus jamaicensis*, at 399, is roughly the same repertoire size at the lower end of mammals, including primates (Niimura et al. 2014). We also show that some birds have very small OR repertoires, such as crows in the genus *Corvus*, that, consistent with evidence from the morphological features of the crow olfactory system, likely have a poor sense of smell (Kondoh et al. 2011). In between these high and low OR repertoire extremes are ever-changing OR ancestral character states and branches experiencing OR gene family birth and death rates. Included among these branches is a high rate of death during the early diverging lineages of Neoaves, about 60-70 million years ago. Through evolutionary time, OR expansions and contractions of various degrees appear frequently in the phylogeny, showing the high turnover of a gene family undergoing the birth and death model of evolution. We show that nocturnality is an ecological factor that increases OR counts during evolution. We also find that increased OR

counts are associated with a larger olfactory bulb, further suggesting that OR counts can be used as a proxy for reliance of a species on smell.

Although we have characterized bird OR genomic repertoires in this study, not all OR genes will be functional or relevant to the olfactory system (Maßberg and Hatt 2018). In mammals, many ORs are expressed in tissues outside of the olfactory system, including roles in environmental responses in the skin and chemotaxis in sperm (Maßberg and Hatt 2018). Therefore, future gene expression studies of the bird olfactory epithelium can pinpoint which ORs within the genomic repertoire may be involved in olfaction. Finally, even for bird ORs expressed in the olfactory epithelium, it is unclear what odors cause a response in bird ORs. This is particularly true for the gamma-c ORs, which have no clear orthology to other vertebrate classes. For gamma-c ORs, binding properties are entirely unknown and cannot be compared with mammalian orthologs that may have known response odors. The subfamily-specific birds and deaths across the phylogeny are often difficult to explain using only bird ecology and behavior. Functional work in the future will allows us to better make sense of births and deaths across the tree, for example, why hummingbirds, parrots, and passerines have few alpha ORs. Our study provides the genomic data to further investigate the individual ORs within our counts, to better understand how birds use smell in their ecology and behavior.

References

Audubon J.J. (1826). Account of the habits of the turkey buzzard, *Vultur aura*, particularly with the view of exploding the opinion generally entertained of its extraordinary power of smelling. *Edinburgh New Philosophical Journal* 2:172–184.

Bang B. G., Cobb S. (1968). The size of the olfactory bulb in 108 species of birds. *The Auk* 85:55–61.

Beichman A. C., Koepfli K. P., Li G., Murphy W., Dobrynin P., Kliver S., Tinker M. T., Murray M. J., Johnson J., Lindblad-Toh K., Karlsson E. K., Lohmueller K. E., Wayne R. K. (2019). Aquatic adaptation and depleted diversity: a deep dive into the genomes of the sea otter and giant otter. *Molecular Biology and Evolution* 36:2631–2655.

Billerman S. M., Keeney B. K., Rodewald P. G., Schulenberg T. S. (2022). Birds of the World. Cornell Laboratory of Ornithology, Ithaca, NY USA. https://birdsoftheworld.org/bow/home

Bonadonna F., Gagliardo A. (2021). Not only pigeons: avian olfactory navigation studied by satellite telemetry. *Ethology Ecology, & Evolution* 33:273–289.

Bravo G. A., Schmitt C. J., Edwards S. V. (2021). What have we learned from the first 500 avian genomes? *Annual Review of Ecology, Evolution, and Systematics* 52:611–639.

Buck L., Axel R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175–187.

Buitron D., Nuechterlein G. L. (1985). Experiments on olfactory detection of food caches by black-billed magpies. *Condor* 87:92–95.

Cobb S. (1959). A note on the size of the avian olfactory bulb. *Epilepsia* 1:394–402.

Corfield J. R., Price K., Iwaniuk A. N., Gutierrez-Ibañez C., Birkhead T., Wylie D. R. (2015). Diversity in olfactory bulb size in birds reflects allometry, ecology, and phylogeny. *Frontiers in Neuroanatomy* 9:102.

Driver R. J., Balakrishnan C. N. (2021). Highly contiguous genomes improve the understanding of avian olfactory receptor repertoires. *Integrative & Comparative Biology* 61:1281–1290.

Faircloth B. C., McCormack J. E., Crawford N. G., Harvey M. G., Brumfield R. T., Glenn T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717–726.

Glusman G., Bahar A., Sharon D., Pilpel Y., White J., Lancet D. (2000). The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mammalian Genome* 11:1016–1023.

Goujon M., McWilliam H., Li W., Valentin F., Squizzato S., Paern J., Lopez R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* 38:W695–W699.

Gregory T. R., Andrews C. B., McGuire J. A., Witt C. C. (2009). The smallest avian genomes are found in hummingbirds. *Proceedings of the Royal Society B* 276:3753–3757.

Gregory T.R. (2022). Animal Genome Size Database. http://www.genomesize.com

Grothe B. (2018). How the barn owl computes auditory space. *Trends in Neurosciences* 41:115–117.

Grubb, T. C. (1972). Smell and foraging in shearwaters and petrels. *Nature* 237:404–405.

Haring E., Däubi B., Pinsker W., Kryukov A., Gamauf A. (2012). Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae) – a first survey based on museum specimens. *Journal of Zoological Systematics and Evolutionary Research* 50:230–246.

Healy S., Guilford T. (1990). Olfactory-bulb size and nocturnality in birds. *Evolution* 44:339–346.

Hill A. (1905). Can birds smell? *Nature* 71:318–319.

Höglund J., Mitkus M., Olsson P., Lind O., Drews A., Bloch N. I., Kelber A., Strandh M. (2019). Owls lack UV-sensitive cone opsin and red oil droplets, but see UV light at night: retinal transcriptomes and ocular media transmittance. *Vision Research* 158:109–119.

Jarvis E. D., Mirarab S., Aberer A. J., Li B., Houde P., Li C., Ho S. Y., Faircloth B. C., Nabholz B., Howard J. T., Suh A., Weber C. C., da Fonseca R. R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M. S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W. C., Ray D., Green R. E., Bruford M. W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E. P., Bertelsen M. F., Sheldon F. H., Brumfield R. T., Mello C. V., Lovell P .V., Wirthlin M., Schneider M. P., Prosdocimi F., Samaniego J. A., Vargas Velazquez A. M., Alfaro-Núñez A., Campos P. F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D. M., Zhou Q., Perelman P., Driskell A. C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F. E., Braun M., Fjeldsa J., Orlando L., Barker F. K., Jønsson K. A., Johnson W., Koepfli K. P., O'Brien S., Haussler D., Ryder O. A., Rahbek C., Willerslev E., Graves G. R., Glenn T. C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S. V., Stamatakis A., Mindell D. P., Cracraft J., Braun E. L., Warnow T., Jun W., Gilbert M. T., Zhang G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.

Joseph L., Toon A., Schirtzinger E. E., Wright T. F., Schodde R. (2012). A revised nomenclature and classification for family-group taxa of parrots (Psittaciformes). *Zootaxa* 3205:26–40.

Katoh K., Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.

Khan I., Yang Z., Maldonado E., Li C., Zhang G., Gilbert M. T. P., Jarvis E. D., O'Brien S. J., Johnson W. E., Antunes A. (2015). Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Molecular Biology and Evolution* 32:2832–2843.

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63.

Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111–120.

Kondoh D., Nashimoto M., Kanayama S., Nakamuta N., Taniguchi K. (2011). Ultrastructural and histochemical properties of the olfactory system in the Japanese jungle crow, *Corvus macrorhynchos*. *Journal of Veterinary Medical Science* 8:1007.

Kruska D. (1988). Effects of domestication on brain structure and behavior in mammals. *Human Evolution* 3:473–485.

Kück P., Meusemann K. (2010). FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution* 56:1115–1118.

Laurin M., Reisz R. R. (1995). A reevaluation of early amniote phylogeny. *Zoological Journal of the Linnean Society* 113:165–223.

Letunic I., Bork P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:W256–W259.

Librado P., Vieira F. G., Rozas J. (2012). Badirate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28:279–281.

Maßberg D., Hatt H. 2018. Human olfactory receptors: novel cellular functions outside of the nose. *Physiological Reviews* 98:1739–1763.

McRae J. F., Mainland J. D., Jaeger S. R., Adipietro K. A., Matsunami H., Newcomb R. D. (2012). Genetic variation in the odorant receptor OR2J3 is associated with the ability to detect the "grassy" smelling odor, cis-3-hexen-1-ol. *Chemical Senses* 37:585–593.

Mendes F. K., Vanderpool D., Fulton B., Hahn M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36:5516–5518.

Mindell D. P., Fuchs J., Johnson J. A. (2018). Phylogeny, taxonomy, and geographic diversity of diurnal raptors: Falconiformes, Accipitriformes, and Cathartiformes. In: Sarasola J., Grande J., Negro J. (eds) Birds of Prey. Springer, Cham. https://doi.org/10.1007/978-3-319-73745-4_1

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., von Haeseler A., Lanfear R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534.

Molina-Morales M., Castro J., Albaladejo G., Parejo D. (2020). Precise cache detection by olfaction in a scatter-hoarder bird. *Animal Behaviour* 167:185–191.

Montague M. J., Li G., Gandolfi B., Khan R., Aken B. L., Searle S. M. J., Minx P., Hillier L. W., Koboldt D. C., Davis B. W., Driscoll C. A., Barr C. S., Blackistone K., Quilez J., Lorente-Galdos B., Bonet-Marques T., Alkan C., Thomas G. W. C., Hahn M. W., Menotti-Raymond M., O'Brien S. J., Wilson R. K., Lyons L. A., Murphy W. J., Warren W. C. (2014). Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proceedings of the National Academy of Sciences* 111:17230–17235.

Niimura Y., Nei M. (2005). Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences* 102:6039–6044.

Niimura Y., Nei M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2:e708.

Niimura Y. (2009). On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biology and Evolution* 1:34–44.

Niimura Y. (2013). Identification of olfactory receptor genes from mammalian genome sequences. *Methods in Molecular Biology* 1003:39–49.

Niimura Y., Matsui A., Touhara K. (2014). Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Research* 24:1485–1496.

Oliveros C. H., Field D. J., Ksepka D. T., Barker F. K., Aleixo A., Andersen, M. J., Alström P., Benz B. W., Braun E. L., Braun M. J., Bravo G. A., Brumfield R. T., Chesser R. T., Claramunt S., Cracraft J., Cuervo A. M., Derryberry E. P., Glenn T. C., Harvey M. G., Hosner P. A., Joseph L., Kimball R. T., Mack A. L., Miskelly C. M., Peterson A. T., Robbins M. B., Sheldon F. H., Silveira L. F., Smith B. T., White N. D., Moyle R. G., Faircloth B. C. (2019). Earth history and the passerine superradiation. *Proceedings of the National Academy of Sciences* 116:7916–7925.

Owre O. T., Northington P. O. (1961). Indication of the sense of smell in the turkey vulture, *Cathartes aura* (Linnaeus), from feeding tests. *American Midland Naturalist* 66:200–205.

Paradis E., Schliep K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analysis in R. *Bioinformatics* 35:526–528.

Pinheiro J., Bates D., DebRoy S., Sarkar D., R Core Team. (2021). nlme: linear and nonlinear mixed effects models. R package version 3.1-144 https://CRAN.R-project.org/package=nlme.

Prum R. O., Berv J. S., Dornburg A., Field D. J., Townsend J. P., Moriarty Lemmon E., Lemmon A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:569–573.

Revell L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223.

Saito H., Chi Q., Zhuang H., Matsunami H., Mainland J. D. (2009). Odor coding by a mammalian receptor repertoire. *Science Signaling* 2:ra9.

Schuppe, E. R., Cantin L., Chakraborty M., Biegler M. T., Jarvis E. R., Chen C. C., Hara E., Bertelsen M. F., Witt C. C., Jarvis E. D., Fuxjager M. F. (2022). Forebrain nuclei linked to woodpecker territorial drum displays mirror those that enable vocal learning in songbirds. *PLoS Biology* 20:e3001751.

Silva M. C., Chibucos M., Munro J. B., Daugherty S., Coelho M. M., Silva J. C. (2020). Signature of adaptive evolution in olfactory receptor genes in Cory's shearwater supports molecular basis for smell in procellariiform seabirds. *Scientific Reports* 10:543.

Sin S. Y. W., Cloutier A., Nevitt G., Edwards S. V. (2022). Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* 220:iyab210.

Steiger S. S., Fidler A. E., Valcu M., Kempenaers B. (2008). Avian olfactory receptors gene repertoires: evidence for a well-developed sense of smell in birds? *Proceedings of the Royal Society B* 275:2309–2317.

Steiger S. S., Kuryshev V. Y., Stensmyr M. C., Kempenaers B., Mueller J. C. (2009). A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group γ in birds. *BMC Genomics* 10:446.

Strotmann J., Wanner I., Krieger J., Raming, K. Breer, H. (1992). Expression of odorant receptors in spatially restricted subsets of chemosensory neurons. *Neuroreport* 3:1053–1056.

Suh, A. (2016). The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta* 45:50–62.

Tixier-Boichard M., Bed'hom B., Rognon X. (2011). Chicken domestication: from archeology to genomics. *Comptes Rendus Biologies* 3:197–204.

Van Huynh A., Rice A. M. (2021). Odor preferences in hybrid chickadees: implications for reproductive isolation and asymmetric introgression. *Behavioral Ecology and Sociobiology* 75:129.

Vandewege M. W., Mangum S. F., Gabaldón T., Castoe T. A., Ray D. A., Hoffmann F. G. (2016). Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biology and Evolution* 8:470–480.

Vieira F. G., Sánchez-Gracia A., Rozas J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biology* 8:R235.

Wikelski M., Quetting M., Cheng Y., Fiedler W., Flack A., Gagliardo A., Salas R., Zannoni N., Williams J. (2021). Smell of green leaf volatiles attracts white storks to freshly cut meadows. *Scientific Reports* 11:12912.

Wilman H., Belmaker J., Simpson J., de la Rosa C., Rivadeneira M. M., Jetz W. (2014). EltonTraits 1.0: species-level foraging attributes of the world's bird and mammals. *Ecology* 95:2027.

Wink, M. (2018). Phylogeny of Falconidae and phylogeography of peregrine falcons. *Ornis Hungarica* 26:27–37.

Yokosuba M., Hagiwara A., Saito T. R., Tsukahara N., Aoyama M., Wakabayashi Y., Sugita S., Ichikawa M. (2009). Histological properties of the nasal cavity and olfactory bulb of the Japanese jungle crow *Corvus macrorhynchos*. *Chemical Senses* 34:581–593.

Zelenitsky D. K., Therrien F., Ridgely R. C., McGee A. R., Witmer L. M. (2011). Evolution of olfaction in non-avian theropod dinosaurs and birds. *Proceedings of the Royal Society B* 278:3625–3634.

Zhan X., Pan S., Wang J., Dixon A., He J., Muller M. G., Ni P., Hu L., Liu Y., Hou H., Chen Y., Xia J., Luo Q., Xu P., Chen Y., Liao S., Cao C., Gao S., Wang Z., Yue Z., Li G., Yin Y., Fox N., Wang J., Bruford M. W. (2013). Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics* 45:563–566.

III. FUNCTIONAL CHARACTERIZATION OF OLFACTORY RECEPTORS IN THE CONTEXT OF THEIR RADIATION IN BIRDS

Abstract

Olfaction plays a critical role in animal behavior and ecology. In birds, olfaction is used in foraging, kin recognition, and mate choice. Odorants are detected by olfactory receptors (ORs), however ORs also function outside of the olfactory system in tissues throughout the body. Gene expression studies of the olfactory epithelium (OE) can inform researchers about which ORs are involved in olfaction. Such studies have occurred in reptiles and mammals, but have only occurred recently in birds, and in a limited number of species. Here, we perform the first formal measurement of OR expression in the OE across the bird phylogeny, targeting four species that span avian diversity and represent diverse ecology and behavior. We successfully detected the set of ORs from the genomic repertoire with expression in the olfactory system (OE) and pectoralis muscle tissues. Our results show that the majority of the genomic OR repertoire is expressed in the bird OE, including the large bird-specific gamma-c OR subfamily. We show that some gamma-c ORs are highly expressed in the OE relative to other bird ORs, and that many gamma-c ORs are present in the OE. In addition to indicating which ORs in birds are used in olfaction, my study will provide a framework for future functional assays pinpointing the odors perceived by birds.

Introduction

Olfaction is essential for survival and reproduction in many animals. It plays a central role in foraging, avoiding predation, kin recognition, and territorial behavior. In vertebrates, air or waterborne odor molecules are detected with olfactory receptors (ORs) a gene family of G protein-coupled receptors expressed in the olfactory sensory neurons of the olfactory epithelium (OE, Buck and Axel 1991, Strotmann et al. 1992). To accommodate the incredible variety of odors in nature, ORs constitute the largest gene family in vertebrates, with over 1,000 genes in some mammals and over 300 genes in some birds (Niimura et al. 2014, Niimura and Nei 2005, Chapter II).

The number of ORs in a species' genome can be used to derive total genomic repertoire counts (Niimura et al. 2014), but not all of the genomic repertoire will be functional or relevant to the olfactory system (Maßberg and Hatt 2018). Many ORs are expressed in tissues outside of the olfactory system. Such ORs play diverse roles including regulating environmental responses in the skin and chemotaxis in sperm (Maßberg and Hatt 2018). Within this context of a complex gene family, understanding the function (or lack thereof) of specific ORs is a major challenge. Gene expression studies of the olfactory epithelium can distinguish ORs that likely bind odorants from ORs with other physiological roles and non-functional pseudogenes. Expression studies of the OE have occurred in all vertebrate classes, including in fish, amphibians, reptiles, and mammals (Ressler et al. 1993, Marchand et al. 2004, Komakov et al. 2008, Kishida et al. 2019). However, OE expression studies in birds have lagged behind other vertebrates (Sin et al. 2022).

Birds are the most speciose class of terrestrial vertebrates, inhabiting nearly all land environments. Among birds there is high diversity of social structures and foraging strategies, yet birds were long thought to rely on visual rather than olfactory signals (Audubon 1826, Hill

1905). Recent behavioral work in birds has shown important roles for olfaction in foraging, locating nest sites, seed caching behavior, and species recognition, among other behaviors (Buitron and Nuechterlein 1985, Molina-Morales et al. 2020, Bonadonna and Gagliardo 2021, Wikelski et al. 2021, Van Huynh and Rice 2021). Additionally, specific bird species rely on a highly specialized olfactory system for foraging, including *Cathartes aura* (turkey vulture) and many seabirds (order Procellariformes, Owre and Northington 1961, Grubb 1972, Bonadonna and Gagliardo 2021).

To add to the recent surge of interest in how olfaction influences bird behavior, we showed that birds have many more OR genes in their genomes than previously realized (Driver and Balakrishnan 2021, see Chapter 1, Chapter 2). Genomic analysis divides bird species' OR repertoires into three phylogenetic subgroups: alpha, gamma, and gamma-c ORs (Niimura and Nei 2005, Steiger et al. 2009, Driver and Balakrishnan 2021). The alpha and gamma OR subgroups are shared across tetrapods: chicken alpha and gamma ORs form phylogenetic clades with alpha and gamma ORs from amphibians, reptiles, and mammals (Niimura and Nei 2005, Steiger et al. 2009, Vandewege et al. 2016). This illustrates a degree of sequence conservation in the OR repertoire of these subfamilies despite at least 315 million years of divergence between mammalian and bird lineages (Lauren and Reisz 1995). Contrastingly, the gamma-c OR subfamily is only present in birds (Niimura and Nei 2005, Steiger et al. 2009, Driver and Balakrishnan 2021). Previous studies show that the gamma-c OR subfamily was the most abundant OR clade in most species (Steiger et al. 2009, Khan et al. 2015). For example, the gamma-c subfamily constituted over 85% of all OR genes in the zebra finch (60 total gamma-c ORs) and chicken (303 total gamma-c ORs, Driver and Balakrishnan 2021). Phylogenetic analyses of OR repertoires containing multiple bird species reveal that gamma-c ORs cluster into

species-specific clades as opposed to showing clear orthologous relationships among species (Zhan et al. 2013, Silva et al. 2020), suggesting possible species-specific roles for the gamma-c. Gamma-c ORs within a species also have shorter phylogenetic terminal branch lengths compared to alpha and gamma ORs, showing a high degree of sequence similarity between gamma-c genes (Steiger et al. 2009, Silva et al. 2020). However, we cannot discern the functional roles of such ORs in smell without expression studies of the bird OE.

Expression studies of the OE have not occurred in birds until recently, with only two studies published this year (Luo et al. 2022, Sin et al. 2022). In the Leach's storm-petrel (*Oceanodroma leucorhoa*) the OE expressed over 30 different ORs from the 61 OR genomic repertoire, nearly all at low expression levels (Sin et al. 2022). Only two ORs were "highly" expressed relative to the other ORs, and neither were gamma-c ORs (Sin et al. 2022). In black-crowned night heron (Nycticorax nycticorax) the OE expressed 61 ORs of the 93 OR genomic repertoire, and again most ORs were lowly expressed (Luo et al. 2022). Little egret (Egretta garzetta), also expressed ORs at low levels in the OE, with 132 ORs present (Luo et al. 2022). However, for these three bird species, only short-read Illumina-based genome assemblies are available (Luo et al. 2022, Sin et al. 2022). Therefore, the total count of the genomic repertoire may be underestimated in these species (Driver and Balakrishnan 2021). Indeed, the little egret expresses 132 ORs but had a detectable genomic repertoire of only 108 ORs, providing strong evidence of an incomplete genomic count in these studies.

To properly understand the portion of the genomic repertoire expressed in the OE, expression levels need to be compared to species with long-read assemblies (Driver and Balakrishnan 2021). Additionally, previous studies looked at either single bird species or at

species within the same bird family (Luo et al. 2022, Sin et al. 2022), and therefore it is still unknown how expression vary when examining multiple bird orders. Given the dynamic birth and death rates of ORs across the bird phylogeny (Chapter II), it is possible that expression is also dynamic, and the portion of the OR repertoire that is relevant to smell may change between species. We hypothesize that bird express a subset of their genomic OR repertoire in the OE, and that the subset of ORs expressed varies across different species. These undetected ORs would represent either nonfunctional ORs or ORs with potentially unexplored and unknown functions in other tissues.

Methods

Sample collection

To determine the location OE and specific OE regions (the anterior, middle, and posterior conchae), we referenced morphological descriptions and images of the maxilla (Yokosuba et al. 2009, Danner et al. 2017). We originally practiced dissections on bird carcasses donated by the North Carolina Museum of Natural Sciences. In this unique dissection, the maxilla is cut transversely through the nares and then from this incision the sides of the maxilla are cut proximally towards the lores. There are now three cuts in the maxilla, one transverse and distal, the other two sagittal from the nares to the lores. From this, the proximal half of the maxilla can be lifted up from the nares, exposing the tissue in the maxilla. We sampled as much tissue as possible in this part of the maxilla, and tried to sample from all three regions of the conchae, and placed immediately in microcentrifuge tubes on dry ice. Following sample collection, samples were stored in -80 C freezers. In the case of the hummingbird, maxillas were cut off at the lores,

stored on dry ice and at -80 C, and dissection occurred at the time of extraction. We obtained pectoralis muscle at the same time, following olfactory epithelium sampling.

We obtained olfactory epithelia from four bird species: chicken (*Gallus gallus*), Anna's hummingbird (*Calypte anna*), zebra finch (*Taeniopygia guttata*), and brown-headed cowbird (*Molothrus ater*). In total, we obtained four OE samples from chicken and cowbird, and five OE samples from hummingbird and zebra finch. We obtained three pectoralis samples from chicken, zebra finch, and cowbird, but we did not obtain pectoralis for hummingbird. I personally sampled the chickens immediately following a routine dispatch in the laboratory of Dr. Ken Anderson at the Prestage Department of Poultry Science at North Carolina State University. The chickens were 21-week old hyline W-36 white leghorn hens. I personally collected the zebra finch samples from he laboratory of Dr. Richard Mooney in the Department of Neurobiology at the Duke University School of Medicine. All zebra finches were adult females from separate parents. Dr. Christopher Clark at the Department of Evolution, Ecology, and Organismal Biology at the University of California Riverside collected the Anna's hummingbird maxilla, and I performed the olfactory epithelium dissections (permits USFWS MB-087454 and CDFW SC-006598 to Christopher Clark). Dr. Marc Schmidt at the Department of Biology at the University of Pennsylvania collected and dissected the cowbirds. All brown-headed cowbirds were adult males. All four species were sampled from captive populations, including the domesticated chicken and zebra finch.

RNA extractions and sequencing

To extract RNA from the olfactory epithelium and pectoralis tissue, we cut a small amount of tissue (roughly 2x2 cm) from each sample, and cut samples on dry ice. We immediately transferred tissue to RNAzol RT (RNAzol® RT Brochure, 2017) and dissolved the sample with a homogenizer. We then added water to precipitate DNA, protein, and polysaccharides, and we centrifuged to remove these. We also added 4-bromoanisole for phase separation, and we performed this optional step of the protocol twice. We then precipitated the isolated RNA with ethanol, washed with isopropanol, and solubilized in water. We tested RNA concentration and purity using a Nanodrop, and tested for RNA quality and integrity using a BioAnalyzer at the Brody Integrative Genomics Core in the Department of Pathology & Laboratory medicine at East Carolina University.

RNA quality was examined by the 4200 TapeStation (Agilent Technologies, Santa Clara, CA), with RNA integrity number (RIN) of samples ranged from 6 to 10. RNA concentration was determined by the Qubit Fluorometric Quantitation (Thermo Fisher, Waltham, MA), with 150 ng of RNA samples used for each NGS library preparation. Stranded cDNA libraries were prepared using the TruSeq Stranded LT mRNA kit (Illumina, San Diego, CA) in accordance with the manufacturer's protocol using the poly-adenylated RNA isolation. Sequencing of paired-end reads (100 bp × 2) was performed by pooling all the samples together on the NextSeq 2000 system with a P3 200 cycles reagent. Raw sequence reads were de-multiplexed and trimmed for adapters by the on-instrument DRAGEN GenerateFastQ pipeline (v3.7.4).

Read mapping

We mapped reads using the Spliced Transcripts Alignment to a Reference (STAR) aligner (Dobin et al. 2013). We were interested in OR expression specifically, so we generated the STAR reference genome not from the available species' genome assemblies, but from our previously established genomic OR repertoires of each species (Chapter II). We found the genomic OR repertoires for chicken, hummingbird, zebra finch, and cowbird as described previously (Driver and Balakrishnan 2021, Chapter II). From our final curated OR alignments, we used custom R scripts and bedtools to extract nucleotides from the associated genome (R core team, Quinlan and Hall 2010). We generated the reference genome of OR sequences without using a GTF reference annotation. We then mapped reads to the genomic OR repertoires using STAR default parameters.

Counting and differential expression

We counted the number of reads in output SAM files using the dplyr package in R (Wickham et al. 2022). To measure gene expression, we converted raw counts to counts per million (CPM). CPM is the total number of counts for a given locus divided by the total number of counts in the sample, and then multiplying by one million, which controls for sequencing depth of the sample. We analyzed differential gene expression using the *limma* and *edgeR* packages in R (Robinson et al. 2010, Ritchie et al. 2015). We used the TMM method to normalize expression data (Robinson and Oshlack 2010). We did not filter genes with low expression due to previous reports of many bird ORs showing low expression levels (Luo et al. 2022, Sin et al. 2022). A standard linear model with "tissue" (either pectoralis "PEC" or olfactory epithelium "OE") as the independent variable was used for testing within chicken, zebra finch, and brown-headed cowbird. We only

obtained OE tissue from hummingbird so we used only three species in the differential expression analyses. We adjusted *P* values for multiple testing using the Benjamini-Hochberg correction. We also ran a student's t-test comparing CPM values between OE and pectoralis samples for chicken, zebra finch, and cowbird, as an alternative way to measure differential expression from a relatively small number of overall genes. For mapping to phylogenetic trees, we used trees created as described previously, using maximum likelihood methods in IQ-TREE (Minh et al. 2020). We overlayed expression heatmap plots to the phylogeny using the gheatmap function in *ggtree* in R (Yu 2020).

Results

ORs found in tissues

We sequenced whole-mRNA transcriptomes from the OE of four bird species and from the PEC of three species. Across all four species, we detected 590 expressed ORs out of 667 genomic ORs from Chapter II (Fig. 1, 88.46% of genomic ORs showed expression). Zebra finch was the only species that had its entire genomic OR repertoire expressed in the OE. Brown-headed cowbird expressed 136 of 137 ORs expressed in the OE (99.28%). Anna's hummingbird also had a high proportion of its ORs expressed in the OE (99 of 109, 90.83%). Although chicken had the highest total number of OR genes expressed in the OE, with 286 ORs, chicken also had the largest genomic OR repertoire of the species sampled, and had the lowest overall proportion of OR expressed (286 of 352, 81.25%).

There was a large amount of variation between samples, even within the same species and tissue (Fig. 2). For all four species, individual variation was high, with specific samples

consistently showing higher OR expression than other samples. This was not due to different ORs being expressed between samples, but rather, consistent high or low expression across the entire OR repertoire. For example, the one zebra finch sample had an average OR expression of 1.07 log CPM, whereas another sample had an average of -0.20 log CPM. Variation was high in all species, for example hummingbird had one sample with an average of 0.414 log CPM OR expression, and another sample had an average expression of -0.80 log CPM. Variation in log CPM between and within species for OE is visualize in figure 2. In addition to variation in expression, there was high variation in samples between total number of ORs expressed in OE. This was highest in the chicken, with one sample expressing 246 ORs, whereas another OE sample expressed only 9 ORs (Fig. 2). We saw a similar but less extreme version of this variation in other species, including hummingbird, with one OE sample containing 100 ORs and another sample containing only 3 ORs (Fig. 2).

Contrary to mammals that express ORs with tissue-specific roles across the body (Maßberg and Hatt 2018), all ORs that were expressed in the OE were also expressed in pectoralis, so that no ORs were expressed exclusively in the pectoralis. Zebra finch had the largest number of OR genes expressed in the pectoralis, with 46 total (66.67 % of 69 genomic OR). Brown-headed cowbird expressed 35 ORs in the pectoralis (25.55% of 137 genomic ORs). The chicken had the smallest OR repertoire in the pectoralis, with only 17 ORs (4.83% of 352 genomic ORs).

Expression in the OE included ORs from the alpha, gamma, and gamma-c subfamilies. The ORs with the highest expression levels in Anna's hummingbird zebra finch and cowbird were in the gamma-c subfamily. The chicken had at least one gene in all three subfamilies that showed high

expression levels, although most abundant OR in the chicken (as well as the most abundant OR in this study) was a gamma-c OR. All OR subfamilies were also present in the pectoralis across the three samples species, although with fewer representatives.

Differential expression

Overall, few ORs were differentially expressed between tissue comparisons, showing that the majority of ORs expressed in both tissues have similar expression levels following TMM normalization. However, fold changed tended to be in one direction, the higher expression in the OE. Due to the large variance between OE samples, these differential expression results were not significance. In all cases, OE samples had the highest levels of gene expression, and for ORs expressed in both tissues, OE expression was on average 266 times higher than in pectoralis in zebra finch, 40 times higher than in pectoralis in cowbird, and 26 times higher than in pectoralis in chicken. In zebra finch, of 46 total ORs expressed in both tissues there were five differentially expressed (DE) ORs between tissues (Fig 1., red asterisks). These consisted of one alpha OR (t = 14.27, *P-adj.* < 0.01), one gamma OR (t = 13.49, *P-adj.* < 0.01), and three gamma-c ORs (t = 13.27, *P-adj.* = 0.02; t = 14.12, *P-adj.* = 0.03; t = 12.40, *P-adj.* = 0.03). Four of the ORs were more highly expressed in OE compared to PEC, however, the alpha OR was more highly expressed in PEC compared to OE. This was the only OR to show this pattern in our dataset. In the cowbird, of the 35 ORs expressed in both tissues, a single gamma-c OR showed increased expression in OE compared to PEC (t = 14.48, *P-adj.* = 0.02). In the chicken, of 17 ORs expressed in both tissues, one gamma-c OR showed higher expression in OE (t = 16.90, *P-adj* =

0.01). In our student's T-test comparing OR expression between OE and pectoralis within species, we did not find any significant differences.

Discussion

Most genomic ORs are expressed in OE

We successfully detected OR expression in both the OE and pectoralis muscle tissues in three bird species, the chicken, zebra finch and brown-headed cowbird, and in the OE tissue of Anna's hummingbird. This is the first study of OR expression levels in the OE for the bird orders represented here, including Galliformes, Trochiliformes, and Passeriformes. These three orders represent diverse lineages within the bird phylogeny- the Galloanseres, including Galliformes, separated from the Neoaves, including Trochiliformes and Passeriformes, 85 to 90 million years ago, and is one of the earliest diverging lineages within the extant birds. We show that the majority of genomic ORs are expressed in the OE in both Galloanseres and Neoaves species, illustrating that the majority of genomic ORs are involved in the olfactory system, and that this role is preserved across the phylogeny. Therefore, genomic OR counts across the bird phylogeny are likely relevant to the ecology and behavior of many bird species. These results agree with previous studies that showed that the majority of the genomic OR repertoires were also expressed in the OE of Leach's storm-petrel, black-crowned night heron, and little egret (Luo et al. 2022, Sin et al. 2022). However, these genomic OR repertoires were determined by surveying short-read Illumina-based assemblies, that we have shown to undercount the number of genomic ORs (see Chapter I, Driver & Balakrishnan 2021). For example, the little egret expressed more ORs in the OE than were detected in the genome (Luo et al. 2022). Here, we present the first

study comparing OR expression in the OE to the more reliable genomic OR counts from long read assemblies, and we continue to show that the majority of ORs are expressed in the OE.

In the zebra finch, we found that all genomic ORs were expressed in the OE. Similarly, in the cowbird, we detected the expression of 136 of the 137 genomic ORs. This suggests that either the entire intact genomic OR repertoire of these species is functional and relevant to the olfactory system, or that we are still undercounting the genomic OR repertoires of these species, despite using long read genomes (see Chapter II). These expression results support the possibility that despite being highly contiguous, there are still problematic areas of long read assemblies, and that additional ORs may be present in these problem regions. ORs clusters in mammals and birds are flanked by repeat regions, thereby making the assembly of these regions particularly difficult (Glusman et al. 2000, Vandewege, Driver). Therefore, even current technologies may not resolve these regions. Additional surveys could be performed to extract putative ORs from our RNA-seq data that are not based on the previously determined genomic OR repertoires. These searches may pull out unique ORs not detected in the genomic repertoire. However, given the high sequence similarity of the gamma-c ORs, it may be difficult to assign reads to particular ORs with no genomic reference, as sequence differences may be between alleles as opposed to different genes. Conversely, in hummingbird and particularly the chicken, there are also genomic ORs absent from the OE and muscle, indicating that a portion of the genomic OR repertoire was either transcriptionally inactive in the individual birds we sampled, or expressed in other tissues. It is unclear what role these unexpressed ORs may play in birds, although it is likely that these ORs serve some function as their genomic sequences maintain an open reading frame (Chapter II). It is also possible that the expression of ORs, particularly in the OE, is dynamic and

responsive to odorants in the environment, and that these ORs would be “turned on” in response to particular stimuli, which were not implemented in this study.

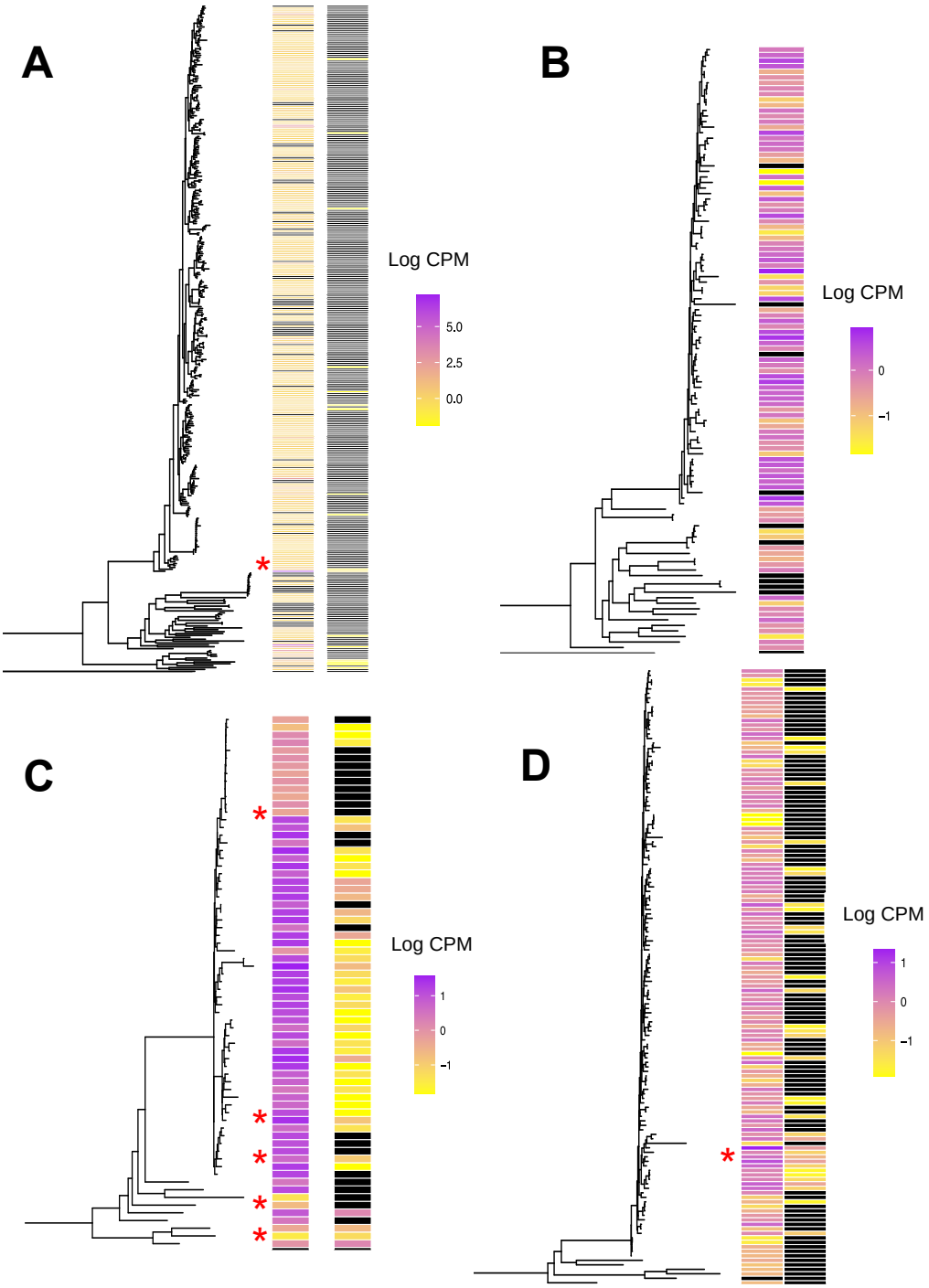

Fig 3.1. OR gene expression in OE and pectoralis in log CPM. Left columns in each panel are OE, right columns pectoralis. Asterisks show differentially expressed ORs between tissue types. In B, the single column shows OE expression. Species genomic OR repertoires are depicted in phylogenetic trees. Each tip is one OR, and corresponding OE and pectoralis expression levels are shown next to the OR. (A) chicken *Gallus gallus*, (B) Anna's hummingbird *Calypte anna*, (C) zebra finch *Taeniopygia guttata*, (D) brown-headed cowbird *Molothrus ater*.

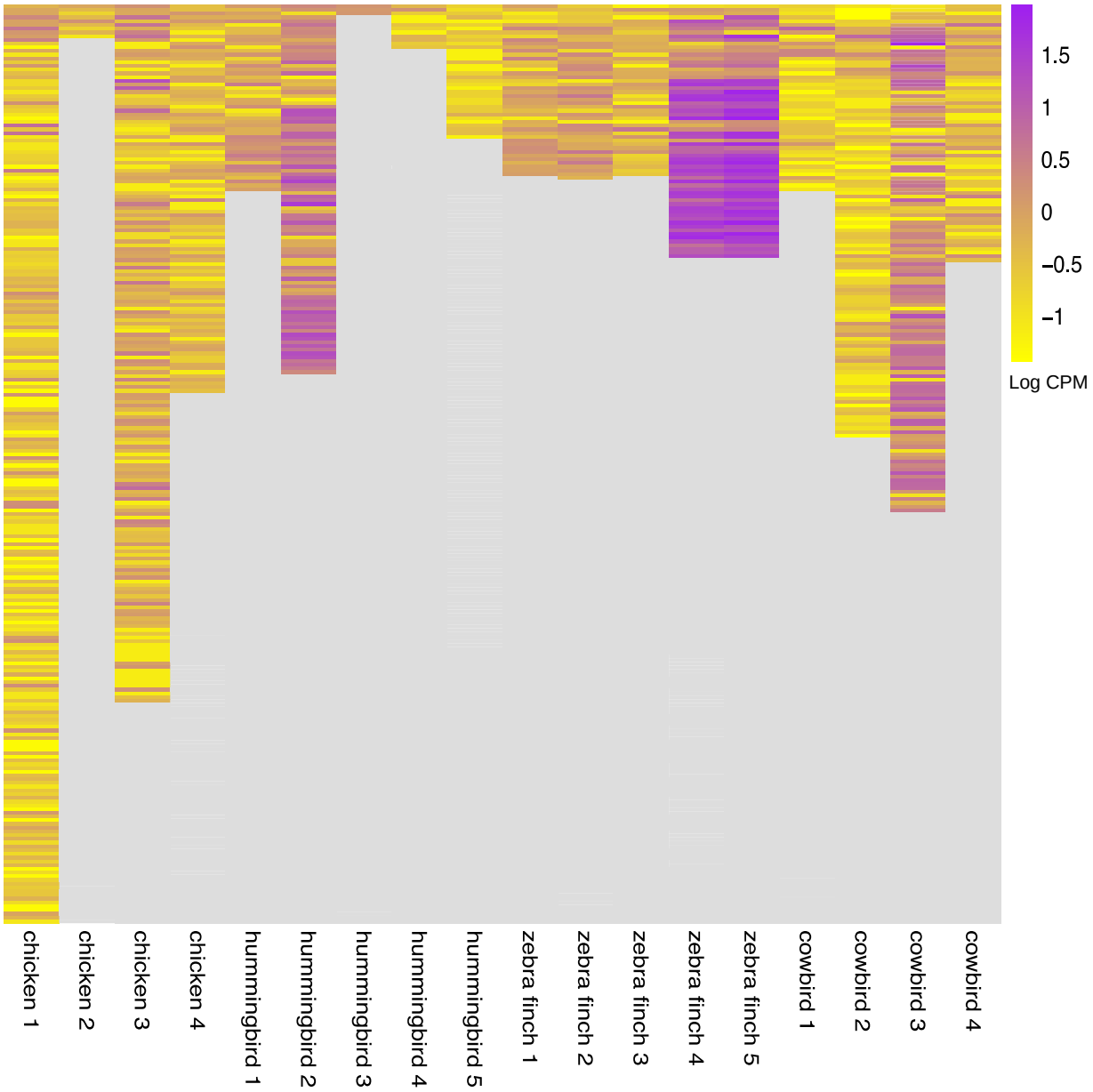

Fig 3.2. OR expression levels and the total number of ORs expressed varied substantially between OE samples within species. Each column represents one OE sample from one of the four species included (chicken, hummingbird, zebra finch, cowbird). Colors show the log CPM count for individual OR genes, represented by each cell. Cell colors give expression levels for a particular OR, and cells in the same row do not necessarily correspond to the same OR, especially between different species. Zebra finch and hummingbird differences in expression between OE samples are especially strong.

High expression levels of ORs, including gamma-c ORs

Compared to previous studies of the bird OE, we found relatively high expression of numerous ORs, including gamma-c ORs. Previous studies of OR expression in the bird OE showed that although a large number of ORs may be present (for example, 132 OE expressed ORs in little egret), that the majority of these ORs are expressed at low levels. For example, in little egret, all expressed ORs were below 1.5 TPM (read counts divided by length of each gene in kilobases), and all night heron ORs were expressed below 2 TPM, except one OR at 3.0 TPM (Luo et al. 2022). Additionally, only two ORs detected in the storm-petrel OE were expressed above 1.0 Log CPM (Sin et al. 2022). Of the two highly expressed ORs in the storm-petrel, one was in the alpha subfamily (OR5-11), and one was a member of the gamma subfamily (OR6-6, Sin et al. 2022). Although gamma-c ORs were present in the storm-petrel OE, all were expressed at low levels (OR family 14, Sin et al. 2022). Low OR expression levels are also reported in mammals, including humans (Olender et al. 2016). Each olfactory sensory neuron expresses only one OR, meaning that expression of any individual OR is restricted to a subset of the total number of olfactory sensory neurons, decreasing overall OR expression levels (Lomvardas et al. 2006).

These previous results are consistent with our findings in hummingbird and cowbird, where all ORs expressed in the OE were relatively low when averaged across samples. In hummingbird all ORs were below expressed 1.0 Log CPM, and in cowbird only one OR was expressed above 1.0 Log CPM. However, across our zebra finch samples, we found that 32 of the ORs had expression levels above 1.0 Log CPM. Of these 32 ORs, 31 were in the gamma-c subfamily, and one OR was in the gamma subfamily. This is the highest expression level

reported for gamma-c ORs, and also shows that this expression level is consistent across a substantial fraction of the total zebra finch gamma-c OR repertoire, providing strong support that gamma-c ORs are integral in the olfactory system.

In the chicken, we also detected higher OR expression levels than previous studies, with 18 chicken ORs expressed above 1.0 Log CPM when averaged across all chicken OE samples. In contrast to the zebra finch, these 18 ORs were diverse across OR subfamily type, including three alpha ORs, two gamma ORs, and 15 gamma-c ORs. This shows that across highly divergent lineages of birds, the expression levels of subfamilies differ substantially. This is consistent with genomic patterns that show reduced numbers of alpha and gamma ORs but increased numbers of gamma-c ORs in passerines (Chapter II), whereas Galloanseres maintains high levels of all subfamilies (Chapter II). The chicken therefore may rely on all subfamilies to detect odors, whereas zebra finch is more dependent on gamma-c. Whether gamma-c in zebra finch has replaced the functional roles of odor detection provided by alpha and gamma in chicken, or if zebra finch is simply detecting different odors, is unknown. In the chicken, the most highly expressed OR, a gamma-c OR (genomic coordinates CM000108.5_1785570_1786508) was the most highly expressed OR across all ORs and all species in our study, at 7.19 Log CPM. This is the most highly expressed bird OR ever reported, and the functional relevance of this OR could be investigated in future analyses.

Few differentially expressed ORs and high variance between samples

We detected relatively few differentially expressed OR when comparing OE and pectoralis muscle tissues within species. Differentially expressed ORs included a single gamma-c OR in

chicken and cowbird, and five ORs (one alpha, one gamma, three gamma-c) in zebra finch. All of these ORs, except one, showed higher expression in the OE as opposed to the pectoralis when measuring differential expression using the *limma* and *edgeR* packages following TPM normalization of the counts. We also performed student's T-tests to compare our two tissues, but this did not show any significant differences. The zebra finch had the greatest overlap in expression between tissues, with 42 ORs present in both tissues and not showing differential expression between the tissues. These results suggest that for ORs present in both tissues, expression is similar, and that ORs may function in both tissues, perhaps in different functional roles. Alternatively, the ORs may function similarly across tissues, and such as performing essential "housekeeping" roles that are consistent and uniform across tissue types. For these ORs, a functional role in the olfactory system is therefore unclear despite expression in the OE.

The lack of differential expression found in our study is due to the large amount of variation within OE samples of the same species. This high level of variation within the same species and tissue may mask true levels of differential expression. It is unclear why we have some individual OE samples that express all ORs at higher levels than other OE samples, even following correction for sequencing depth. There are several possibilities, including possible sampling and RNA extraction differences or errors. We performed dissections as uniformly as possible and in each case freshly sacrificed birds were dispatched and dissected in the same manner. We performed RNA extractions on different dates but we were consistent, with minimal time between removal from -80 C storage and dissolving in RNAzol. It is possible that between individuals, different parts of the OE were sampled in the final tissue sent for sequencing. The bird OE is divided into three sections, the anterior, middle, and posterior conchae (Danner et al. 2017). Although there is no evidence as to which region of the OE expresses more or fewer ORs,

Sin et al. specifically sampled from the anterior conchae (Sin et al 2022). We sampled from the OE generally, and did not target a particular region, therefore, slight differences in the region of the OE used for each sample may account for some of the variation that we observed.

Alternatively, the variation seen between samples could reflect real biological variation between the individual birds of the same species. Specific individuals may express more ORs or be more sensitive to odorants than others. Although we controlled sex and age in the within species comparisons in this study, it is possible that other genetic factors cause different individuals of the same species to express ORs in the OE at different rates. It is also possible that OR expression in the OE is highly dynamic and dependent on some type of external stimuli. Zebra finch and chicken both showed variation between samples, and individuals from both species were sacrificed a sterile laboratory setting, and individuals came from the same enclosures. However, it is possible odorant stimuli in the air were slightly different between when each individual was sampled, and that the different birds were responding to different concentrations of odorants in the air that varied slightly between sampling efforts. These small effects could also explain the variation in number of ORs expressed, which in addition to expression levels, also varied substantially, particularly in chicken and hummingbird (Fig. 2). Future studies in more controlled settings, as well as examining variation in relation to genetic background, could help resolve the reason for this high variation. In turn, these studies could more accurately characterize differential expression between tissues after carefully controlling for this variation. Finally, our differential expression methods relied on *limma* and *edgeR*, methods used traditionally to analyze differential expression across transcriptome-wide data. Here, we apply these methods to a small set of target genes, the genomic OR repertoire for each species. It is unclear whether this alters the differential expression methods substantially.

However, a t-test performed on the expression data comparing tissue types also did not detect differential expression, so although there may be issues with differential expression, variation between samples remains a major issue.

In the zebra finch, we detected one alpha OR that was more highly expressed in the pectoralis muscle than in the OE. This was the only instance in our dataset of the pectoralis showing significantly higher expression of an OR than the in the OE. Although this alpha OR was also present in the zebra finch OE, this expression pattern presents the interesting possibility of a bird OR with a primary function outside of olfaction, and potentially a function that is muscle specific. Further expression studies of the muscle, OE, and other tissues would help us understand the functional role of this OR in birds. Additionally, because this is an alpha OR, there are likely orthologous and paralogous relationships between the zebra finch OR and mammalian and reptilian alpha ORs (Steiger et al. 2009). It may be possible to match known expression levels or odor binding properties of the orthologous mammalian ORs to this alpha OR, to see if muscle expression is a consistent role in this alpha OR over evolutionary time.

Conclusion

We have shown that across multiple bird species and orders, that the majority of ORs found in bird genomes are expressed in the olfactory epithelium, solidifying the connection between genomic OR repertoire size and a species' reliance on olfaction. We show that most ORs are lowly expressed, with a few exceptions. We also show for the first time that many gamma-c ORs are expressed in the OE, and that gamma-c ORs are often the most highly expressed ORs in the OE. This is the first time that a large gamma-c OR repertoire was shown to

be expressed in the OE, and we also report highest expression levels of gamma-c detected in the bird OE. Gamma-c ORs are bird-specific and OE expression studies in other vertebrate classes do not provide information about the functional role of gamma-c. We show the strongest evidence to date that this expansive bird OR subfamily has duplicated and retained duplications due to a relevance of this OR subfamily to olfaction. This expression study, the first of its kind in birds to look widely across the bird phylogeny, shows the importance of ORs and gamma-c ORs across distantly related bird lineages. By successfully detecting the expression of many ORs in the bird OE, these data will facilitate future work to select ORs for functional ("deorphanization") experiments to identify the specific odorants that bird ORs can detect (see Saito et al. 2009). In addition to characterizing expression in the OE of bird species, we have implicated which ORs in the genomic repertoire are involved in olfaction, allowing for subsequent work to select and functionally test the unknown binding properties of bird ORs.

References

Audubon J.J. (1826). Account of the habits of the turkey buzzard, *Vultur aura*, particularly with the view of exploding the opinion generally entertained of its extraordinary power of smelling. *Edinburgh New Philosophical Journal* 2:172–184.

Bonadonna F., Gagliardo A. (2021). Not only pigeons: avian olfactory navigation studied by satellite telemetry. *Ethology Ecology, & Evolution* 33:273–289.

Buck L., Axel R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175–187.

Buitron D., Nuechterlein G. L. (1985). Experiments on olfactory detection of food caches by black-billed magpies. *Condor* 87:92–95.

Danner R. M., Gulson-Castillo E. R., James H. F., Dzielski S. A., Frank D. C., Sibbald E. T., Winkler D. W. (2017). Habitat-specific divergence of air conditioning structures in bird bills. *The Auk* 134:65–75.

Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M.,

Gingeras T. R. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Driver R. J., Balakrishnan C. N. (2021). Highly contiguous genomes improve the understanding of avian olfactory receptor repertoires. *Integrative & Comparative Biology* 61:1281–1290.

Glusman G., Bahar A., Sharon D., Pilpel Y., White J., Lancet D. (2000). The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mammalian Genome* 11:1016–1023.

Grubb, T. C. (1972). Smell and foraging in shearwaters and petrels. *Nature* 237:404–405.

Hill A. (1905). Can birds smell? *Nature* 71:318–319.

Khan I., Yang Z., Maldonado E., Li C., Zhang G., Gilbert M. T. P., Jarvis E. D., O'Brien S. J., Johnson W. E., Antunes A. (2015). Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Molecular Biology and Evolution* 32:2832–2843.

Kishida T., Go Y., Tatsumoto S., Tatsumi K., Kuraku S., Toda M. (2019). Loss of olfaction in sea snakes provides new perspectives on the aquatic adaptation of amniotes. *Proceedings of the Royal Society B* 286:20191828.

Kolmakov N. N., Kube M., Reinhardt R., Canario A. V. M. (2008). Analysis of the goldfish *Carassius auratus* olfactory epithelium transcriptome reveals the presence of numerous non-olfactory GPCR and putative receptors for progestin pheromones. *BMC Genomics* 9:429.

Laurin M., Reisz R. R. (1995). A reevaluation of early amniote phylogeny. *Zoological Journal of the Linnean Society* 113:165–223.

Lomvardas S., Barnea G., Pisapia D. J., Mendelsohn M., Kirkland J., Axel R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* 126:403–413.

Luo H., Luo S., Fang W., Lin Q., Chen X, Zhou X. (2022). Genomic insight into the nocturnal adaptation of the black-crowned night heron (*Nycticorax nycticorax*). *BMC Genomics* 23:683.

Maßberg D., Hatt H. 2018. Human olfactory receptors: novel cellular functions outside of the nose. *Physiological Reviews* 98:1739–1763.

Marchand J. E., Yang X., Chikaraishi D., Krieger J., Breer H., Kauer J. S. (2004). Olfactory receptor gene expression in tiger salamander olfactory epithelium. *Journal of Comparative Neurology* 474:453–467.

Minh B. Q., Schmidt H. A., Chernomor O., Schrempf D., Woodhams M. D., von Haeseler A., Lanfear R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534.

Molina-Morales M., Castro J., Albaladejo G., Parejo D. (2020). Precise cache detection by olfaction in a scatter-hoarder bird. *Animal Behaviour* 167:185–191.

Niimura Y., Matsui A., Touhara K. (2014). Extreme expansion of the olfactory receptor gene

repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Research* 24:1485–1496.

Niimura Y., Nei M. (2005). Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences* 102:6039–6044.

Olender T., Keydar I., Pinto J. M., Tatarskyy P., Alkelai A., Chien M., Fishilevich S., Restrepo D., Matsunami H., Gilad Y., Lancet D. (2016). The human olfactory transcriptome. *BMC Genomics* 17:619.

Owre O. T., Northington P. O. (1961). Indication of the sense of smell in the turkey vulture, *Cathartes aura* (Linnaeus), from feeding tests. *American Midland Naturalist* 66:200–205.

Quinlan A. R., Hall I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Ressler K. J., Sullivan S. L., Buck L. B. (1993). A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* 73:597–609.

Ritchie M. E., Phipson B., Wu D., Hu Y., Law C., Shi W., Smyth G. K. (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43:e47.

RNAzol® RT Brochure, (2017). Molecular Research Center, Inc. Cincinnati, OH.

Robinson M. D., McCarthy D. J., Smyth G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Robinson M. D., Oshlack A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11:R25.

Saito H., Chi Q., Zhuang H., Matsunami H., Mainland J. D. (2009). Odor coding by a mammalian receptor repertoire. *Science Signaling* 2:ra9.

Silva M. C., Chibucos M., Munro J. B., Daugherty S., Coelho M. M., Silva J. C. (2020). Signature of adaptive evolution in olfactory receptor genes in Cory's shearwater supports molecular basis for smell in procellariiform seabirds. *Scientific Reports* 10:543.

Sin S. Y. W., Cloutier A., Nevitt G., Edwards S. V. (2022). Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* 220:iyab210.

Steiger S. S., Kuryshev V. Y., Stensmyr M. C., Kempenaers B., Mueller J. C. (2009). A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group γ in birds. *BMC Genomics* 10:446.

Strotmann J., Wanner I., Krieger J., Raming, K. Breer, H. (1992). Expression of odorant receptors in spatially restricted subsets of chemosensory neurons. *Neuroreport* 3:1053–

1056.

Vandewege M. W., Mangum S. F., Gabaldón T., Castoe T. A., Ray D. A., Hoffmann F. G. (2016). Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biology and Evolution* 8:470–480.

Wickham H., François R., Henry L., Müller K. (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.10. https://CRAN.R-project.org/package=dplyr

Wikelski M., Quetting M., Cheng Y., Fiedler W., Flack A., Gagliardo A., Salas R., Zannoni N., Williams J. (2021). Smell of green leaf volatiles attracts white storks to freshly cut meadows. *Scientific Reports* 11:12912.

Yokosuba M., Hagiwara A., Saito T. R., Tsukahara N., Aoyama M., Wakabayashi Y., Sugita S., Ichikawa M. (2009). Histological properties of the nasal cavity and olfactory bulb of the Japanese jungle crow *Corvus macrorhynchos*. *Chemical Senses* 34:581–593.

Yu, G. (2020). Using ggtree to visualize data on tree-like structures *Current Protocols in Bioinformatics* 69:e96.

Zhan X., Pan S., Wang J., Dixon A., He J., Muller M. G., Ni P., Hu L., Liu Y., Hou H., Chen Y., Xia J., Luo Q., Xu P., Chen Y., Liao S., Cao C., Gao S., Wang Z., Yue Z., Li G., Yin Y., Fox N., Wang J., Bruford M. W. (2013). Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics* 45:563–566.