Original Research Article

Independent origins of the Golgi phosphoprotein 3 oncoprotein gene

Juan C. Opazo[1,2,3,*], Michael W. Vandewege[4], Javier Gutierrez[4], Kattina Zavala[2], Luis Vargas-Chacoff[1,5,6], Francisco J. Morera[1,7], Gonzalo A. Mardones[1,8,9,*]

[1]Integrative Biology Group, Universidad Austral de Chile, Valdivia, Chile.
[2]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.
[3]Millennium Nucleus of Ion Channel-Associated Diseases (MiNICAD), Valdivia, Chile.
[4]Department of Biology, Eastern New Mexico University, Portales, New Mexico, USA.
[5]Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile.
[6]Centro Fondap de Investigación de Altas Latitudes (IDEAL), Universidad Austral de Chile, Valdivia, Chile.
[7]Applied Biochemistry Laboratory, Instituto de Farmacología y Morfofisiología, Facultad de Ciencias Veterinarias, Universidad Austral de Chile, Valdivia, Chile.
[8]Department of Physiology, School of Medicine, Universidad Austral de Chile, Valdivia, Chile.
[9]Center for Interdisciplinary Studies of the Nervous System (CISNe), Universidad Austral de Chile, Valdivia, Chile.

***Correspondence:**
Juan C. Opazo
jopazo@gmail.com
Gonzalo A. Mardones
gonzalo.mardones@uach.cl

Keywords: Birds, Cancer, Gene Family Evolution, Golgi apparatus, GOLPH3, GOLPH3L, intrinsically disordered region, Vps74.

Abstract

Golgi phosphoprotein 3 (GOLPH3) is considered the first oncoprotein of the Golgi apparatus. It was identified as an evolutionarily conserved protein upon its discovery about 20 years ago, but its function remains puzzling in normal and cancer cells. The GOLPH3 gene is part of a group of genes that also includes the GOLPH3L gene. Because cancer has deep roots in multicellular evolution, studying the evolution of the GOLPH3 gene family in non-model species represents an opportunity to identify new model systems that could help better understand the biology behind this group of genes. The main goal of this study is to explore the evolution of the GOLPH3 gene family in birds as a starting point to understand the evolutionary history of this oncoprotein. We identified a repertoire of three GOLPH3 genes in birds. We found duplicated copies of the GOLPH3 gene in all main groups of birds other than paleognaths, and a single copy of the GOLPH3L gene. We suggest there were at least three independent origins for GOLPH3 duplicates. Amino acid divergence estimates show that most of the variation is located in the N-terminal region of the protein. Our transcript abundance estimations show that one paralog is highly and ubiquitously expressed, and the others were variable. Our results are an example of the significance of understanding the evolution of the GOLPH3 gene family, especially for unraveling its structural and functional attributes.

1 Introduction

Golgi phosphoprotein 3 (GOLPH3) is a highly conserved protein of the Golgi apparatus (Wu et al., 2000; Bell et al., 2001) considered the first oncoprotein of this subcellular compartment (Sechi et al., 2015). The GOLPH3 gene family comprises the conserved GOLPH3 gene and the GOLPH3L gene found only in vertebrates (Wu et al., 2000; Bell et al., 2001). Despite the vast amount of empirical evidence demonstrating the contribution of GOLPH3 to tumorigenesis and cancer, a full understanding of its molecular role has not yet emerged. This is mainly due to multiple functions attributed to GOLPH3 (Sechi et al., 2015), including, the sorting of Golgi glycosyltransferases (Ali et al., 2012; Isaji et al., 2014; Pereira et al., 2014), the modulation of focal adhesion dynamics (Arriagada et al., 2019), induction of membrane curvature (Rahajeng et al., 2019), and an intriguing function for a Golgi protein, regulating mitochondrial function (Nakashima-Kamimura et al., 2005). Because cancer has deep evolutionary roots that arise as a consequence of the multicellularity (Boddy et al., 2020), and is widespread across animals (Albuquerque et al., 2018), studying the evolution of the GOLPH3 gene family in non-model species can provide significant information for a comparative oncology approach, which is emerging as an integrative field to tackle cancer (Boddy et al., 2020).

The availability of whole-genome sequences opens an opportunity to understand the evolution of gene families. The annotation of gene repertoires in different species has revealed that copy number variation is an important source of variability that should be considered when making functional comparisons (Abegglen et al., 2015; Sulak et al., 2016). Phylogenetic reconstructions show that the evolution of gene families follows complex pathways, including gene gain and losses and independent origins, making it challenging to perform direct interspecies comparisons. Thus, understanding the variability of gene repertoires and their duplicative history represents an essential piece of information to understand the biological functions associated with a group of genes and make biologically meaningful comparisons. Today, the GOLPH3 gene family is viewed as a group of genes containing two paralogs (GOLPH3 and GOLPH3L) with 1:1 orthologs among most vertebrate species (Wu et al., 2000; Bell et al., 2001). Among amniotes, it is suggested that birds are less susceptible to cancer in contrast to mammals (Ratcliffe, 1933; Lombard and Witte, 1959; Effron et al., 1977; Pang et al., 2012; Møller et al., 2017; Boddy et al., 2020); although this information should be taken with caution given sampling bias (Hochberg and Noble, 2017). Thus, the study of genes associated with cancer in birds could provide clues about the genetic bases associated with this difference and suggest additional model systems that could help to understand the biology of the GOLPH3 gene family.

The main goal of this study was to understand the evolutionary history of the GOLPH3 gene family in birds. We took advantage of whole genome sequences in representative species of all main lineages of birds to understand the evolutionary pathways that gave rise to GOLPH3 paralogs. According to our assessment, we identified a repertoire of three GOLPH3 genes in birds. We found duplicated copies of the GOLPH3 gene in all main groups of birds other than paleognaths, and a single copy of the GOLPH3L gene that would be derived from the common ancestor of all birds. In the case of the GOLPH3 gene, our gene tree suggests at least three independent origins for the duplicated copies, in the ancestor of Galliformes and Anseriformes, in the ancestor of Anseriformes, and the ancestor of Neoaves. Divergence estimates between duplicated genes showed that most of the variation is located in the N-terminal region of the protein. Our transcript abundance estimations showed that one paralog was highly and ubiquitously expressed, while the others were variable. Our evolutionary analyses suggest a more

complex than anticipated evolutionary history of the GOLPH3 gene family, a scenario that could have implications for cancer.

2 Material and Methods

2.1 DNA sequences and phylogenetic analyses

We performed searches for GOLPH3 sequences in avian genomes in the National Center for Biotechnology Information (NCBI) (Sharma et al., 2018) and the Ensembl v.102 databases (Yates et al., 2020). We retrieved orthologs and paralogs from the NCBI (Sharma et al., 2018) using the chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and mallard (*Anas platyrhynchos*) sequences using the program blast (blastn) (Altschul et al., 1990) against the non-redundant database (nr) with default parameters. Additionally, we also retrieved sequences from the Ensembl v.102 database (Yates et al., 2020). In cases where sequences are not complete we manually annotated them. To do so, we first identified the genomic fragment containing the GOLPH3 gene in Ensembl v.102 (Yates et al., 2020) or NCBI databases (Geer et al., 2010). Once identified, genomic fragments were extracted, including flanking genes. After extraction, we manually annotated GOLPH3 genes by comparing known exon sequences from a species that share a common ancestor most recently in time to the species of which the genomic piece is being annotated using the program Blast2seq v2.5 (Tatusova and Madden, 1999) with default parameters. Accession numbers and details about the taxonomic sampling are available in Supplementary Table S1.

We performed separate phylogenetic analyses for GOLPH3 and GOLPH3L paralogs. Amino acid sequences were aligned using MAFFT v.7 (Katoh and Standley, 2013), allowing the program to choose the alignment strategy (L-INS-i in both cases). Nucleotide alignments were generated using the amino acid alignments as templates using the software PAL2NAL (Suyama et al., 2006). We used the proposed model tool of IQ-Tree v.1.6.12 (Kalyaanamoorthy et al., 2017) to select the best-fitting model of codon substitution, which selected MGK+F1X4+R3 for GOLPH3 and MGK+F3X4+G4 for GOLPH3L. This approach uses a more realistic description of the evolutionary process at the protein-coding sequence level by incorporating the genetic code structure in the model. We used the maximum likelihood method to obtain the best trees using the program IQ-Tree v1.6.12 (Trifinopoulos et al., 2016). We assessed support for the nodes using three strategies: a Bayesian-like transformation of aLRT (aBayes test) (Anisimova et al., 2011), SH-like approximate likelihood ratio test (SH-aLRT) (Guindon et al., 2010) and the ultrafast bootstrap approximation (Hoang et al., 2018). In each case (GOLPH3 and GOLPH3L), we carried out 25 independent runs to explore the tree space, and the tree with the highest likelihood score was chosen. In both cases, GOLPH3 and GOLPH3L sequences from crocodiles and turtles were used as outgroups (Supplementary Table S1).

2.2 Molecular Evolution analysis

To measure variation in functional constraint among the GOLPH3L genes and to test for evidence of positive selection, we estimated the omega parameter (d_N/d_S), using a maximum-likelihood approach (Goldman and Yang, 1994) implemented in the CODEML module of the program PAML v.4.8a (Yang, 2007). We implemented branch-site models, which explore changes in the omega parameter for a set of sites in a specific branch of the tree to assess changes in their selective regime (Yang and dos Reis, 2011). In this case, we conducted two separate analyses. In the first, the ancestral branch of Galliformes was labeled as the foreground branch, while in the second, the branch leading to manakins was labeled as a foreground branch. We compared the modified model

A (Yang et al., 2005; Zhang et al., 2005; Yang and dos Reis, 2011), in which some sites are allowed to change to an omega value > 1 in the foreground branch, with the corresponding null hypothesis of neutral evolution using a Likelihood Ratio Test (LRT). Three starting omega values (0.5, 1, and 2) were used to check the existence of multiple local optima. The Bayes Empirical Bayes (BEB) method was used to identify sites under positive selection (Nielsen and Yang, 1998; Yang et al., 2000).

2.3 Secondary structure and disordered region prediction and protein structure homology modeling

Secondary structure and disordered region prediction was performed using the PredictProtein server (https://predictprotein.org/)(Yachdav et al., 2014). Multiple sequence alignment was performed using MAFFT v.7 server (https://mafft.cbrc.jp/alignment/server/) with default parameters (Katoh et al., 2019). Multiple sequence alignment editing was performed using Jalview software v.2.11.1.3 (Waterhouse et al., 2009). Protein structure homology modeling was performed using the SWISS-MODEL server (https://swissmodel.expasy.org/)(Waterhouse et al., 2018). Structural figures were prepared with PyMOL Molecular Graphics System, Version 2.0.6 Schrödinger, LLC.

2.4 Transcript abundance analyses

GOLPH3 transcript abundance was measured in chicken (*Gallus gallus*) and mallard (*Anas platyrhynchos*). We collected three RNASeq libraries from brain, heart, intestine, kidney, liver, lung, spleen, ovary, and testis from each species gathered from the NCBI Short Read Archive (SRA) (Leinonen et al., 2011). Accession numbers can be found in Supplemental Table S2. Reference transcript sequences were collected from Ensembl v.102 (Yates et al., 2020), and we only included the longest transcript for each gene. For each library, adapters were removed using Trimmomatic 0.38 (Bolger et al., 2014), and reads were filtered for quality using the parameters HEADCROP:5, SLIDINGWINDOW:5:30, and MINLEN:50. We mapped quality filtered paired-end RNAseq reads back to reference sequences using Bowtie 1.2.2 (Langmead et al., 2009) and default parameters of RSEM (Li and Dewey, 2011). Normalization of raw read counts for each species was performed using the estimateSizeFactors and estimateDispersions functions in DESeq2 v1.26 (Love et al., 2014).

3 Results and Discussion

Understanding the evolution of gene families is an active research area in evolutionary biology. Reconstructing the sister group relationship among gene family members represents a fundamental piece of information to understand homologous relationships, duplicative history, and their modes of evolution (Nei and Rooney, 2005; Glover et al., 2019). Further, the availability of whole-genome sequences opens an opportunity to discover new genes (Himmel et al., 2020; Opazo et al., 2020). Today, the GOLPH3 gene family is understood as a group of genes containing two paralogs (GOLPH3 and GOLPH3L) with 1:1 orthologs among most vertebrate species (Wu et al., 2000; Bell et al., 2001). However, our comparative analyses suggest that its evolutionary history is more complicated than previously thought. The evolution of the GOLPH3 paralog in birds is characterized by independent duplication events, whereas the evolutionary history of GOLPH3L seems to be less dynamic, being possible to identify 1:1 orthologs in all bird species.

3.1 Independent duplication events characterize the evolution of the GOLPH3 gene family member in birds

Our gene tree did not significantly deviate from the most updated phylogenetic hypotheses for the main group of birds (Prum et al., 2015; Kimball et al., 2019; Feng et al., 2020; Kuhl et al., 2021)(Fig. 1), suggesting that GOLPH3 was present in the ancestor of the group as a single copy gene. We recovered a clade containing GOLPH3 sequences from paleognaths (ostriches, tinamous, and allies) sister to GOLPH3 sequences from all other birds (Fig. 1). Further, we recovered the sister group relationship of the GOLPH3 sequences from Galliformes (chickens, pheasants, and allies) and Anseriformes (ducks, swans, and allies), in turn, this clade was recovered sister to GOLPH3 sequences from Neoaves (Fig. 1).

We found a single copy gene, located on chromosome Z, in most paleognaths species, except in the white-throated tinamou (*Tinamus guttatus*), where two copies were identified (Fig. 1), suggesting that this species independently gave rise to a second GOLPH3 copy located on chromosome W. The location of these genes on sexual chromosomes, and given the sex-determination system of birds (Chue and Smith, 2011), indicates that only females (ZW) can express both paralogs. In the case of Neoaves, our tree topology is not well resolved, being difficult to anticipate details regarding the duplicative history of the GOLPH3 paralog in this group (Fig. 1). However, for a diversity of species (e.g., zebra finch, common canary, kakapo), we found duplicated copies on different chromosomes, suggesting that the duplication event that gave rise to them occurred in the ancestor of the group. Similar to the case of the white-throated tinamou (*Tinamus guttatus*), we found duplicated copies in the killdeer (*Charadrius vociferus*) that were recovered sister to each other (Fig. 1), suggesting that they arouse as a product of a species-specific gene duplication event.

The evolutionary history of the GOLPH3 gene in the clade that includes Galliformes and Anseriformes followed a more complicated evolutionary pathway (Fig. 1). According to our assessment, we found a repertoire of two copies in species belonging to both groups (Fig. 1); however, our gene tree suggests that the events that gave rise to them followed a pattern of gene birth-and-death (Hughes and Nei, 1989)(Fig. 2). The reconciliation of the gene tree with the species tree suggests that the last common ancestor of Anseriformes and Galliformes, which lived 80 million years ago approximately (Hedges et al., 2015), had a single copy gene that underwent a duplication event (Fig. 2), giving rise to a repertoire of two GOLPH3 copies (Fig. 2). One of the copies was retained in Galliformes (Fig. 2; GOLPH3.1$_{GA}$; red lineage), but lost in Anseriformes (Fig. 2; red lineage). This GOLPH3 gene copy is located on chromosome W. The other copy, also originated in the ancestor of Galliformes and Anseriformes, was also retained in Galliformes (Fig. 3; GOLPH3.2$_{GA}$; pink lineage) and is located on chromosome Z. In the last common ancestor of Anseriformes, this copy underwent a duplication event giving rise to two copies (Fig. 3; GOLPH3.2.1$_{A}$, purple lineage and GOLPH3.2.2$_{A}$, light purple lineage). Like other cases, in Anseriformes, one of the copies is located on chromosome Z, while the other on chromosome W; the exception is the mallard (*Anas platyrhynchos*) in which is found on chromosome 22.

Thus, the main groups of birds possess gene repertoires with different evolutionary origins (Fig. 2). Most paleognaths retained the ancestral condition of a single gene copy, whereas Galliformes, Anseriformes, and Neoaves possess duplicated copies that were originated independently (Fig. 2). Anseriformes and Neoaves gave rise to their repertoire in the ancestor of each group (Fig. 2), while Galliformes retained copies originated in the ancestor of Galliformes and Anseriformes (Fig. 2). The independent origin of gene families in different groups is not an unusual event during the evolutionary process (Goodman et al., 1987; Kriener et al., 2001;

Hoffmann et al., 2010; Opazo and Zavala, 2018; Gramzow et al., 2020); however, it should be taken into account when making comparisons because non-orthologous genes - i.e., genes with different evolutionary origin - are being compared. Our results also highlight the importance of manual curation in defining the composition of gene families. The description of new genes is also not uncommon (Wichmann et al., 2016; Himmel et al., 2020; Opazo et al., 2020), and their discovery could be attributed to their presence in non-model species and/or the absence of appropriate evolutionary analyses. The presence of species with different gene repertoires represents an opportunity to understand the evolutionary fate of duplicated genes (Lynch and Conery, 2000) and the biological functions associated with a group of genes. This phenomenon, variation in gene copy number, has been associated with differences in susceptibility to diseases in different taxonomic groups. For example, in the African elephant (*Loxodonta africana*), it has been claimed that an expansion in the number of TP53 "the guardian of the genome" gene copies could help to explain the lower risk of developing cancer in this large and long-lived animal (Abegglen et al., 2015; Sulak et al., 2016). Similarly, in whales there are also an expansion of gene families related to cancer, and an accelerated rate of evolution in genomic regions enriched with pathways involved in cancer (Tollis et al., 2019; Tejada-Martinez et al., 2021). Further evidence comes from bats, a group in which the lifespan exceeds the expectation based on their body size (de Magalhães and Costa, 2009). In this group, it has been documented the expansion of several genes, for example, FBXO31, which is related to cell cycle arrest and response to DNA damage diminishing the probability of developing cancer (Santra et al., 2009; Seim et al., 2013; Zhang et al., 2013). Thus, the expanded repertoire of GOLPH3 genes in birds could be part of a set of genomic traits that account for their lower susceptibility to cancer than mammals.

3.2 Molecular divergence between duplicated copies of GOLPH3 genes

According to our analyses, the divergence values between duplicated GOLPH3 copies are low. In Galliformes, the divergence values ranged from 1.79% to 4.76%. However, by checking the amino acid alignment, we realized that most of the observed differences are in the first ~50 amino acids of the N-terminal region of the protein (Fig. 3). By estimating amino acid sequence divergence for the N- and C- terminal regions separately, we observed that the divergence values for the C-terminal region ranged from 1.79% to 2.47%, while for the N-terminal region, which represents only ~1/6 of the amino acid sequence, ranged from 15.69% to 17.65%. In this group of birds we found eleven amino acid positions in the alignment that unequivocally distinguish between both paralogs (Fig. 3). Six of them are in the N-terminal region, while the others in the C-terminal portion of the molecule (Fig. 3). In Anseriformes (Fig. 4), the divergence values range from 1.69% to 4.49%, similar to those estimated for Galliformes. Also, most of the observed differences are in the N-terminal region of the molecule with divergence values ranging from 5.66% to 9.62%. In the case of the C-terminal region, the values varied from 0.8% to 2.9%. In this group of birds there are three amino acid positions that unequivocally distinguish both paralogs (Fig. 4), one of them is located on the N-terminal region of the molecule, whereas the other two in the C-terminal region (Fig. 4). In Neoaves we found the same evolutionary pattern as described for Anseriformes and Galliformes, i.e. most of the amino acid replacements are found in the N-terminal portion of the protein. The divergence values for the whole protein ranged from 1.36% to 3.72%, for the N-terminal region varies from 3.92% to 15.38%, whereas for the C-terminal region goes from 0.41% to 2.06%. Unlike the previous cases we did not identify amino acid sites in the alignment that distinguish both paralogs (Fig. 5).

These differences in amino acid sequence divergence for the N- and C- terminal regions of the GOLPH3 paralogs could have arisen as a consequence of different structural and functional constraints during its evolution. Secondary structure prediction indicates that the region comprising the first ~40-60 amino acids of GOLPH3 is disordered in a variety of organisms (e.g., yeast, fruit fly, spotted gar, human) that share a common ancestor more than a billion of years ago (Fig. 6A and Supplementary Figure 1). Accordingly, only the crystal structures of N-terminal truncation variants of GOLPH3 and Vps74 (GOLPH3 in yeasts) have been solved (Schmitz et al., 2008; Wood et al., 2009). Both structures are remarkably similar (backbone atom root mean square deviation of ~1.0 Å), consisting of a single globular domain that is predominantly α-helical, with a central four-helix bundle surrounded by solvent-exposed loops, and eight amphipathic helices (Schmitz et al., 2008; Wood et al., 2009). The overall structure of the N-terminal truncated GOLPH3 protein is unique, with no strong structural homology to known protein folds, resulting so far challenging to predict its function based on its structure. Protein structure homology modeling of GOLPH3.1$_{GA}$ and GOLPH3.2$_{GA}$ of chicken (Supplementary Figure 2) showed that of the divergent amino acids in the C-terminal region only L255 in GOLPH3.1$_{GA}$ and Q256 in GOLPH3.2$_{GA}$ are non-conservative (Fig. 3) (Barnes and Gray, 2003). The position of Q256 is predicted to be exposed at the surface of GOLPH3.2$_{GA}$, like it is for Q260 in human GOLPH3. However, it is intriguing the variant L255 in GOLPH3.1$_{GA}$ because the preferred position of leucine residues is buried in protein hydrophobic cores. None of the divergent amino acids of the C-terminal region in Anseriformes and Neoaves are structurally disfavored. The C-terminal region of GOLPH3 is sufficient for GOLPH3 physical interaction with the membrane of the Golgi apparatus (Wu et al., 2000). This interaction is mediated by a series of highly conserved residues that are postulated to interact with phosphates groups and the inositol ring of phosphatidylinositol 4-phosphate located in the cytosolic leaflet of the Golgi membrane (Wood et al., 2009), set of residues that are also conserved in both copies of GOLPH3 in birds (Figs. 3-5 and Supplementary Figures 1 and 3). In contrast, the N-terminal disordered region has no known function. Intriguingly, some proteins containing disordered regions have the capacity to undergo liquid-liquid phase separation that could result in their partitioning in functional biomolecular condensates also known as membrane-less compartments (Wu et al., 2000; Borcherds et al., 2020). However, it is unknown whether GOLPH3 has this capacity. In any case, the distinct amino acid sequence divergence values for the N-terminal disordered region of GOLPH3 suggest a much flexible functional role. Thus, it will be important to determine whether this domain contributes in the functions of GOLPH3 as oncoprotein.

3.3 Evolution of GOLPH3L paralog

In contrast to GOLPH3, GOLPH3L is largely uncharacterized. Although the amino acid sequences of human GOLPH3 and GOLPH3L are 78% similar (65% identical), it has been suggested that GOLPH3L antagonizes the functions of GOLPH3 (Ng et al., 2013). Despite this, other reports suggest a similar function to GOLPH3 for GOLPH3L in some types of cancer (Kunigou et al., 2011; Sotgia et al., 2012; Feng et al., 2015, 2017; He et al., 2017). The evolutionary history of GOLPH3L followed a different trajectory in comparison to the GOLPH3 gene (Fig. 7). In this case, our gene tree recovered the main groups of birds according to the most updated organismal phylogenies (Prum et al., 2015; Kimball et al., 2019; Feng et al., 2020; Kuhl et al., 2021); nevertheless, it was not possible to define the relationships among them (Fig. 7). We will assume that the lack of variability does not allow us to resolve their sister-group relationship, instead of more complex evolutionary scenarios invoking gene duplications and reciprocal loss in the

ancestor of the main groups of birds. Thus, according to our results the GOLPH3L gene was present in the ancestor of birds as a single copy gene (Fig. 2), and this gene was inherited by all descendant lineages (Fig. 2). Thus, GOLPH3L genes in different bird species are 1:1 orthologs. Amino acid divergence values show a similar trend as we described for GOLPH3, i.e., the N-terminal portion of the protein is more divergent than the C-terminal region (Fig. 8). In the case of Galliformes, the divergence values for the N-terminal part of the molecule ranged from 5.71% to 17.14%, while for the C-terminal region varies from 3.28% to 6.97%. In Anseriformes, the values for the N-terminal region ranged from 7.89% to 26.32%, whereas for the C-terminal portion varies from 2.05% to 4.92%. In the case of Neoves, the evolutionary trend is the same, although the values for the N-terminal region are higher. Thus, the values for the N-terminal part of the molecule ranged from 20% to 43.34%, while for the C-terminal region varied from 4.10% to 17.55%. The secondary structure prediction indicates that this region in GOLPH3L, although shorter than in GOLPH3, is also disordered (Fig. 6B).

One thing that seems interesting is the number of changes accumulated in the branch leading to Galliformes and to manakins (Fig. 7). This phenomenon could be indicative of an acceleration of the rate of fixation of amino acid changing mutations in the ancestors of both groups. To test this hypothesis, we estimated the omega value (d_N/d_S), i.e., the ratio of non-synonymous (d_N) to synonymous substitutions (d_S), in the branches leading to both groups. In brief, if non-synonymous substitutions are neutral, then the rate of fixation of d_N and d_S will be very similar, and $d_N/d_S \approx 1$. Under negative selection, most non-synonymous substitutions are deleterious, and $d_N/d_S < 1$. Finally, under positive selection non-synonymous (d_N) replacements are advantageous and will be fixed at a greater rate than synonymous substitutions (d_S) and in consequence $d_N/d_S > 1$ (Yang and Bielawski, 2000). According to our analyses, in the ancestor of Galliformes the model in which the omega value was estimated from the data was not significantly different from the model in which the omega value was fixed to 1 (neutral evolution)(LRT=0.142, $P > 0.05$). On the other hand, in the case of manakins the model in which the omega value was estimated from the data (d_N/d_S=6.2) was significantly different from the null hypothesis of neutral evolution (LRT=7.19, $P < 0.01$), indicating that the rate of fixation of non-synonymous substitutions (d_N) is higher in comparison to the neutral expectation (d_S) and suggesting an event of positive selection in the ancestor of manakins. According to the Bayes Empirical Bayes (BEB) approach five sites (152C/G, 191R, 230A, 233R and 263G) were inferred under positive selection with a posterior probability higher than 0.95. All of them are located in the C-terminal region of the protein. Given the limited understanding of the biological functions associated with the GOLPH3 gene family, in particular of the GOLPH3L gene (Sechi et al., 2015), it is challenging to explain the consequences of positive selection in GOLPH3L in this group of birds. However, it could be interesting to carry out functional assays in which the performance of manakins GOLPH3L protein is compared to the one in other birds.

3.4 Expression pattern of GOLPH3 gene family members

Our next step was to investigate the expression pattern of the GOLPH3 gene family members, especially for the duplicated copies derived independently in different groups of birds. To do this, we mapped RNASeq reads to reference gene sequences in the chicken (*Gallus gallus*) and mallard (*Anas platyrhynchos*) and examined transcript abundance in a panel of nine tissues (Fig. 9). It is important to say that in chicken, the duplicated GOLPH3 copies are located on chromosome Z ($GOLPH3.2_{GA}$) and W ($GOLPH3.1_{GA}$), while GOLPH3L is on chromosome 25. Therefore, females (ZW) can express all gene family members, whereas males (ZZ) can only express

370 GOLPH3.2$_{GA}$ and GOLPH3L. This situation is somewhat similar to the allelic trichromacy
371 observed in New World monkeys, where some females (XX) possess trichromatic color vision due
372 to a polymorphism of an opsin gene located on chromosome X, while males (XY) are all
373 dichromatic (Jacobs, 1996). The case is different in mallard, as one copy is located on chromosome
374 Z (GOLPH3.2.1$_{A}$) but GOLPH3.2.2$_{A}$ and GOLPH3L are autosomal genes, so both sexes can
375 potentially express all paralogs.
376 In both species, the paralog located on the Z chromosome (GOLPH3.2$_{GA}$ in chicken and
377 GOLPH3.2.1$_{A}$ in mallard) was highly and ubiquitously expressed across all tissues (Fig. 9). We
378 collected chicken libraries from both male and female tissues, and unfortunately the sex of the
379 individuals for some tissues was not declared (Supplemental Table S2). As such GOLPH3.1$_{GA}$
380 exhibited variable expression from mixed sex sampling (Fig 9A). Thus, GOLPH3.1$_{GA}$ was highly
381 expressed in the brain and ovary where all libraries were constructed from female individuals (Fig
382 9A). Although we do not know the sex of the individuals for the intestine libraries of the chicken,
383 based on the consistent expression of the gene located on chromosome W, we can presume that
384 they were all from female individuals (Fig. 9A). As a validation of what we mentioned above,
385 GOLPH3.1$_{GA}$ was not expressed in all known male tissue libraries, which was most noticeable in
386 the male specific testes (Fig. 9A). All autosomal paralogs in both species were variably expressed
387 among and within tissues (Fig 9). In the case of the chicken GOLPH3L, we recovered expression
388 in all tissues but at variable levels, from low values in the brain and heart to higher values in the
389 intestine and kidney (Fig. 9A). By contrast GOLPH3L was universally expressed in all mallard
390 tissues (Fig. 9B). GOLPH3.2.2$_{A}$ in mallard was not expressed in the brain, liver and testes, but
391 highly expressed in the ovary, spleen and intestine. In both species, all GOLPH3 paralogs were
392 consistently expressed in the intestine (Fig. 9).
393 In humans, both paralogs are ubiquitously expressed across all tissues (Uhlén et al., 2015),
394 suggesting that they are required for the maintenance of basic cellular functions (Eisenberg and
395 Levanon, 2013); however, GOLPH3L seems to be expressed at lower levels. Similarly, the relative
396 expression levels of GOLPH3 and GOLPH3L in several mammalian cell lines with epithelial,
397 fibroblast, myeloid and neuronal characteristics, and in a variety of tissues from mice also indicates
398 that GOLPH3 is also ubiquitously expressed at higher levels than GOLPH3L (Ng et al., 2013).
399 Further, GOLPH3L is expressed more in cells with secretory epithelial characteristics (Ng et al.,
400 2013), suggesting a distinct function for this gene family member. The expression pattern observed
401 in the mallard is similar to what is observed in model species (Fig. 9B). One of the GOLPH3
402 duplicates (GOLPH3.2.1$_{A}$) is expressed in all examined tissues at high levels, while GOLPH3L is
403 also expressed in all tissues, but at lower levels. The expression of the other duplicate
404 (GOLPH3.2.2$_{A}$) is variable, including tissues in which it is not detected (Fig. 9). The case of the
405 chicken seems to be more dissimilar. In this species one of the GOLPH3 duplicates (GOLPH3.2$_{GA}$)
406 possesses an expression pattern similar to the human GOLPH3 (Fig. 9A), however the other two
407 copies seem to follow a specific expression pattern (Fig. 9A).
408
409 **4 Conclusions**
410 Our study shows that the evolution of the GOLPH3 gene family followed a more
411 complicated evolutionary pathway than previously thought. Although the history of the GOLPH3L
412 paralog is according to the current knowledge, the one of GOLPH3 is not. The most exciting thing
413 about the evolution of GOLPH3 in birds is that they possess extra GOLPH3 gene copies never
414 described before, and that all main groups independently originated their repertoire. In other
415 words, they do not have the same evolutionary origin, and in consequence, they are not 1:1

orthologs and are not directly comparable. Most paleognaths retained the ancestral condition of a single gene copy, whereas Galliformes, Anseriformes, and Neoaves possess duplicated copies originated independently. Thus, birds represent a natural experiment of gene copy number variation (Albertson et al., 2009), that in addition to the differences in expression of individuals of different sex, could help us improve our understanding of the biological functions associated with the GOLPH3 gene family. Our results also highlight the power of manually curating genetic data to define gene repertoires, and the reconciliation of gene trees with species trees (Goodman et al., 1979) to understand the duplicative history of gene families to perform biologically meaningful comparisons (Gabaldón, 2008; Glover et al., 2019). Finally, the conservation of the N-terminal portion of GOLPH3 paralogs as a disordered region for more than a billion of years of evolution and the fact that it displays a higher degree of divergence among species, compared to the C-terminal portion, strongly suggests that it performs an essential, specialized and adapted cellular function conserved in distantly related species like yeasts and humans that remains to be elucidated.

Conflict of Interest

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

Author contributions

JCO and GAM designed the study. MWV, JG, KZ, JCO, GAM collected and/or analyzed data. JCO and GAM wrote the manuscript. MWV, LV-C, FJM reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 1210471) and Millennium Nucleus of Ion Channels Associated Diseases (MiNICAD), Iniciativa Científica Milenio, Ministry of Economy, Development and Tourism from Chile to JCO, the US Dept. of Education HSI-STEM grant P031C110114-15 to MWV, Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 1180957) to FJM and LVC, Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 1211481) to GAM and Vicerrectoría de Investigación, Desarrollo y Creación Artística of Universidad Austral de Chile (JCO, FJM, LVC, GAM).

Supplementary material

The supplementary material for this article can be found online at:

454 **Figure legends**
455 **Figure 1.** Maximum likelihood tree showing sister group relationships among GOLPH3 genes of
456 birds. Numbers above the nodes correspond to support values from the ultrafast bootstrap routine.
457 GOLPH3 sequences from crocodiles and turtles were used as outgroups (not shown). The scale
458 denotes substitutions per site and colors represent gene lineages.
459
460 **Figure 2.** An evolutionary hypothesis regarding the evolution of the Golgi phosphoprotein 3
461 (GOLPH3) gene family in birds. According to our results, the last common ancestor of birds had
462 a repertoire of two GOLPH3 genes (GOLPH3 and GOLPH3L). The GOLPH3L gene was inherited
463 by all species of birds and maintained as a single copy gene (brown lineage). The evolution of the
464 GOLPH3 gene is more complicated. The gene present in the bird ancestor was inherited by
465 paleognaths (ostriches, tinamous and allies) and maintained as a single copy gene in most species.
466 In the case of Neoaves (zebra finches, kakapos and allies), the ancestral GOLPH3 gene underwent
467 a duplication event in the ancestor of the group giving rise to a repertoire of two genes (blue and
468 light blue gene lineages). In the last common ancestor of Anseriformes (ducks, swans and allies)
469 and Galliformes (chickens, pheasants and allies) the ancestral GOLPH3 gene underwent a
470 duplication event giving rise to a repertoire of two genes, one of the copies was retained in
471 Galliformes (red lineage), but lost in Anseriformes (red lineage). The other copy originated in the
472 ancestor of Galliformes and Anseriformes, was also retained in Galliformes (pink lineage),
473 however, in the last common ancestor of Anseriformes underwent a duplication event giving rise
474 to the repertoire of two copies (purple and light purple lineages). We use the subscripts to indicate
475 the ancestor in which the gene repertoires were originated, N = the ancestor of Neoaves, A = the
476 ancestor of Anseriformes and GA = the galliform/anseriform ancestor. Silhouette images were
477 obtained from PhyloPic (http://phylopic.org/).
478
479 **Figure 3.** Alignment of Golgi phosphoprotein 3 (GOLPH3) amino acid sequences from chicken
480 (*Gallus gallus*), ring-necked pheasant (*Phasianus colchicus)*, turkey (*Meleagris gallopavo*) and
481 japanese quail (*Coturnix japonica*). Amino acids positions that differentiate between paralogs are
482 in pink and red. The N-terminal region of the protein is shaded.
483
484 **Figure 4.** Alignment of Golgi phosphoprotein 3 (GOLPH3) amino acid sequences from mallard
485 (*Anas platyrhynchos*), Tufted duck (*Aythya fuligula)*, swan goose (*Anser cygnoides*) and ruddy
486 duck (*Oxyura jamaicensis*). Amino acids positions that differentiate between paralogs are in purple
487 and light purple. The N-terminal region of the protein is shaded.
488
489 **Figure 5.** Alignment of Golgi phosphoprotein 3 (GOLPH3) amino acid sequences from lance
490 tailed manakin (*Chiroxiphia lanceolata*), kakapo (*Strigops habroptilus*), New caledonian crow
491 (*Corvus moneduloides*) and zebra finch (*Taeniopygia guttata*). The N-terminal region of the
492 protein is shaded.
493
494 **Figure 6.** Secondary structure and disordered region predictions of GOLPH3 (A) and GOLPH3L
495 (B) in human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), Western clawed
496 frog (*Xenopus tropicalis*), coelacanth (*Latimeria chalumnae*), spotted gar (*Lepisosteus oculatus*),
497 thorny skate (*Amblyraja radiata*), fruit fly (*Drosophila melanogaster*), roundworm
498 (*Caenorhabditis elegans*) and brewer yeast (*Saccharomyces cerevisiae*).
499

Figure 7. Maximum likelihood tree showing sister group relationships among GOLPH3L genes of birds. Numbers above the nodes correspond to support values from the ultrafast bootstrap routine. GOLPH3L sequences from crocodiles and turtles were used as outgroups (not shown). The scale denotes substitutions per site.

Figure 8. Alignment of Golgi phosphoprotein 3L (GOLPH3L) amino acid sequences from mallard (*Anas platyrhynchos*), Tufted duck (*Aythya fuligula)*, black swan (*Cygnus atratus*), ruddy duck (*Oxyura jamaicensis*), chicken (*Gallus gallus*), ring-necked pheasant (*Phasianus colchicus)*, turkey (*Meleagris gallopavo*), japanese quail (*Coturnix japonica*), lance tailed manakin (*Chiroxiphia lanceolata*), kakapo (*Strigops habroptilus*), New caledonian crow (*Corvus moneduloides*) and zebra finch (*Taeniopygia guttata*). The N-terminal region of the protein is shaded.

Figure 9. Transcript abundance measurements of GOLPH3 paralogs across a panel of nine tissues in A) chicken (*Gallus gallus*) and B) mallard (*Anas platyrhynchos*). The mean and standard deviation are plotted for three replicates per tissue. In many cases the standard deviation was smaller than the size of the circle representing the mean and appears hidden.

520 **References**

521 Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., et al. (2015).
522 Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular
523 Response to DNA Damage in Humans. *JAMA* 314, 1850–1860.

524 Albertson, R. C., Cresko, W., Detrich, H. W., 3rd, and Postlethwait, J. H. (2009). Evolutionary
525 mutant models for human disease. *Trends Genet.* 25, 74–81.

526 Albuquerque, T. A. F., Drummond do Val, L., Doherty, A., and de Magalhães, J. P. (2018).
527 From humans to hydra: patterns of cancer across the tree of life. *Biol. Rev. Camb. Philos.*
528 *Soc.* 93, 1715–1734.

529 Ali, M. F., Chachadi, V. B., Petrosyan, A., and Cheng, P.-W. (2012). Golgi phosphoprotein 3
530 determines cell binding properties under dynamic flow by controlling Golgi localization of
531 core 2 N-acetylglucosaminyltransferase 1. *J. Biol. Chem.* 287, 39564–39577.

532 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local
533 alignment search tool. *J. Mol. Biol.* 215, 403–410.

534 Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., and Gascuel, O. (2011). Survey of
535 branch support methods demonstrates accuracy, power, and robustness of fast likelihood-
536 based approximation schemes. *Syst. Biol.* 60, 685–699.

537 Arriagada, C., Luchsinger, C., González, A. E., Schwenke, T., Arriagada, G., Folch, H., et al.
538 (2019). The knocking down of the oncoprotein Golgi phosphoprotein 3 in T98G cells of
539 glioblastoma multiforme disrupts cell migration by affecting focal adhesion dynamics in a
540 focal adhesion kinase-dependent manner. *PLoS One* 14, e0212321.

541 Barnes, M. R., and Gray, I. C. (2003). *Bioinformatics for Geneticists*. Wiley.

542 Bell, A. W., Ward, M. A., Blackstock, W. P., Freeman, H. N., Choudhary, J. S., Lewis, A. P., et
543 al. (2001). Proteomics characterization of abundant Golgi membrane proteins. *J. Biol.*
544 *Chem.* 276, 5152–5165.

545 Boddy, A. M., Harrison, T. M., and Abegglen, L. M. (2020). Comparative Oncology: New
546 Insights into an Ancient Disease. *iScience* 23, 101373.

547 Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
548 sequence data. *Bioinformatics* 30, 2114–2120.

549 Borcherds, W., Bremer, A., Borgia, M. B., and Mittag, T. (2020). How do intrinsically
550 disordered protein regions encode a driving force for liquid-liquid phase separation? *Curr.*
551 *Opin. Struct. Biol.* 67, 41–50.

552 Chue, J., and Smith, C. A. (2011). Sex determination and sexual differentiation in the avian
553 model. *FEBS J.* 278, 1027–1034.

554 de Magalhães, J. P., and Costa, J. (2009). A database of vertebrate longevity records and their
555 relation to other life-history traits. *J. Evol. Biol.* 22, 1770–1774.

556 Effron, M., Griner, L., and Benirschke, K. (1977). Nature and rate of neoplasia found in captive
557 wild mammals, birds, and reptiles at necropsy. *J. Natl. Cancer Inst.* 59, 185–198.

558 Eisenberg, E., and Levanon, E. Y. (2013). Human housekeeping genes, revisited. *Trends Genet.*
559 29, 569–574.

560 Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., et al. (2020). Dense
561 sampling of bird diversity increases power of comparative genomics. *Nature* 587, 252–257.

562 Feng, Y., He, F., Wu, H., Huang, H., Zhang, L., Han, X., et al. (2015). GOLPH3L is a Novel
563 Prognostic Biomarker for Epithelial Ovarian Cancer. *J. Cancer* 6, 893–900.

564 Feng, Y., He, F., Yan, S., Huang, H., Huang, Q., Deng, T., et al. (2017). The Role of GOLPH3L
565 in the Prognosis and NACT response in Cervical Cancer. *J. Cancer* 8, 443–454.

566 Gabaldón, T. (2008). Large-scale assignment of orthology: back to phylogenetics? *Genome*
567 *Biology* 9, 235. doi:10.1186/gb-2008-9-10-235.

568 Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., et al. (2010). The NCBI
569 BioSystems database. *Nucleic Acids Res.* 38, D492–6.

570 Glover, N., Dessimoz, C., Ebersberger, I., Forslund, S. K., Gabaldón, T., Huerta-Cepas, J., et al.
571 (2019). Advances and Applications in the Quest for Orthologs. *Mol. Biol. Evol.* 36, 2157–
572 2164.

573 Goldman, N., and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-
574 coding DNA sequences. *Molecular Biology and Evolution*.
575 doi:10.1093/oxfordjournals.molbev.a040153.

576 Goodman, M., Czelusniak, J., Koop, B. F., Tagle, D. A., and Slightom, J. L. (1987). Globins: a
577 case study in molecular phylogeny. *Cold Spring Harb. Symp. Quant. Biol.* 52, 875–890.

578 Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E., and Matsuda, G. (1979).
579 Fitting the Gene Lineage into its Species Lineage, a Parsimony Strategy Illustrated by
580 Cladograms Constructed from Globin Sequences. *Systematic Biology* 28, 132–163.
581 doi:10.1093/sysbio/28.2.132.

582 Gramzow, L., Lobbes, D., Innard, N., and Theißen, G. (2020). Independent origin of MIRNA
583 genes controlling homologous target genes by partial inverted duplication of antisense-
584 transcribed sequences. *Plant J.* 101, 401–419.

585 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010).
586 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
587 performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.

588 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015). Tree of life reveals
589 clock-like speciation and diversification. *Mol. Biol. Evol.* 32, 835–845.

590 He, S., Niu, G., Shang, J., Deng, Y., Wan, Z., Zhang, C., et al. (2017). The oncogenic Golgi
591 phosphoprotein 3 like overexpression is associated with cisplatin resistance in ovarian
592 carcinoma and activating the NF-κB signaling pathway. *J. Exp. Clin. Cancer Res.* 36, 137.

593 Himmel, N. J., Gray, T. R., and Cox, D. N. (2020). Phylogenetics Identifies Two Eumetazoan
594 TRPM Clades and an Eighth TRP Family, TRP Soromelastatin (TRPS). *Mol. Biol. Evol.* 37,
595 2034–2044.

596 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2:
597 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522.

598 Hochberg, M. E., and Noble, R. J. (2017). A framework for how environment contributes to
599 cancer risk. *Ecol. Lett.* 20, 117–134.

600 Hoffmann, F. G., Storz, J. F., Gorr, T. A., and Opazo, J. C. (2010). Lineage-specific patterns of
601 functional diversification in the alpha- and beta-globin gene families of tetrapod vertebrates.
602 *Mol. Biol. Evol.* 27, 1126–1138.

603 Hughes, A. L., and Nei, M. (1989). Nucleotide substitution at major histocompatibility complex
604 class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U. S. A.* 86, 958–
605 962.

606 Isaji, T., Im, S., Gu, W., Wang, Y., Hang, Q., Lu, J., et al. (2014). An oncogenic protein Golgi
607 phosphoprotein 3 up-regulates cell migration via sialylation. *J. Biol. Chem.* 289, 20694–
608 20705.

609 Jacobs, G. H. (1996). Primate photopigments and primate color vision. *Proc. Natl. Acad. Sci. U.*
610 *S. A.* 93, 577–581.

611 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017).
612 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14,
613 587–589.

614 Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence
615 alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166.

616 Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
617 improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

618 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., et al. (2019).
619 A Phylogenomic Supertree of Birds. *Diversity* 11, 109.

620 Kriener, K., O'hUigin, C., and Klein, J. (2001). Independent origin of functional MHC class II
621 genes in humans and New World monkeys. *Hum. Immunol.* 62, 1–14.

622 Kuhl, H., Frankl-Vilches, C., Bakker, A., Mayr, G., Nikolaus, G., Boerno, S. T., et al. (2021). An
623 Unbiased Molecular Approach Using 3'-UTRs Resolves the Avian Family-Level Tree of
624 Life. *Mol. Biol. Evol.* 38, 108–127.

625 Kunigou, O., Nagao, H., Kawabata, N., Ishidou, Y., Nagano, S., Maeda, S., et al. (2011). Role of
626 GOLPH3 and GOLPH3L in the proliferation of human rhabdomyosarcoma. *Oncol. Rep.* 26,
627 1337–1342.

628 Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient
629 alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

630 Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database
631 Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–21.

632 Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data
633 with or without a reference genome. *BMC Bioinformatics* 12. doi:10.1186/1471-2105-12-
634 323.

635 Lombard, L. S., and Witte, E. J. (1959). Frequency and types of tumors in mammals and birds of
636 the Philadelphia Zoological Garden. *Cancer Res.* 19, 127–141.

637 Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
638 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

639 Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes.
640 *Science* 290, 1151–1155.

641 Møller, A. P., Erritzøe, J., and Soler, J. J. (2017). Life history, immunity, Peto's paradox and
642 tumours in birds. *J. Evol. Biol.* 30, 960–967.

643 Nakashima-Kamimura, N., Asoh, S., Ishibashi, Y., Mukai, Y., Shidara, Y., Oda, H., et al. (2005).
644 MIDAS/GPP34, a nuclear gene product, regulates total mitochondrial mass in response to
645 mitochondrial dysfunction. *J. Cell Sci.* 118, 5357–5367.

646 Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene
647 families. *Annu. Rev. Genet.* 39, 121–152.

648 Ng, M. M., Dippold, H. C., Buschman, M. D., Noakes, C. J., and Field, S. J. (2013). GOLPH3L
649 antagonizes GOLPH3 to determine Golgi morphology. *Mol. Biol. Cell* 24, 796–808.

650 Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid
651 sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.

652 Opazo, J. C., and Zavala, K. (2018). Phylogenetic evidence for independent origins of GDF1 and
653 GDF3 genes in anurans and mammals. *Sci. Rep.* 8, 13595.

654 Opazo, J. C., Zavala, K., Vandewege, M. W., and Hoffmann, F. G. (2020). Phylogenetic
655 diversification of sirtuin genes with a description of a new family member. *bioRxiv*.

656 doi:10.1101/2020.07.17.209510.

657 Pang, V. F., Chang, P.-H., Wang, F.-I., Chin, S.-C., Jeng, C.-R., Liu, C.-H., et al. (2012).
658 Spontaneous neoplasms in zoo mammals, birds, and reptiles in Taiwan – a 10-year survey.
659 *Anim. Biol. Leiden Neth.* 62, 95–110.

660 Pereira, N. A., Pu, H. X., Goh, H., and Song, Z. (2014). Golgi phosphoprotein 3 mediates the
661 Golgi localization and function of protein O-linked mannose β-1,2-N-
662 acetlyglucosaminyltransferase 1. *J. Biol. Chem.* 289, 14762–14770.

663 Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., et al.
664 (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA
665 sequencing. *Nature* 526, 569–573.

666 Rahajeng, J., Kuna, R. S., Makowski, S. L., Tran, T. T. T., Buschman, M. D., Li, S., et al.
667 (2019). Efficient Golgi Forward Trafficking Requires GOLPH3-Driven, PI4P-Dependent
668 Membrane Curvature. *Dev. Cell* 50, 573–585.e5.

669 Ratcliffe, H. L. (1933). Incidence and Nature of Tumors in Captive Wild Mammals and Birds.
670 *Am. J. Cancer* 17, 116–135.

671 Santra, M. K., Wajapeyee, N., and Green, M. R. (2009). F-box protein FBXO31 mediates cyclin
672 D1 degradation to induce G1 arrest after DNA damage. *Nature* 459, 722–725.

673 Schmitz, K. R., Liu, J., Li, S., Setty, T. G., Wood, C. S., Burd, C. G., et al. (2008). Golgi
674 localization of glycosyltransferases requires a Vps74p oligomer. *Dev. Cell* 14, 523–534.

675 Sechi, S., Frappaolo, A., Belloni, G., Colotti, G., and Giansanti, M. G. (2015). The multiple
676 cellular functions of the oncoprotein Golgi phosphoprotein 3. *Oncotarget* 6, 3493–3506.

677 Seim, I., Fang, X., Xiong, Z., Lobanov, A. V., Huang, Z., Ma, S., et al. (2013). Genome analysis
678 reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. *Nat.*
679 *Commun.* 4, 1–8.

680 Sharma, S., Ciufo, S., Starchenko, E., Darji, D., Chlumsky, L., Karsch-Mizrachi, I., et al. (2018).
681 The NCBI BioCollections Database. *Database* 2018. doi:10.1093/database/bay006.

682 Sotgia, F., Whitaker-Menezes, D., Martinez-Outschoorn, U. E., Salem, A. F., Tsirigos, A., Lamb,
683 R., et al. (2012). Mitochondria "fuel" breast cancer metabolism: fifteen markers of
684 mitochondrial biogenesis label epithelial cancer cells, but are excluded from adjacent
685 stromal cells. *Cell Cycle* 11, 4390–4401.

686 Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., et al. (2016). TP53 copy
687 number expansion is associated with the evolution of increased body size and an enhanced
688 DNA damage response in elephants. doi:10.7554/eLife.11994.

689 Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein
690 sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34,

691 W609–12.

692 Tatusova, T. A., and Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing
693 protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174, 247–250.

694 Tejada-Martinez, D., de Magalhães, J. P., and Opazo, J. C. (2021). Positive selection and gene
695 duplications in tumour suppressor genes reveal clues about how cetaceans resist cancer.
696 *Cold Spring Harbor Laboratory*, 2020.01.15.908244. doi:10.1101/2020.01.15.908244.

697 Tollis, M., Robbins, J., Webb, A. E., Kuderna, L. F. K., Caulin, A. F., Garcia, J. D., et al. (2019).
698 Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an
699 Assembly for the Humpback Whale (Megaptera novaeangliae). *Mol. Biol. Evol.* 36, 1746–
700 1763.

701 Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast
702 online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–5.

703 Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al.
704 (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.

705 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018).
706 SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids*
707 *Res.* 46, W296–W303.

708 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009).
709 Jalview Version 2--a multiple sequence alignment editor and analysis workbench.
710 *Bioinformatics* 25, 1189–1191.

711 Wichmann, I. A., Zavala, K., Hoffmann, F. G., Vandewege, M. W., Corvalán, A. H., Amigo, J.
712 D., et al. (2016). Evolutionary history of the reprimo tumor suppressor gene family in
713 vertebrates with a description of a new reprimo gene lineage. *Gene* 591, 245–254.

714 Wood, C. S., Schmitz, K. R., Bessman, N. J., Setty, T. G., Ferguson, K. M., and Burd, C. G.
715 (2009). PtdIns4P recognition by Vps74/GOLPH3 links PtdIns 4-kinase signaling to
716 retrograde Golgi trafficking. *J. Cell Biol.* 187, 967–975.

717 Wu, C. C., Taylor, R. S., Lane, D. R., Ladinsky, M. S., Weisz, J. A., and Howell, K. E. (2000).
718 GMx33: a novel family of trans-Golgi proteins identified by proteomics. *Traffic* 1, 963–
719 975.

720 Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., et al. (2014).
721 PredictProtein--an open resource for online prediction of protein structural and functional
722 features. *Nucleic Acids Res.* 42, W337–43.

723 Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24,
724 1586–1591.

725 Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation.

Trends Ecol. Evol. 15, 496–503.

Yang, Z., and dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* 28, 1217–1228.

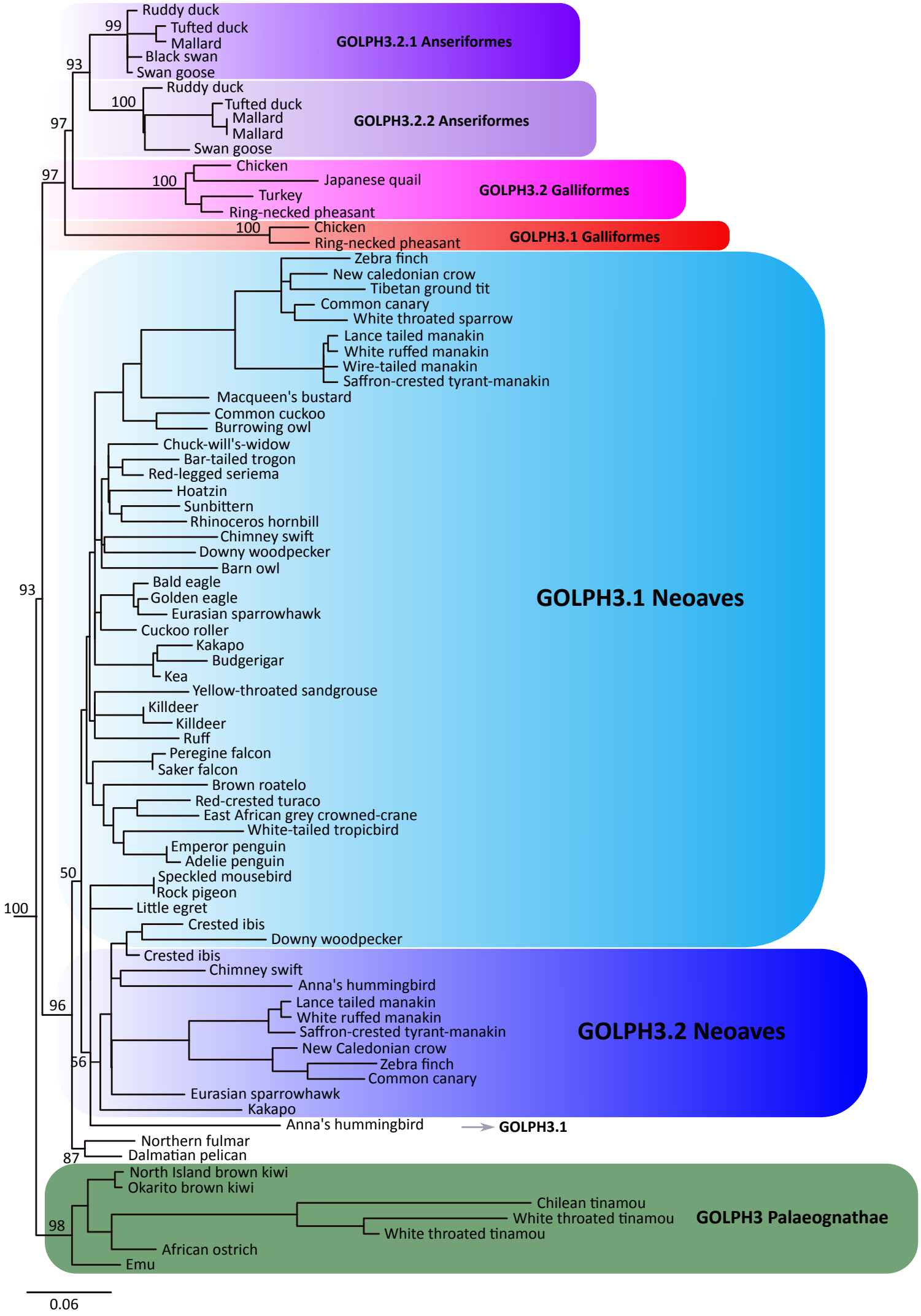
Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.

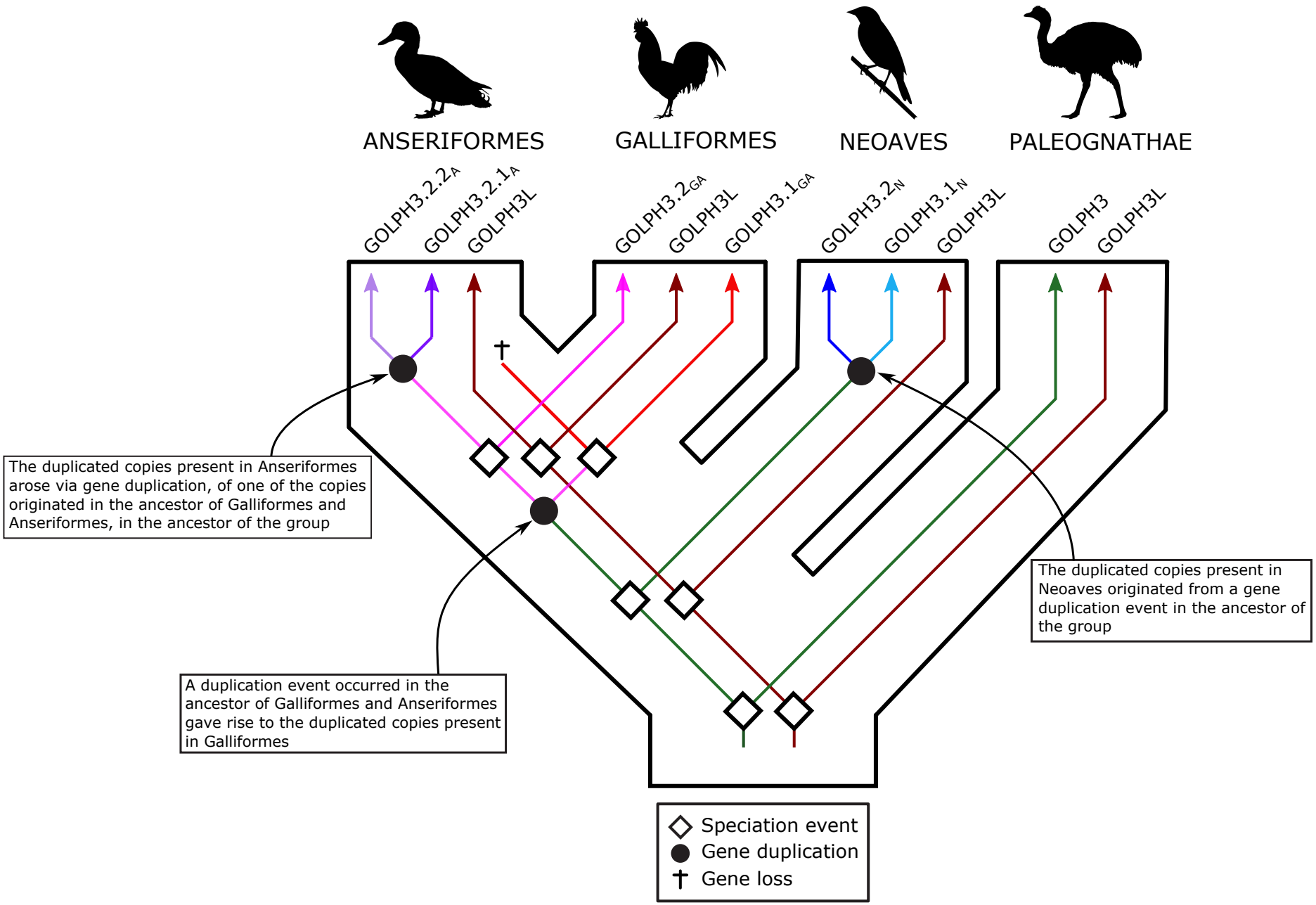
Yang, Z., Wong, W. S. W., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K. A., Fang, X., et al. (2013). Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339, 456–460.

Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479.


GOLPH3.2.1 Anseriformes
Ruddy duck
Tufted duck
Mallard
Black swan
Swan goose
99
93
GOLPH3.2.2 Anseriformes
Ruddy duck
Tufted duck
Mallard
Mallard
Swan goose
100
97
GOLPH3.2 Galliformes
Chicken
Japanese quail
Turkey
Ring-necked pheasant
100
97
GOLPH3.1 Galliformes
Chicken
Ring-necked pheasant
100
GOLPH3.1 Neoaves
Zebra finch
New caledonian crow
Tibetan ground tit
Common canary
White throated sparrow
Lance tailed manakin
White ruffed manakin
Wire-tailed manakin
Saffron-crested tyrant-manakin
Macqueen's bustard
Common cuckoo
Burrowing owl
Chuck-will's-widow
Bar-tailed trogon
Red-legged seriema
Hoatzin
Sunbittern
Rhinoceros hornbill
Chimney swift
Downy woodpecker
Barn owl
Bald eagle
Golden eagle
Eurasian sparrowhawk
Cuckoo roller
Kakapo
Budgerigar
Kea
Yellow-throated sandgrouse
Killdeer
Killdeer
Ruff
Peregine falcon
Saker falcon
Brown roatelo
Red-crested turaco
East African grey crowned-crane
White-tailed tropicbird
Emperor penguin
Adelie penguin
Speckled mousebird
Rock pigeon
Little egret
Crested ibis
Downy woodpecker
93
50
100
GOLPH3.2 Neoaves
Crested ibis
Chimney swift
Anna's hummingbird
Lance tailed manakin
White ruffed manakin
Saffron-crested tyrant-manakin
New Caledonian crow
Zebra finch
Common canary
Eurasian sparrowhawk
Kakapo
Anna's hummingbird
96
66
GOLPH3.1
Northern fulmar
Dalmatian pelican
87
GOLPH3 Palaeognathae
North Island brown kiwi
Okarito brown kiwi
Chilean tinamou
White throated tinamou
White throated tinamou
African ostrich
Emu
98
0.06


ANSERIFORMES
GALLIFORMES
NEOAVES
PALEOGNATHAE
GOLPH3.2.2A
GOLPH3.2.1A
GOLPH3L
GOLPH3.2GA
GOLPH3L
GOLPH3.1GA
GOLPH3.2N
GOLPH3.1N
GOLPH3L
GOLPH3
GOLPH3L
The duplicated copies present in Anseriformes arose via gene duplication, of one of the copies originated in the ancestor of Galliformes and Anseriformes, in the ancestor of the group
A duplication event occurred in the ancestor of Galliformes and Anseriformes gave rise to the duplicated copies present in Galliformes
The duplicated copies present in Neoaves originated from a gene duplication event in the ancestor of the group
Speciation event
Gene duplication
Gene loss

N-terminal

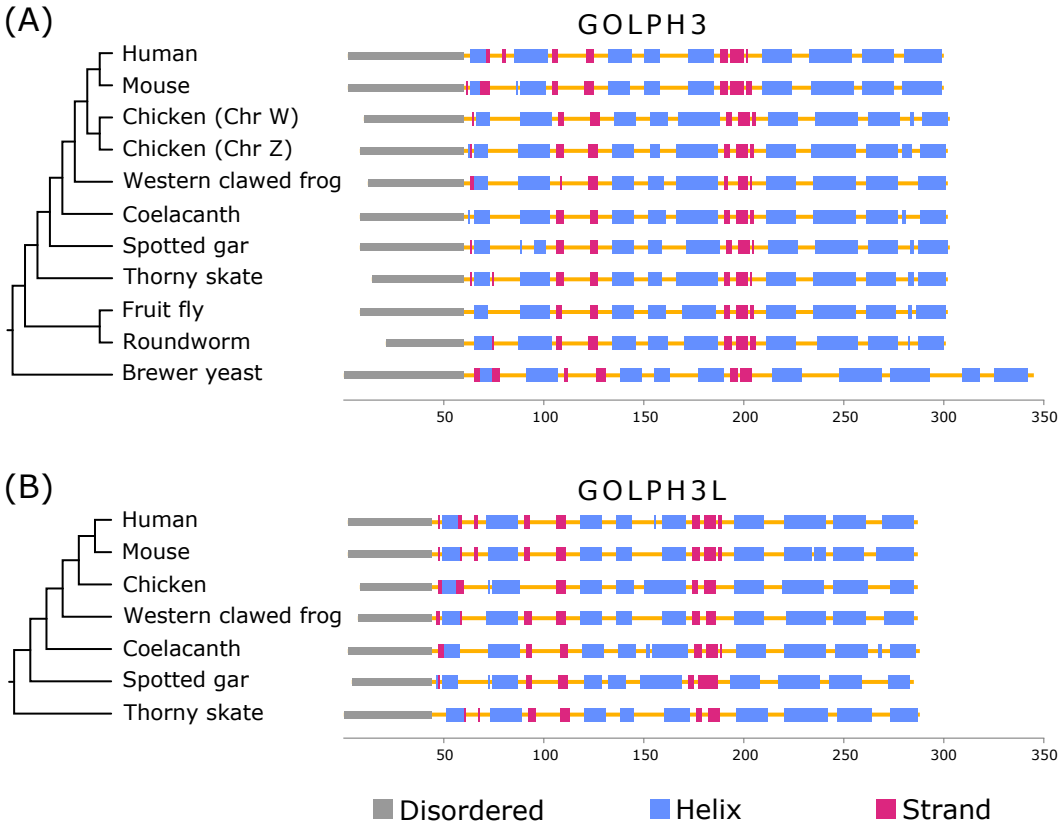
```
                                  1         10        20        30        40        50        60        70
Chicken GOLPH3.2GA                MTSLTQRSSGLVQRRTEASRSAAADKERGAGGGSEDEGRREEPGDDEKGDSKETRLTLMEEVLLLGLKDREGYT
Ring-necked pheasant GOLPH3.2GA   MTSLTQRSSGLVQRRTEASRSAAADKERGAGGGPEDEGRREEPGDDEKGDSKETRLTLMEEVLLLGLKDREGYT
Turkey GOLPH3.2GA                 -------------------------------------------------------------------GYT
Japanese quail GOLPH3.2GA         MTSLTQRSSGLVQRRTEASRSAAADKERGAGGGPEDEGRREEPGDDEKGDSKETRLTLMEEVLLLGLKDREGYT

Chicken GOLPH3.1GA                MTSLTQRCSGLVQRRTEASRS-AADEERGARGGPEDEGCRDELGDDEKGDCKQTRLTLMEEVLLLGLKDREGYT
Ring-necked pheasant GOLPH3.1GA   MTSLTQRCSGLVQRRTEASRS-AADEERGARGGLENEGRRDELGDDEKGDSKQTRLTLMEEVLLLGLKDREGYT

                                  80        90        100       110       120       130       140
Chicken GOLPH3.2GA                SFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQNW
Ring-necked pheasant GOLPH3.2GA   SFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQNW
Turkey GOLPH3.2GA                 SFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQNW
Japanese quail GOLPH3.2GA         SFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQNW

Chicken GOLPH3.1GA                SFWNDCISSGLRGCMLIELSLRGRLQLEACGMRRKSLLTRKVICKSDVPTGDVLLDEALKHIKETQPPETVQNW
Ring-necked pheasant GOLPH3.1GA   SFWNDCISSGLRGCMLIELSLRGRLQLEACGMRRKSLLTRKVICKSDVPTGDVLLDEALKHIKETQPPETIQNW

                                  150       160       170       180       190       200       210       220
Chicken GOLPH3.2GA                IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV
Ring-necked pheasant GOLPH3.2GA   IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV
Turkey GOLPH3.2GA                 IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV
Japanese quail GOLPH3.2GA         IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV

Chicken GOLPH3.1GA                IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV
Ring-necked pheasant GOLPH3.1GA   IELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDKWV

                                  230       240       250       260       270       280       290   294
Chicken GOLPH3.2GA                NDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Ring-necked pheasant GOLPH3.2GA   NDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Turkey GOLPH3.2GA                 NDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Japanese quail GOLPH3.2GA         NDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK

Chicken GOLPH3.1GA                NDPHRMDKRLLALIYLAHASDVLENAFAPLLDELYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Ring-necked pheasant GOLPH3.1GA   NDPHRMDKRLLALIYLAHASDVLENAFAPLLDELYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
```

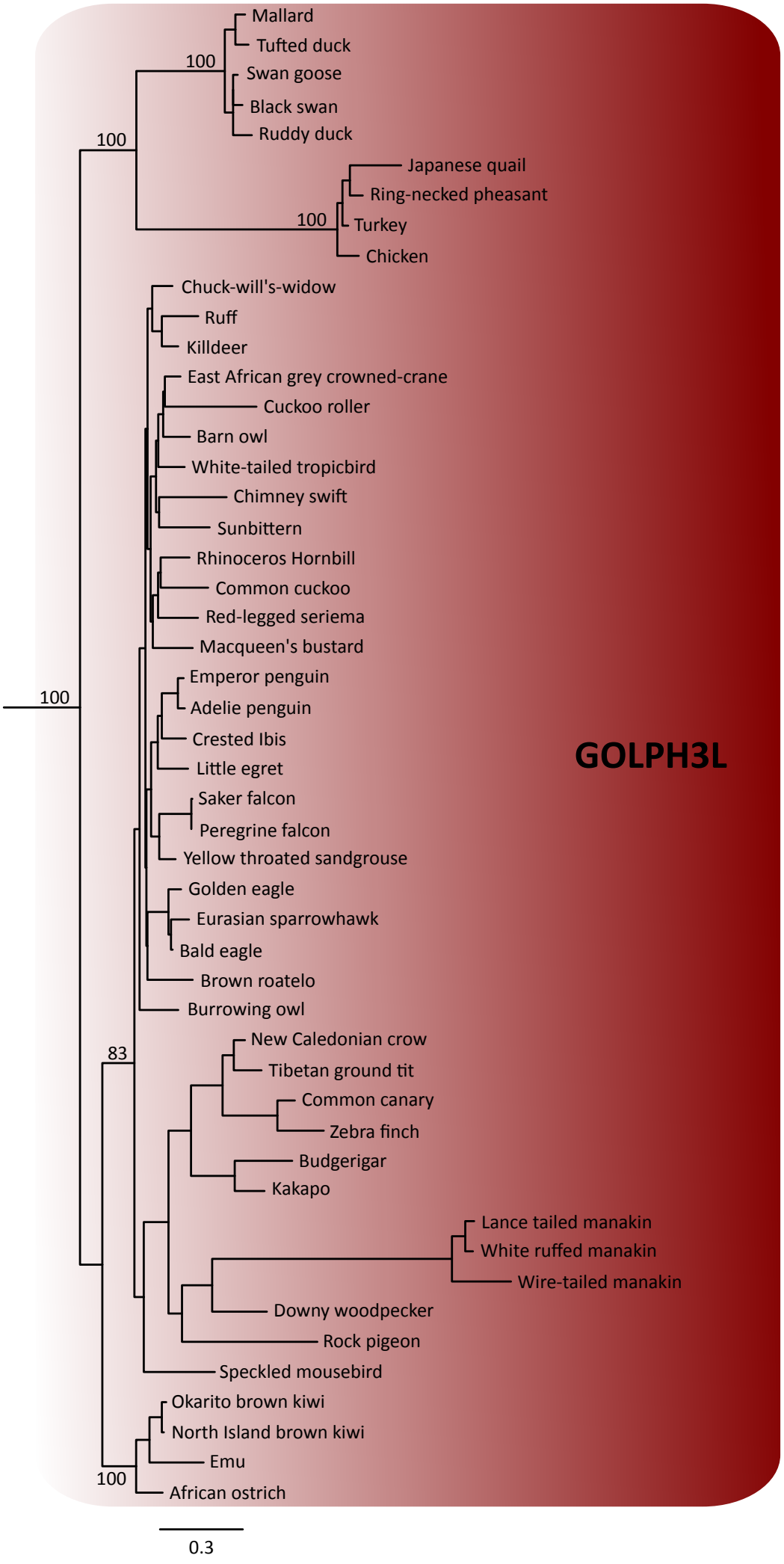
N-terminal

```
                                 1         10        20        30        40        50        60        70
Mallard GOLPH3.2.1A              MTSLTQRSSGLVQRRTEASRS-AAAADKERGAGGGPEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Tufted duck GOLPH3.2.1A          MTSLTQRSSGLVQRRTEASRS-AAAADKERGAGGGPEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Swan goose GOLPH3.2.1A           MTSLTQRSSGLVQRRTEASRS-AAAADKERGAGGGPEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Ruddy duck GOLPH3.2.1A           MTSLTQRSSGLVQRRTEASRS-AAAADKERGAGGGPEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG

Mallard GOLPH3.2.2A              MTSLTQRSSGLVQRRTEASRS-AAAADKEWGAGGGLEDEGHRDELGDDEKGDSKETRLTLMEEVLLLGLKDREG
Tufted duck GOLPH3.2.2A          MTSLTQRSSGLVQRRTEASRS-AAAADKEWGAGGGLEDEGHRDELGDDVKGDSKETRLTLMEEVLLLGLKDREG
Swan goose GOLPH3.2.2A           MTSLTQRSSGLVQRRTEASRSAAAAAAKERGTGGGLEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDRXG
Ruddy duck GOLPH3.2.2A           MTSLTQRNSGLVQRRTEASRSAAAAADKERGAGGGLEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG

                                         80        90       100       110       120       130       140
Mallard GOLPH3.2.1A              YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Tufted duck GOLPH3.2.1A          YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Swan goose GOLPH3.2.1A           YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Ruddy duck GOLPH3.2.1A           YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ

Mallard GOLPH3.2.2A              YTSFWNDCISSGLRGCMLIELALRGRLQLEACGIRRKSLLTRKVICKSDTSTGDVLLDEALKHIKETQPPETVQ
Tufted duck GOLPH3.2.2A          YTSFWNDCISSGLRGCMLIELALRGRLQLEACGIRRKSLLTRKVICKSDTSTGDVLLDEALKHIKETQPPETVQ
Swan goose GOLPH3.2.2A           YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDTSTGDVLLDEALKHIKETQPPETVQ
Ruddy duck GOLPH3.2.2A           YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDTSTGDVLLDEALKHIKETQPPETVQ

                                        150       160       170       180       190       200       210       220
Mallard GOLPH3.2.1A              NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Tufted duck GOLPH3.2.1A          NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Swan goose GOLPH3.2.1A           NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Ruddy duck GOLPH3.2.1A           NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK

Mallard GOLPH3.2.2A              NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Tufted duck GOLPH3.2.2A          NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Swan goose GOLPH3.2.2A           NWIELLS-------------------------------------------------------------------
Ruddy duck GOLPH3.2.2A           NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK

                                            230       240       250       260       270       280       290   296
Mallard GOLPH3.2.1A              WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Tufted duck GOLPH3.2.1A          WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Swan goose GOLPH3.2.1A           WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Ruddy duck GOLPH3.2.1A           WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK

Mallard GOLPH3.2.2A              WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRMRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Tufted duck GOLPH3.2.2A          WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRMRQLLDLDPEVECMKANTNEVLWAVVAAFTK
Swan goose GOLPH3.2.2A           ------------------------------------------------------------------------------
Ruddy duck GOLPH3.2.2A           WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
```

N-terminal

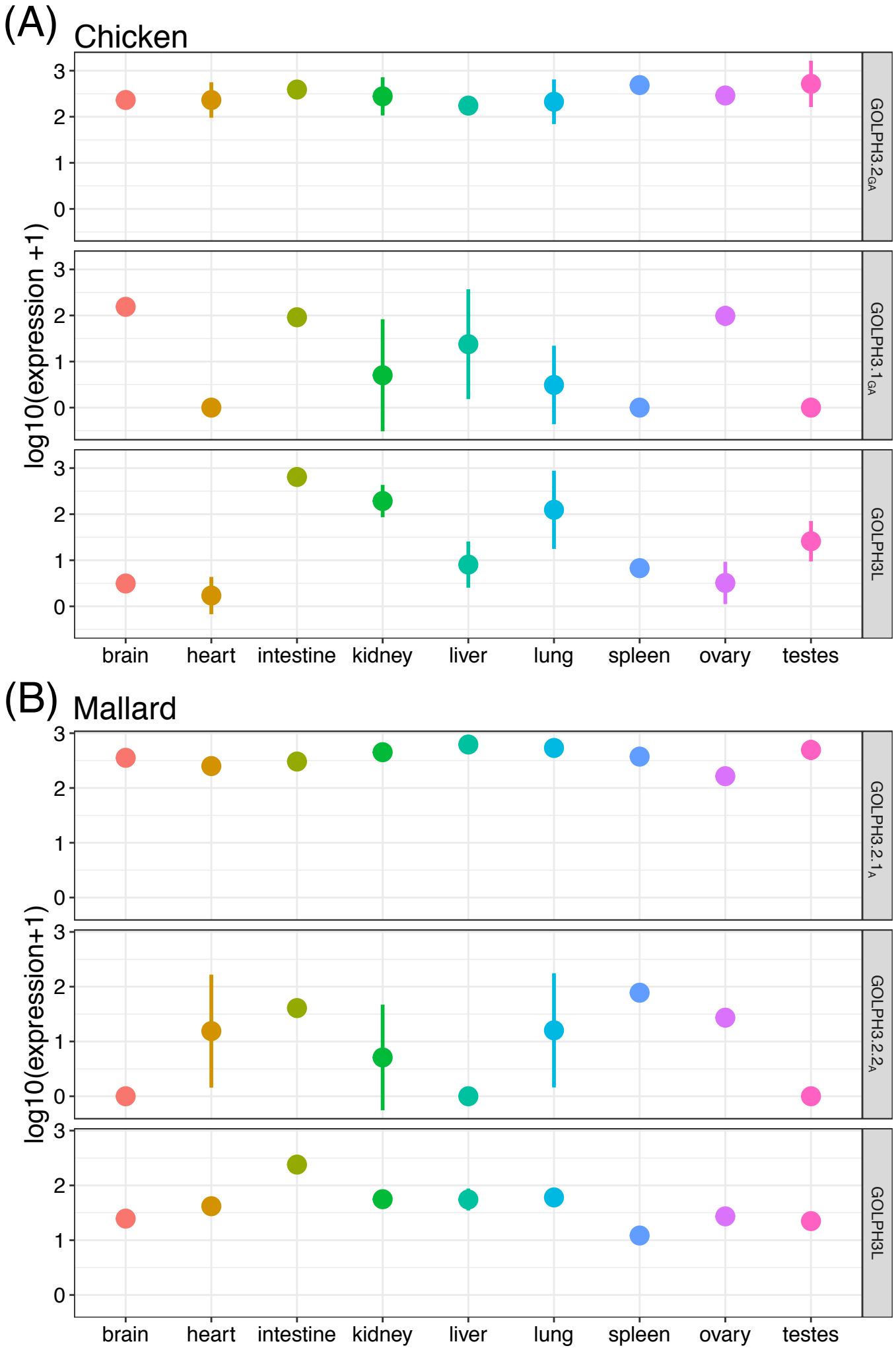
```
                                                  1         10        20        30        40        50        60        70
Lance tailed manakin GOLPH3.1N   MTSLTQRSSGLVQRRTEASRS--AAADKERGAGGGSDDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Kakapo GOLPH3.1N                 MTSLTQRSSGLVQRRTEASRS--AAADKERGAGGGSEDEGRRDESGDDEKGDSKETRLTLMEEVLLLGLKDREG
New Caledonian crow GOLPH3.1N    MTSLTQRSSGLVQRRTEASRSAAAAADKERGAGGGPEDEGRRDESGDDEKGDSKETRLTLMEEVLLLGLKDREG
Zebra finch GOLPH3.1N            MTSLTQRSSGLVQRRTEASRS-AAAADKERGAGGGPEDEGRRDESGDDEKGDSKETRLTLMEEVLLLGLKDREG

Lance tailed manakin GOLPH3.2N   MTSLTQRSSGLVQRRTEASRS--AAADKDREAGGGSEDGGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Kakapo GOLPH3.2N                 MTSLTQRSSGLVQRRTEASRS--TAAGKERGAGGGPEDEGRRDDPGDDEKGDSKEIRLTLMEEVLLLGLKDREG
New Caledonian crow GOLPH3.2N    MTSLTQRSSGLVQRRTEASRS--AAADKEREAGGGSEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG
Zebra finch GOLPH3.2N            MTSLTQRSSGLVQRRTEAFRN--AAAAKEREAGGSSEDEGRRDEPGDDEKGDSKETRLTLMEEVLLLGLKDREG

                                                 80        90       100       110       120       130       140
Lance tailed manakin GOLPH3.1N   YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Kakapo GOLPH3.1N                 YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
New Caledonian crow GOLPH3.1N    YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Zebra finch GOLPH3.1N            YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ

Lance tailed manakin GOLPH3.2N   YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Kakapo GOLPH3.2N                 YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLSRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
New Caledonian crow GOLPH3.2N    YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ
Zebra finch GOLPH3.2N            YTSFWNDCISSGLRGCMLIELALRGRLQLEACGMRRKSLLTRKVICKSDAPTGDVLLDEALKHIKETQPPETVQ

                                 150       160       170       180       190       200       210       220
Lance tailed manakin GOLPH3.1N   NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Kakapo GOLPH3.1N                 NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
New Caledonian crow GOLPH3.1N    NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Zebra finch GOLPH3.1N            NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK

Lance tailed manakin GOLPH3.2N   NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Kakapo GOLPH3.2N                 NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVVDK
New Caledonian crow GOLPH3.2N    NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK
Zebra finch GOLPH3.2N            NWIELLSGETWNPLKLHYQLRNVRERLAKNLVEKGVLTTEKQNFLLFDMTTHPLTNNNIKQRLIKKVQEAVLDK

                                          230       240       250       260       270       280       290     296
Lance tailed manakin GOLPH3.1N   WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKTNTNEVLWAVVAAFTK
Kakapo GOLPH3.1N                 WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANTNEVLWAVVAAFTK
New Caledonian crow GOLPH3.1N    WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKTNTNEVLWAVVAAFTK
Zebra finch GOLPH3.1N            WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKTNTNEVLWAVVAAFTK

Lance tailed manakin GOLPH3.2N   WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANMNEVLWAVVAAFTK
Kakapo GOLPH3.2N                 WVNDPHRMDKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEAECMKANTNEVLWAVVAAFTK
New Caledonian crow GOLPH3.2N    WVNDPHRMEKRLLALVYLAHASDVLENAFAPLLDEQYDLATKRVRQLLDLDPEVECMKANMNEVLWAVVAAFTK
Zebra finch GOLPH3.2N            WVNDPHRMEKRLLALVYLAHASDVLENAFSPLLDEQYDLATKRVRQLLDLDPEVECMKANMNEVLWAVVAAFTK
```


(A)
GOLPH3
Human
Mouse
Chicken (Chr W)
Chicken (Chr Z)
Western clawed frog
Coelacanth
Spotted gar
Thorny skate
Fruit fly
Roundworm
Brewer yeast
50
100
150
200
250
300
350
(B)
GOLPH3L
Human
Mouse
Chicken
Western clawed frog
Coelacanth
Spotted gar
Thorny skate
50
100
150
200
250
300
350
Disordered
Helix
Strand


GOLPH3L
Mallard
Tufted duck
Swan goose
Black swan
Ruddy duck
Japanese quail
Ring-necked pheasant
Turkey
Chicken
Chuck-will's-widow
Ruff
Killdeer
East African grey crowned-crane
Cuckoo roller
Barn owl
White-tailed tropicbird
Chimney swift
Sunbittern
Rhinoceros Hornbill
Common cuckoo
Red-legged seriema
Macqueen's bustard
Emperor penguin
Adelie penguin
Crested Ibis
Little egret
Saker falcon
Peregrine falcon
Yellow throated sandgrouse
Golden eagle
Eurasian sparrowhawk
Bald eagle
Brown roatelo
Burrowing owl
New Caledonian crow
Tibetan ground tit
Common canary
Zebra finch
Budgerigar
Kakapo
Lance tailed manakin
White ruffed manakin
Wire-tailed manakin
Downy woodpecker
Rock pigeon
Speckled mousebird
Okarito brown kiwi
North Island brown kiwi
Emu
African ostrich
100
100
100
100
83
100
0.3

N-terminal

```
                                  1         10        20        30        40        50        60        70        80        90
Mallard GOLPH3L                   MTTLTHRGRRSEAGGSVGKRLDGEEE---ERDP--EEDDDSKGTRLTLMEEVLLLGLKDK-GYTSFWNDCISSGLRGGILIELALRGRIQLEPLTA
Tufted duck GOLPH3L               MTTLTHRGRRTEAGGSVGKRVDGEE----ERDL---EDDDSKGTRLTLMEEVLLLGLKDKQGYTSFWNDCISSGLRGGILIELALRGRIQLEPLTA
Black swan GOLPH3L                MTTLTHRGRRTEAGGSTGKKLDDEEEAAAERDP---EDDDSKGTRLTLMEEVLLLGLKDKQGYTSFWNDCISPGLRGGILVELALRGRIQLEPLTT
Ruddy duck GOLPH3L                MTTLTHRGRRGDAGGSVGKKLDDEEEAAAERDP---EDDDSKGTRLTLMEEVLLLGLKDKQGYTSFWNDCISPGLRGGILVELALRGRIQLEPLTA

Chicken GOLPH3L                   MTTLTRRGWRGD------KRTEPEEDAAGDREQPDDDSDNSKATRLTLMEEVLLLGLQDKEGYPSFWNDCLSPGLRGGILIELALRGRIHLEPLTA
Ring-necked pheasant GOLPH3L      MTTLTRRGWRGD------KRTEPEEDVAADRELPDDDSDNSKAPRLTLMEEVLLLGLQDKEGYPSFWNDCLSPGLRGGILIELALRGRIHLEPLTA
Turkey GOLPH3L                    MTTLTRRGWRGD------KRTEPEEDAAADRELPDDDSDNSKAPRLTLMEEVLLLGLQDKEGYPSFWNDCLSPGLRGGILIELALRGRIHLEPLTA
Japanese quail GOLPH3L            MTTLTHRGWRGD------KRAEPEEDVAADKELPDDDSDNSKAPRLTLMEEVLLLGLQDKEGYPSFWNDCLSPGLRGGILIELALRGRIHLEPLTA

Lance tailed manakin GOLPH3L      MTTLTHRARRGDPGGGSAKRPDGEEDGAADREPGEEEPDDSKDIRLTLMEEVLLLGLKDKEGYTSFWNDCISSGLRGGILIELALRGRIRLEPLAL
Kakapo GOLPH3L                    MTTLTHRSRRAE-GRSTDRKPTGEEDAAADRDLNDEDPDDSKDIRLTLMEEVLLLGLKDKEGYTSFWNDCISSGLRGGILIELAFRGRIRLEPPSM
New Caledonian crow GOLPH3L       MTTLTHRGRRVE-GRNSDKKPDGEEDAAADRDLNEDDPDDSKDIRLTLMEEVLLLGLKDKEGYTSFWNDCISSGLRGGILIELALRGRIRLEPLSL
Zebra finch GOLPH3L               MTTLTHRGRRAE-GRSADRRTEGEEEAAADGELSQEHPEDSKDIRLTLMEEVLLLGLKDKEGYTSFWNDCISSGLRGGILIELALRGRIRLEPPAL

                                      100       110       120       130       140       150       160       170       180       190
Mallard GOLPH3L                   RKKRLLDRKVHLKSDAPTGDILLDETLRHIKATEPAETVQTWIELLTGETWNPFKLQYQLRNVRERIAKNLVEKGILTTEKQNFLLFDMTTHPVSN
Tufted duck GOLPH3L               RKKRLLDRKVHLKSDAPTGDILLDETLRHIKATEPAETVQTWIELLTGETWNPFKLQYQLRNVRERIAKSLVEKGILTTEKQNFLLFDMTTHPVSN
Black swan GOLPH3L                RKKRLLDRKVHLKSDTPTGDILLDETLRHIKATEPAETVQTWIELLTGETWNPFKLQYQLRNVRERIAKNLVEKGILTTEKQNFLLFDMTTHPVSN
Ruddy duck GOLPH3L                RKKRLLDRKVHLKSDTPTGDILLDETLRHIKATEPAETVQTWIELLTGETWNPFKLQYQLRNVRERIAKSLVEKGILTTEKQNFLLFDMTTHPVSN

Chicken GOLPH3L                   RKKRLLDRKVLLKSAAPTGDVLLDETLRHIKASESTETVQTWIELLTGETWNPFKLQYQLRNVRERVAKGLVEKGILTTGKQSFLLFDMTTHPICD
Ring-necked pheasant GOLPH3L      RKKRLLDRKVLLKSAAPTGDVLLDETLRHIKASESMETVQTWIELLTGETWNPFKLQYQLRNVRERVAKGLVEKGILTTGKQSFLLFDMTTHPICD
Turkey GOLPH3L                    RKKRLLDRKVLLKSAAPTGDVLLDETLRHIKASESTETVQTWIELLTGETWNPFKLQYQLRNVRERVAKGLVEKGILTTGKQSFLLFDMTTHPVCD
Japanese quail GOLPH3L            RKKRLLDRKVLLKSAAPTGDVLLDETLRHIKAAESIETVQTWIELLTGETWNPFKLQYQLRNVREHVAKGLVEKGILTTGKQSFLLFDMTTHPVCD

Lance tailed manakin GOLPH3L      RRKRLLDRKVLLKSDTPTGDVLLDETLRHIKGSECPETVPTWIELLTGETWNPFKLQFQLRNVRERLAKALVEKGILGTEKQNFLLFDMTTHPLSN
Kakapo GOLPH3L                    RKKRLLERKVLLKSDAPTGDVLLDETLRHIKATEAAESVQTWIELLTGETWNPFKLQYQLRNVRERVAKALVEKGILTTEKQNFLLFDMTTHPVSN
New Caledonian crow GOLPH3L       RKKRLLERKVLLKSDAPTGDVLLDETLRHIKATESAETVQTWIELLTGETWNPFKLQYQLRNVRERLAKALVEKGILTTEKQNFLLFDMTTHPVSN
Zebra finch GOLPH3L               RRKRLLERKVLLKSDAPTGDVLLDETLRHIKATEPAETVQSWIELLTGETWNPFKLQYQLRNVRERLAKALVEKGILTTEKQNFLLFDMTTHPVSN

                                          200       210       220       230       240       250       260       270       280     286
Mallard GOLPH3L                   AAEKQRLLKRLQEGVLERWAGDPHRMDRRTLALLVLAHSSDVLENAFSSLTDERYEAAMNRTKEVLDADPEVEAAKNR-GTEMIWAILAAFNKA
Tufted duck GOLPH3L               AAEKQRLLKRLQEGVLERWAGDPHRMDRRTLALLVLAHSSDVLENAFSSLTDDRYEVAMNRTKEVLDADPEVEAAKNR-DTEMIWAILAAFNKA
Black swan GOLPH3L                AAEKQRLLKRLQEGVLERWAGDPHRMDRRTLALLVLAHSSDVLENAFSSLADDRYEVAMNRTKEVLDADPEVEAAKGR-GTEMIWAILAAFNKA
Ruddy duck GOLPH3L                AAEKQRLLKRLQEGVLERWVGDPHRMDRRTLALLVLAHSSDVLENAFSSLEDDRYKVAMNRTKEVLDADPEVEAAKGR-GTEMIWAILAVFNKA

Chicken GOLPH3L                   TTEKQRLLKKLQGSVLERWTGDLRRMDHRLLALLVLAHASDVLEKALGSLEDTQYEMAMSRAKDVLEANAELEAAQGR-GTEVIWAVLAAFSRA
Ring-necked pheasant GOLPH3L      TTEKQRLLRKLQDSLLERWTGDPRRMDRRLLALLVLAHASDVLEKALGGLEDTQYEMAMSRVRDVLEADAELEAAQGK-GTEVIWAVLAAFNRA
Turkey GOLPH3L                    TTEKQRLLRKLQDSVLERWMGDPRRMDRRLLALLVLAHASDVLEKALGSLEDTQYEMAMSRAKDVLEADAELEAAQGR-GTEVIWAVLAAFSRA
Japanese quail GOLPH3L            TTEKQRLLRKLQSSLLERWTGDVRRMERRLLALLVLAHTSDVLEKALGGLEDTQYELAMSRAKDVLEADTELEAAQGR-GTEVIWAVLAALSRA

Lance tailed manakin GOLPH3L      GAEKLRLQRRLQESVLERWARDPRRLEPRTLALLVLAHSSDVLEGVFGALADDRYELALARARDLLDLDPELEAAKARGGCEMIWAVLAAFNKG
Kakapo GOLPH3L                    GTEKQRLVRRLQESVLERWVNDPHRMDRRTLALLVLAHSSDVLENVFASLADDKYDVAMNRTKDLLDMDPEVEAAKAK-GTEMVWAVLAAFNKS
New Caledonian crow GOLPH3L       ASEKQRLVKKLQESVLERWVNDPQRMERRTLALLVLAHASDVLENVFASLADDKYDVAMNRTKDLLDMDPEVEAAKAR-GTEMIWAVLAAFNKS
Zebra finch GOLPH3L               ASEKQRLLKKLQESVLERWVNEPQRMERRTLALLVLAHASDVLENVFGSLADDKYDVAMNRTKDLLDMDPEVEAAKGK-GTEMIWAVLAAFNKS
```


(A)
Chicken
log10(expression +1)
GOLPH3.2GA
GOLPH3.1GA
GOLPH3L
brain
heart
intestine
kidney
liver
lung
spleen
ovary
testes
(B)
Mallard
log10(expression+1)
GOLPH3.2.1A
GOLPH3.2.2A
GOLPH3L
brain
heart
intestine
kidney
liver
lung
spleen
ovary
testes