











Genome Assembly and Evolutionary Analysis of the Mandarin Duck *Aix galericulata* Reveal Strong Genome Conservation among Ducks

Chen Siang Ng ^{1,2,3,4,†}, Cheng-Kuo Lai ^{5,6,†}, Huei-Mien Ke ⁵, Hsin-Han Lee ⁵, Chih-Feng Chen ^{4,7}, Pin-Chi Tang ^{4,7}, Hsu-Chen Cheng ^{4,8}, Meiyeh J. Lu ⁵, Wen-Hsiung Li ^{4,5,9,*}, and Isheng Jason Tsai ^{5,*}

¹Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

²Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan

³Bioresource Conservation Research Center, National Tsing Hua University, Hsinchu, Taiwan

⁴The iEGG and Animal Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

⁵Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

⁶Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan

⁷Department of Animal Science, National Chung Hsing University, Taichung, Taiwan

⁸Department of Life Science, National Chung Hsing University, Taichung, Taiwan

⁹Department of Ecology and Evolution, University of Chicago, Illinois, USA

*Corresponding authors: E-mails: ijtsai@gate.sinica.edu.tw; whli@uchicago.edu.

†These authors contributed equally to this work.

Accepted: 25 May 2022

Abstract

The mandarin duck, *Aix galericulata*, is popular in East Asian cultures and displays exaggerated sexual dimorphism, especially in feather traits during breeding seasons. We generated and annotated the first mandarin duck *de novo* assembly, which was 1.08 Gb in size and encoded 16,615 proteins. Using a phylogenomic approach calibrated with fossils and molecular divergences, we inferred that the last common ancestor of ducks occurred 13.3–26.7 Ma. The majority of the mandarin duck genome repetitive sequences belonged to the chicken repeat 1 (CR1) retroposon CR1-J2_Pass, which underwent a duck lineage-specific burst. Synteny analyses among ducks revealed infrequent chromosomal rearrangements in which breaks were enriched in LINE retrotransposons and DNA transposons. The calculation of the d_N/d_S ratio revealed that the majority of duck genes were under strong purifying selection. The expanded gene families in the mandarin duck are primarily involved in olfactory perception as well as the development and morphogenesis of feather and branching structures. This new reference genome will improve our understanding of the morphological and physiological characteristics of ducks and provide a valuable resource for functional genomics studies to investigate the feather traits of the mandarin duck.

Significance

The mandarin duck is known for its flamboyant plumages in males and nesting in tree holes. We present the first *de novo* genome assembly and annotation for the species. By comparing 12 representative bird genomes, we found that the mandarin duck genome exhibits many features that are conserved across multiple duck genomes, and our test for selection showed that the entire lineage was under strong negative selection genome-wide. We identified several gene families that were expanded in the mandarin duck that may play a role in its feather morphology such as branching structures. This new reference genome will promote advances in our understanding of feather diversification and sexual dimorphism in ducks.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Key words: mandarin duck, bird genomics, *de novo* assembly, genome annotation, gene family evolution, synteny.

Introduction

The mandarin duck (*Aix galericulata*), a species in the family Anatidae in the order Anseriformes, is a perching duck that mainly inhabits the East Palearctic (Carboneras and Kirwan 2020). Mandarin ducks are highly social animals that usually breed in wooded areas near shallow lakes, marshes, ponds, swamps, rivers, and fast-flowing streams, where they nest in cavities in tree trunks (Carboneras and Kirwan 2020). In traditional Chinese, Japanese, and Korean cultures, mandarin ducks are imagined to be life-long couples, symbolizing love, romance, devotion, affection, and fidelity, and are often featured in oriental arts and poetry and other forms of the literature. Because of their beautiful and elaborate plumage, the mandarin duck is also a popular aviary bird and common in zoos. It is closely related to the North American wood duck (*Aix sponsa*) (Livezey 1991; Johnson and Sorenson 1999; Liu et al. 2014); however, hybridization between these species has rarely been found in the wild (Johnsgard 1968).

Mandarin ducks are in fact serially monogamous in which they engage to reproduce but may pair with different individuals in the next mating season (Sun Y-H, personal communications). Male mandarin ducks have a bright yellow–orange or red bill, large white crescent above the eye, purple breast, and long orange whiskers on their reddish face, whereas female ducks have a gray bill and a white eye-ring (Shurtleff and Savage 1996). Both males and females have crests, which are purple and more pronounced in males. The male has two orange sail-feathers—elongated tertial feathers—on its wings. The flamboyant plumage of the male is crucial to its elaborate courtship ritual, in which the male raises its crest and orange sails, then displays ritual drinking and preening behind the sail. Males molt after the breeding season into eclipse plumage. After molting, they are brown and gray in color and look similar to the females, except for their yellow–orange or red bill. Seasonal plumage changes in males may be regulated by the interaction between sexual hormones and daylight durations, but the underlying molecular mechanisms remain largely unknown.

The colors of the mandarin duck's plumage during their breeding seasons attract lots of attention. The male mandarin duck's plumage is red, orange, brown, copper, green, and purple. This array of colors comes from both the feathers' pigments and their structures (Burns et al. 2017; Caro et al. 2017; Price-Waldman and Stoddard 2021). Pigment-based coloration is the result of a specific wavelength range of diffusely scattered light being absorbed (Shawkey and D'Alba 2017). Carotenoids frequently give rise to red and yellow hues in birds, and black and brown plumages are associated with eumelanin and pheomelanin

pigments (McGraw et al. 2005; Mills and Patterson 2009; Galvan and Solano 2016). Structural coloration, on the other hand, occurs when light is reflected by or experiences constructive interference from quasi-ordered spongy structures in the feather barbs and melanosomes in feather barbules (Prum et al. 1998; Vinther et al. 2010; Shawkey and D'Alba 2017; Babarovic et al. 2019). While these pigments are understood from a physics perspective, the molecular, cellular, and genomic mechanisms that shape this extravagant, ornate plumage phenotype are poorly understood. Mandarin ducks possess a large variety of pigmented and structural colors in their plumage, along with highly sexual dimorphism, providing an excellent model to study feather evolution and sexual selection.

Asian populations of mandarin duck are usually migratory, but a resident population was established in Taiwan (Sun et al. 2011). Adult annual survival rates of 74–80% for males and 44–50% for females were recorded in Taiwanese populations (Sun et al. 2011, 2014). Population sizes of mandarin ducks were estimated to be about 20,000 in China, 5,000 in Korea, and 40,000 in Japan using winter data exclusively in early 2000s (Kear 2005). Relatively small numbers (about 300–500 birds) were found in central Taiwan (Sun et al. 2011). Habitat destruction of mandarin duck in Asia was mainly caused by urbanization, road construction, water pollution, and expansion of mining, agriculture and other industry (Kear 2005; Lever 2013). Besides habitat losses, nest desertion (72.8%), and human disturbance (11.1%) have been reported as main causes for reproductive failures (Deng et al. 2011). Although Mandarin ducks are not considered as globally threatened, many aspects of its breeding and population ecology remain little understood.

New genomic tools have opened doors to studying the evolution, ecology, and population genomics of nonmodel animals (Thomson et al. 2010; Ekblom and Galindo 2011; Ekblom and Wolf 2014). Many bird genome assemblies have been generated in recent years (Jarvis et al. 2014, 2015; O'Brien et al. 2014; Zhang et al. 2014a, 2014b; Eory et al. 2015; Koepfli et al. 2015; Jarvis 2016; Feng et al. 2020; Bravo et al. 2021), but there is still no reference genome sequence for the mandarin duck. Having a reference genome is important because it will facilitate studies on the evolution of feathers in Anseriformes, particularly as mandarin ducks can be easily kept in captivity. Another species of *Aix* is the wood duck (*A. sponsa*), which also has similar plumage, habitat, courtship, and nesting behaviors. Also, the reference genome of *A. galericulata* can facilitate the ecological and evolutionary genetic studies of this species. In this study, we sequenced, assembled, and annotated the genome of a female mandarin duck from

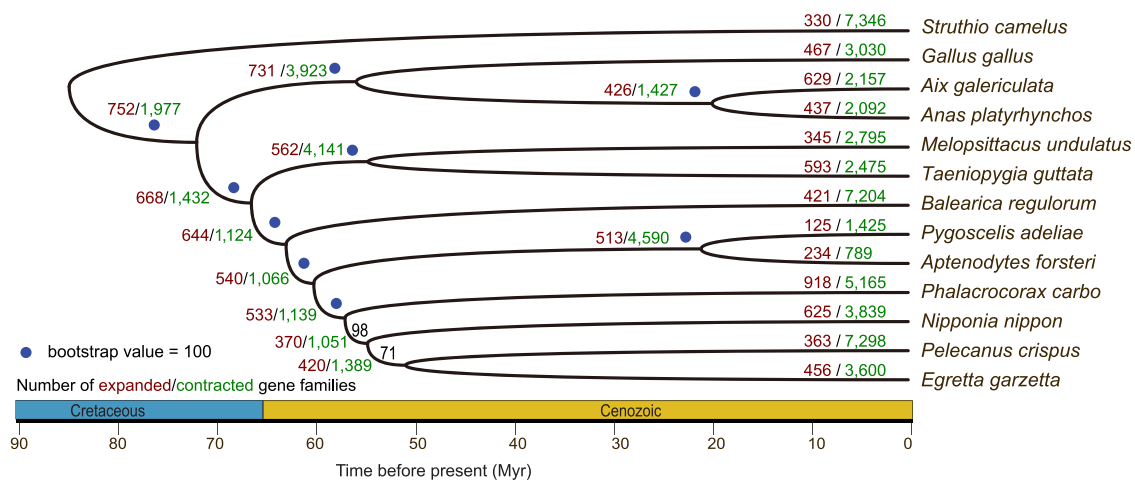


Fig. 1.—A species tree of 13 bird species constructed from 3,232 single-copy orthologs. Bootstrap support values were from the ASTRAL (Zhang et al. 2018a) inferred species phylogeny based on coalescence approach.

Taiwan using Illumina paired-end (PE) and mate-pair (MP) sequencing and reference scaffolding using the mallard duck genome. We then reconstructed the phylogenetic relationships among ducks and some other birds, analyzed the evolution of the mandarin duck's gene families, and detected signatures of selection in its genome.

Results and Discussions

Genome Assembly and Annotation

The mandarin duck has an estimated genome size of 1.085 Gb based on GenomeScope (Ranallo-Benavidez et al. 2020). We generated a total of 186 \times and 44 \times genome coverage of PE and MP Illumina reads, respectively, from a 1-year-old female mandarin duck (supplementary table 1). An initial assembly was produced using Masurca (Zimin et al. 2013). This assembly was iteratively corrected, re-scaffolded, and its gaps closed with REAPR (Hunt et al. 2013), SSPACE (Boetzer et al. 2011), and GapFiller (Nadalin et al. 2012), respectively, yielding a final 1.08 Gb assembly of 5,267 scaffolds with N50 = 6.3 Mb (supplementary table 2). A total of 16,615 protein-coding gene models were annotated using the MAKER2 pipeline aided by reference bird protein homology support and transcriptome sequencing from regenerating feathers (supplementary table 1). The predicted gene models of the assembly was estimated to be 91.2% complete based on BUSCO (Benchmarking Universal Single-Copy Orthologs) (Seppey et al. 2019) assessment, which is comparable to the completeness of other sequenced bird species (supplementary table 2). Additionally, 90.7% (15,074) of gene models could be functionally assigned with a gene ontology term using eggno-mapper (Huerta-Cepas et al. 2017). Of these, 87.6%, 85.7%, and 92.3% of the gene models were orthologous to mallard

duck (14,105), chicken (13,784), and at least one other bird species, respectively, indicating that mandarin duck genes have been largely conserved among these bird species.

Orthologue clustering was inferred using OrthoFinder (Emms and Kelly 2019) with 12 representative bird genomes. With ostrich (*Struthio camelus*) as the outgroup, OrthoFinder yielded 27,436 orthologous groups (OGs). Using 3,232 single-copy orthologues, we constructed a maximum-likelihood phylogeny based on a concatenated alignment of these single-copy orthologues and a coalescent-based phylogeny of individual gene trees (Zhang et al. 2018a). Both phylogenies produced the same topology showing the mandarin duck's close relationship with the mallard duck (fig. 1). Incorporating MCMCtree (Yang 2007) with five fossil and one molecular calibration, we calculated that ducks' most recent common ancestor occurred 13.3 to 26.7 Ma, during the Cenozoic (fig. 1). This is consistent with previous estimates based on mitochondrial genomes (Sun et al. 2017).

Characteristics of the Mandarin Duck Genome

To understand the evolution of the mandarin duck genome structure, synteny relationships among the mandarin duck, mallard duck, and chicken were inferred using DAGchainer (Haas et al. 2004). Based on 13,595 ortholog pairs, 95.6% and 93.4% of the mallard duck genome were in synteny with those of the mandarin duck and chicken, respectively, which recapitulated their evolutionary relationships. A total of 1,244 scaffolds were found to display one-to-one relationships with mallard duck chromosomes, and we scaffolded the mandarin duck assembly with the *Anas platyrhynchos* genome as the reference (Li et al. 2021) using Ragtag (Alonge et al. 2021) to produce a final reference genome in which 98.7% of the assembly was

contained in 33 chromosomes (supplementary table 2) including both sex chromosomes. The distribution of single-copy orthologs between the two duck genomes shows that chromosome synteny has been well preserved and that inter-chromosomal rearrangements were rare in ducks (fig. 2a).

The mandarin duck has an average heterozygosity of 0.23% (one heterozygous single nucleotide polymorphism every 422 bp). Pairwise sequentially Markovian coalescent (PSMC) analysis (Li and Durbin 2011) based on these biallelic SNP densities revealed an expansion followed by a sharp decline in effective population size \sim 25,866 and 19,763 years ago (fig. 2b). The habitat of mandarin duck could be affected as Late Pleistocene environments became drier during the Last Glacial Maximum (LGM, ca. 21–18 kya) in eastern Asia (Li et al. 2019; Dennell et al. 2020), causing depopulation of mandarin ducks. For comparison,

domesticated mallard ducks exhibited a similar pattern, but its population peaked much earlier—around 162,045 years ago (fig. 2b). Global mandarin duck populations nowadays are declared to be healthy, but reliable and long-term measures of its populations are required to evaluate properly the effects of habitat changes. This reference genome provides an essential material for population genomic studies to learn about reproductive ecology and population biology of mandarin duck and its closely related species.

In addition, we identified long runs of homozygosity regions in the mandarin duck autosomes, which appeared to be distributed randomly across the chromosomes (fig. 2c), consistent with patterns reflecting recent bottlenecks, possibly due to anthropogenic selective pressures (Zhang et al. 2018b) or inbreeding in the wild. Genes enriched in the large loss of heterozygosity region (LOH) of Chr 1 (fig. 2c; supplementary table 3) include female gonad

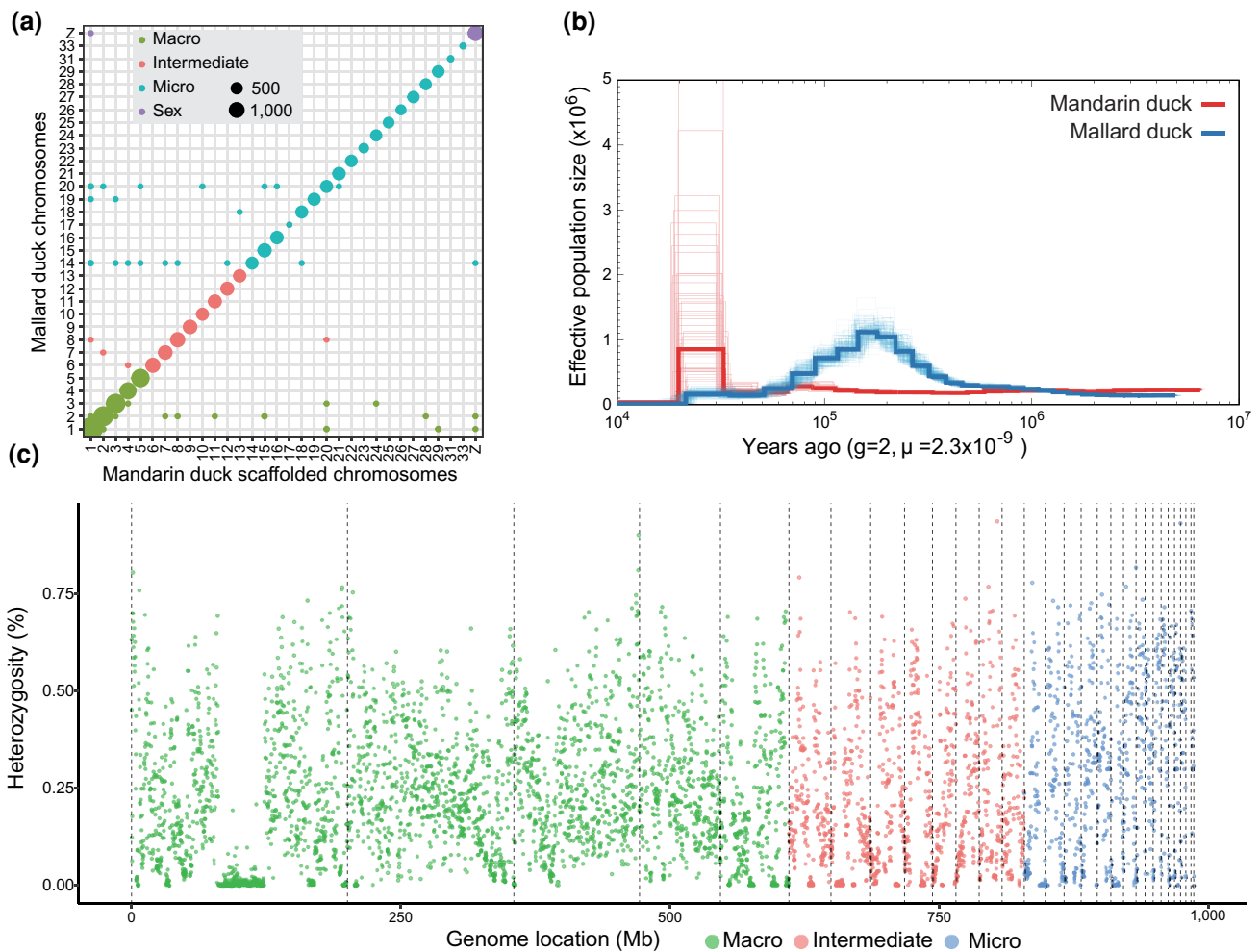


FIG. 2.—Characteristics of the mandarin duck genome. (a) Mapping of one-to-one orthologs between mandarin duck and mallard duck assemblies. The clustering pattern shows that most reference-guided chromosomes in mandarin duck can be unambiguously assigned to mallard duck chromosomes. The dot size indicates numbers of one-to-one orthologs between the two ducks and different color denote different chromosome types. (b) Past effective population sizes of the two ducks. (c) Heterozygosity of mandarin genome along 200 kb nonoverlapping windows. The sex chromosome was excluded because female reads were used and contained only one copy of chromosome.

development (GO:0008585). Three homologs of roundabout (ROBO; Mandarin_00247000, Mandarin_00247100, and Mandarin_00247200) family were identified in tandem at this LOH and were involved in the follicular development of hen ovary (Qin et al. 2015), suggesting that they may be under strong positive selection in the Taiwanese population of mandarin ducks. One member, *ROBO2*, in the *ROBO* family has also been found to have undergone highly selective sweeps in chickens (Fan et al. 2013), implying similar selection patterns in Galloanserae.

Transposable Element Dynamics Were Similar Amongst Ducks

Like other bird species, long interspersed elements (LINEs) were the most abundant repeat class, making up 5.2% of the mandarin duck assembly, followed by simple repeats and other transposon families (fig. 3a). Such a markup of repeats is similar to what has been observed in other bird species such as chicken and zebra-finch (Chalopin et al. 2015). Consistent with other duck species, the chicken repeat 1 (CR1) retroposon CR1-J2_Pass is the most abundant LINE family in mandarin duck that was retained after a duck lineage-specific burst (Li et al. 2021). CR1-J2_Pass is not distributed uniformly across the genome and clusters (defined by >10% repeat family content in 100 kb windows) are present in macrochromosomes (fig. 3b). These clusters are found to be enriched in chromosome centers and arms (fig. 3c). Strikingly, we observed a much higher TE composition in chromosome W compared to other chromosomes (47.1% vs. 5.1%; fig. 3b and c), which is consistent to the 46.5% reported in the mallard duck genome (Li et al. 2021). In addition, we did not detect CR1-J2_Pass repeat clusters in microchromosomes or minichromosomes (fig. 3b), suggesting that centromeric repeats were likely different in these chromosomes and may play a role in the spatial segregations of these chromosomes, as in other birds (Zlotina et al. 2019; Li et al. 2021).

Genomic rearrangements, particularly inversions, have played important roles in the local adaptation and speciation of birds (Jiang et al. 2021; Li et al. 2021; Zhu et al. 2021). The orientations of syntenic regions were further classified into inversions for each pairwise genome comparison. Large-scale studies of chromosome inversion were done in passerine birds (Knief et al. 2016; Hooper and Price 2017). The 35.5% of genomes were inversions between wild ducks and chickens, indicating that this kind of rearrangement events may also be prevalent in non-passerine birds (fig. 4a). In contrast, only 5.7% were identified to be inversion regions between the two duck species (supplementary table 4), and the breaks between synteny regions contained significantly longer DNA and LINE transposon elements (Wilcoxon rank-sum test; $P < 0.001$, fig. 4b), suggesting that these inversions were recent and

that these repeats play a role in mediating genome rearrangement events. The scarcity of genomic inversions observed between duck genomes suggests that newly arisen inversions may get purged or reduced in frequency as a consequence of genetic drift and/or selection (Hoffmann and Rieseberg 2008).

The average ratio of nonsynonymous to synonymous substitution rates (d_N/d_S) between 9,956 mandarin and mallard duck single-copy orthologs was 0.2 (fig. 5a), suggesting strong purifying selection on duck proteins. Fifty-two genes displayed signals of balancing or positive selection ($d_N/d_S > 1$; supplementary table 5). In addition, we compared the selective pressure of 3,232 single-copy orthologs using PAML (Yang 2007) between duck species (mandarin and mallard ducks) and a group of 11 other birds used in the species phylogeny (fig. 1). The distribution of d_N/d_S was significantly lower in the duck group compared with the other bird lineages (Wilcoxon rank-sum test, $P < 0.001$, fig. 5b) because of approximately three times higher synonymous substitution rates (d_S median 0.102 in the 2 ducks vs. 0.038 in the 11 other birds), which was also observed previously (Lanfear et al. 2010). Despite the rates of both d_S and nonsynonymous substitutions (d_N) in the duck group were estimated to be higher than those of the 11 other birds (fig. 5b), the overall lower d_N/d_S suggests that the proteome of the duck lineage has been better maintained by purifying selection since it diverged from other bird groups. It is interesting to further investigate if the proteome of Antidae is especially conserved.

Specialization of Mandarin Duck Proteins

By detecting using HMMER (Mistry et al. 2013) and comparing the copy number distribution of protein domains (Pfam) among 13 birds, we inferred the specific domains in the mandarin duck (fig. 6). A total of 287 expanded and 154 constructed protein domains were found in the mandarin duck compared with the other 12 birds (supplementary table 6). The expanded cytochromes P450 (CYPs) genes were observed in the mandarin duck, which is consistent with the inhibited formation of aflatoxin B1 in turkeys (Rawal and Coulombe 2011). In contrast, domestic ducks are more sensitive to aflatoxins and less able to detoxify and eliminate this mycotoxin than chickens (Ostrowski-Meissner 1983). The P450 expansion in mandarin ducks may imply that they are exposed to aflatoxin B1 in their environment and thus adapted to the compound. In addition, we found that the mandarin duck is enriched in neurotransmitter sodium symporter (SNF) domains, which are associated with the re-uptake and recycling of neurotransmitters (Shi et al. 2008). The BRCA1 C terminus (BRCT) domain—which is a protein–protein interaction motif that binds phosphorylated proteins in the DNA damage response system (Wu et al. 2010)—was lost in the mandarin duck. Losses or mutations of BRCT impair the function of

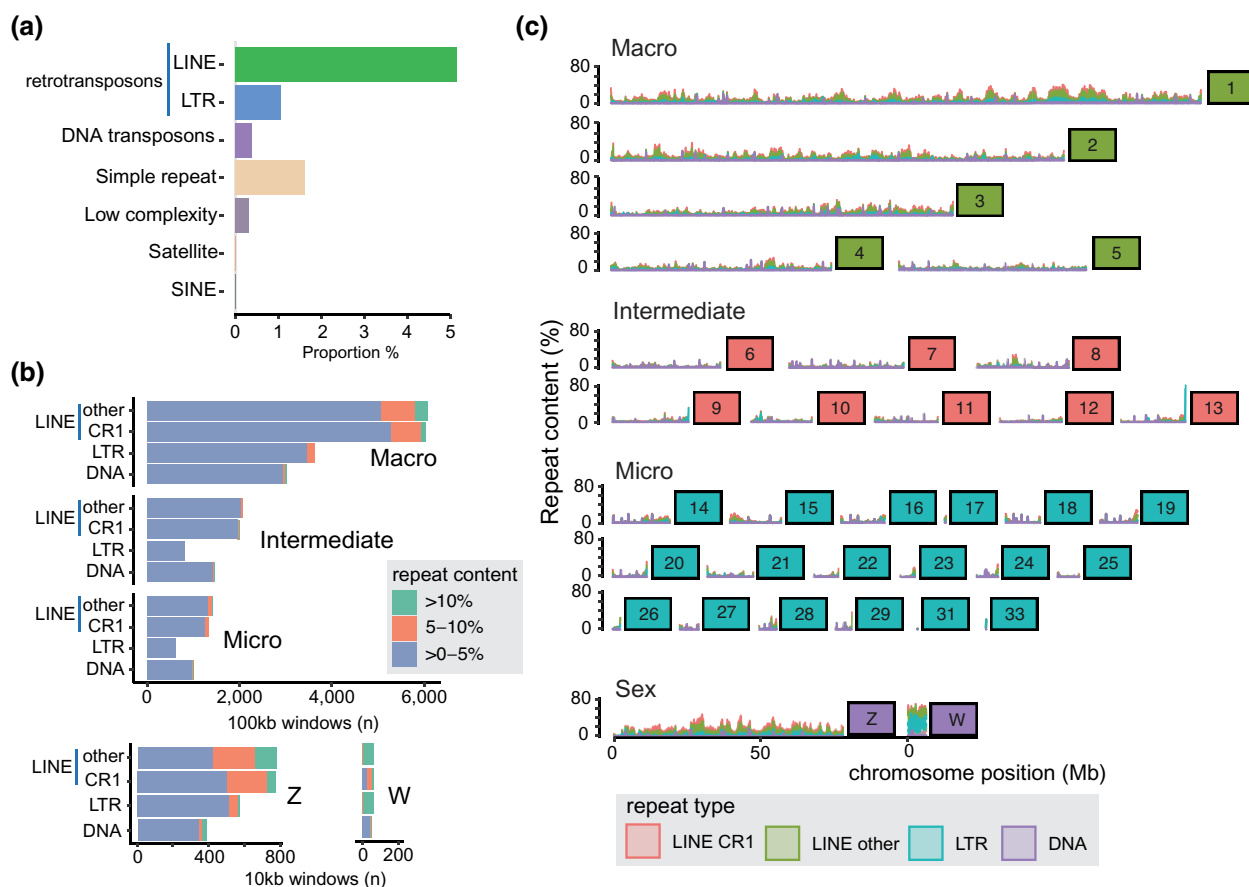


FIG. 3.—Transposon elements in the mandarin duck genome. (a) Repeat contents in the mandarin duck assembly calculated using RepeatMasker (Tarailo-Graovac and Chen 2009). (b) Number of 100 kb windows containing repeats and their proportions categorized by chromosome types. Colors denote different proportions. (c) Coverage of three transposon element types along 100 kb nonoverlapping windows. Different colors denote types of repeats.

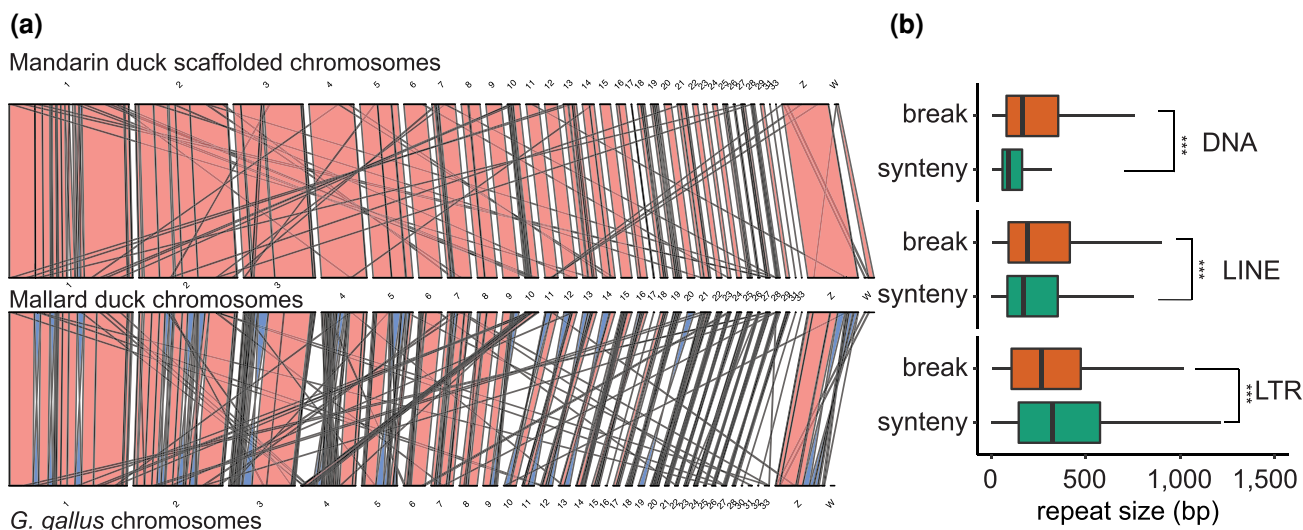


FIG. 4.—Synteny and repeats. (a) The synteny relationships between mandarin, mallard duck, and chicken reveal a closer evolutionary relationship between the two ducks. Blue color indicates inversions. (b) The DNA transposons and LINES were significantly higher in the break region of synteny between the two ducks, but not LTR retrotransposons. *** $P < 0.001$.

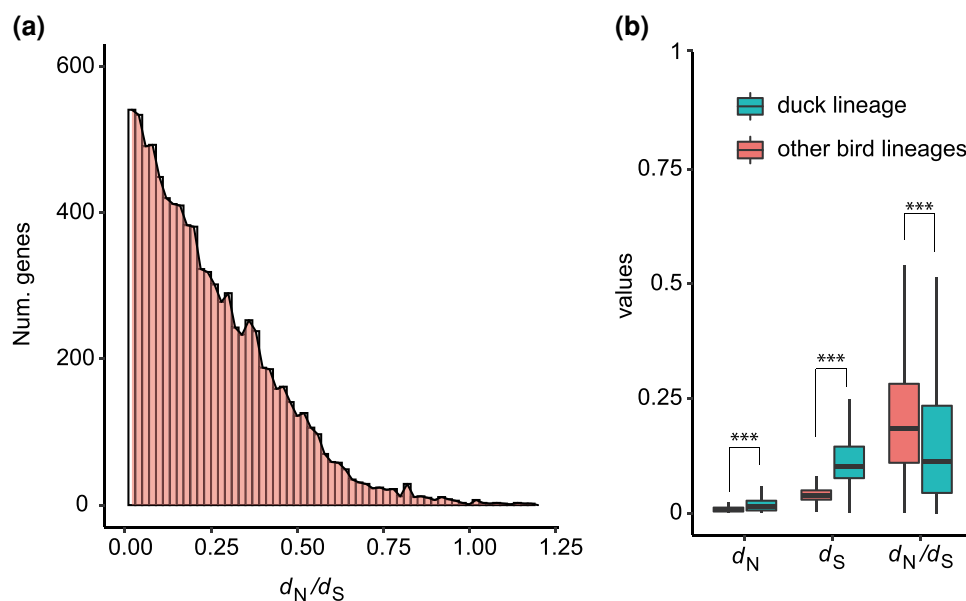


Fig. 5.— d_N/d_S of orthologues between the two ducks and in the duck lineage. (a) The distribution of d_N/d_S between Mandarin and mallard duck single-copy orthologues. (b) The distribution of d_N , d_S , and d_N/d_S between the duck lineage and other 11 birds. Different colors denote different branch types: the blue color represents the duck group (the branch of Mandarin and mallard duck); the red color denotes the median of 11 other bird lineages in the species phylogeny. *** $P < 0.001$.

BRCA1 (Vallon-Christersson et al. 2001; Williams et al. 2003), so it is interesting to further investigate if mandarin duck evolved a different mechanism of DNA repair if the loss of BRCT is confirmed to be fixed in this species.

A total of 629 OGs were found to be expanded in the mandarin duck by a Bayesian method (Liu et al. 2011; fig 1). Similar numbers were obtained in the duck lineage using a maximum-likelihood approach implemented in CAFÉ (supplementary fig. 1) (Hahn et al. 2005). The enrichment of GO terms from the expanded families revealed several features that are associated with the lifestyle and morphogenesis of feathers in mandarin ducks (supplementary table 7). Fifteen genes belonging to the GO term “development of primary male sexual characteristics” (GO:0046546) were enriched, which might partially explain the extreme sexual dimorphism in mandarin ducks. Eighty-three and 27 genes belonging to GO terms related to actin cytoskeleton organization (GO:0030036) and striated muscle contraction (GO:0006941), respectively, were enriched, indicating that cell movement might have adaptive fitness in this lineage. At least eight genes in the GO term “development of primary male sexual characteristics” (GO:0046546) have known functions in feather development. The beta-catenin pathway modulates epithelial morphogenesis, initiates feather follicle development in embryonic skin, and is required for the subsequent morphogenesis of the feather bud (Mandarin_00850000 and Mandarin_00029800) (Noramly et al. 1999; WidELITZ et al. 2000).

On the branch leading to the duck lineages, which includes the mandarin duck and mallard duck (*Anas platyrhynchos*), a total of 146 OGs were significantly expanded (supplementary fig. 1). Among them, a total of 22 orthologues responsible for olfactory receptor were enriched in the GO term involving detection of the chemical stimulus involved in the perception of smell (supplementary table 8). GO terms related to vertebrate eye function were also enriched. These findings indicate that smell and the visual perception of colors may contribute to the sexual dimorphism in this lineage, or these sensory cues are important for foraging and/or food selection in ducks.

Genes and Gene Families Related to Development of Feather and Branching Structure

One of the iconic features in mandarin ducks is exaggerated feather traits especially in males (Shurtleff and Savage 1996) and expanded gene families may provide more genetic materials for generating unique plumage characters. At least one gene involved in male sexual characteristics development has a central role in hair development in mammals and may be important in feather morphogenesis of mandarin duck since some molecules and signaling pathway in development of hair and feather are somewhat similar (Dhouailly 2009; Lowe et al. 2015). TBX3 plays important roles in initial stages of mammary placode morphogenesis (Eblaghie et al. 2004; Carroll and Capocchi 2015), its regulatory mutations also cause radially

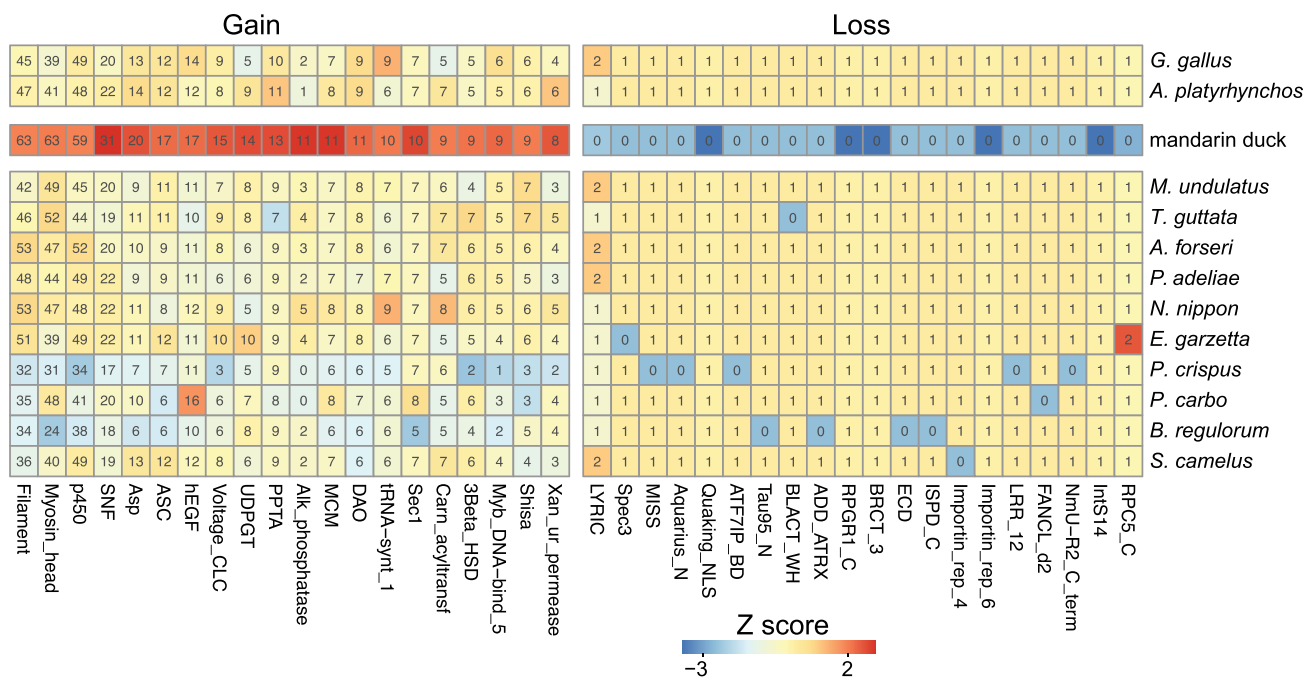


FIG. 6.—Protein family (pfams) dynamics in the mandarin duck. The top 20 most expanded and reduced pfams are shown.

asymmetric deposition of pigment in the growing hair of horses (Mandarin_01039200, Mandarin_01039300, and Mandarin_01029700) (Imsland et al. 2016). Furthermore, we also sought to identify several additional genes that may be involved in feather development in mandarin ducks. Matrix metalloproteinases (MMP) play essential roles in both proliferation of mesenchymal cells and invagination of epithelium during feather morphogenesis (Mandarin_01257700 and Mandarin_01618700) (Jiang et al. 2011). Bone morphogenetic protein 6 (BMP6) may play a role in the apoptosis of keratinocytes in the secondary feather follicles morphogenesis in *Anser cygnoides* (Mandarin_00326000 and Mandarin_00325900) (Wu et al. 2019).

The enriched GO term of branching involved in mammary gland duct morphogenesis (GO:0060444; supplementary table 7) is also particularly interesting, because both feathers and mammary glands undergo extensive branching (Widelitz et al. 2007; Chuong et al. 2014). Some genes in this GO term are known to be involved in feather development or morphogenesis. The WNT signaling is involved in multiple critical roles on feather morphogenesis (Mandarin_01257700 and Mandarin_01030300) (Widelitz 2008). Expressions of Eph receptors and ephrins are found in chicken feather morphogenesis (Mandarin_01261900 and Mandarin_00927000) (Ji et al. 2021) and they are involved in boundary stabilization during feather morphogenesis (Suksaweang et al. 2012). Although several other genes involved in mammary gland duct morphogenesis have no known functions in feather development, they play essential roles of

mammary gland development and could be important for developing the extravagant ornate plumage of mandarin duck. Impaired Discoidin domain receptor 1 tyrosine kinase can cause hyperproliferation and abnormal branching of mammary ducts (Mandarin_01602200 and Mandarin_01643800) (Vogel et al. 2001). Promyelocytic leukemia protein (PML) is required for functional differentiation of mammary epithelia for ductal and alveolar development (Mandarin_00706700, Mandarin_00707100, Mandarin_00707000, and Mandarin_00706700) (Li et al. 2009a, 2009b); there were four PML copies in tandem on chromosome 11 presumably as a result of duplication. Functional validations using RCAS and/or lentivirus transgenic systems (Hughes 2004; Scott and Lois 2005) are required to verify these findings.

Conclusion

The mandarin duck genome that we present in this study will be useful for functional gene mapping of plumage traits to identify interesting genes or gene families, as well as for genome comparisons in evolutionary studies of Anatidae. We also found that the mandarin duck went through many gene expansions and a population bottleneck in its evolutionary history, warranting further investigations into the wild population genetic diversity of this iconic bird species.

Materials and Methods

Extraction and Sequencing of Genomic DNA and RNA

The female mandarin duck (ID#: NCHU-2012, NCBI: txid8832) used in this study was a captive and undomesticated individual which had been derived from a wild population and kept alive in the aviary of National Chung Hsing University (NCHU), Taichung, Taiwan, when the blood was drawn (supplementary fig. 1). The use of mandarin ducks for genomic research was approved by the Institutional Animal Care and Use Committee (IUCAC) of NCHU (no. 100-88, December 28, 2011). Genomic DNAs were extracted from peripheral venous blood using the QIAGEN-Gentra Puregene Cell Kit (Qiagen, Venlo, The Netherlands). The purified DNA was assessed for purity and quality by NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA), Qubit (Invitrogen Corp., Carlsbad, CA, USA), and gel electrophoresis. High-quality genomic DNA was then selected for PE library preparation following the manufacturer's protocol (Illumina Inc., San Diego, CA, USA). Seven 185 bp PE libraries with insert size ~238 bp totaling 201.5 Gb were constructed using the TruSeq DNA Preparation Kit with the standard protocol (Illumina) and sequenced by Illumina HiSeq 2000 (supplementary table 1). Four MP 150 bp libraries of various insert sizes (2, 4, 6, and 9 kb) totaling 48.8 Gb were constructed using the Nextera Mate Pair Sample Preparation Kit and sequenced by Illumina HiSeq 2000. Sequencing of genomic DNA was performed at the NGS High Throughput Genomics Core of Biodiversity Research Center, Academia Sinica, Taiwan (supplementary table 1).

The feather follicle tissue of sail feathers was incubated at 4°C overnight for penetration by RNALater solution and then transferred to -20°C before further isolation of total RNA. Epithelium was dissected from the follicle tissue and separated from the mesenchyme in Calcium-Magnesium-Free Saline (CMFS 2×) on ice (Chuong 2000). Total RNA from feather epithelium was extracted using Trizol reagent (Invitrogen, Waltham, MA, USA). Libraries were prepared using the standard protocol (Wu et al. 2015). mRNA sequencing was performed using HiSeq 2000 or NextSeq 500 at the University of Southern California Epigenome Center. mRNA-seq library preparation and sequencing were performed as described (Wu et al. 2018).

Assembly and Annotation of the Mandarin Duck Genome

Genome size was estimated from one lane of Illumina sequence reads using GenomeScope (Vurture et al. 2017). PE and MP reads were quality and adaptor trimmed using Trimmomatic (version 0.35; parameter: 2:30:10 LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15

MINLEN:36) (Bolger et al. 2014) and NextClip (version 1.1; parameter -m 50) (Leggett et al. 2014), respectively. An initial assembly was produced with both PE and MP reads using Masurca (Zimin et al. 2013). Redundant scaffolds were removed based on MUMMER4 (Marcais et al. 2018) alignments of the assembly itself. Regions indicative of mis-assembly were detected by remapping 9 kb MP reads to the assembly using REAPR (version 1.0.16) (Hunt et al. 2013). These regions were broken and re-scaffolded using all the MP reads with SSPACE (version 2.1) (Boetzer et al. 2011). Gaps within scaffolds were closed with Illumina reads using GapCloser (Nadalin et al. 2012) and reference scaffolded against the *Anas platyrhynchos* genome as the reference (Li et al. 2021) using Ragtag (Alonge et al. 2021). The mitochondrial genome was assembled separately using MITObim (Hahn et al. 2013) with the published *A. galericulata* mitochondrion (GenBank: KJ169568) as the seed sequence.

The MAKER2 (version 2.31.9) pipeline (Holt and Yandell 2011) was used to generate the protein gene predictions from the mandarin duck assembly. MAKER2 combined four sources as evidence hints and invoked the gene predictors Augustus (Stanke et al. 2006) and SNAP, which were iteratively trained to generate a final set of gene annotations. The three sources of evidence hints were: (1) transcripts that were assembled from RNA-seq reads using Trinity (Haas et al. 2013) (using a reference-guided approach; version 2.5.1), and filtered using MIKADO (version 1.1) (Venturini et al. 2018); (2) proteomes from representative reference bird species downloaded from Ensembl (Aken et al. 2017) birds (version 87); (3) de novo predictions from BUSCO; and (4) predictions from BRAKER1 (Hoff et al. 2016) with RNA-seq reads as guides. The amino acid sequences from the predicted gene models were functionally annotated using Blast2GO (Conesa et al. 2005) for product descriptions and gene ontology terms using eggno-mapper (Huerta-Cepas et al. 2017).

To characterize repetitive elements, a consensus library was built by merging the output from RepeatModeler (Flynn et al. 2020), TransposonPSI (version 1.0.0; <https://github.com/NBISweden/TransposotrGnPSI>), LTRharvest (Ellinghaus et al. 2008), and EDTA (version 2.0.0) using USEARCH (Ou et al. 2019). The detailed parameters were described in Berriman et al. (2018). The CR1 retroposon sequences that was reported to be expanded in several duck lineages (Li et al. 2021) was searched in the merged repeat library and updated. Repetitive elements of the assembly were masked using Repeatmasker with the merged library.

Comparative Genome Analysis

We downloaded the amino acid and nucleotide sequences of five bird species from Ensembl (Release 101)—*Anas platyrhynchos*, *Gallus gallus*, *Melopsittacus undulatus*,

S. camelus, and *Taeniopygia guttata*—and seven species from NCBI—*Aptenodytes forsteri* GCF_000699145.1, *Balearica regulorum* GCF_000709895.1, *Egretta garzetta* GCF_000687185.1, *Nipponia nippon* GCF_000708225.1, *Pelecanus crispus* GCF_000687375.1, *Phalacrocorax carbo* GCF_000699105.1, and *Pygoscelis adeliae* GCF_000699105.1. Functional protein domains (Pfam) in each bird species were annotated using HMMER against the Pfam database (version 31) (Punta et al. 2012) with an e-value of <0.001 . The predicted results were further converted into a copy number distribution matrix, and the matrix was converted to a z-score to compute significance. Mandarin duck z-scores >1.96 were inferred to be significantly enriched; mandarin ducks with z-scores <-1.96 were inferred to be significantly reduced.

Orthologs and gene families were assigned using OrthoFinder (version 2.2.7) (Emms and Kelly 2019). For the 3,232 orthogroups that contained only a single-copy gene from each bird species, MAFFT (version 7.310) (Kato et al. 2002) and TranslatorX (version 1.1) (Abascal et al. 2010) was used to construct protein and codon sequence alignments, respectively. A maximum-likelihood phylogeny with 500 bootstrap replicates was constructed from alignments of each orthogroup or a concatenated alignment of all single-copy orthogroups using RAxML (version 8.2.3) (Stamatakis 2014). A consensus species tree was inferred using ASTRAL-III (version 5.6.3) (Zhang et al. 2018a). Divergence times for this species tree were inferred using MCMCTree in the PAML package (version 4.9j) (Yang 2007) using fossil calibrations—Stem Galloanserae 66.7–72.1 Ma (Field et al. 2020), stem Gruiformes 53.9–66.5 Ma (Musser et al. 2019), stem Sphenisciformes 60.5–72.1 Ma (Ksepka et al. 2006; Slack et al. 2006), Stem Threskiornithidae (Threskiornithidae–Pelecanidae/Ardeidae split) 53.9–66.5 Ma (Mayr and Bertelli 2011; Smith and Ksepka 2015), and Psittacopasserae (Psittaciformes and Passeriformes split) 53.9–66.5 Ma (Mayr 2008, 2015). Minimal and maximal ages of fossils mentioned above were referred as described (Kimball et al. 2019). We also used a molecular calibration for the *Pygoscelis/Aptenodytes* split 19.06–25.19 Ma (Vianna et al. 2020). After excluding three gene families with >100 gene copies, 15,022 gene family dynamics across the 13 bird species in this study were inferred using CAFÉ (version 4.2.1; lambda command) (Hahn et al. 2005) and an alternative Bayesian implementation estimation (unlink option) (Liu et al. 2011) with the time-calibrated species tree.

To infer past effective population size and heterozygosity, Illumina genomic reads were first aligned to the assembly using BWA (version 0.7.17-r1188) (Li and Durbin 2010) and heterozygous SNPs were predicted using BCFtools (version 1.10.2) (Li 2011). Heterozygosity along 200 kb nonoverlapping windows was estimated using BEDTools

(version 2.27) (Quinlan and Hall 2010). Population size history between mandarin and mallard were inferred using PSMC tools (version 0.6.5-r67, parameter: $g = 2$, $u = 2.3e \times 10^{-9}$) (Nadachowska-Brzyska et al. 2016); the generation of duck per year and the mutation of duck per site per year were referred by the previously measured mutation rates in birds (Smeds et al. 2016). Synonymous (d_S) and nonsynonymous (d_N) substitutions rates were calculated using PAML (version 4.9e, parameter: runmode = 2, seqtype = 1, CodonFreq = 2, fix_omega = 0) (Yang 2007).

Supplementary Material

Supplementary data are available from *Genome Biology and Evolution* online.

Acknowledgments

This research was funded by the Ministry of Science and Technology, Taiwan (grant nos. MOST 107-2311-B-007-008 -MY3 and MOST 110-2628-B-007-001 to C.S.N., MOST 104-2621-B-001-003-MY3, MOST 107-2311-B-001-016-MY3 and MOST110-2311-B-001-035 to W.-H.L.). The funders had no role in the study's design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Yu-Ching Liu for her help to deposit genome assembly and raw data to NCBI and Profs Cheng-Ming Choung, Randall Wideltz, Ping Wu, and Yuan-Hsun Sun for discussions.

Author Contributions

C.S.N., C.-F.C., P.-C.T., H.-C.C., and W.-H.L. conceived this research. C.S.N., I.J.T., and W.-H.L. designed the study. C.-F.C. and P.-C.T. collected and provided essential samples. C.-K.L., H.-M.K., H.-H.L. and I.J.T. performed most bioinformatic analyses. C.S.N., I.J.T., and W.-H.L. interpreted the data and wrote the initial draft of the manuscript. All authors contributed to its revision and approved the final version.

Data Availability

The sequencing data and the mandarin duck reference were deposited into the NCBI and associated under BioProject accession PRJNA787744. The alignments used for the PAML, gene alignments, and tree used to construct the species phylogeny are available in <https://github.com/ishengtsai/mandarinduckgenomedata> with explanation of each file provided.

Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38:W7–W13.

- Aken BL, et al. 2017. Ensembl 2017. *Nucleic Acids Res.* 45: D635–D642.
- Alonge M, et al. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *bioRxiv*.
- Babarovic F, et al. 2019. Characterization of melanosomes involved in the production of non-iridescent structural feather colours and their detection in the fossil record. *J R Soc Interface.* 16:20180921.
- Berriman M, Cogburn LA, Tsai JJ. 2018. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protoc Ex.* doi:10.1038/protex.2018.054.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bravo GA, Schmitt CJ, Edwards SV. 2021. What have we learned from the first 500 avian genomes? *Annu Rev Ecol Evol Syst.* 52:611–639.
- Burns KJ, McGraw KJ, Shultz AJ, Stoddard MC, Thomas DB. 2017. Advanced methods for studying pigments and coloration using avian specimens. In: Webster MS, editor. *The extended specimen: emerging frontiers in collections-based ornithological research.* Boca Raton (FL): CRC Press.
- Carboneras C, Kirwan GM. 2020. Mandarin Duck (*Aix galericulata*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA and de Juana E, editors. *Birds of the world.* Ithaca (NY): Cornell Lab of Ornithology.
- Chalopin D, Naville M, Plard F, Galiana D, Volff JN. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol.* 7:567–580.
- Chuong CM. 2000. Skin morphogenesis. Embryonic chicken skin explant cultures. *Methods Mol Biol.* 136:101–106.
- Chuong CM, Bhat R, Widelitz RB, Bissell MJ. 2014. SnapShot: branching morphogenesis. *Cell* 158:1212–1212.e1.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Deng QX, et al. 2011. Conspecific brood parasitism and nesting biology of mandarin ducks (*Aix Galericulata*) in Northeastern China. *Wilson J Ornithol.* 123:479–485.
- Dennell R, Martinon-Torres M, de Castro JMB, Xing G. 2020. A demographic history of Late Pleistocene China. *Quat Int.* 559:4–13.
- Dhouailly D. 2009. A new scenario for the evolutionary origin of hair, feather, and avian scales. *J Anat.* 214:587–606.
- Eblaghie MC, et al. 2004. Interactions between FGF and Wnt signals and *Tbx3* gene expression in mammary gland initiation in mouse embryos. *J Anat.* 205:1–13.
- Ekblom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb)* 107:1–15.
- Ekblom R, Wolf JB. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl.* 7:1026–1042.
- Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* 9:18.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Eory L, et al. 2015. Avianbase: a community resource for bird genomics. *Genome Biol.* 16:21.
- Fan WL, et al. 2013. Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol Evol.* 5:1376–1392.
- Feng S, et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature* 587:252–257.
- Field DJ, Benito J, Chen A, Jagt JWM, Ksepka DT. 2020. Late Cretaceous neornithine from Europe illuminates the origins of crown birds. *Nature* 579:397–401.
- Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117: 9451–9457.
- Galvan I, Solano F. 2016. Bird integumentary melanins: biosynthesis, forms, function and evolution. *Int J Mol Sci.* 17:520.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.
- Haas BJ, Delcher AL, Wortman JR, Salzberg SL. 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20:3643–3646.
- Hahn C, Bachmann L, Chevreaux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Res.* 41:e129.
- Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15:1153–1160.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769.
- Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst.* 39:21–42.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 12:491.
- Hooper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat Ecol Evol.* 1: 1526–1534.
- Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol.* 34:2115–2122.
- Hughes SH. 2004. The RCAS vector system. *Folia Biol (Praha).* 50: 107–119.
- Hunt M, et al. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* 14:R47.
- Imsland F, et al. 2016. Regulatory mutations in *TBX3* disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nat Genet.* 48:152–158.
- Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Jarvis ED, et al. 2015. Phylogenomic analyses data of the avian phylogenomics project. *Gigascience* 4:4.
- Jarvis ED. 2016. Perspectives from the Avian Phylogenomics Project: questions that can be answered with sequencing all genomes of a vertebrate class. *Annu Rev Anim Biosci.* 4:45–59.
- Ji GG, et al. 2021. A gene co-expression network analysis of the candidate genes and molecular pathways associated with feather follicle traits of chicken skin. *J Anim Breed Genet.* 138:122–134.
- Jiang F, et al. 2021. A chromosome-level genome assembly of *Cairina moschata* and comparative genomic analyses. *BMC Genomics* 22: 581.
- Jiang TX, Tuan TL, Wu P, Widelitz RB, Chuong CM. 2011. From buds to follicles: matrix metalloproteinases in developmental tissue remodeling during feather morphogenesis. *Differentiation* 81:307–314.
- Johnson KP, Sorenson MD. 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence. *Auk* 116:792–805.

- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kear J. 2005. Ducks, geese and swans. Volume 1: General chapters, and species accounts (Anhima to Salvadorina). Oxford (UK): Oxford University Press.
- Kimball RT, et al. 2019. A phylogenomic supertree of birds. *Diversity-Basel* 11:109.
- Knief U, et al. 2016. Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* 17:199.
- Koepfli KP, Paten B, Genome K, O'Brien SJ. 2015. The genome 10K project: a way forward. *Annu Rev Anim Biosci.* 3:57–111.
- Ksepka DT, Bertelli S, Giannini NP. 2006. The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics.* 22:412–441.
- Lanfear R, Ho SY, Love D, Bromham L. 2010. Mutation rate is linked to diversification in birds. *Proc Natl Acad Sci U S A.* 107:20423–20428.
- Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera long mate pair libraries. *Bioinformatics* 30:566–568.
- Lever C. 2013. The mandarin duck. London (UK): Bloomsbury Publishing.
- Li W, et al. 2009a. PML depletion disrupts normal mammary gland development and skews the composition of the mammary luminal cell progenitor pool. *Proc Natl Acad Sci U S A.* 106:4725–4730.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
- Li F, et al. 2019. Heading north: Late Pleistocene environments and human dispersals in central and eastern Asia. *PLoS ONE* 14(5): e0216433.
- Li J, et al. 2021. A new duck genome reveals conserved and convergently evolved chromosome architectures of birds and mammals. *Gigascience* 10:giaa142.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
- Li W, Rich T, Watson CJ. 2009b. PML: a tumor suppressor that regulates cell fate in mammary gland. *Cell Cycle* 8:2711–2717.
- Liu L, Yu L, Kalavacharla V, Liu Z. 2011. A Bayesian model for gene family evolution. *BMC Bioinform.* 12:426.
- Liu G, Zhou L, Li B, Zhang L. 2014. The complete mitochondrial genome of *Aix galericulata* and *Tadorna ferruginea*: bearings on their phylogenetic position in the Anseriformes. *PLoS ONE* 9: e109701.
- Livezey BC. 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe Anatini) based on comparative morphology. *Auk* 108:471–507.
- Lowe CB, Clarke JA, Baker AJ, Haussler D, Edwards SV. 2015. Feather development genes and associated regulatory innovation predate the origin of Dinosauria. *Mol Biol Evol.* 32:23–28.
- Marcais G, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 14:e1005944.
- Mayr G. 2008. Phylogenetic affinities of the enigmatic avian taxon *Zygodactylus* based on new material from the early Oligocene of France. *J Syst Palaeontol.* 6:333–344.
- Mayr G. 2015. A reassessment of Eocene parrotlike fossils indicates a previously undetected radiation of zygodactyl stem group representatives of passerines (Passeriformes). *Zoolog Scr.* 44:587–602.
- Mayr G, Bertelli SA. 2011. A record of Rhynchaeites (Aves, Threskiornithidae) from the early Eocene Fur Formation of Denmark, and the affinities of the alleged parrot Mopsitta. *Palaeobiodivers Palaeoenvir.* 91:229–236.
- McGraw KJ, Safran RJ, Wakamatsu K. 2005. How feather colour reflects its melanin content. *Funct Ecol.* 19:816–821.
- Mills MG, Patterson LB. 2009. Not just black and white: pigment pattern development and evolution in vertebrates. *Semin Cell Dev Biol.* 20:72–81.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121.
- Musser G, Ksepka DT, Field DJ. 2019. New material of Palaeocene-Eocene Pellornis (Aves: Gruiformes) clarifies pattern and timing of the extant gruiform radiation. *Diversity* 11:102.
- Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol Ecol.* 25:1058–1072.
- Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinf.* 13(Suppl 14):S8.
- Noramly S, Freeman A, Morgan BA. 1999. Beta-catenin signaling can initiate feather bud development. *Development* 126:3509–3521.
- O'Brien SJ, Haussler D, Ryder O. 2014. The birds of genome10K. *Gigascience* 3:32.
- Ostrowski-Meissner HT. 1983. Effect of contamination of diets with aflatoxins on growing ducks and chickens. *Trop Anim Health Prod.* 15:161–168.
- Ou S, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20:275.
- Price-Waldman R, Stoddard MC. 2021. Avian coloration genetics: recent advances and emerging questions. *J Hered.* 112:395–416.
- Prum RO, Torres RH, Williamson S, Dyck J. 1998. Coherent light scattering by blue feather barbs. *Nature* 396:28–29.
- Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
- QCaro T, Stoddard MC, Stuart-Fox D. 2017. Animal coloration: production, perception, function and application. *Phil Trans R Soc Lond B Biol Sci.* 372:20170047.
- Qin N, et al. 2015. New insights into implication of the SLIT/ROBO pathway in the prehierarchal follicle development of hen ovary. *Poult Sci.* 94:2235–2246.
- QSmith ND, Ksepka DT. 2015. Five well-supported fossil calibrations within the “Waterbird” assemblage (Tetrapoda, Aves). *Palaeontol Electron.* 18:1.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 11:1432.
- Rawal S, Coulombe RA, Jr. 2011. Metabolism of aflatoxin B1 in turkey liver microsomes: the relative roles of cytochromes P450 1A5 and 3A37. *Toxicol Appl Pharmacol.* 254:349–354.
- RCarroll LS, Capecchi MR. 2015. Hoxc8 initiates an ectopic mammary program by regulating Fgf10 and Tbx3 expression and Wnt/beta-catenin signaling. *Development* 142:4056–4067.
- Rohnsgard PA. 1968. Some putative Mandarin Duck hybrids. *Bull Br Ornithol Club.* 88:140–148.
- RZhang Z, et al. 2018b. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *Gigascience* 7:gij027.
- Scott BB, Lois C. 2005. Generation of tissue-specific transgenic birds with lentiviral vectors. *Proc Natl Acad Sci U S A.* 102:16443–16447.
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol.* 1962:227–245.

- Shawkey MD, D'Alba L. 2017. Interactions between colour-producing mechanisms and their effects on the integumentary colour palette. *Phil Trans R Soc Lond B Biol Sci.* 372:20160536.
- Shi L, Quick M, Zhao Y, Weinstein H, Javitch JA. 2008. The mechanism of a neurotransmitter:sodium symporter–inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol Cell.* 30:667–677.
- Shurtleff LL, Savage C. 1996. *The wood duck and the mandarin: the northern wood ducks.* Berkeley (CA): University of California Press.
- Slack KE, et al. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol Biol Evol.* 23:1144–1155.
- Smeds L, Qvarnstrom A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* 26:1211–1218.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* 7(Suppl 1): S11.11–18.
- Suksaweang S, Jiang TX, Roybal P, Chuong CM, Widelitz R. 2012. Roles of EphB3/ephrin-B1 in feather morphogenesis. *Int J Dev Biol.* 56:719–728.
- Sun YH, et al. 2011. Sex ratio and survival of mandarin ducks in the Tachia river of central Taiwan. *Waterbirds* 34:509–513.
- Sun YH, et al. 2014. The seasonal home range and movements of mandarin ducks *Aix galericulata* on tributaries of the Tachia river, central Taiwan. *Forktail* 30:41–44.
- Sun Z, et al. 2017. Rapid and recent diversification patterns in Anseriformes birds: inferred from molecular phylogeny and diversification analyses. *PLoS ONE* 12:e0184529.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform.* Chapter 4:Unit 4.10.
- Thomson RC, Wang IJ, Johnson JR. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Mol Ecol.* 19:2184–2195.
- Vallon-Christersson J, et al. 2001. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families. *Hum Mol Genet.* 10:353–360.
- Venturini L, Caim S, Kaithakottil GG, Mapleson DL, Swarbreck D. 2018. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* 7:gij093.
- Vianna JA, et al. 2020. Genome-wide analyses reveal drivers of penguin diversification. *Proc Natl Acad Sci U S A.* 117: 22303–22310.
- Vinther J, Briggs DE, Clarke J, Mayr G, Prum RO. 2010. Structural coloration in a fossil feather. *Biol Lett.* 6:128–131.
- Vogel WF, Aszodi A, Alves F, Pawson T. 2001. Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. *Mol Cell Biol.* 21:2906–2917.
- Vurtture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202–2204.
- Widelitz RB. 2008. Wnt signaling in skin organogenesis. *Organogenesis* 4:123–133.
- Widelitz RB, Jiang TX, Lu J, Chuong CM. 2000. Beta-catenin in epithelial morphogenesis: conversion of part of avian foot scales into feather buds with a mutated beta-catenin. *Dev Biol.* 219:98–114.
- Widelitz RB, Veltmaat JM, Mayer JA, Foley J, Chuong CM. 2007. Mammary glands and feathers: comparing two skin appendages which help define novel classes during vertebrate evolution. *Semin Cell Dev Biol.* 18:255–266.
- Williams RS, et al. 2003. Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *J Biol Chem.* 278:53007–53016.
- Wu P, et al. 2015. Topographical mapping of alpha- and beta-keratins on developing chicken skin integuments: functional interaction and evolutionary perspectives. *Proc Natl Acad Sci U S A.* 112: E6770–E6779.
- Wu P, et al. 2018. Multiple regulatory modules are required for scale-to-feather conversion. *Mol Biol Evol.* 35:417–430.
- Wu P, et al. 2019. The balance of Bmp6 and Wnt10b regulates the telogen-anagen transition of hair follicles. *Cell Commun Signal.* 17:16.
- Wu J, Lu LY, Yu X. 2010. The role of BRCA1 in DNA damage response. *Protein Cell* 1:117–123.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Zhang G, et al. 2014a. Comparative genomic data of the Avian Phylogenomics Project. *Gigascience* 3:26.
- Zhang G, et al. 2014b. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018a. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19:153.
- Zhu F, et al. 2021. Three chromosome-level duck genome assemblies provide insights into genomic variation during domestication. *Nat Commun.* 12:5932.
- Zimin AV, et al. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.
- Zlotina A, et al. 2019. Heterochromatic regions in Japanese quail chromosomes: comprehensive molecular-cytogenetic characterization and 3D mapping in interphase nucleus. *Chromosome Res.* 27: 253–270.

Associate editor: Esther Betran