Molecules and Cells

Whole Genome Analysis of the Red-Crowned Crane Provides Insight into Avian Longevity

HyeJin Lee[1,8], Jungeun Kim[1,8], Jessica A. Weber[2,8], Oksung Chung[3], Yun Sung Cho[3], Sungwoong Jho[1], JeHoon Jun[3], Hak-Min Kim[4,5], Jeongheui Lim[6], Jae-Pil Choi[1], Sungwon Jeon[4,5], Asta Blazyte[4,5], Jeremy S. Edwards[7,*], Woon Kee Paek[6,*], and Jong Bhak[1,3,4,5,*]

[1]Personal Genomics Institute, Genome Research Foundation, Cheongju 28160, Korea, [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA, [3]Clinomics, Ulsan 44919, Korea, [4]KOGIC, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea, [5]Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Korea, [6]National Science Museum, Ministry of Science and ICT, Daejeon 34143, Korea, [7]Chemistry and Chemical Biology, UNM Comprehensive Cancer Center, University of New Mexico, Albuquerque, NM 87131, USA, [8]These authors contributed equally to this work.
*Correspondence: jsedwards@salud.unm.edu (JSE); paekwk@naver.com (WKP); jongbhak@genomics.org (JB)
https://doi.org/10.14348/molcells.2019.0190
www.molcells.org

The red-crowned crane (*Grus japonensis*) is an endangered, large-bodied crane native to East Asia. It is a traditional symbol of longevity and its long lifespan has been confirmed both in captivity and in the wild. Lifespan in birds is known to be positively correlated with body size and negatively correlated with metabolic rate, though the genetic mechanisms for the red-crowned crane's long lifespan have not previously been investigated. Using whole genome sequencing and comparative evolutionary analyses against the grey-crowned crane and other avian genomes, including the long-lived common ostrich, we identified red-crowned crane candidate genes with known associations with longevity. Among these are positively selected genes in metabolism and immunity pathways (*NDUFA5, NDUFA8, NUDT12, SOD3, CTH, RPA1, PHAX, HNMT, HS2ST1, PPCDC, PSTK CD8B, GP9, IL-9R,* and *PTPRC*). Our analyses provide genetic evidence for low metabolic rate and longevity, accompanied by possible convergent adaptation signatures among distantly related large and long-lived birds. Finally, we identified low genetic diversity in the red-crowned crane, consistent with its listing as an endangered species, and this genome should provide a useful genetic resource for future conservation studies of this rare and iconic species.

Keywords: genome, longevity, red-crowned crane

INTRODUCTION

The red-crowned crane (*Grus japonensis*) is one of the rarest cranes in the world and is a symbol of longevity in East Asia. It has a maximum lifespan of 30 years in the wild and 65 years in captivity (Rasmussen and Engstrom, 2004), which is substantially longer than the average lifespan recorded in 144 other avian species (Supplementary Table S1). It is also one of the largest avian species (Chen et al., 2012), measuring on average 150-158 cm long with a 220-250 cm wingspan (del Hoyo et al., 1996), and weighing on average 8.9 kg (John and Dunning, 2008) (ranging from 4.8 to 10.5 kg). Since avian body sizes are known to be positively correlated to lifespan and negatively correlated to metabolic rate (McKechnie and Wolf, 2004; Scholander et al., 1950), it has previously been suggested that the longevity of the red-crowned crane is related to its low metabolic rate (Speakman, 2005).

Received 26 August, 2019; revised 31 October, 2019; accepted 18 December, 2019; published online 14 January, 2020

eISSN: 0219-1032

There are currently two primary red-crowned crane populations, including a non-migratory population in northern Japan and a migratory continental group that ranges across Southeastern Russia, Northeastern China, Eastern Mongolia, and Korea (IUCN, 2017). It relies on wetlands for breeding and the loss and pollution of these habitats has resulted in severe population declines (Wang et al., 2011). It has been classified as threatened since 2000 and is classified as 'endangered' on the International Union for Conservation of Nature Red List (Yu et al., 2001). With only estimated 3,000 individuals in 2017, coupled with long generational lengths (12 years), it is one of the most endangered avian species (IUCN, 2017). Conservation genetic studies have primarily focused on microsatellite and mitochondrial markers, and an absence of genomic-level sequence data has precluded more comprehensive demographic modeling.

To infer the population history and evolutionary adaptations of this endangered species, we sequenced the first red-crowned crane whole genome and compared it to 18 other avian species. The Avian Phylogenomics Project (http://avian.genomics.cn/) previously published 48 avian reference genomes (Zhang et al., 2014), including the grey-crowned crane (*Balearica regulorum*) and the common ostrich (*Struthio camelus*). The grey-crowned crane belongs to the family of *Gruidae* together with the red-crowned crane, though the body mass of this crane (weighing up to 4 kg, height up to 1.4 m) (Walkinshaw, 1973) is smaller than that of red-crowned crane and the lifespan, particularly in captivity, is shorter (Wasser and Sherman, 2010). Conversely, the common ostrich is a member of the infraclass *Palaeognathae* and is deeply divergent from the red-crowned crane (111 million years ago, Mya) (Hedges et al., 2015). It is the largest Aves (up to 156.8 kg, up to 1.7 m in height) with a lifespan closer to the red-crowned crane (maximum longevity of 50 years in captivity) (Wasser and Sherman, 2010). To identify common evolutionary signatures associated with lifespan in birds, we compared our red-crowned crane genome to the genomes of the common ostrich (large/long-lived, but belonging to different family) and grey-crowned crane (belonging to the same family, but small/short-lived).

MATERIALS AND METHODS

Sample preparation, genomic DNA extraction and sequencing

The protocol for blood sample preparation was carried out in accordance with guidelines of Korean Association for Bird Protection under the Cultural Heritage Administration (Korea) permit. All experimental protocols were approved by the Genome Research Foundation. All methods were carried out in accordance with relevant guidelines and regulations. A blood sample was secured from a single female red-crowned crane collected upstream of Gunnam dam, Gyeonggi-do, Republic of Korea (Lat = 38°06'24.8"N and Long = 127°01'17.1"E). Genomic DNA was extracted using a QIAamp DNA Mini Kit (Qiagen, USA) following the manufacturer's instructions. The DNA concentration was measured using the Qubit dsDNA assay kit (Invitrogen, USA) and an Infinite 200 PRO Nanoquant system (Tecan, Germany). Fragmentation of high-molecular weight genomic DNA was carried out with a Covaris S2 Ultrasonicator (Covaris, USA), generating 500 bp fragments. Whole-genome shotgun libraries were prepared using a TruSeq library sample prep kit (Illumina, USA). Aliquots were analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) to determine the library concentration and size. Sequencing was performed on an Illumina HiSeq 2000 sequencer (Illumina), using the TruSeq Paired-End Cluster Kit v3 (Illumina) and the TruSeq SBS HS Kit v3 (Illumina) for 200 cycles.

Red-crowned crane genome assembly by reference mapping

Low quality DNA reads with phred quality scores < 20 and/or ambiguous nucleotides 'N' ratios > 10% were filtered out. Clean reads were aligned to the grey-crowned crane genome sequence using BWA-aln 0.6.2 (Li and Durbin, 2009) at default settings. Polymerase chain reaction duplicates from the reads were removed using 'rmdup' command of SAMtools 0.1.18 (Li et al., 2009) at default settings. Complete consensus sequences were determined by SAMtools mpileup, Bcftools view, and SAMtools vcfutils.pl vcf2fq pipelines from the SAMtools 0.1.19 suite (Li et al., 2009). The consensus sequence included all alignment depths. Single nucleotide variant (SNV) calling was conducted based on the consensus genome sequences (McKenna et al. 2010; Van der Auwera et al., 2013). Indels were called using the GATK 3.3 UnifiedGenotyper with the -dcov 1000 option. The indels were marked by VariantFiltration in GATK 3.3 with the following criteria: (1) hard to validate, MQ0 ≥ 4 && ((MQ0 / (1.0 × DP)) > 0.1); (2) quality filter, QUAL < 10.0; (3) depth filter, DP < 5. The SnpEff 3.3 software was used to predict the effects of the indels. Heterozygous SNV rate was used as the nucleotide diversity value. The average avian nucleotide diversity was calculated by averaging the values of the American flamingo (0.00372), Anna's hummingbird (0.00288), common ostrich (0.00176), downy woodpecker (0.00455), grey-crowned crane (0.00202), peregrine falcon (0.00112), white-throated tinamou (0.00560), and white-tailed tropicbird (0.00162). Coding sequence (CDS) were selected by filtering out the following conditions: (1) there is an ambiguous nucleotide 'N' in a CDS; (2) there is a premature stop codon in a CDS. Among the 14,173 CDSs, 13,407 CDSs were selected and used in subsequent analyses. *K*-mers (*K* = 17) were obtained using the JELLYFISH 2.2.4 (Marcais and Kingsford, 2011) program with the command "jellyfish count -m 17 -o output -c 32 -s 8589934592 -t 64". We calculated the genome size from the total *K*-mer frequencies.

Orthologous gene clustering

The protein sequences of the red-crowned crane were predicted using the gene model of the grey-crowned crane. We downloaded 18 high quality avian reference genomes from the GigaDB dataset (Zhang et al., 2014) (http://gigadb.org/dataset/101000). Orthologous gene clusters were constructed using OrthoMCL 2.0.9 (Li et al., 2003). Self-to-self BLASTP searches were conducted for all the protein sequences of the red-crowned crane and reference genomes with an E-value of 1E-5. The Markov clustering algorithm was used to define

the cluster structure between the proteins with inflation values (-I) of 1.5. Lastly, the OrthoMCL program was used to determine orthologous gene clusters with an e-value exponent cutoff value of $1.0 \times e^{-5}$ and 'percentMatchCutoff' value of 50%.

Evolutionary analyses

We applied the tree topology and divergence times of the 18 avian reference genomes represented from Jarvis et al. (2015) to the positively selected gene (PSG) analysis. The divergence time of the red-crowned crane and grey-crowned crane was taken from TimeTree database (Hedges et al., 2015). The white-throated tinamou's divergence time was also taken from TimeTree (Hedges et al., 2015). The multiple sequence alignment of orthologous genes was constructed using PRANK alignment program package (Loytynoja and Goldman, 2010), and the rates of synonymous (d_S) and non-synonymous substitutions (d_N) and the d_N/d_S ratio (nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site, ω) were estimated using the CODEML program in PAML 4.5 (Yang, 1997). The branch-site test of positive selection was conducted with a 10% false discovery rate criterion. The red-crowned and common ostrich were used as the foreground branch for each analysis accordingly, respectively, and the other species were used as background branches. Functional effects of the species-specific amino acid (SSAA) changes were predicted using PolyPhen-2 (Adzhubei et al., 2010) and PROVEAN (Choi et al., 2012) programs using the default cutoff values. Gene enrichment analysis was conducted with the database for annotation, visualization, and integrated discovery (DAVID) database (Huang et al., 2008) and network drawn by Cytoscape (ver. 3.6.1) (Shannon et al., 2003).

Demographic history

Whole genome sequencing data can be used to accurately estimate the population history (Li and Durbin, 2011; Osada, 2014). We inferred the demographic history of the red-crowned crane using a pairwise sequentially Markovian coalescent (PSMC) analysis with scaffolds of length $\geq$ 50 kb. For grey-crowned crane, we downloaded illumina short reads from National Center for Biotechnology Information (Acc. PRJNA212879) and used in the PSMC analysis. The PSMC analysis was performed with 100 bootstrapping rounds, with 9.59×10^{-9} substitutions per site per generation and generation time of 12.3 years, as previously reported (IUCN, 2017).

RESULTS AND DISCUSSION

The red-crowned crane genome

Genomic DNA from a red-crowned crane individual was sequenced using the Illumina HiSeq2000 platform. We produced a total of 989 million paired-end reads with a 100 bp read length and an insert size of 216 bp (Supplementary Table S2, Supplementary Fig. S1). After trimming the low-quality reads, we obtained 750 M (75.86%) reads, yielding 75 Gb and a depth of coverage of 66x. A *K*-mer analysis with K = 17-21 was performed and the red-crowned crane's genome size was estimated to be 1.146 Gb (Supplementary Table S3, Supplementary Fig. S2), which is comparable to other avian species (0.96-2.2 Gb) (Zhang et al., 2014). Its genome was assembled by mapping the short reads to the grey-crowned crane (*Balearica regulorum*) reference (Zhang et al., 2014) (Supplementary Table S4), which has a 20 Mya divergence time (Hedges et al., 2015). A total of 610 M red-crowned crane reads (81.08%) were mapped to the 1.13 Gb grey-crowned crane reference (Supplementary Table S4). We obtained 94.1% genomic regions with more than 10× sequencing depth (Supplementary Fig. S3). The reads covered 99.77% of CDS region with a mapping depth of at least 1×, and 99.02% of the CDS regions had higher than a 5× sequencing depth (Supplementary Table S4). The total GC content was 41.75% (Supplementary Table S4). Using the grey-crowned crane reference genome, approximately 36 million SNVs and 3.6 million small insertions/deletions (indels) were identified (Supplementary Table S5). These SNVs were comprised of 34,733,024 (96%) homozygous and 1,433,466 (4%) heterozygous SNVs, yielding a nucleotide diversity of 1.29×10^{-3} nucleotide substitutions per site; which is lower than nucleotide diversity of the grey-crowned crane (1.88×10^{-3}). Compared with a median nucleotide diversity of 2.49×10^{-3} in least concern avian species (Zhang et al., 2014), the red-crowned crane possesses a low level of nucleotide diversity and is genetically consistent with an endangered species. The red-crowned crane consensus sequences, covering a total of 14,173 genes (Supplementary Table S4), were created by substituting the red-crowned crane SNVs into the grey-crowned crane reference genome (see Materials and Methods section).

Comparative evolutionary analyses of the red-crowned crane

To investigate the environmental adaptations of the red-crowned crane and the genetic basis of its longevity (Table 1) (Ji and DeWoody, 2017), we analyzed the red-crowned crane consensus sequence compared to 18 other avian genomes (Supplementary Table S6). Among these genomes, we identified a total of 31,795 orthologous genes, 7,455 of which were conserved across all 18 of the avian genomes (Fig. 1A). Of these, 5,878 orthologous families had one-to-one relationships among the 19 avian genomes. We used a phylogenetic tree topology which was constructed by Bayesian phylogenomic analysis (Zhang et al., 2014) (Fig. 1B) and estimated the divergence time with one-to-one orthologous genes. It estimated the divergence time between the red-crowned and grey-crowned cranes to be 20 Mya. The reported divergence time between the red-crowned crane and the common ostrich (used as an outgroup in this study) is roughly 111 Mya (Hedges et al., 2015).

Avian body sizes are known to be positively correlated with lifespan (Supplementary Table S1) and negatively correlated with metabolic rate (Zhang et al., 2014). In order to investigate the contribution of genes in these pathways to the longevity of the red-crowned crane, PSGs were identified among the 5,878 one-to-one orthologues shared by all 19 species. A total of three and 162 genes were identified as putative PSGs in the red-crowned crane by applying the branch-site model and branch model, respectively (Supplementary Tables S7

Table 1. Reference genomes and biological traits

	Latin name	Common name	Abbreviation	Height (cm)	Body weight (kg)	Metabolic rate (W/kg)	Lifespan (y)	
							Wild	Captivity
1	*Phoenicopterus ruber ruber*	American flamingo[c]	PHORU	120-145	2.8	15.254[d]	-	-
2	*Calypte anna*	Anna's hummingbird[b]	CALAN	10-11	0.004	-	8.5	-
3	*Haliaeetus leucocephalus*	Bald eagle[b]	HALEU	70-102	5.6 (female) 4.1 (male)	-	-	-
4	*Melopsittacus undulatus*	Budgerigar (parakeet)[b]	MELUN	18	0.03-0.04	9.8[e]	-	21.0
5	*Cuculus canorus*	Common cuckoo[b]	CUCCA	32-34	0.11-0.13	0.838[d]	12.9	-
6	*Struthio camelus australis*	Common ostrich[b]	STRCA	170-280	63-145	6.305[d]	-	70.0
7	*Picoides pubescens*	Downy woodpecker[b]	PICPU	14-17	0.02-0.03	0.383[d]	11.9	-
8	*Aptenodytes forsteri*	Emperor penguin[b]	LEPDI	110-130	23	42.871[d]	-	23.4
9	*Balearica regulorum gibbericeps*	Grey-crowned crane[c]	BALRE	100	3.5	-	-	27.2
10	*Ophisthocomus hoazin*	Hoatzin[b]	OPHHO	65	0.8	-	-	-
11	*Charadrius vociferus*	Killdeer[b]	CHAVO	23-27	0.09	0.416[d]	10.9	-
12	*Egretta garzetta*	Little egret[b]	EGRGA	55-65	0.35-0.55	-	22.34	-
13	*Anas platyrhynchos domestica*	Peking duck[b]	ANAPL	50-76	1.6-2.3	4.068[d]	23.4	-
14	*Falco peregrinus*	Peregrine falcon[b]	FALPE	34-58	0.33-1.5	-	-	-
15	*Gallus gallus*	Chicken[a]	GALGA	30-45	1-3	6.005[d]	-	30.0
16	*Columba livia*	Rock pigeon[b]	COLLI	32-37	0.36	1.714[d]	-	-
17	*Phaethon lepturus*	White-tailed tropicbird[c]	PHALE	71-80	0.33	-	-	-
18	*Tinamus guttatus*	White-throated tinamou[b]	THGUT	32-36	-	-	-	-
19	*Grus japonensis*	Red-crowned crane	GRJAP	150-158	8.9	-	30.0	65.0

-, data not available.
[a]Sanger, [b]high-coverage genomes, [c]low-coverage genomes.
Basal metabolic rate values were obtained from the literature [d]McKechnie and Wolf (2004), [e]Ji and DeWoody (2017).

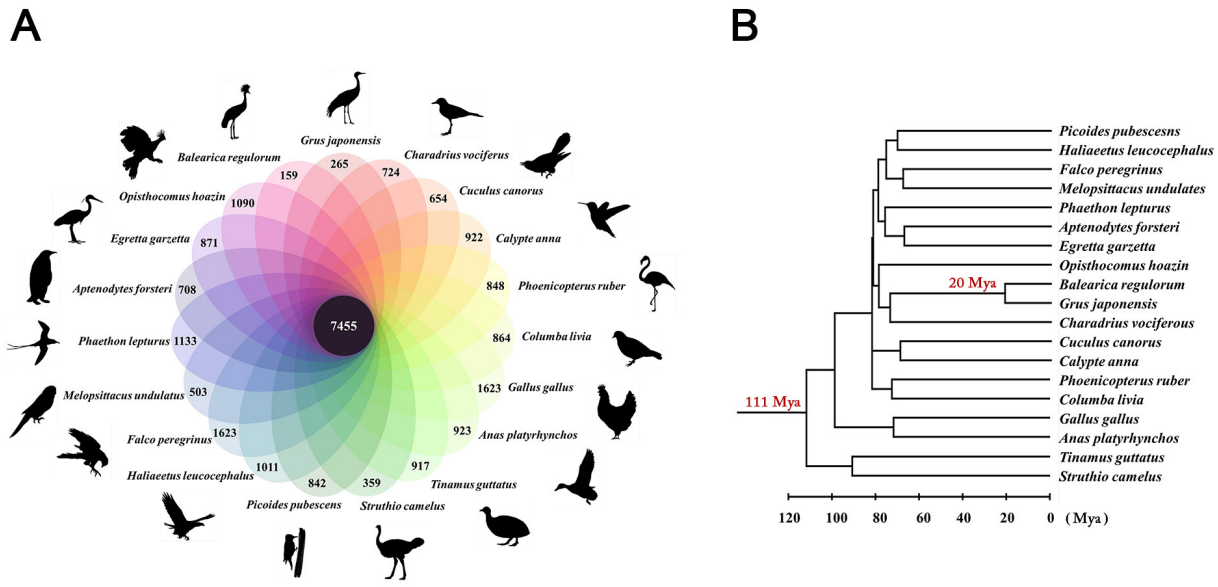

Fig. 1. Orthologous gene clusters of the red-crowned crane compared to the other avian species. (A) The Venn diagram shows the number of unique (number in the ovals) and shared (the number in the center black circle) gene families analyzed by OrthoMCL (Li et al., 2003). It was drawn manually using the online photo editor (https://pixlr.com/editor/). (B) Divergence time of the red-crowned-crane and grey-crowned crane (Hedges et al., 2015).

and S8). With the PSG set, a functional enrichment analysis of the red-crowned crane PSGs was conducted using the DAVID functional annotation tool (Huang et al., 2008) for Gene Ontology (GO) categories (Fig. 2A, Supplementary Table S9) and the Fisher's exact tests for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Fig. 2B, Supplementary Table S10). Two statistically enriched GO terms included "ribonucleoprotein complex biogenesis" (GO:0022613; *EBNA1BP2*, *PHAX*, *PDCD11*, *LSM6*, and *FTSJ2*, $P = 3.7 \times e^{-3}$) and "ncRNA processing" (GO:0034470; *PDCD11*, *LSM6*, *RG9MTD1*, *FTSJ2*, and *RPP14*, $P = 4.3 \times e^{-3}$) (Fig. 2A).

Telomere length has been proposed to be a 'molecular clock' that underlies organismal aging (Collins, 2008; Collins and Mitchell, 2002). Interestingly, the *PHAX* (phosphorylated adaptor for RNA export) gene plays an essential role in the human telomerase RNA (hTR) export to Cajal body (Boulon et al., 2004) enabling hTR maturation and telomerase localization (Egan and Collins, 2012; Schmidt and Cech, 2015). In addition, numerous reports indicate that immune functions are associated with aging (Accardi and Caruso, 2018) and we identified eight red-crowned crane PSGs (*CD8B, GP9, IL-9R, PTPRC, LOC102094583, LOC104630894, LOC104643762*, and *LOC104641139*) which were significantly enriched in immunity associated pathways (*P* value < 0.01); including hematopoietic cell lineage (KO04640), cell adhesion molecules (KO04514), primary immunodeficiency (KO05340), antigen processing and presentation (KO04612), and T cell receptor signaling pathway (KO04660) (Fig. 2B).

We further investigated red-crowned crane PSGs involved in metabolic pathways, which are associated with longevity (Speakman, 2005), by searching against the KEGG database (Kanehisa and Goto, 2000). We could identify seven red-crowned crane PSGs (*HNMT*, *HS2ST1*, *NDUFA5*, *NDUFA8*, *NUDT12*, *PPCDC*, and *PSTK*) that were predicted by the branch-site model (Table 2). Of these, two genes were associated with energy metabolism, including *NDUFA5* (ω average: 0.19652, ω red-crowned crane: 2.2221) and *NDUFA8* (ω average: 0.13284, ω red-crowned crane: 1.1281). *NDUFA5* and *NDUFA8* encode the Q-module and ND1 module of NADH-ubiquinone oxidoreductase (Complex 1; EC 1.6.5.3), respectively (Supplementary Fig. S4). Complex 1 functions in the transfer of electrons derived by NADH to the respiratory chain, and is a major source of damaging reactive oxygen species (ROS) that are known to be involved in aging and energy metabolism processes (Stroud et al., 2016). Previous studies of *NDUFA5* mutants in *Caenorhabditis elegans* have shown this gene to be involved in lifespan extension in the nematode and the regulation of lifespan in yeast (Chen et al., 2012; Valentine and Valentine, 2014). Furthermore, it was reported that expression levels of both *NDUFA5* and *NDUFA8* are decreased in the mitochondria of older humans (Ghosh et al., 2011).

Comparative genetic analyses of metabolic genes among long-lived Aves

The common ostrich has a lifespan comparable to the red-crowned crane (Wasser and Sherman, 2010), and it is possible that the longevity of these species could be due in part to convergent polymorphisms affecting the rate of energy metabolism. To identify any shared adaptations for longevity-related physiological characteristics, we compared the PSGs of the red-crowned crane with that of the common ostrich. We identified 36 and 76 ostrich PSGs using branch-site and branch model, respectively (Supplementary Tables S11 and S12). Six of the ostrich PSGs (*KIAA1009*, *NOVA1*, *SOD3*, *SOX14*, *TMEM80*, and *UBXN11*) were also positively

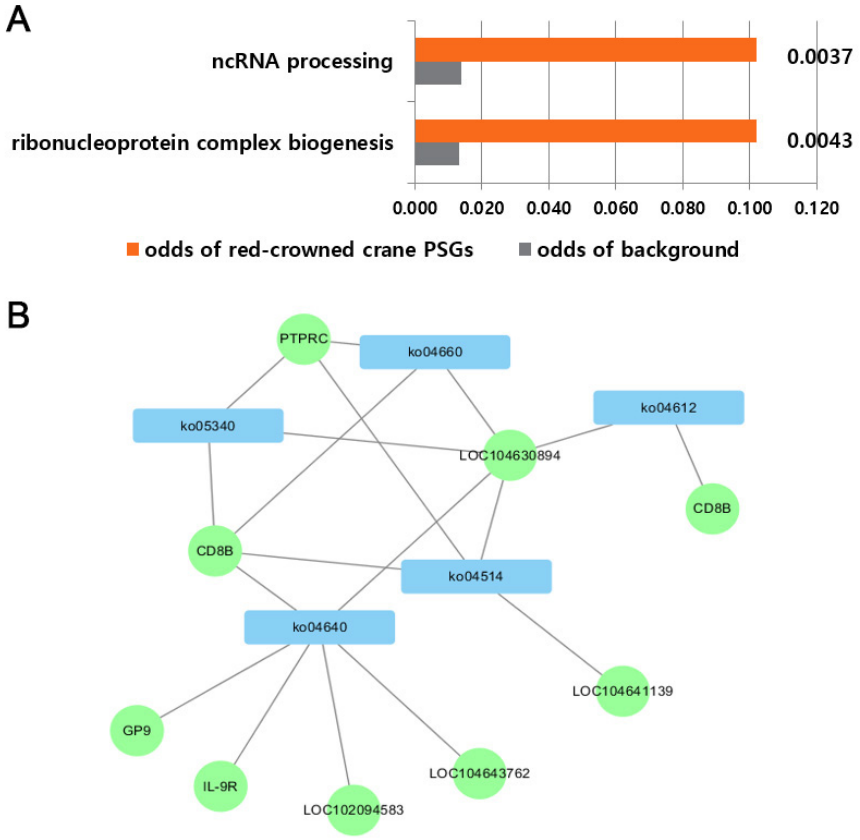

Fig. 2. Functional analysis of the red-crowned crane PSGs. (A) Red-crowned crane PSGs enriched in GO categories. The values to the right of the bar represent the p-values from the enrichment test analyzed on the DAVID website (Huang et al., 2008). (B) Network diagram of the red-crowned crane PSGs associated with KEGG pathways. The blue rectangles represent pathway IDs from the KEGG database and the green circles represent the red-crowned crane PSGs. Each edge represents the association between the pathways and the genes.

Table 2. PSGs involved in pathways related to metabolism in the red-crowned crane

KEGG pathways	Branch model PSGs	d_N/d_S of all avian means	d_N/d_S of the red-crowned crane
Histidine metabolism	*HNMT*	0.3162	1.5981
Oxidative phosphorylation	*NDUFA8*	0.1328	1.1281
	NDUFA5	0.1965	2.2221
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	*HS2ST1*	0.0437	1.1207
Metabolic pathways	*NUDT12*	0.0304	1.0047
	NDUFA8	0.1328	1.1281
	NDUFA5	0.1965	2.2221
	PPCDC	0.1664	1.7130
Nicotinate and nicotinamide metabolism	*NUDT12*	0.0304	1.0047
Pantothenate and CoA biosynthesis	*PPCDC*	0.1664	1.7130
Selenocompound metabolism	*PSTK*	0.3001	1.1039

selected in the red-crowned crane (Table 3). *SOD3* is associated with the superoxide metabolic process (GO:0006801) GO molecular function (Ashburner et al., 2000). The free-radical theory of aging is one of the major hypotheses of biological aging, and suggests that aging is a result of accumulated damage from ROS (Speakman, 2005). Organisms have a variety of protection mechanisms to mitigate the deleterious effects of the ROS (Held, 2012), and *SOD3* is an antioxidant enzyme that plays a critical role in converting the superoxide anion radical into hydrogen peroxide (Supplementary Fig.

Table 3. PSGs shared in both red-crowned crane and common ostrich

PSGs	Statistics of PSGs prediction of red-crowned crane				Statistics of PSGs prediction of common ostrich			
	Branch models		Branch-site models		Branch models		Branch-site models	
	M0:one-ratio[a]	M1:free-ratio[b]	2ΔL[c]	*P* value	M0:one-ratio[a]	M1:free-ratio[b]	2ΔL[c]	*P* value
KIAA1009	0.4014	1.0295	-	-	-	-	10.8273	0.0155
NOVA1	0.1016	1.5792	-	-	0.1016	1.8746	7.9132	0.0489
SOD3	0.1454	1.2380	-	-	0.1454	1.4134	-	-
SOX14	0.0186	1.4472	-	-	0.0186	1.6574	-	-
TMEM80	0.2138	2.1448	-	-	0.2138	2.6535	-	-
UBXN11	0.4847	2.1552	-	-	0.4847	1.3068	-	-

-, no statistical significance.
[a]M0 denotes d_N/d_S for all the species in the study.
[b]M1 denotes d_N/d_S for the lineage leading to the target species.
[c]2ΔL: Likelihood ratio tests (LRT) were used to detect positive selection.

S5A) (Gentschew et al., 2013; Held, 2012). In addition, the overexpression of *SOD3* results in increased cell growth and proliferation, and the inhibition of apoptosis (Mendez et al., 2005). We speculate that the *SOD3* gene product may help protect the tissues and organs from oxidative stress, consequently increasing longevity.

The SSAAs with possible function-change can provide evolutionary information to given species. We hypothesized that the common genes with function altering SSAAs in both red-crowned crane and common ostrich are probably associated with longevity by comparing common ostrich SSAAs to the other 18 avian species with an average lifespan. We could identify 10,126 SSAAs for the common ostrich from 3,551 genes (Supplementary Table S13). Among them, 2,001 genes contained at least one function altering SSAA (in a total of 3,698 ostrich SSAAs). Interestingly, 435 genes with ostrich function-altering SSAAs were shared with the red-crowned crane, suggesting the possibility of some level of convergent evolution (Supplementary Table S14). The permutation test of these genes identified an enrichement of "cilium morphogenesis" (*P* value of $1.2 \times e^{-4}$) and "centriole" (*P* value $1.2 \times e^{-2}$). Notably, cilium assembly and dis-assembly are unique mechanisms that regulate cell physiology during proliferation and morphogenesis (Eggenschwiler and Anderson, 2007; Loncarek and Bettencourt-Dias, 2017; Stroud et al., 2016), and are associated with replicative aging such as the aging of skin and the hemopoietic system (Tkemaladze and Chichinadze, 2005). To further examine the role of these genes in aging, we queried the red-crowned crane and common ostrich predicted function-altering PSGs in the GeneAge database (Tacutu et al., 2013), and identified a total of eleven genes (*BUB1B, MDM2, ERCC8, RPA1, NCOR2, CTH, MTBP, SOD3, NCOR2, CIT*, and *NCOR1*) for the red-crowned crane and seven genes in the common ostrich (*ALG6, TKT, CTH, MDH2, GCLC, SDHB*, and *COQ3*). Interestingly, the only common gene, *CTH* (*Cystathionine gamma lyase*), encodes a cytoplasmic enzyme in the trans-sulfuration pathway that converts cystathione derived from methionine into cysteine (Supplementary Fig. S5B). Glutathione is a well-known anti-oxidant known to contribute to longevity, and glutathione synthesis in the liver is dependent upon the availability of cysteine. Higher levels cystathionine are associated with aging and produce an augmented oxidant effect in elderly humans (Hernanz et al., 2000). In addition, it was reported that RNAi-mediated knockdown of *CTH-1* and *CTH-2* in *C. elegans* caused median lifespan reductions (up to ~15%) (Hine et al., 2015). These biological data observed in other models raise the intriguing possibility that similar mechanisms may increase longevity in the red-crowned crane and common ostrich.

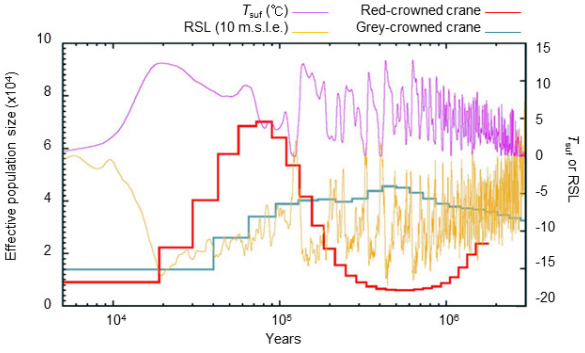

Fig. 3. Demographic history of the red-crowned crane and grey-crowned crane. PSMC demographic modeling of the red-crowned crane, using a generation time (g=12.3 years) and mutation rate per generation time ($\mu=9.66\times10^{-9}$), and the grey-crowned crane, with a generation time (g=15.1) and mutation rate ($\mu=1.18\times10^{-8}$). T_{suf}, atmospheric surface air temperature; RSL, relative sea level; 10 m.s.l.e., 10 m sea level equivalent.

Genetic diversity and population history

A PSMC analysis was used to model the historical population size fluctuations of the red-crowned crane and the closely related grey-crowned crane (Fig. 3) (Li and Durbin, 2011). We hypothesized that historical population size of these two cranes helps understand crane's adaptation associated with environmental condition. Based on the genomic data of the grey-crowned crane (Balericinae) and red-crowned crane (Gruinae), we estimated effective population size (N_e) (Fig. 3). We discovered a striking difference in the effective pop-

ulation size between the two species during last glacial age. While red-crowned crane's effective population size appears to have peaked about 80,000 years ago in the last glacial age, the grey-crowned crane effective population size was at its maximum roughly 500,000 years ago. The N_e of the grey-crowned crane increased from 1 Mya, and peaked at 0.4 Mya, with an estimated N_e ~4.4×10^4. Conversley, we observed a N_e decrease in the red-crowned crane from the late Pliocene to the early Pleistocene (~0.4 Mya). This is consistent with the fossil records, which is biased towards the Gruinae subfamily (Krajewski et al., 2010). Thus, we can infer that the environment was less favorable for the Gruinae at this time. The N_e of the red-crowned crane was increased at late Plestocene (0.7-0.1 Mya), at the end of last ice age. The population size of the red-crowned crane is predicted to be N_e ~7.0×10^4, and was the highest roughly 80,000 years ago (Kya). While it is possible that the red-crowned crane's northern habitat naturally suppresses their effective population size, it is also known that habitat loss and pollution have contributed to severe population declines.

In conclusion, we generated and analyzed the first whole genome sequence data for the red-crowned crane. Evolutionary genomic analyses of 19 avian species, including the long-lived red-crowned crane and common ostrich, yielded PSGs and function-altering SSAA candidate genes associated with longevity. Functional annotations and enrichment analyses were conducted using GO, KEGG pathways, and GeneAge databases. Both PSGs and function-altering SSAA analyses showed candidate genes responsible for longevity in red-crowed crane by using bioinformatic analysis and discussed longevity-related biological and physiological features previously identified in both human and animal models. Demographic modeling of the red-crowned crane revealed low genetic diversity and a trend of population decline. Taken together, these findings further highlight the importance of continued monitoring and management of this endangered species. Despite the SNV based genetic differences predicted by PSGs and SSAAs which provided us with some speculative interpretations on how such variations affected certain longevity related genes in the red-crowned crane species, it should be noted that the distinct phenotypic differences between the grey-crowned crane reference and whole genome sequences of red-crowned crane that are mapped to the reference may not be fully explained by SNPs only. SNPs are usually functionally relevant in expressed genes and longevity as a biological phenomenon is a whole species evolution and regulation problem that involves numerous regulatory mechanisms of transcriptional, translational, and even epigenetic factors which are often associated with the genome structure itself. Therefore, it will be absolutely necessary to produce a high quality red-crowned crane de novo assembly in the future to compare the structural and copy number variations between the two species' references with long DNA read generation methods such as PacBio SMRT and Oxford Nanopore sequencing. Nevertheless, we hope the consensus genome presented here will be a valuable resource for aging research and future conservation genetic studies of this iconic species.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

Disclosure

O.C., Y.S.C., and J.J. are employees and J.B. is one of the founders of Clinomics Ltd. All other authors have no potential conflicts of interest to disclose.

AVAILABILITY OF DATA AND MATERIALS

Whole-genome sequence data were deposited in the SRA database at NCBI with BioSample accession number SAMN07580860. The data can be accessed via reference number SRX3148473 or through BioProject accession number PRJNA400839. Currently, another genome project of the red-crowned crane was reported (PRJNA400839).

ACKNOWLEDGMENTS

This study was supported by PGI of Genome Research Foundation and Clinomics Ltd. internal research funds and the No.10075262 of National Center for standard Reference Data. This research was also supported by Ulsan National Institute of Science & Technology (UNIST) internal research funds. And the National Research Foundation of Korea (2013M3A9A5047052 and 2017M3A9A5048999) research funds.

The authors thank many people not listed as authors who provided analyses, data, feedback, samples, and encouragement. Especially, thanks to Taehyung Kim and Byungchul Kim.

ORCID

HyeJin Lee	https://orcid.org/0000-0002-8937-4142
Jungeun Kim	https://orcid.org/0000-0002-6576-5456
Jessica A. Weber	https://orcid.org/0000-0001-8405-4396
Oksung Chung	https://orcid.org/0000-0001-5003-4071
Yun Sung Cho	https://orcid.org/0000-0003-4490-8769
Sungwoong Jho	https://orcid.org/0000-0002-7488-6268
JeHoon Jun	https://orcid.org/0000-0002-6558-1091
Hak-Min Kim	https://orcid.org/0000-0001-6066-2469
Jeongheui Lim	https://orcid.org/0000-0003-0045-3037
Jae-Pil Choi	https://orcid.org/0000-0002-7048-7753
Sungwon Jeon	https://orcid.org/0000-0002-2729-9087
Asta Blazyte	https://orcid.org/0000-0001-7309-1482
Jeremy S. Edwards	https://orcid.org/0000-0003-3694-3716
Woon Kee Paek	https://orcid.org/0000-0002-3178-9134
Jong Bhak	https://orcid.org/0000-0002-4228-1299

REFERENCES

Accardi, G. and Caruso, C. (2018). Immune-inflammatory responses in the elderly: an update. Immun. Ageing *15*, 11.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248-249.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25-29.

Boulon, S., Verheggen, C., Jady, B.E., Girard, C., Pescia, C., Paul, C., Ospina, J.K., Kiss, T., Matera, A.G., Bordonne, R., et al. (2004). PHAX and CRM1 are

required sequentially to transport U3 snoRNA to nucleoli. Mol. Cell *16*, 777-787.

Chen, W., Ma, J., Zhang, H., Li, D., and Zhang, X. (2012). Behavioural alterations in domestication process: comparative studies between wild, captive and inbred red-crowned cranes (*Grus japonensis*). J. Anim. Vet. Adv. *11*, 2711-2715.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One *7*, e46688.

Collins, K. (2008). Physiological assembly and activity of human telomerase complexes. Mech. Ageing Dev. *129*, 91-98.

Collins, K. and Mitchell, J.R. (2002). Telomerase in the human organism. Oncogene *21*, 564-579.

del Hoyo, J., Elliott, A., and Sargatal, J. (1996). Handbook of the Birds of the World (Barcelona: Lynx Edicions).

Egan, E.D. and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. RNA *18*, 1747-1759.

Eggenschwiler, J.T. and Anderson, K.V. (2007). Cilia and developmental signaling. Annu. Rev. Cell Dev. Biol. *23*, 345-373.

Gentschew, L., Flachsbart, F., Kleindorp, R., Badarinarayan, N., Schreiber, S., and Nebel, A. (2013). Polymorphisms in the superoxidase dismutase genes reveal no association with human longevity in Germans: a case–control association study. Biogerontology *14*, 719-727.

Ghosh, S., Lertwattanarak, R., Lefort, N., Molina-Carrion, M., Joya-Galeana, J., Bowen, B.P., Garduno-Garcia Jde, J., Abdul-Ghani, M., Richardson, A., DeFronzo, R.A., et al. (2011). Reduction in reactive oxygen species production by mitochondria from elderly subjects with normal and impaired glucose tolerance. Diabetes *60*, 2051-2060.

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

Held, P. (2012). An Introduction to Reactive Oxygen Species: Measurement of ROS in Cells. Application Guide (Winooski: BioTek Instruments).

Hernanz, A., Fernandez-Vivancos, E., Montiel, C., Vazquez, J.J., and Arnalich, F. (2000). Changes in the intracellular homocysteine and glutathione content associated with aging. Life Sci. *67*, 1317-1324.

Hine, C., Harputlugil, E., Zhang, Y., Ruckenstuhl, C., Lee, B.C., Brace, L., Longchamp, A., Trevino-Villarreal, J.H., Mejia, P., Ozaki, C.K., et al. (2015). Endogenous hydrogen sulfide production is essential for dietary restriction benefits. Cell *160*, 132-144.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

IUCN (International Union for Conservation of Nature) (2017). The IUCN Red List of Threatened Species. IUCN Data, https://www.iucnredlist.org/species/22692167/93339099.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., et al. (2015). Phylogenomic analyses data of the avian phylogenomics project. GigaScience *4*, 4.

Ji, Y. and DeWoody, J.A. (2017). Relationships among powered flight, metabolic rate, body mass, genome size, and the retrotransposon complement of volant birds. Evol. Biol. *44*, 261-272.

John, B. and Dunning, J. (2008). CRC Handbook of Avian Body Masses (Florida: CRC Press).

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27-30.

Krajewski, C., Sipiorski, J.T., and Anderson, F.E. (2010). Complete mitochondrial genome sequences and the phylogeny of cranes (gruiformes: gruidae). Auk *127*, 440-452.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493-496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178-2189.

Loncarek, J. and Bettencourt-Dias, M. (2017). Building the right centriole for each cell type. J. Cell Biol. *217*, 823-835.

Loytynoja, A. and Goldman, N. (2010). webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics *11*, 579.

Marcais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764-770.

McKechnie, A.E. and Wolf, B.O. (2004). The allometry of avian basal metabolic rate: good predictions need good data. Physiol. Biochem. Zool. *77*, 502-521.

McKenna, A.H., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297-1303.

Mendez, J.I., Nicholson, W.J., and Taylor, W.R. (2005). SOD isoforms and signaling in blood vessels: evidence for the importance of ROS compartmentalization. Arterioscler. Thromb. Vasc. Biol. *25*, 887-888.

Osada, N. (2014). Extracting population genetics information from a diploid genome sequence. Front. Ecol. Evol. *2*, 7.

Rasmussen, P.C. and Engstrom, R.T. (2004). Threatened birds of Asia: the Birdlife international red data book. Auk *121*, 619-622.

Schmidt, J.C. and Cech, T.R. (2015). Human telomerase: biogenesis, trafficking, recruitment, and activation. Genes Dev. *29*, 1095-1105.

Scholander, P.F., Hock, R., Walters, V., and Irving, L. (1950). Adaptation to cold in arctic and tropical mammals and birds in relation to body temperature, insulation, and basal metabolic rate. Biol. Bull *99*, 259-271.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498-2504.

Speakman, J.R. (2005). Body size, energy metabolism and lifespan. J. Exp. Biol. *208*, 1717-1730.

Stroud, D.A., Surgenor, E.E., Formosa, L.E., Reljic, B., Frazier, A.E., Dibley, M.G., Osellame, L.D., Stait, T., Beilharz, T.H., Thorburn, D.R., et al. (2016). Accessory subunits are integral for assembly and function of human mitochondrial complex I. Nature *538*, 123.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhaes, J.P. (2013). Human ageing genomic resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027-D1033.

Tkemaladze, J.V. and Chichinadze, K.N. (2005). Centriolar mechanisms of differentiation and replicative aging of higher animal cells. Biochemistry (Mosc) *70*, 1288-1303.

Valentine, R.C. and Valentine, D.L. (2014). Human Longevity: Omega-3 Fatty Acids, Bioenergetics, Molecular Biology, and Evolution (Florida: CRC Press).

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1-11.10.33.

Walkinshaw, L. (1973). Cranes of the World (New York: Winchester Press).

Wang, Z., Li, Z., Beauchamp, G., and Jiang, Z. (2011). Flock size and human disturbance affect vigilance of endangered red-crowned cranes (*Grus japonensis*). Biol. Conserv. *144*, 101-105.

Wasser, D.E. and Sherman, P.W. (2010). Avian longevities and their interpretation under evolutionary theories of senescence. J. Zool. *280*, 103-155.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555-556.

Yu, J., Liu, J., and Jin, W. (2001). Analysis of the environment feature of breeding area and endangered factors of red-crowned crane in China. Chin. Geogr. Sci. *11*, 186-191.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science *346*, 1311-1320.