Marine Genomics 42 (2018) 63–66

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

De novo assembly and annotation of the blood transcriptome of the southern giant petrel *Macronectes giganteus* from the South Shetland Islands, Antarctica

Bo-Mi Kim[a], Do-Hwan Ahn[a], Jeong-Hoon Kim[b], Jin-Woo Jung[b], Jae-Sung Rhee[c,d,*], Hyun Park[a,e,**]

[a] Unit of Polar Genomics, Korea Polar Research Institute, Incheon 21990, South Korea
[b] Department of Polar Life Science, Korea Polar Research Institute, Incheon 21990, South Korea
[c] Department of Marine Science, College of Natural Sciences, Incheon National University, Incheon 22012, South Korea
[d] Research Institute of Basic Sciences, Incheon National University, Incheon 22012, South Korea
[e] Polar Sciences, University of Science & Technology, Yuseong-gu, Daejeon 34113, South Korea

ARTICLE INFO

Keywords:
Southern giant petrel
Macronectes giganteus
Blood
Transcriptome
Avian immunity

ABSTRACT

The southern giant petrel is a large Procellariiform seabird of the southern oceans and has a circumpolar habitat. In this study, we generated the first high-quality *de novo* assembly of the blood transcriptome of the southern giant petrel (*Macronectes giganteus*) using the Illumina Miseq platform. A total of 28.7 million raw reads were obtained and assembled using the Oases assembly platform, resulting in 27,989 contigs with an N50 value of 1,044 bp. We performed functional gene annotations using Gene Ontology (GO), Eukaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes pathway analyses. As one of the top consumers in the southern oceans, *M. giganteus* feeds on carrion and carcasses, unlike most other Procellariiformes. However, geographical isolation is not an absolute defense against parasites or pathogens. We detected many genes that are critically involved in classic innate immunity. In terms of the GO terms analyzed, many genes were assigned to the subcategories of response to stimuli and immune system processes. These numbers were higher than those in the whole blood of lipopolysaccharide (LPS)-injected greenfinches and blood lymphocytes of the Chinese goose but lower than those found in the whole blood of the cinereous vulture. This genomic information will be useful for checking the immune status of southern giant petrels without sacrifice, as the species is vulnerable.

1. Introduction

The southern giant petrel *Macronectes giganteus* is the largest avian predator/scavenger of the Southern Ocean, found on the sub-Antarctic islands, the Antarctic Peninsula, the Malvinas (Falkland Islands), Patagonia (southern Chile), and Argentina (Patterson et al., 2008). The bird normally has a long life, females lay single eggs, the annual reproductive output is low, and sexual maturity is deferred (Hamer et al., 2002). Over the past decade, anthropogenic activities, habitat destruction, introduced predators, and the exploitation of fisheries have increased, threatening the *M. giganteus* population (Quintana et al., 2006). The bird is listed as vulnerable in the Red Data List because of continuous declines in the total breeding population (IUCN, 2004), although several colonies have irregularly exhibited increased breeding (Huin and Reid, 2005). Evaluation of the health status of the bird via gene expression analysis may be valuable, as it requires only a small volume of blood or piece of tissue and sacrifice is unnecessary.

Blood is very sensitive to endogenous and exogenous stimulation. Blood circulates; therefore, analyses of molecular and biochemical parameters may provide useful information about responses to parasites or pathogens. Whole-transcriptome profiling has been used successfully to estimate immune system homeostasis and transcriptional changes after experimental challenges (Meitern et al., 2014; Chung et al., 2015; Tariq et al., 2015; Videvall et al., 2015; Watson et al., 2017). Blood transcriptome profiling and/or analyses of target gene/pathway-specific expression can evaluate homeostasis at the individual level, improving the disease resistance and protection of avian populations. The immune homeostasis of *M. giganteus* requires monitoring. The bird feeds on fish and seal carcasses (Maynard, 2003; BirdLife International, 2009). In a previous study, blood stable isotope levels revealed that *M.*

* Correspondence to: Jae-Sung Rhee, Department of Marine Science, College of Natural Sciences, Incheon National University, Incheon 22012, South Korea
** Correspondence to: Hyun Park, Unit of Polar Genomics, Korea Polar Research Institute, Incheon 21990, South Korea
E-mail addresses: jsrhee@inu.ac.kr (J.-S. Rhee), hpark@kopri.re.kr (H. Park).

https://doi.org/10.1016/j.margen.2018.05.003
Received 30 March 2018; Received in revised form 6 May 2018; Accepted 12 May 2018
Available online 19 May 2018
1874-7787/

Table 1
Characteristics of the *M. giganteus* blood transcriptome sequencing project: MIxS standard data.

Item	Description
Investigation type	Eukaryote transcriptome
Project name	*Macronectes giganteus* blood transcriptome sequencing
Geographical location, name	King George Island (South Shetland Islands), Antarctica
Geographical location	62°14′14.0″S, 58°45′59.8″W
Collection date	21 Jan 2017
Environment (biome)	OMIT_0002267 (Antarctic region)
Environment (feature)	ENVO_00000098 (island)
Environment (material)	ENVO_01000428 (rocky shore)
Tissue type	Blood
Developmental stage	Adult
Sequencing method	Pyrosequencing
Sequencing platform	Illumina Miseq
Assembly	Oases (ver. 0.2.08)
Finishing strategy	Contig
Bioproject number	PRJNA445761
Data accessibility	SRR6902605

giganteus chicks were preferentially fed carrion of fish, penguins, and seals (Raya Rey et al., 2012). Moreover, despite their geographical isolation, birds in the Antarctic remain under threat of parasitic or pathogenic infection (Barbosa and Palacios, 2009). In fact, among birds, *M. giganteus* has the highest recorded number of parasites/pathogens (Barbosa and Palacios, 2009). *Salmonella* sp. and avian cholera (caused by the bacterium *Pasteurella multocida*) normally found in scavenging species have been detected in giant petrels and skuas (Leotta et al., 2003; Shearn-Boschler et al., 2008; Barbosa and Palacios, 2009).

Here, using basic bioinformatic tools, we derive a new blood transcriptome resource for work on *M. giganteus*. The information will facilitate future investigations of blood-specific pathways in other Procellariiformes or in birds in general, with a particular focus on innate immunity and pathogen exposure.

2. Description of the data

2.1. Sample collection and Illumina sequencing

A blood specimen (ca. 1 mL) was collected from an *M. giganteus* individual on King George Island, South Shetland Islands, Antarctica (Table 1). The species was identified by reference to morphological characteristics and the mitochondrial cytochrome oxidase subunit 1 (*CO1*) sequence. The blood sample was immediately transferred to an RNAlater vial (QIAGEN, Valencia, CA, USA) and stored at –20°C until RNA extraction. Total RNA was extracted using an RNeasy® Micro Kit (QIAGEN) according to the manufacturer's protocol and stored in RNAstable (Biometrica, San Diego, CA, USA). The quality and quantity of the total RNA was checked using the 2100 Bioanalyer (Agilent Technologies, Santa Clara, CA, USA). High-quality mRNA (2 μg) was

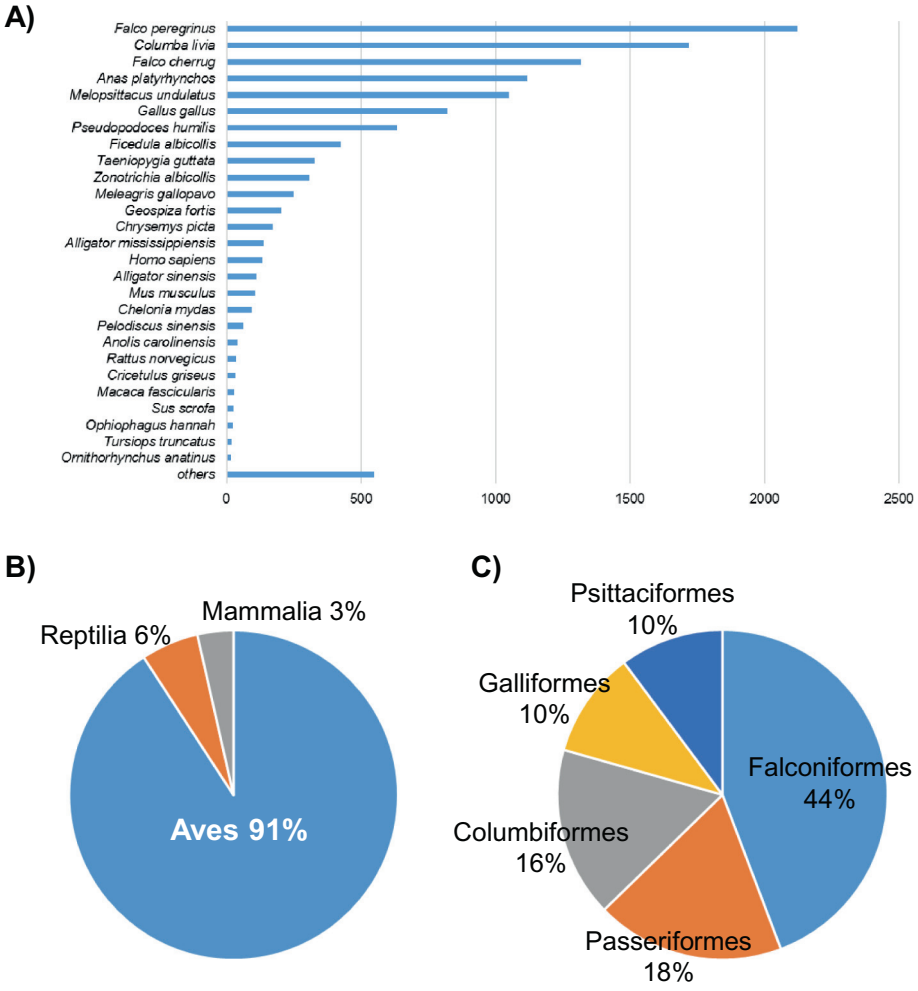

Fig. 1. Numbers of major BLAST hits matched to *Macronectes giganteus* blood transcripts at the phylum and species levels. Each number is the number of orthologous gene families shared by the indicated genomic database.

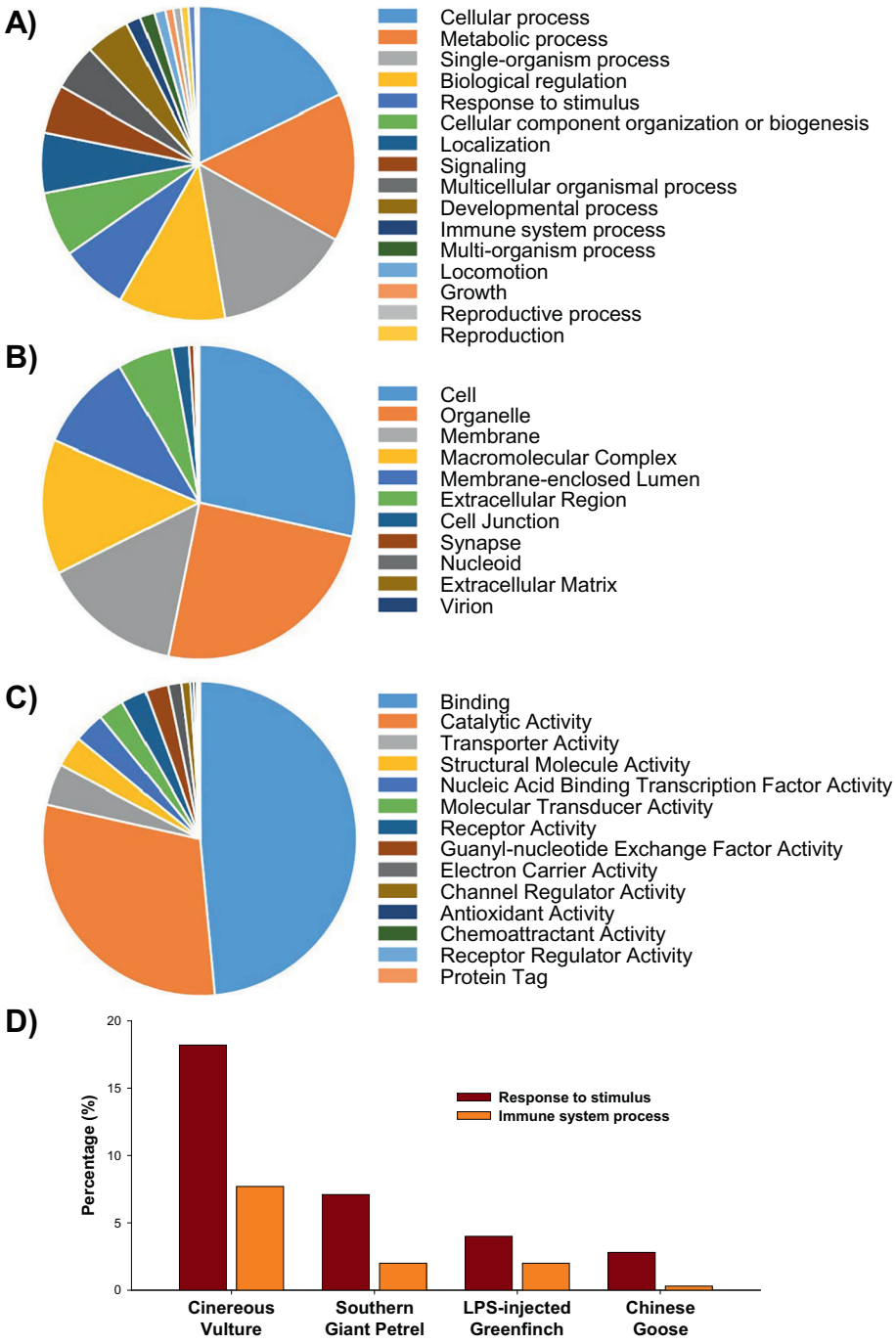

Fig. 2. Gene Ontology (GO) analyses: A) biological processes, B) cellular components, and C) molecular functions enriched in the blood transcriptome of *Macronectes giganteus*. D) A comparison of the compositions (percentages) of two subcategories: response to stimuli and immune system processes of avian blood transcriptomes.

used to generate a double-stranded cDNA library using poly A selection (Kim et al., 2017). A NuGEN Encore Complete RNA-Seq Library System (NuGEN, San Carlos, CA, USA) was used to construct paired-end libraries (PE500) of sheared cDNA (500 bp fragments) that were sequenced on an Illumina MiSeq® System platform (300 × 2 paired-end reads). We trimmed the index and adaptor sequences using Trimmomatic (Bolger et al., 2014) and removed low-quality reads using the FASTX toolkit (Gordon and Hannon, 2010) with the parameters set at –t, 20; –l, 70; and –Q, 33.

2.2. De novo assembly

Sequencing produced 28.7 million reads. We removed low-quality reads (average quality score < 10), adapters, linkers, and PCR primers via quality filtering. Then we assembled high-quality reads *de novo* using Oases (ver. 0.2.08; Schulz et al., 2012) using the default parameters. TransDecoder was used to identify coding regions. Unclustered transcripts and the longest sequences were considered unigenes. We found 27,989 contigs of an average length of 899 bp and an N50 length of 1,044 bp (Supplementary Table 1). All unigenes were searched against the NCBI nonredundant (NR) database using BLASTx with an E-value threshold of 1.00E-04 to identify functional transcripts. The overall sequencing quality and the *de novo* assembly parameters (i.e., average read size, longest length, gene density, and GC ratio) were comparable to those of previous avian blood transcriptomes (Meitern et al., 2014; Chung et al., 2015; Tariq et al., 2015; Videvall et al., 2015;

Watson et al., 2017).

2.3. Transcriptome annotation

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of all contigs were performed using the Blast2GO sequence annotation tool (ver. 4.0; Conesa et al., 2005). The specific GO composition of each category is presented as a Level 2 percentage. After aligning the contigs, we analyzed three principal categories (biological processes, cellular components, and molecular function) using default parameters. The BLAST search and functional domain annotation by InterProScan of the Blast2GO software package assigned 2,194 contigs to at least one GO term (Supplementary Tables 2–4). Finally, the assembled data were arranged in terms of read length, gene annotation, GenBank number, E-value, species, and species accession number. We calculated the mRNA expression level using the reads per kilobase of the transcriptome per million mapped reads (RPKM) method (Mortazavi et al., 2008).

The principal BLAST hits indicated that about 2,123 *M. giganteus* contigs exhibited sequence similarities to transcripts of the peregrine falcon (*Falco peregrinus*), and 1,719 exhibited similarities to transcripts of the rock pigeon (*Columba livia*; Fig. 1A). Of the top hits, 90.8% and 5.7% were homologous to transcripts from the classes Aves and Reptilia, respectively (Fig. 1B and C). Thus, sample preparation and sequencing had been successful; the raw read assembly was undoubtedly avian. Our reference transcriptome of the southern giant petrel will be valuable for studying avian disease, comparative blood compositions, and blood-specific gene expression and also for protecting *M. giganteus*.

The diverse GO assignments analyzed by Blast2GO showed that blood performs many complex biological functions. The vast majority of transcripts in the biological processes category were assigned to cellular processes (18%), metabolic processes (15%), and single-organism processes (14%; Fig. 2A). Of the GO terms dealing with cellular components, many genes were categorized as cell (28%) and organelle (25%) components (Fig. 2B). In terms of the molecular function category, many genes were classified as binding (48%), followed by catalytic activity (30%; Fig. 2C). It is interesting that a considerable number of unigenes were categorized as response to stimuli (7.1%) and immune system processes (2%) potentially associated with blood-borne immunity. Although the proportions of these categories among the biological processes were lower than those in the whole blood of the cinereous vulture *Aegypius monachus* (18.2% and 7.7%; Chung et al., 2015), they were higher than those in the whole blood of the LPS-injected greenfinch *Carduelis chloris* (4% and 2%; Meitern et al., 2014) and the blood lymphocytes of the Chinese goose *Anser cygnoides* (2.8% and 0.3%; Tariq et al., 2015; Fig. 2D). The *M. giganteus* transcriptome included sequences of major innate immune system proteins, such as NF-κB (KEGG pathway number: 04064), Toll-like receptor (TLR; #04620), and JAK-STAT (#04630), which suggests that canonical vertebrate immune pathways are expressed in *M. giganteus* blood and serve to defend against exogenous challenges to the immune system. Transcriptional data on genes associated with the immune system and mapped to the various pathways are appended in Supplementary Table 5. Blood is constantly circulated to transport oxygen, nutrients, and waste products and is exposed to many endogenous and exogenous molecules, including those of pathogens. Thus, diverse transcriptional pathways are active in blood.

We present the first whole-blood transcriptome of the southern giant petrel and a brief comparison of pathway genes related to the immune system in this bird and other avian blood transcriptomes. As the *M. giganteus* blood transcriptome covers the essential gene repertoire, we believe that our data will be useful for exploring immune homeostasis of southern giant petrels in the field. They will also make it possible to analyze the sensitivities and response patterns of the annotated genes.

3. Data deposition

The raw RNA-seq data have been deposited in the NCBI Sequence Read Archive (SRA; accession number SRR6902605) under bioproject number PRJNA445761. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GGLK00000000. The version described in this paper is the first version, GGLK01000000.

Acknowledgments

This work was supported by the grant Polar Genome 101 Project (PE18080) funded by the Polar Research Institute of South Korea and by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant no. NRF-2017R1A6A1A06015181).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2018.05.003.

References

Barbosa, A., Palacios, M.J., 2009. Health of Antarctic birds: a review of their parasites, pathogens and diseases. Polar Biol. 32, 1095–1115.
BirdLife International, 2009. Southern Giant Petrel - BirdLife Species Factsheet. Data Zone. Retrieved 15 Jul 2009.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Chung, O., Jin, S., Cho, Y.S., Lim, J., Kim, H., Jho, S., Kim, H.M., Jun, J., Lee, H., Chon, A., Ko, J., Edwards, J., Weber, J.A., Han, K., O'Brien, S.J., Manica, A., Bhak, J., Paek, W.K., 2015. The first whole genome and transcriptome of the cinereous vulture reveals adaptation in the gastric and immune defense systems and possible convergent evolution between the Old and New World vultures. Genome Biol. 16, 215.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Gordon, A., Hannon, G., 2010. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. Unpublished. Available online at: http://hannonlabcshledu/fastx_toolkit.
Hamer, K.C., Schreiber, E.A., Burger, J., 2002. Breeding biology, life histories and life history-environment interactions in seabirds. In: Schreiber, E.A., Burger, J. (Eds.), Biology of Marine Birds. CRC Press, Boca Raton, pp. 217–261.
Huin, N., Reid, T., 2005. Census of the giant-petrel population of the Falkland Island. Falkland Conservation, Falkland Island.
IUCN, 2004. 2004 IUCN red list of threatened species. http://www.iucnredlist.org.
Kim, B.-M., Kim, K., Choi, I.-Y., Rhee, J.-S., 2017. Transcriptome response of the Pacific oyster, *Crassostrea gigas* susceptible to thermal stress: a comparison with the response of tolerant oyster. Mol. Cell. Toxicol. 13, 105–113.
Leotta, G.A., Rivas, M., Chinen, I., Vigo, G.B., Moredo, F.A., Coria, N., Wolcott, M.J., 2003. Avian cholera in a southern giant petrel (*Macronectes giganteus*) from Antarctica. J. Wildl. Dis. 39, 732–735.
Maynard, B.J., 2003. Shearwaters, petrels, and fulmars (Procellariidae). In: Hutchins, Michael, Jackson, Jerome A., Bock, Walter J., Olendorf, Donna (Eds.), Grzimek's Animal Life Encyclopedia. 8 Birds I. Tinamous and Ratites to Hoatzins, 2nd ed. Gale Group, Farmington Hills, MI, pp. 123–133 Joseph E. Trumpey, Chief Scientific Illustrator.
Meitern, R., Andreson, R., Hõrak, P., 2014. Profile of whole blood gene expression following immune stimulation in a wild passerine. BMC Genomics 15, 533.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat. Methods 5, 621–628.
Patterson, D.L., Woehler, E.J., Croxall, J.P., Cooper, J., Poncet, S., Peter, H.-U., Hunter, S., Fraser, W.R., 2008. Breeding distribution and population status of the northern giant petrel *Macronectes halli* and the southern giant petrel *M. giganteus*. Mar. Ornithol. 36, 115–124.
Quintana, F., Punta, G., Copello, S., Yorio, P., 2006. Population status and trends of southern giant petrels (*Macronectes giganteus*) breeding in North Patagonia, Argentina. Polar Biol. 30, 53–59.
Raya Rey, A., Polito, M., Archuby, D., Coria, N., 2012. Stable isotopes identify age- and sex-specific dietary partitioning and foraging habitat segregation in southern giant petrels breeding in Antarctica and southern Patagonia. Mar. Biol. 159, 1317–1326.
Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092.
Shearn-Boschler, V., Green, D.E., Converse, K.A., Docherty, D.E., Thiel, T., Geisz, H.N., Fraser, W.R., Patterson-Fraser, D.P., 2008. Cutaneous and diphtheritic avian poxvirus infection in a nestling southern giant petrel (*Macronectes giganteus*) from Antarctica. Polar Biol. 31, 569–573.
Tariq, M., Chen, R., Yuan, H., Liu, Y., Wu, Y., Wang, J., Xia, C., 2015. *De novo* transcriptomic analysis of peripheral blood lymphocytes from the Chinese goose: gene discovery and immune system pathway description. PLoS One 10, e0121015.
Videvall, E., Cornwallis, C.K., Palinauskas, V., Valkiūnas, G., Hellgren, O., 2015. The avian transcriptome response to malaria infection. Mol. Biol. Evol. 32, 1255–1267.
Watson, H., Videvall, E., Andersson, M.N., Isaksson, C., 2017. Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. Sci. Rep. 7, 44180.