



## Screening of Genetic Variations in Korean Native Duck using Next-Generation Resequencing Data

Eunjin Cho<sup>1</sup>, Minjun Kim<sup>2</sup>, Hyo Jun Choo<sup>3</sup> and Jun Heon Lee<sup>1,2†</sup>

<sup>1</sup>Department of Bio-AI Convergence, Chungnam National University, Daejeon 34134, Republic of Korea

<sup>2</sup>Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Republic of Korea

<sup>3</sup>Poultry Research Institute, National Institute of Animal Science, Rural Development Administration, Pyeongchang 25342, Republic of Korea

**ABSTRACT** Korean native ducks (KNDs) continue to have a high preference from consumers due to their excellent meat quality and taste characteristics. However, due to low productivity and fixed plumage color phenotype, it could not secure a large share in the domestic market compared to imported species. In order to improve the market share of KNDs, the genetic characteristics of the breed should be identified and used for improvement and selection. Therefore, this study was conducted to identify the genetic information of colored and white KNDs using next-generation resequencing data and screening for differences between the two groups. As a result of the analysis, the genetic variants that showed significant differences between the colored and white KND groups were mainly identified as mutations related to tyrosine activity. The variants were located in the genes that affect melanin synthesis and regulation, such as *EGFR*, *PDGFRA*, and *DDR2*, and these were reported as the candidate genes related to plumage pigmentation in poultry. Therefore, the results of this study are expected to be useful as a basis for understanding and utilizing the genetic characteristics of KNDs for genetic improvement and selection of white broiler KNDs.

(Key words: Korean native duck, next-generation sequencing, genome sequence assembly, genetic variants)

## INTRODUCTION

The introduction of next-generation sequencing (NGS) technology in the early 2000s brought about dramatic changes in the field of biotechnology research (Mardis, 2013). This tendency led to a decrease in analysis cost and an increase in accuracy over time, and as a result, it provides a possibility of genome analysis for large-scale populations (Hu et al., 2021). In particular, NGS technology has been utilized in various species for *de novo* genome sequencing and DNA resequencing (Park and Kim, 2016). Also, the accumulated sequences are applied to identify the genetic markers such as single nucleotide polymorphism (SNP) and insertion and deletion (InDel).

Ducks are one of the most useful protein sources in the poultry industry, following chickens. The breast meat of ducks is close to red unlike chicken, so it has similar sensory characteristics to beef and pork (Chartrin et al., 2006). Also, it has a higher unsaturated fatty acid content than other meats, so consumer interest is continuously increasing (Onk et al., 2019).

The Korean native ducks (KNDs) are a pure breed that was bred from wild migratory ducks by the National Institute of Animal Science (NIAS) in Korea to be used in duck farming, then improved to be used as broiler ducks. Although it has a lower growth rate and productivity than imported broiler ducks, it has a high content of useful fatty acids. Also, the tenderness and shearing power of the meat are high, so it has a high preference of consumers (Hong et al., 2012; Kim et al., 2012). However, in the case of native ducks, the most of plumage color in the group is almost fixed to black-brown, so there are limitations to their use as meat, and in the end, the Pekin species imported from overseas accounts for 90% of the domestic broiler duck market. Therefore, to improve the market share of KNDs for broiler use, it is necessary to understand the genetic background of white KNDs and utilize this to introduce genetic improvement and selection systems.

Based on this background and necessity, this study was conducted to identify the genomic information of KNDs using genomic data of colored and white KNDs and search for the candidate genes associated with plumage color.

<sup>†</sup> To whom correspondence should be addressed : junheon@cnu.ac.kr

## MATERIALS AND METHODS

### 1. Samples

The KND samples were provided by the NIAS in Korea. The KND population consisted of two groups: colored KND ( $n = 2$ ) and white KND ( $n = 2$ ). The blood samples were collected from the brachial vein and used for extraction of genomic DNA (gDNA). gDNA was extracted using the PrimePrep Genomic DNA Extraction Kit from Blood (GeNetBio, Daejeon, Korea), following the manufacturer's instructions. The purity and concentration of the extracted gDNA were confirmed using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The final gDNA was stored at  $-20^{\circ}\text{C}$  until use.

### 2. Whole-Genome Resequencing and Quality Control

The gDNA library was constructed according to the NGS library preparation workflow (TruSeq Nano DNA library prep; Illumina, San Diego, CA, USA), and the raw sequence data were acquired through Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA), resulting in the generation of 151 bp paired-end reads. The raw sequence reads were pre-processed following the quality control (QC) procedure to remove unusable reads. The sequences with less than 10% of their total lengths were removed.

### 3. Read Alignment and Variant Calling

The procedure of calling variants from raw sequence data was performed in accordance with the GATK best practice workflow. All clean reads were aligned to the chromosome-level reference genome of the *Anas platyrhynchos* (assembly version ZJU1.0; accession number GCF\_015476345.1) downloaded from the NCBI RefSeq database using the "BWA-MEM" (Li, 2013) with default parameters. Following the alignment step, the SAM format data were converted to BAM format and indexed by the "SAMtools" (Li et al., 2009), and duplicated reads were filtered using the "MarkDuplicates" program in the "GATK" tool (version 4.1.4.1) (McKenna et al., 2010). Since no available variant information was provided for the present version of the reference genome, the recalibration

step was performed slightly differently from the existing method. The first variant calling step was performed using the "HaplotypeCaller", and the output variants were filtered as following criteria:  $\text{QD} < 2.0$ ,  $\text{FS} > 60.0$ ,  $\text{MQ} < 40.0$ ,  $\text{MQRankSum} < -12.5$ ,  $\text{ReadPosRankSum} < -8.0$ . Then, the recalibration step was performed using the filtered variants, and the second variant calling step proceeded with recalibrated data. In the end, the output data was filtered using the same criteria as the previous filtering step.

### 4. Variant Annotation and Functional Enrichment Analysis

The variant annotation for the final variants of four KNDs was performed using the "SnpEff" program (Cingolani et al., 2012). After the annotation step, the common differential variants were distinguished to identify the genetic difference in genomes between colored and white KNDs. The workflow was as follows: we first collected the variants with the same genotype in each group and then determined the differential variants between the two groups. Subsequently, these common differential variants were used for functional enrichment analysis. The Gene Ontology (GO) analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the gene sets containing the common differential variants were conducted using the "g:Profiler" online web server (<https://biit.cs.ut.ee/gprofiler/gost>) (Raudvere et al., 2019). The conditions for filtering significant analysis results were as follows: significance threshold was  $P < 0.05$ , minimum term size was 10 and maximum term size was 350.

## RESULTS AND DISCUSSION

### 1. Genome Resequencing of KNDs

NGS resequencing was performed on four birds from each KND group. Table 1 summarizes the results of resequencing. The output indicated that more than 132,124,410 of 151 bp paired-end short reads were acquired from each bird, with total sequence lengths exceeding 39 Gb. The sequencing quality was high, with a Q30 ratio of more than 90%. The average sequencing depth of coverage was found to be approximately 35X.

**Table 1.** Raw sequence data from four Korean native ducks

Group	Sample ID	Total sequence (Gb)	No. of raw reads	Q30 (%)	Depth (X)
Colored	KND01	43.65	144,549,912	91.94	36
	KND02	39.91	132,124,410	90.21	33
White	KND03	43.92	145,453,098	91.76	36
	KND04	41.81	138,448,726	92.07	34

KND, Korean native duck.

2. Detection of Genetic Variants and Annotation

A total of 3,713,890 SNPs and 795,089 InDels were identified in the colored KNDs, and a total of 3,682,977 SNPs and 784,006 InDels were identified in the white KNDs. For the common variants between the two groups, a total of 1,887,587 SNPs and 419,135 InDels were identified (Table 2).

Table 3 shows the variant effects by functional class as the results of variant annotation using "SnEff". More than 16 million effects were identified in the colored and white KNDs, and among them, the effect of mutations common to the two groups was confirmed to be about half. In addition,

**Table 2.** The types of identified variants from Korean native ducks

Group	Total No. of variants	No. of variants by type (%)		
		SNP	Insertion	Deletion
Colored	4,508,979	3,713,890 (82.36)	370,351 (8.21)	424,738 (9.43)
White	4,466,983	3,682,977 (82.45)	363,574 (8.14)	420,432 (9.41)
Common	2,306,722	1,887,587 (81.83)	193,679 (8.39)	225,456 (9.78)

SNP, single nucleotide polymorphism.

**Table 3.** The variant effects by functional class of Korean native ducks confirmed through SnEff software

Group	Total No. of effects	No. of effects by functional class (%)		
		Missense	Nonsense	Silent
Colored	16,941,195	49,713 (29.02)	318 (0.19)	121,276 (70.79)
White	16,762,624	49,544 (29.31)	286 (0.17)	119,194 (70.52)
Common	8,620,138	26,738 (29.77)	160 (0.18)	62,902 (70.05)

about 30% of the confirmed variants were found to be non-synonymous mutations (missense and nonsense).

Table 4 indicates the results aggregated by region where the effects of the variants shown in Table 3 were located on the genome. According to these results, the region with the most variants was in the intron, followed by the variants in the upstream, downstream, and intergenic regions. Although there were relatively few variants of more than 2% in the exon region, it was confirmed that some variants with a high impact of the variant effect, such as frameshift or exon loss variants, existed in the exon region.

3. Functional Enrichment Analysis for Common Differential Variants

A total of 1,248,986 variants with the same genotype in the colored KND group were identified, and a total of 1,253,989 variants were identified in the white KND group. Among these variants, 753,175 variants existed in common within the two groups, and ultimately, a total of 210,937 common differential variants were confirmed between the colored and white KNDs. The putative variant impacts of

**Table 4.** The variation effects by region of Korean native ducks confirmed through SnEff software

Group	No. of effects by region (%)						
	Upstream	5'UTR	Exon	Intron	3'UTR	Downstream	Intergenic
Colored	1,693,903 (9.99)	85,502 (0.50)	392,666 (2.32)	11,314,645 (66.79)	281,306 (1.66)	1,632,471 (9.63)	1,500,783 (8.86)
White	1,666,497 (9.94)	82,721 (0.49)	383,796 (2.29)	11,198,150 (66.81)	276,501 (1.65)	1,615,519 (9.64)	1,499,624 (8.95)
Common	878,877 (10.19)	46,249 (0.54)	202,041 (2.35)	5,715,304 (66.30)	141,990 (1.65)	844,124 (9.79)	770,418 (8.94)

each variant were classified as follows: high (186), moderate (993), low (3,012), and modifier (206,746). These variants participated in 7,983 genes and these genes were used for functional enrichment analysis.

As a result of analysis using 7,982 genes, a total of 337 GO terms (42 molecular function terms, 258 biological process terms, and 37 cellular component terms) and 2 KEGG pathways (Wnt signaling pathway and MAPK signaling pathway) were observed ( $P < 0.05$ ) (Fig. 1A). To further increase the significance of the results, the term size (the number of genes associated with a given GO term) was adjusted between 10 and 350, resulting in a total of four terms in the molecular function category, 14 terms in the biological process category, eight terms in the cellular component category, and one KEGG pathway were significantly identified (Fig. 1B).



**Fig. 1.** The result of the functional enrichment analysis for common differential variants confirmed through g:Profiler. (A) is the result of total terms based on the threshold  $P < 0.05$ . (B) is the result of filtered terms based on the term size less than 350 genes.

Among the identified GO terms, the significantly related terms with the KND phenotype were as follows: transmembrane receptor protein tyrosine kinase activity (GO:0004714), protein tyrosine kinase activity (GO:0004713), peptidyl-tyrosine phosphorylation (GO:0018108), and peptidyl-tyrosine modification (GO:0018212). All identified significant GO terms were related to tyrosine activity, and according to previous studies, they have been reported to be related to the mechanism of melanin present in muscles or feathers (Li et al., 2012; Kulikova, 2021; Zhang et al., 2022; Xu et al., 2023). In addition, some of the genes such as *EGFR*, *PDGFRA*, and *DDR2* included in each term category were confirmed to affect tissue pigmentation during melanin synthesis and regulation (Quillen et al., 2012; Hulsman et al., 2014; Reger de Moura et al., 2020). Therefore, the GO terms and genes presented as a result of this study suggest the possibility and necessity of follow-up research to explain the genetic background of plumage color formation in KND.

## SUMMARY

In this study, we used NGS resequencing data of colored and white KND to screen the genetic characteristics of KND by comparing variants present in the two groups. In addition, through the comparison results, the putative candidate gene information that affects the plumage color of KND was confirmed. This result could be significant in understanding the genetic characteristics of KND and establishing a research foundation for the genetic improvement of white KNDs.

## ACKNOWLEDGMENTS

This study was supported by the research project (No. RS-2023-00225185) of the Rural Development Administration, Republic of Korea.

## ORCID

Eunjin Cho

<https://orcid.org/0000-0003-4800-1603>

Minjun Kim

<https://orcid.org/0000-0002-8173-8431>

Hyo Jun Choo

<https://orcid.org/0000-0002-7747-5077>

Jun Heon Lee

<https://orcid.org/0000-0003-3996-9209>

## REFERENCES

- Chartrin P, Méteau K, Juin H, Bernadet MD, Guy G, Larzul C, Rémignon H, Mourot J, Duclos MJ, Baéza E 2006 Effects of intramuscular fat levels on sensory characteristics of duck breast meat. *Poult Sci* 85(5):914-922.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2):80-92.
- Hong EC, Choo HJ, Kang BS, Kim CD, Heo KN, Lee MJ, Hwangbo J, Suh OS, Choi HC, Kim HK 2012 Performance of growing period of large-type Korean native ducks. *Kor J Poult Sci* 39:143-149.
- Hu T, Chitnis N, Monos D, Dinh A 2021 Next-generation sequencing technologies: an overview. *Hum Immunol* 82(11):801-811.
- Hulsman Hanna LL, Sanders JO, Riley DG, Abbey CA, Gill CA 2014 Identification of a major locus interacting with MC1R and modifying black coat color in an F2 Nellore-Angus population. *Genet Sel Evol* 46(1):1-8.
- Kim HK, Kang BS, Hwangbo J, Kim CD, Heo KN, Choo HJ, Park DS, Suh OS, Hong EC 2012 The study on growth performance and carcass yield of meat-type Korean native ducks. *Kor J Poult Sci* 39:45-52.
- Kulikova IV 2021 Molecular mechanisms and gene regulation of melanic plumage coloration in birds. *Russ J Genet* 57:893-911.
- Li H 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
- Li S, Wang C, Yu W, Zhao S, Gong Y 2012 Identification of genes related to white and black plumage formation by RNA-Seq from white and black feather bulbs in ducks. *PLoS One* 7(5):e36592.
- Mardis ER 2013 Next-generation sequencing platforms. *Annu Rev Anal Chem* 6:287-303.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
- Onk K, Yalcintan H, Sari M, Isik SA, Yakan A, Ekiz B 2019 Effects of genotype and sex on technological properties and fatty acid composition of duck meat. *Poult Sci* 98(1):491-499.
- Park ST, Kim J 2016 Trends in next-generation sequencing and a new era for whole genome sequencing. *Int Neurol J* 20(Suppl 2):S76.
- Quillen EE, Bauchet M, Bigham AW, Delgado-Burbano ME, Faust FX, Klimentidis YC, Mao X, Stoneking M, Shriver MD 2012 OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. *Hum Genet* 131:1073-1080.
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J 2019 g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47(W1):W191-W198.
- Reger de Moura C, Prunotto M, Sohail A, Battistella M, Jouenne F, Marbach D, Lebbe C, Fridman R, Mourah S 2020 Discoidin domain receptors in melanoma: potential therapeutic targets to overcome MAPK inhibitor resistance. *Front Oncol* 10:1748.
- Xu M, Tang S, Liu X, Deng Y, He C, Guo S, Qu X 2023 Genes influencing deposition of melanin in breast muscle of the Xuefeng black bone chicken based on bioinformatic analysis. *Genome* 66(8):212-223.
- Zhang P, Cao Y, Fu Y, Zhu H, Xu S, Zhang Y, Li X, Sun G, Jiang R, Han R, Li H, Li G, Tian Y, Liu X, Kang X, Li D 2022 Revealing the regulatory mechanism of lncRNA-LMEP on melanin deposition based on high-throughput sequencing in Xichuan Chicken skin. *Genes* 13(11):2143.