



Identification of Key Candidate Genes for Beak Length Phenotype by Whole-Genome Resequencing in Geese

Jianhua Huang[†], Cong Wang[†], Jing Ouyang, Hongbo Tang, Sumei Zheng, Yanpeng Xiong, Yuren Gao, Yongfei Wu, Luping Wang, Xueming Yan* and Hao Chen*

College of Life Science, Jiangxi Science and Technology Normal University, Nanchang, China

OPEN ACCESS

Edited by:

Junwu Ma,
Jiangxi Agricultural University, China

Reviewed by:

Kent M. Reed,
University of Minnesota Twin Cities,
United States
Fenghua Lyu,
China Agricultural University, China

*Correspondence:

Xueming Yan
xuemingyan@hotmail.com
Hao Chen
haochen-jxau@hotmail.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Veterinary Science

Received: 02 January 2022

Accepted: 15 February 2022

Published: 15 March 2022

Citation:

Huang J, Wang C, Ouyang J, Tang H,
Zheng S, Xiong Y, Gao Y, Wu Y,
Wang L, Yan X and Chen H (2022)
Identification of Key Candidate Genes
for Beak Length Phenotype by
Whole-Genome Resequencing in
Geese. *Front. Vet. Sci.* 9:847481.
doi: 10.3389/fvets.2022.847481

The domestic goose is an important economic animal in agriculture and its beak, a trait with high heritability, plays an important role in promoting food intake and defending against attacks. In this study, we sequenced 772 420-day-old Xingguo gray geese (XGG) using a low-depth (~1×) whole-genome resequencing strategy. We detected 12,490,912 single nucleotide polymorphisms (SNPs) using the standard GATK and imputed with STITCH. We then performed a genome-wide association study on the beak length trait in XGG. The results indicated that 57 SNPs reached genome-wide significance levels for the beak length trait and were assigned to seven genes, including *TAPT1*, *DHX15*, *CCDC149*, *LGI2*, *SEPSECS*, *ANAPC4*, and *Slc34a2*. The different genotypes of the most significant SNP (top SNP), which was located upstream of *LGI2* and explained 7.24% of the phenotypic variation in beak length, showed significant differences in beak length. Priority-based significance analysis concluded that *CCDC149*, *LGI2*, and *SEPSECS* genes in the most significant quantitative trait locus interval were the most plausible positional and functional candidate genes for beak length development in the XGG population. These findings not only enhance our understanding of the genetic mechanism of the beak length phenotype in geese, but also lay the foundation for further studies to facilitate the genetic selection of traits in geese.

Keywords: Xingguo gray geese, beak length, whole-genome resequencing, GWAS, *LGI2*

INTRODUCTION

As a globally important poultry species, geese provide abundant amounts of meat, eggs, and feather products with relatively high economic value for human consumption, especially in China. Currently, China is the leading producer of geese because they have the greatest numbers of geese raised worldwide. They are also the most abundant source of goose breeds in the world (1). The Xingguo gray goose (XGG) is a famous medium-sized gray goose breed in China, which has been included in the conservation list of livestock and poultry genetic resources by China's Ministry of Agriculture. XGG is so named because of its gray feathers and Xingguo County origin. Furthermore, it is known for its high-quality meat and possesses qualities geared toward excellent growth performance, including fast growth, tolerance to roughage, and strong stress resistance.

It is generally believed that facial phenotypes in animals are important because of their close relationship with identification and their morphology is strongly influenced by genetic factors

(2). The beak is one of the major structures formed during facial development and it is involved in many significant biological functions, such as predation, vocalization, mating, and grooming. These functions are necessary for adaptation through natural selection and diversifying evolution. Underlying genetic variations might explain the diversity in beak shape, which is a key phenotypic trait. For example, it was found that the *ALXI* and *HMGA2* genes had a major influence on beak shape and size among Darwin's finches (3, 4). In addition, a study on the morphological development of beaks in chickens and ducks suggested that localized cell proliferation mediated by *BMP4* is likely to modulate the size of beaks (5). However, to date, such reports on the beak length of birds, such as domestic geese, have been relatively scarce.

A genome-wide association study (GWAS) is a powerful method for identifying the genetic variations responsible for particular phenotypes. Similarly, the combination of whole-genome resequencing (WGR) technology is expected to provide higher marker density and reveal a wider range of genetic variation than other molecular marker technologies, and enhance the ability to identify the genetic basis of phenotypic traits (6). The increasing availability of WGR provides an effective means for revealing the genetic mechanism of phenotypic traits in detail, as well as a good means for animal protection and breeding. So far, GWAS based on whole-genome resequencing data has been fully applied in many livestock and poultry animals, including pigs (7), cattle (8), chickens (9), and ducks (10).

In general, goose beaks are long, wide, and nearly flat, mainly consisting of upper and lower jaws, with blunt rounded ends, along with numerous horizontal ridges on the edges that are convenient for filtering water and crushing food when eating in water (Supplementary Figure S1). As it is the main organ for ingesting nutrients, studying the genetic mechanism of beak development is of great significance for improving the growth performance and breeding management of geese. In this study, we performed a GWAS on the beak length trait in XGG, based on SNP data obtained by whole-genome resequencing. Through this study, we hope to identify the candidate genes affecting the development of goose beak length, reveal the genetic mechanisms involved, and provide a reference for the study of similar traits in birds.

MATERIALS AND METHODS

Ethics Statement

All procedures used in this study, and those involving animals, complied with the guidelines for the care and utility of experimental animals, which were established by China's Ministry of Agriculture. The ethics committee of Jiangxi Science & Technology Normal University approved this study.

Experimental Animals

XGG is a nationally protected livestock and poultry breed in China. In the XGG conservation farm, the beaks of the XGG population members show differences in length. To explore the molecular mechanism of beaks, we collected 772 mature XGG samples from animals that were 420-day-old from the

Xingguo Gray Goose Farm in Jiangxi Province. Vernier calipers were used to measure the beak lengths of each individual using three different phenotypic measurement groups under the same criteria. The measurement range of beak length was the straight line distance from the feather on the top of the head to the tip of the upper beak (Figure 1A), which was accurate to 0.01 mm. Additionally, vacuum vascular sampling was used to collect 1 mL of blood from the subalar vein of the experimental birds for DNA extraction.

DNA Extraction and Whole-Genome Resequencing

A modified phenol-chloroform method (11) was used to extract genomic DNA from the fresh blood samples of the experimental birds. The concentration and quality of the DNA samples were assessed using a NanoDrop 2,000 ultra-micro spectrophotometer and 1% agarose gel electrophoresis. Only qualified DNA samples were used to construct paired-end libraries according to standard protocols, with an average insert size of ~350 bp. All libraries were sequenced on the Illumina NovaSeq 6,000 platform to generate raw reads with an average length of ~150 bp. After sequencing, we filtered the raw data to remove adapter sequences and low-quality bases with Trimmomatic (12), resulting in an average depth of ~1× coverage for clean reads.

Variant Calling and Quality Control

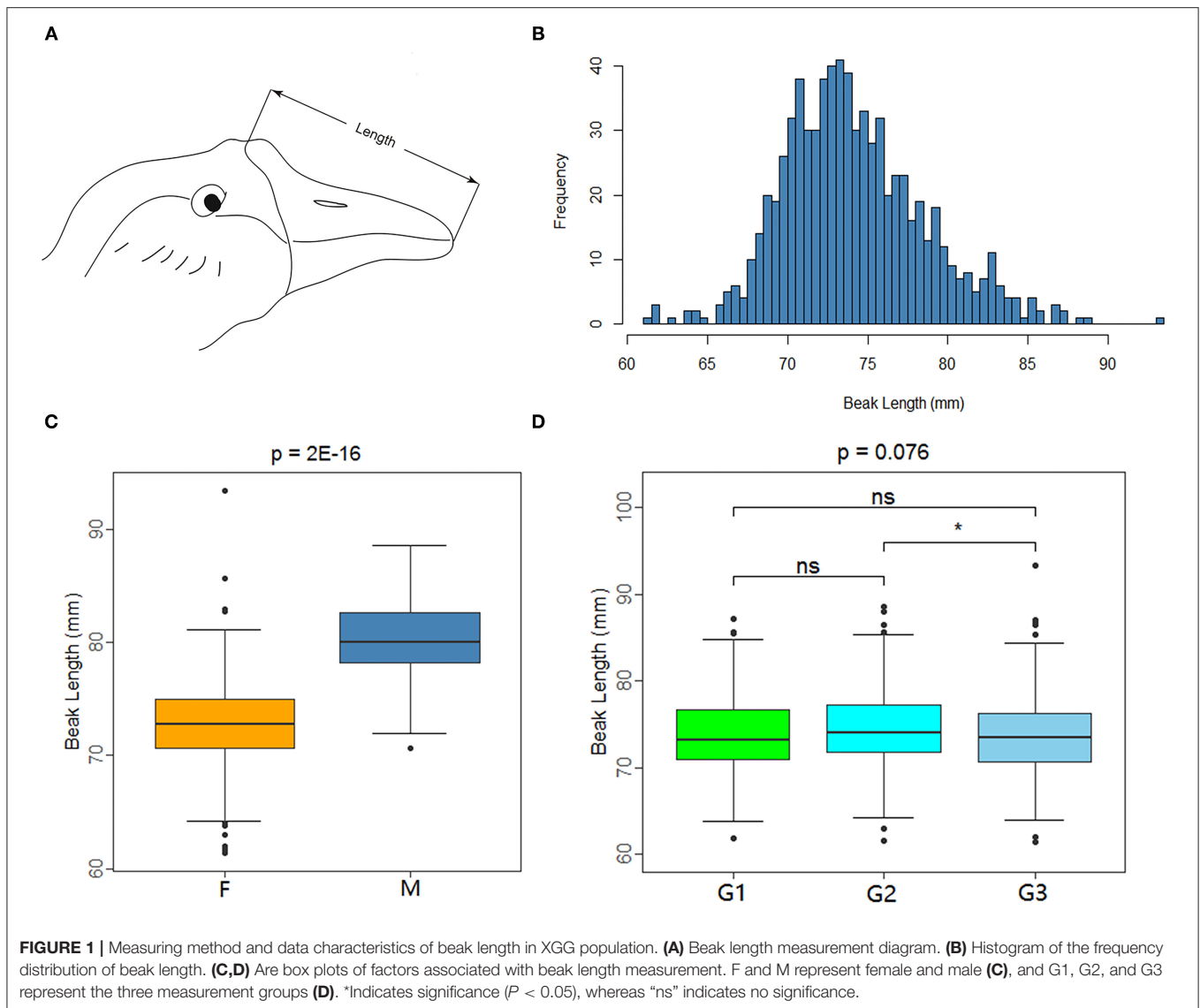
To obtain a high-quality SNP dataset, 150-bp paired-end clean reads were mapped to the XGG reference genome (<https://bigd.big.ac.cn/gwh/>, accession number: GWHBAAW00000000) with Burrows-Wheeler alignment MEM (BWA-MEM) (13), using default parameters. Thereafter, the mapping results were converted into BAM format and sorted using SAMtools (14). Duplicate reads were removed using SENTIEON commercial software (15) and calibrated according to the quality score. After mapping, we performed SNP calling for all samples using GATK v4.0.12 (16) and the output was further filtered using PLINK v1.90 (17). The genotype imputation of each individual SNP was completed using STITCH v1.68 (18), and 12,490,912 SNPs were retained with a minimum allele frequency (MAF) of >0.005 and call rate of >0.95. Further quality control methods were applied to the identified SNPs according to the following standards: those with a MAF of less than 0.02 and a missing genotype rate higher than 0.05 were filtered. After a series of quality controls, we obtained a new dataset that consisted of 772 XGG with a total of 11,513,753 high-confidence whole-genome SNPs. These were used, with measured beak length phenotype data, for subsequent genome-wide association analysis.

Genome-Wide Association Analysis

Here, we used the fastGWA software (19) to implement a GWAS using a mixed linear model as follows:

$$y = x_{\text{snp}}\beta_{\text{snp}} + X_c\beta_c + g + e$$

where y is an $n \times 1$ vector of phenotypes; x_{snp} is a vector of genotype variables that affect β_{snp} ; X_c is the incidence matrix of fixed effects with their corresponding coefficients β_c ; g is a vector



of the total genetic effects with $g \sim N(0, \pi \sigma_g^2)$; π represents the family relatedness matrix (FAM) or the genetic relationship matrix (GRM); e is a vector of residuals with $e \sim N(0, I\sigma_e^2)$; and $\pi \sigma_g^2 + I\sigma_e^2$ is the variance-covariance matrix (V) of y .

To enhance the statistical power and minimize false-positive results, we included goose gender and the phenotype measurement groups as fixed effects into the GWAS model. The association significance threshold between SNPs and phenotypes was obtained by Bonferroni adjustment, and the genome-wide significance threshold was set as $0.05/N$, which is $\sim 4.34E-09$, where N is the number of all SNPs tested. Thereafter, the threshold was extended by 20 times as the chromosome level threshold, namely $1/N$, which is $\sim 8.69E-08$. To avoid population stratification misleading the validity of the GWAS results, the quantile-quantile plot (Q-Q plot) was used to evaluate the GWAS results and determine whether the calculated P -value deviated from the hypothesis test as a whole.

The heritability of the whole genome in this study was calculated based on V_g/V_p , where V_g and V_p represent genetic variance and phenotypic variance, respectively, and V_p is the sum of the V_g and environmental variance (V_e). These variances were obtained by constructing a sparse GRM with a cutoff of 0.05 using fastGWA software. Additionally, an ordinary linear model was established to estimate phenotypic variation explained (PVE) by the lead SNP, where PVE is the ratio of the genotype effect variance of the lead SNPs to the sum of the genotype and residual variances in the model.

Identification of Candidate Genes

The SNPs were annotated using snpEff v4.3t (20) based on the XGG reference genome. Significant SNPs were screened according to the defined significance threshold, and related candidate genes were identified based on the results of variant annotations. The quantitative trait locus (QTL) was defined by

extending the position of the top SNP (with the most significant P -value) on either side so that the $-\log_{10}$ (P -value) of all SNPs in this region was higher than the $-\log_{10}$ (P -value) of the top SNP minus 2 units, or expressed by $(-\log_{10} P) - 2$, where P is the P -value of the top SNP. Taking into account the linkage disequilibrium (LD) of the genome and the regulatory patterns of genes, we also defined those genes with significant SNPs within 10 Kb upstream and downstream as candidate genes. Additionally, haplotype blocks containing significantly associated SNPs were analyzed using Haploview software (21) and protein structure predictions were performed at <https://robotta.bakerlab.org/>.

Statistical Analyses

The R language tool (v4.0.2) was used to perform descriptive statistics of the phenotypic data. One-way analysis of variance (ANOVA) was used to analyze the overall data, and the t -test was used among the groups. Statistical significance was set at $P < 0.05$, and results with $P < 0.01$ were considered extremely significant.

RESULTS

The Data on Beak Length in XGG

The beak of XGG is an important structure with different phenotypes. We measured the beak length of 772 XGG individuals from a conservation farm. We then carried out quality control for the phenotypes and removed individuals with missing phenotypes. We eventually retained 749 high-accuracy phenotypic samples (613 females and 136 males). Beak length ranged from 61.36 mm to 93.40 mm, with an average beak length of 74.15 mm (Figure 1B). The beak length of the females was remarkably shorter than that of the males ($p = 2E-16$), indicating that gender had a significant effect on beak length (Figure 1C). In addition, in light of the fact that the beak length was measured by three investigators separately, we analyzed the measurement results by group. The results showed that the measurement results varied between groups, but the differences were not significant (Figure 1D).

The Most Significant QTL Affected Beak Length

The heritability of beak length was estimated as 0.53 in the XGG population. This indicates that beak length had a moderate level of heritability with the potential for genetic selection for this trait. Considering the factors affecting phenotypic traits, we added gender and measurement group as fixed-effect factors in the GWAS analysis model. The Q-Q plot (Supplementary Figure S2) showed that the results of this model were reasonable and reliable, and the false-positive results caused by population stratification had little effect on the integrity of the results.

We identified 232 SNPs above the P -value threshold at the chromosome level ($P = 8.69E-08$) that were potentially affecting beak length trait (Supplementary Table S1). Among those, 57 achieved a P -value threshold at the genome-wide level ($P = 4.34E-09$) (Figure 2A). To identify the causal genes

that affect the beak length of geese, we further analyzed the SNPs that reached the level of genome-wide association for confirmation and fine mapping of the globe locus, and identified the most significant QTL region within a 0.35 Mb interval (Chr 4: 18.95–19.30 Mb) that harbored 51 significant SNPs. Notably, the genotype of the top SNP (Chr 4:19121505, $P = 3.89E-12$) explained 7.24% of the phenotypic variation in beak length, which further illustrated that this region was likely to be a causal candidate region associated with the beak length trait in the XGG population.

Candidate Genes and Mutations of Beak Length

To address the association of SNPs with the beak length trait, we performed region annotation and association analysis of candidate genes on all 232 SNPs that reached the chromosome-wide significance level (Supplementary Table S1). From this, twelve promising candidate genes were annotated, including *TBC1D1*, *BOD1L1*, *TAPT1*, *NCAPG*, *DHX15*, *CCDC149*, *LGI2*, *SEPSECS*, *ZCCHC4*, *ANAPC4*, *Slc34a2*, and *Rbm47*. Among them, seven genes (*TAPT1*, *DHX15*, *CCDC149*, *LGI2*, *SEPSECS*, *ANAPC4*, and *Slc34a2*) reached genome-wide significance levels (Figures 2C,D). The position of *TAPT1* (~16 Mb) was located far from the other six genes (~19 Mb).

To identify the strong candidate genes affecting the beak length of XGG, 57 SNPs at the genome-wide significance level were prioritized. Furthermore, we found that the visually remarkable region was divided into three QTLs, including a most significant QTL identified by the top SNP and two other low-effect QTLs (Figures 2C,D). In the most significant QTL region, we identified 51 SNPs at the genome-wide level involving five candidate genes (*DHX15*, *CCDC149*, *LGI2*, *SEPSECS*, and *ANAPC4*) associated with beak length (Figure 2C). Furthermore, we also assessed the degree of LD between the top SNP and surrounding SNPs. We found a low degree of LD ($r^2 < 0.3$) between the most significant SNPs identified in each QTL, which further confirms that they were different QTLs. However, the top SNP had a high degree of LD with its neighboring SNPs, with an average value of 0.62 (Table 1). We then tested for significant differences between pairwise genotypes of the top SNP (Figure 2B). Remarkable P -values were seen among three genotypes ($P < 0.01$), suggesting that it was likely that the top SNP was likely to be closely linked to the causal mutation affecting beak length. Interestingly, among these 51 SNPs, we found a missense mutation (Chr4: 4_19180040) in *SEPSECS* that resulted in a glutamine to lysine substitution, yet the presence of the mutation did not result in a substantial change in the overall structure of the protein (Supplementary Figure S3). However, the protein encoded by this gene converts O-phosphoserine-tRNA (Sep) into selenocysteinyl-tRNA (Sec) to promote selenoprotein biosynthesis. This process might then affect the function of the brain by participating in neuronal development and further regulating beak development. Furthermore, we conducted haplotype block analysis for all significant SNPs. We found that these SNPs formed LD blocks, which were significantly associated with the beak length trait (Supplementary Figure S4).

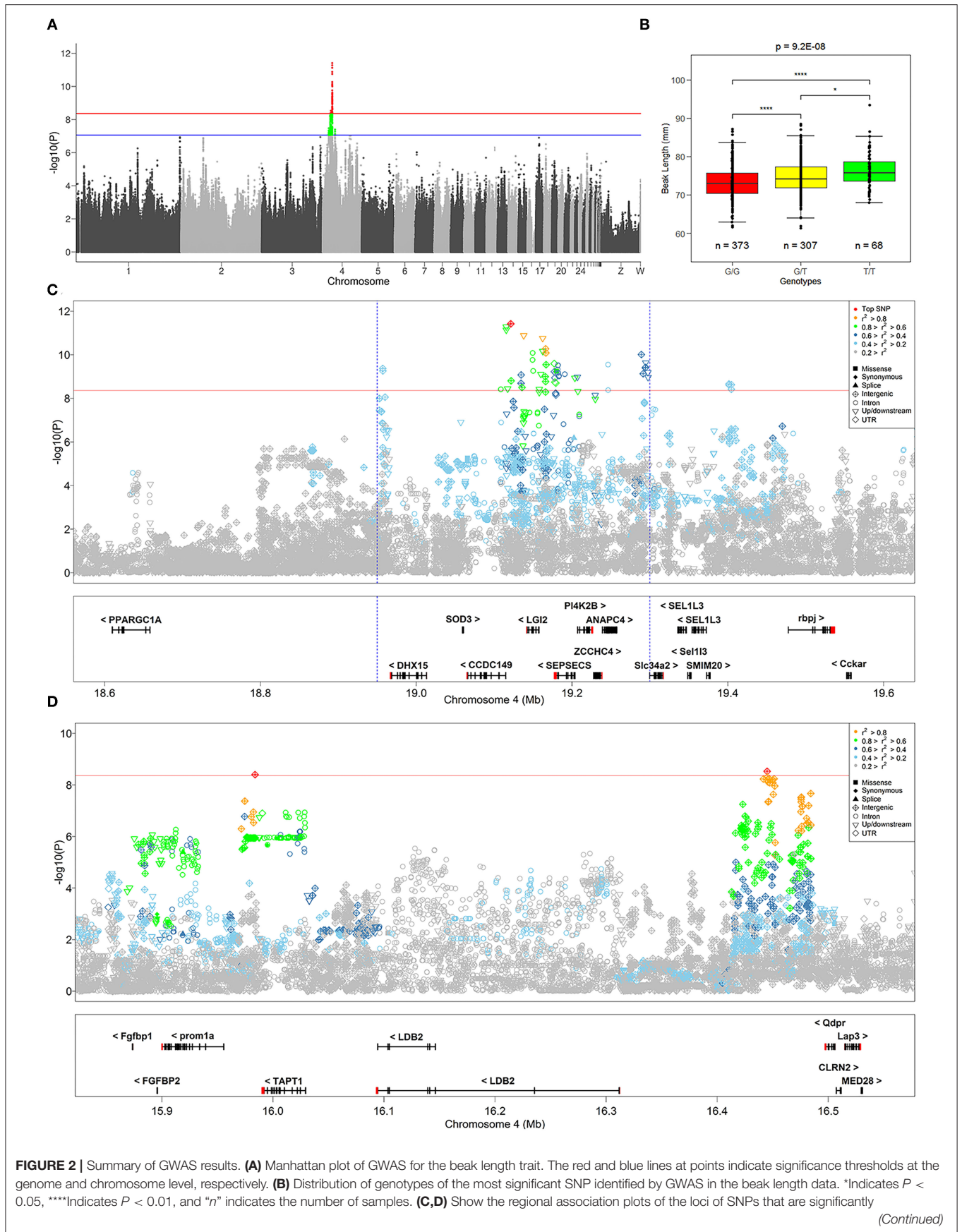


FIGURE 2 | associated with beak length. The red solid line indicates the significance threshold at the genome-wide level, and the defined most significant QTL region is marked with a blue dashed line. The difference in the color and shape of the points are shown as the result of r^2 and variant annotation, respectively. The plot shows the names and locations of the genes, and the transcribed strand is indicated with an arrow. Genes are represented with intronic and exonic regions, and UTR regions are shown in red.

Therefore, it can be inferred from these analyses that beak length is jointly regulated by multiple genes, including *CCDC149*, *LGI2*, and *SEPSECS*.

DISCUSSION

The beak is an important feature of birds, but there are few reports on the genetic mechanism affecting its length. In this study, we studied the beak length of individuals of XGG, a representative local goose breed in China. By studying the variation in beak length in the XGG population, we found that the heritability of beak length was at an intermediate level; thus, it is possible to perform genetic selection for this trait. Based on GWAS analysis of the SNP datasets, seven candidate genes were identified at the genome-wide level, all of which are protein-coding genes. Homologs of the candidate gene *CCDC149* are reported to be expressed ubiquitously in the brain, suggesting that it plays an indispensable role in brain development (22). Although the exact function of this gene is still unclear and its effect on the beak has not been reported, it may have a previously unrecognized role in this process.

Both *LGI2* and *SEPSECS* are essential genes related to neurological disorders (23, 24), and they could affect development of the early nervous system. This in turn could affect the maturation of the brain and, thus, regulate the development of beak. All or most avian beaks consist of premaxillae bones that fuse during facial development in the embryo (25). The neural crest plays a vital role in head and face morphogenesis, as well as in bird beaks (26, 27). Establishment of the facial bones of birds, including the upper and lower jaw, is mediated by the neural crest, and the derived mesenchyme continues to interact with local tissues to form the characteristic patterns of the upper and lower beak (28). Neural crest mesenchyme also confers not only the shape and size of the beak, but also the density of the coracoid bone by regulating bone resorption (29). The beak of geese gradually forms in the direction of slenderness and broadness in the later stages of embryonic development (30). Waterfowl such as geese often need to immerse their heads in the water to catch food or overcome buoyancy when diving; thus, a slender and dense beak is an evolutionary adaptation to the aquatic environment.

The development of beak length is likely regulated by multiple genes and is mainly affected by the intracranial skeleton and osteoblasts. For example, there is evidence that *FGF13* may affect beak morphology by modulating the levels of osteoblasts (31). In comparison, the candidate gene *LGI2* identified in this study is comparable to *FGF13* as a key gene in synaptic transmission and may regulate the function of intracranial bone formation in the nervous system and brain, thereby affecting beak formation. In addition, *LGI2* also acts as a receptor to interact with some inactive ADAM proteins, thereby participating in processes in

the brain (32). This could be another way for *LGI2* to participate in beak formation, because some ADAM proteins with catalytic activity play a role in cranial development by inducing neural crest cells in the Wnt signaling pathway (33).

Selenoprotein encoded by *SEPSECS* is involved in many physiological processes, especially in brain development. There is evidence that the synthesis of neuronal selenoproteins is essential for cerebellar development (34). Moreover, mutations in *SEPSECS* would lead to an overall selenoprotein deficiency, resulting in severe phenotypic variation (35). Previous studies have suggested that the beak and skull are highly integrated structures, and beak shape is linked to signaling molecules (36, 37). Therefore, we hypothesized that *SEPSECS* is involved in the regulation of brain function and plays a role in the regulation of beak length. Delay of telencephalic neurogenesis in the developmental stage is considered a prerequisite for vocalization learning in birds (38). Compared with songbirds, geese cannot expand their telencephalon by delaying telencephalic neurogenesis, which explains why the beaks of geese are better for preying than singing, and this may indicate that beak length is controlled jointly by the brain and neurogenesis.

The genes *TAPT1*, *DHX15*, *ANAPC4*, and *Slc34a2* are mainly related to the regulation of gene expression. *DHX15* is an ATP-dependent RNA helicase implicated in pre-mRNA splicing (39). *ANAPC4* is an E3 ubiquitin ligase that is regulated by the cell cycle, which mediates the ubiquitination and subsequent degradation of target proteins (40). *Slc34a2* is a phosphate transporter that may be involved in the transport of phosphate into cells via Na^+ co-transport, providing sufficient phosphate for normal bone development (41). *TAPT1* is a transmembrane transporter that may be involved in the transduction or transmission of extracellular information needed for the axial skeletal pattern during development (42). In addition, *TAPT1* has also been reported to be involved in the development of craniofacial cartilage, which is related to the differentiation of cranial neural crest cells (43).

Overall, the candidate genes *CCDC149*, *LGI2*, and *SEPSECS* identified in this study are likely to be key genes that affect the beak length phenotype in geese. The brain is the most advanced part of the nervous system and its development determines the formation of the face, which undoubtedly has an important impact on the formation of beaks. Specifically, the key candidate genes identified in this study play a crucial role in the brain and nerves, which provides basic cognition for the formation and development of the beak.

CONCLUSIONS

We conducted a GWAS for the beak length trait in the XGG population, based on SNPs obtained using whole-genome resequencing data. The analyses revealed that 57 SNPs on

TABLE 1 | Summary of significant SNPs that are associated with the beak length trait.

SNPs ^a	P-value	Allele (major/minor)	MAF	LD (r^2) ^b	Variant annotation ^c	Candidate genes ^d
4_19121505	3.89E-12	G/T	0.30	1	Intergenic	<i>CCDC149</i>
4_19115564	5.30E-12	C/T	0.29	0.74	Upstream	<i>CCDC149</i>
4_19115596	7.45E-12	T/C	0.30	0.75	Upstream	<i>CCDC149</i>
4_19138017	1.33E-11	T/A	0.28	0.92	Downstream	<i>LGI2</i>
4_19162352	1.78E-11	G/A	0.30	0.86	Upstream	<i>LGI2</i>
4_19166083	5.43E-11	T/G	0.30	0.86	Intergenic	<i>LGI2</i>
4_19162397	7.03E-11	C/G	0.33	0.75	Upstream	<i>LGI2</i>
4_19166557	8.37E-11	T/G	0.30	0.87	Intergenic	<i>LGI2</i>
4_19149430	8.39E-11	T/C	0.23	0.71	Intron	<i>LGI2</i>
4_19289093	9.90E-11	A/G	0.35	0.51	Intergenic	-
4_19149283	1.74E-10	A/G	0.23	0.71	Intron	<i>LGI2</i>
4_19294723	2.37E-10	T/C	0.34	0.51	Upstream	<i>Slc34a2</i>
4_19177265	2.58E-10	C/T	0.25	0.66	UTR	<i>SEPSECS</i>
4_19246235	2.85E-10	G/A	0.29	0.35	Intron	<i>ANAPC4</i>
4_19166770	2.90E-10	A/G	0.25	0.66	Intergenic	<i>LGI2</i>
4_19182087	3.03E-10	T/A	0.22	0.56	Intron	<i>SEPSECS</i>
4_19181967	3.11E-10	C/T	0.23	0.56	Intron	<i>SEPSECS</i>
4_19294546	3.80E-10	G/C	0.34	0.50	Intergenic	<i>Slc34a2</i>
4_19294090	4.04E-10	G/A	0.34	0.51	Intergenic	<i>Slc34a2</i>
4_18957086	4.33E-10	T/A	0.15	0.37	Intergenic	<i>DHX15</i>
4_19181666	4.78E-10	T/C	0.22	0.56	Intron	<i>SEPSECS</i>
4_19150091	5.56E-10	C/T	0.22	0.67	Intron	<i>LGI2</i>
4_19150098	5.56E-10	A/G	0.22	0.67	Intron	<i>LGI2</i>
4_18957231	5.57E-10	A/G	0.14	0.34	Intergenic	<i>DHX15</i>
4_19180040	5.77E-10	T/G	0.34	0.65	Missense	<i>SEPSECS</i>
4_19177427	6.10E-10	T/C	0.23	0.57	UTR	<i>SEPSECS</i>
4_19297473	6.58E-10	G/A	0.34	0.5	Upstream	<i>Slc34a2</i>
4_19182183	6.73E-10	G/C	0.23	0.56	Intron	<i>SEPSECS</i>
4_19181913	6.87E-10	A/T	0.23	0.56	Intron	<i>SEPSECS</i>
4_19179134	6.89E-10	A/G	0.22	0.56	UTR	<i>SEPSECS</i>
4_19291403	7.54E-10	T/C	0.36	0.5	Intergenic	<i>Slc34a2</i>
4_19190667	7.85E-10	G/A	0.26	0.48	Intron	<i>SEPSECS</i>
4_19160006	7.91E-10	C/A	0.25	0.62	Upstream	<i>LGI2</i>
4_19134771	8.53E-10	A/G	0.16	0.48	Intergenic	<i>LGI2</i>
4_19183481	9.22E-10	G/A	0.23	0.6	Intron	<i>SEPSECS</i>
4_19183064	1.01E-09	A/G	0.22	0.56	Intron	<i>SEPSECS</i>
4_19165746	1.03E-09	A/G	0.33	0.8	Intergenic	<i>LGI2</i>
4_19206949	1.07E-09	G/A	0.24	0.52	Upstream	<i>SEPSECS</i>
4_19297635	1.10E-09	G/A	0.34	0.5	Upstream	<i>Slc34a2</i>
4_19178071	1.16E-09	G/A	0.23	0.56	UTR	<i>SEPSECS</i>
4_19203675	1.18E-09	C/G	0.24	0.63	Upstream	<i>SEPSECS</i>
4_19121824	1.58E-09	A/G	0.28	0.64	Intergenic	<i>CCDC149</i>
4_19166158	1.72E-09	G/A	0.32	0.79	Intergenic	<i>LGI2</i>
4_19178068	1.84E-09	C/A	0.37	0.71	UTR	<i>SEPSECS</i>
4_19179255	1.99E-09	G/C	0.34	0.65	UTR	<i>SEPSECS</i>
4_19134585	2.07E-09	A/G	0.16	0.48	Intergenic	<i>LGI2</i>
4_19401591	2.21E-09	C/T	0.44	0.35	Intergenic	-
4_19404666	2.40E-08	T/G	0.44	0.35	Intergenic	-
4_19401510	2.47E-09	T/G	0.45	0.35	Intergenic	-
4_16445056	2.95E-09	G/A	0.38	0.24	Intergenic	-
4_19135882	3.20E-09	A/T	0.34	0.71	Intergenic	<i>LGI2</i>

(Continued)

TABLE 1 | Continued

SNPs ^a	P-value	Allele (major/minor)	MAF	LD (r^2) ^b	Variant annotation ^c	Candidate genes ^d
4_19117354	3.67E-09	A/G	0.23	0.71	Upstream	<i>CCDC149</i>
4_19158002	3.86E-09	G/A	0.23	0.72	Upstream	<i>LGI2</i>
4_19404633	3.88E-09	C/T	0.45	0.35	Intergenic	-
4_19108634	3.89E-09	G/A	0.27	0.62	Intron	<i>CCDC149</i>
4_15983935	3.99E-09	G/A	0.1	0.13	Intergenic	<i>TAPT1</i>
4_19246383	4.20E-09	T/C	0.3	0.32	Intron	<i>ANAPC4</i>

MAF, minor allele frequency; LD, linkage disequilibrium.

^aSNPs are expressed as chromosome_position.

^b r^2 represents the degree of LD between the top SNP and nearby SNPs.

^cVariation annotation is supported by snpEff (v4.3t) software.

^dThe upstream and downstream influence range of a single gene was extended to 10 Kb to facilitate the detection of associated candidate genes when conducting association analyses. “-”, Indicates that no candidate gene was found.

chromosome 4 were significantly associated with beak length at the genome-wide level, and seven candidate genes located near them were identified, namely *TAPT1*, *DHX15*, *CCDC149*, *LGI2*, *SEPSECS*, *ANAPC4*, and *Slc34a2*. Among them, *CCDC149*, *LGI2*, and *SEPSECS* might be the key genes affecting the beak length of XGG, especially the *LGI2* gene, which may play a vital role in the development of cranial neural crest cells and are highly relevant to the formation of geese beaks. The results of this study lay a foundation for studying the genetic mechanisms of phenotypic traits in geese, and also provide references for the molecular genetics of beak development in other avian species.

DATA AVAILABILITY STATEMENT

All raw resequencing data used in this study have been deposited into the NCBI BioProject database under accession PRJNA678815.

ETHICS STATEMENT

The animal study was reviewed and approved by Jiangxi Science and Technology Normal University.

AUTHOR CONTRIBUTIONS

JH, XY, and HC conceived and designed the study. CW and HC analyzed the data. CW, JH, and HC wrote the paper. JO, HT, SZ, YX, YG, YW, and LW collected data and performed the

sequencing. XY and JO provided funding support. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 31860622 and 32060735).

ACKNOWLEDGMENTS

The authors are grateful to the relevant staff at Xingguo Gray Goose Conservation Farm for their assistance in the housing and care of the animals.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2022.847481/full#supplementary-material>

Supplementary Figure S1 | Image of beak of XGG.

Supplementary Figure S2 | Q-Q plot of beak length trait.

Supplementary Figure S3 | The predicted 3-D structures of SEPSECS corresponding to a substitution of glutamine (A) to lysine (B) at position 420.

Supplementary Figure S4 | Haplotype block analysis of all significant SNPs for beak length trait.

Supplementary Table S1 | Summary of all detected significant SNP loci that associated with beak length trait. The red background showed the SNPs above the genome-wide level.

REFERENCES

- Chen KW, Yang N, Wang GY, Wang ZY, Wang KH, Wang JY, et al. *Animal Genetic Resources in China: Poultry*. Beijing: China Agriculture Press (2011). 513 p.
- Weinberg SM, Roosenboom J, Shaffer JR, Shriver MD, Wysocka J, Claes P. Hunting for genes that shape human faces: initial successes and challenges for the future. *Orthod Craniofac Res*. (2019) 22:207–12. doi: 10.1111/ocr.12268
- Lamichaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*. (2015) 518:371–5. doi: 10.1038/nature14181
- Lamichaney S, Han F, Berglund J, Wang C, Almen MS, Webster MT, et al. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*. (2016) 352:470–4. doi: 10.1126/science.aad8786
- Wu P, Jiang TX, Shen JY, Widelitz RB, Chuong CM. Morphoregulation of avian beaks: comparative mapping of growth zone activities and morphological evolution. *Dev Dyn*. (2006) 235:1400–12. doi: 10.1002/dvdy.20825
- Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations, and practical recommendations. *Mol Ecol*. (2017) 26:5369–406. doi: 10.1111/mec.14264
- Choi JW, Chung WH, Lee KT, Cho ES, Lee SW, Choi BH, et al. Whole-genome resequencing analyses of five pig breeds, including Korean wild

- and native, and three European origin breeds. *DNA Res.* (2015) 22:259–67. doi: 10.1093/dnares/dsv011
8. Jiang J, Gao Y, Hou Y, Li W, Zhang S, Zhang Q, et al. Whole-genome resequencing of holstein bulls for indel discovery and identification of genes associated with milk composition traits in dairy cattle. *PLoS ONE.* (2016) 11:e0168946. doi: 10.1371/journal.pone.0168946
 9. Yang S, Shi Z, Ou X, Liu G. Whole-genome resequencing reveals genetic indels of feathered-leg traits in domestic chickens. *J Genet.* (2019) 98:47. doi: 10.1007/s12041-019-1083-4
 10. Zhang Y, Guo Q, Bian Y, Wang Z, Xu Q, Chang G, et al. Whole genome resequencing of crested traits and expression analysis of key candidate genes in duck. *Gene.* (2020) 729:144282. doi: 10.1016/j.gene.2019.144282
 11. Yu S, Wang G, Liao J, Liu F. Establishment of an efficient and common method for extracting genomic DNA from poultry blood. *Heilongjiang Animal Sci Veter Med.* (2018) 13:195–8. doi: 10.13881/j.cnki.hljxmsy.2017.11.0579
 12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
 13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
 14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
 15. Freed D, Aldana R, Weber JA, Edwards JS. The sentieon genomics tools – A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv.* (2017) 115717. doi: 10.1101/115717
 16. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010) 20:1297–303. doi: 10.1101/gr.107524.110
 17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* (2007) 81:559–75. doi: 10.1086/519795
 18. Davies RW, Flint J, Myers S, Mott R. Rapid genotype imputation from sequence without reference panels. *Nat Genet.* (2016) 48:965–9. doi: 10.1038/ng.3594
 19. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet.* (2019) 51:1749–55. doi: 10.1038/s41588-019-0530-8
 20. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. 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.* (2012) 6:80–92. doi: 10.4161/fly.19695
 21. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* (2005) 21:263–5. doi: 10.1093/bioinformatics/bth457
 22. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* (2014) 13:397–406. doi: 10.1074/mcp.M113.035600
 23. Gu WL, Gibert Y, Wirth T, Elischer A, Bloch W, Meyer A, et al. Using gene-history and expression analyses to assess the involvement of LGI genes in human disorders. *Mol Biol Evol.* (2005) 22:2209–16. doi: 10.1093/molbev/msi214
 24. Puppala AK, French RL, Matthies D, Baxa U, Subramaniam S, Simonovic M. Structural basis for early-onset neurological disorders caused by mutations in human selenocysteine synthase. *Sci Rep.* (2016) 6:32563. doi: 10.1038/srep32563
 25. Bhullar BA, Morris ZS, Sefton EM, Tok A, Tokita M, Namkoong B, et al. A molecular mechanism for the origin of a key evolutionary innovation, the bird beak and palate, revealed by an integrative approach to major transitions in vertebrate history. *Evolution.* (2015) 69:1665–77. doi: 10.1111/evo.12684
 26. Schneider RA, Helms JA. The cellular and molecular origins of beak morphology. *Science.* (2003) 299:565–8. doi: 10.1126/science.1077827
 27. Le Douarin NM, Brito JM, Creuzet S. Role of the neural crest in face and brain development. *Brain Res Rev.* (2007) 55:237–47. doi: 10.1016/j.brainresrev.2007.06.023
 28. Abramyan J, Richman JM. Craniofacial development: discoveries made in the chicken embryo. *Int J Dev Biol.* (2018) 62:97–107. doi: 10.1387/ijdb.170321ja
 29. Ealba EL, Jheon AH, Hall J, Curantz C, Butcher KD, Schneider RA. Neural crest-mediated bone resorption is a determinant of species-specific jaw length. *Dev Biol.* (2015) 408:151–63. doi: 10.1016/j.ydbio.2015.10.001
 30. Li S, Bai S, Qin X, Zhang J, Irwin DM, Zhang S, et al. Comparison of whole embryonic development in the duck (*Anas platyrhynchos*) and goose (*Anser cygnoides*) with the chicken (*Gallus gallus*). *Poult Sci.* (2019) 98:3278–91. doi: 10.3382/ps/pez133
 31. Cheng Y, Gao B, Wang H, Han N, Shao S, Wu S, et al. Evolution of beak morphology in the ground tit revealed by comparative transcriptomics. *Front Zool.* (2017) 14:58. doi: 10.1186/s12983-017-0245-6
 32. Hsia HE, Tushaus J, Brummer T, Zheng Y, Scilabra SD, Lichtenthaler SF. Functions of 'A disintegrin and metalloproteases (ADAMs)' in the mammalian nervous system. *Cell Mol Life Sci.* (2019) 76:3055–81. doi: 10.1007/s00018-019-03173-7
 33. Ji Y, Hao H, Reynolds K, McMahon M, Zhou CJ. Wnt signaling in neural crest ontogenesis and oncogenesis. *Cells.* (2019) 8:1173. doi: 10.3390/cells8101173
 34. Wirth EK, Bharathi BS, Hatfield D, Conrad M, Brielmeier M, Schweizer U. Cerebellar hypoplasia in mice lacking selenoprotein biosynthesis in neurons. *Biol Trace Elem Res.* (2014) 158:203–10. doi: 10.1007/s12011-014-9920-z
 35. Fradejas-Villar N. Consequences of mutations and inborn errors of selenoprotein biosynthesis and functions. *Free Radic Biol Med.* (2018) 127:206–14. doi: 10.1016/j.freeradbiomed.2018.04.572
 36. Bright JA, Marugan-Lobon J, Cobbe SN, Rayfield EJ. The shapes of bird beaks are highly controlled by non-dietary factors. *Proc Natl Acad Sci USA.* (2016) 113:5352–7. doi: 10.1073/pnas.1602683113
 37. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. Bmp4 and morphological variation of beaks in Darwin's finches. *Science.* (2004) 305:1462–5. doi: 10.1126/science.1098095
 38. Charvet CJ, Striedter GF. Developmental modes and developmental mechanisms can channel brain evolution. *Front Neuroanat.* (2011) 5:4. doi: 10.3389/fnana.2011.00004
 39. Yoshimoto R, Kataoka N, Okawa K, Ohno M. Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.* (2009) 37:891–902. doi: 10.1093/nar/gkn1002
 40. Jin L, Williamson A, Banerjee S, Philipp I, Rape M. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell.* (2008) 133:653–65. doi: 10.1016/j.cell.2008.04.012
 41. Marks J. The role of SLC34A2 in intestinal phosphate absorption and phosphate homeostasis. *Pflugers Arch.* (2019) 471:165–73. doi: 10.1007/s00424-018-2221-1
 42. Howell GR, Shindo M, Murray S, Gridley T, Wilson LA, Schimenti JC. Mutation of a ubiquitously expressed mouse transmembrane protein (Tapt1) causes specific skeletal homeotic transformations. *Genetics.* (2007) 175:699–707. doi: 10.1534/genetics.106.065177
 43. Symoens S, Barnes AM, Gistelincck C, Malfait F, Guillemin B, Steyaert W, et al. Genetic defects in TAPT1 disrupt ciliogenesis and cause a complex lethal osteochondrodysplasia. *Am J Hum Genet.* (2015) 97:521–34. doi: 10.1016/j.ajhg.2015.08.009

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Wang, Ouyang, Tang, Zheng, Xiong, Gao, Wu, Wang, Yan and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.