

RESEARCH

Open Access



Origins, timing and introgression of domestic geese revealed by whole genome data

Junhui Wen¹, Haiying Li², Huie Wang^{3,4}, Jincheng Yu⁵, Tao Zhu¹, Jinxin Zhang¹, Xinghua Li¹, Zhihua Jiang⁶, Zhonghua Ning¹ and Lujiang Qu^{1*}

Abstract

Background Geese are among the most important poultry species in the world. The current generally accepted hypothesis is that the European domestic geese originated from greylag geese (*Anser anser*), and Chinese domestic geese have two origins, most of which originated from swan geese (*Anser cygnoides*), and the Yili goose originated from greylag geese. To explain the origin and demographic history of geese, we selected 14 goose breeds from Europe and China and wild populations of swan and greylag geese, and whole genome sequencing data were obtained for 74 samples.

Results Population structure analysis and phylogenetic trees showed that the wild ancestor of Chinese domestic geese, except for Yili, is the swan geese, and the wild ancestor of Chinese Yili and European domestic geese is greylag geese. Analysis of the demographic history suggests that the domestication of Chinese geese occurred ~3499 years ago and that of the European geese occurred ~7552 years ago. Furthermore, gene flow was observed between domestic geese and their wild ancestors. Analysis of introgression showed that Yili geese had been introgressed by Chinese domestic geese, and the body size of Yili geese may be influenced by introgression events of some growth-related genes, including *IGF-1*.

Conclusions Our study provides evidence for the origin of geese at the genome-wide level and advances the understanding of the history of goose domestication and the traits affected by introgression events.

Keywords Domestication, Goose, Introgression, Phylogeny

*Correspondence:

Lujiang Qu
qulij@163.com

¹ Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China

² College of Animal Science, Xinjiang Agricultural University, Urumqi, China

³ Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Alar, Xinjiang, China

⁴ College of Animal Science, Tarim University, Alar, Xinjiang, China

⁵ Liaoning Academy of Agricultural Sciences, Shenyang, China

⁶ Department of Animal Sciences, Washington State University, Pullman, USA

Background

Animal domestication has led to the development of urban communities and expansion of agricultural economies. It is an important event in human history and has contributed to the progression of human civilizations [1]. Domestication is the process by which humans select animals that meet their needs and preferences, causing domesticated animals to differ from their wild ancestors [2]. Domestication syndrome is a common characteristic of domestic animals [3], and includes phenotypical, behavioral, and production variations. These variations have been selected to correlate with human needs, such as meat, eggs, and milk as major sources of high-quality



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

protein. The genetic mechanisms underlying variation due to domestication have been investigated using molecular techniques in a large variety of domesticated animals, including dogs [4], goats [5], cattle [6], chickens [7], and ducks [8].

Domestic geese (family: Anatidae, order: Anseriformes) are economically important poultry [9]. Romanov proposed six centers of domestication, breed formation, and dispersal for domestic geese based on genetic, phenotypic, phylogenetic, and historical analyses: Western Europe, China, Eurasia, Egypt, North America, and Australia [10]. Geese were recorded in ancient Egyptian paintings (2686–1991 BC) [11]. Ancient Greek and Roman written records include mentions of domestic geese. Romans consumed goose meat and eggs, produced fatty liver, and used goose feathers to make goose quills and ornaments [12]. The domestication of geese in China can be traced back

to the Neolithic period (5000 BC), as evidenced by a goose-shaped artifact from that period [13]. In 2022, goose bone fossils were found in a 7000-year-old rice-cultivation village in the lower Yangtze River, China [14]. After a long period of domestication, China has had abundant domestic-geese breeding resources. In 2020, China published a national list of livestock genetic resources, including 30 Chinese goose breeds [15].

The taxonomy and phylogeny of geese have been studied for more than 300 years, and traditional classification is based on behavioral traits, morphological characteristics, and anatomical structure [16]. There are clear morphological differences between the two regions of Europe and China for domestic geese. The Chinese domestic goose is relatively small and light, with high egg production [17]. It has strong legs and wings, and can forage over large distances. This may

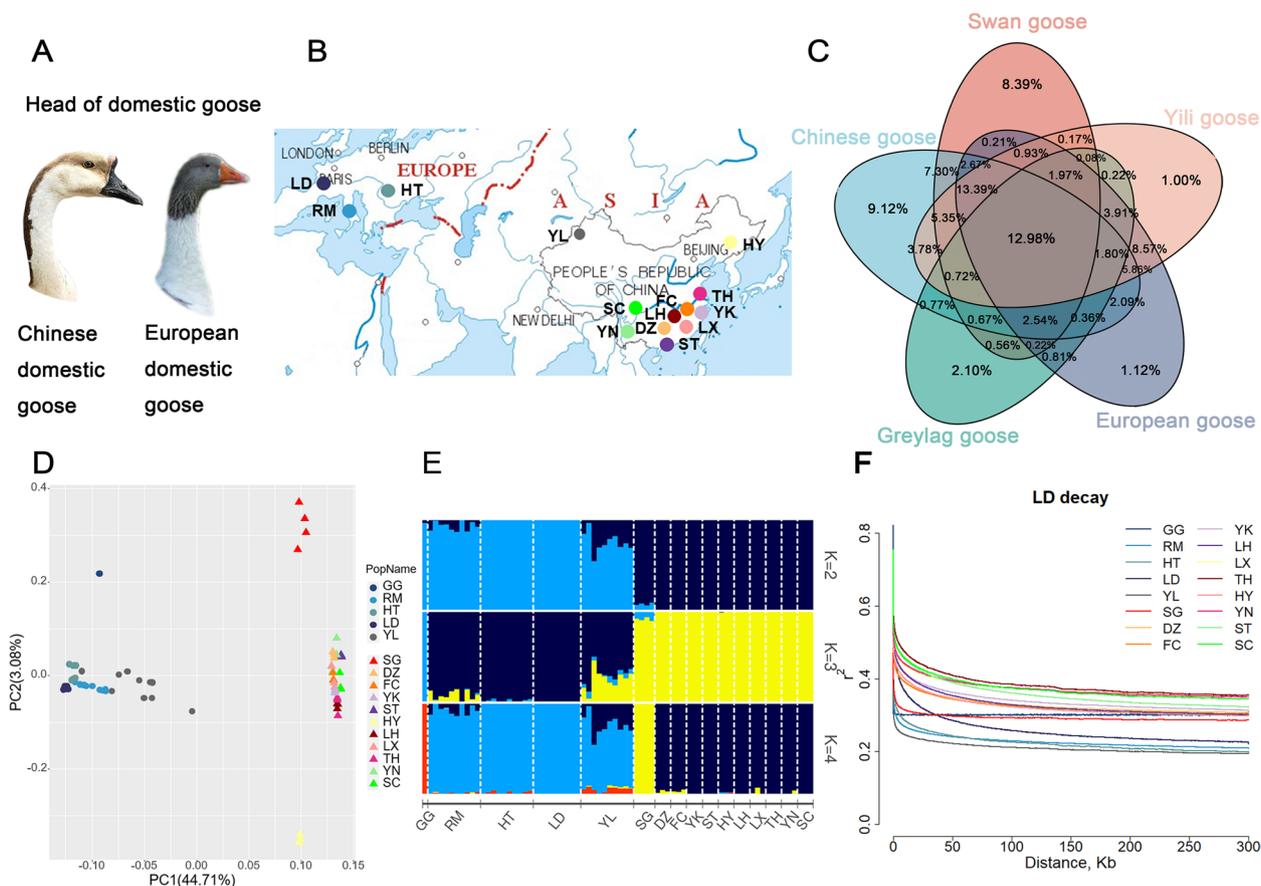


Fig. 1 Experimental setup and population genetic structure. **A** Knob on head of domestic geese, the Chinese domestic goose is on the left, and the European domestic goose is on the right. **B** Location of the samples used for this study. A total of 74 geese, including swan geese ($n = 4$; origin, China), greylag goose ($n = 1$; origin, Europe), 11 Chinese goose breeds ($n = 10$ of YL; $n = 3$, each a different breed), and three European domestic goose breeds ($n = 29$) were used. See Additional file 1: Table S1. **C** Genomic SNPs of five populations. **D** PCA plot of geese breeds. **E** Population genetic structure of geese. The length of each colored segment represents the proportion of the individual genome from the ancestral populations ($K = 2-4$), population names are at the bottom. **F** LD decay patterns

explain the extended distribution of Chinese geese beyond Chinese borders into India and Siberia, which increases the possibility of gene flow events [18]. The Chinese domestic goose has a knob at the base of its beak, while European domestic geese and Yili geese in northwest China do not (Fig. 1A). The European domestic goose is relatively large and heavy, with less egg production than the Chinese geese [18, 19].

It is generally accepted that domestic geese have two origins [9]. The ancestors of domestic geese in Europe are the greylag geese (*Anser anser*). The ancestors of domestic geese in China have two sources. Yili geese are domesticated from greylag geese, whereas all the other local Chinese geese are derived from swan geese (*Anser cygnoides*) [20]. Darwin suggested that the European domestic goose was derived from the greylag, possibly based on plumage patterns [2]. In 1968, Bhatnagar found that the karyotypes of European domestic and European greylag geese were identical, and the karyotypes of Chinese domestic and Asian swan geese were identical [21]. In 1998, Shi et al. performed restriction fragment length polymorphism analyses on mitochondrial DNA from Chinese geese and found that Yili have a different origin than other Chinese geese [22]. In 2011, Li et al. performed shared haplotype and systematic evolution analyses using the 521-bp control region (D-loop) of mitochondrial DNA from 26 Chinese domestic goose and six Landaise geese breeds and found two maternal origins of Chinese domestic geese [23]. Gao et al. performed a phylogenetic analysis using single-copy genes to study the evolution of geese in 2016, and found that wild and domestic geese clustered into a subclade, which is consistent with the hypothesis that domestic geese were domesticated from wild geese [24].

In 2020, Heikkinen et al. showed that domestication and interspecific hybridization of European domestic geese using genome-wide markers [25]. Interspecific genetic introgression, which alters traits, is also a topic of interest in domestication studies. A typical finding of introgression is the historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep [26]. Chen et al. found historical introgression events in East Asian domestic cattle that helped them adapt to high-altitude environments [27]. There are no studies on the mechanisms of introgression and alteration of traits in domestic geese of different origins.

To elucidate the evolutionary history and introgression events of the domestic goose, we analyzed the genomes of 74 geese, including the swan goose, greylag goose, and 14 domestic goose breeds. We identified the origin of Chinese and European domestic geese at the genomic level, and inferred the timing of domestic goose

domestication. Additionally, our study identified substantial evidence of gene flow from swan goose originating from Chinese geese to greylag goose originating from Yili geese, and found that introgression could lead to changes in the body size of Yili geese.

Methods

Sampling

Whole blood samples were obtained from 74 geese, including 11 Chinese domestic goose breeds, three European domestic goose breeds, and their wild ancestors, the swan goose (*A. cygnoides*) (SG, $n = 4$), and the greylag goose (*A. anser*) (GG, $n = 1$), whose data were acquired from a previous study [28]. Domestic geese have two origins. Ten Chinese domestic goose breeds are thought to have originated from the swan goose, including Daozhou grey goose (DZ, $n = 3$), Fengcheng grey goose (FC, $n = 3$), Yongkang grey goose (YK, $n = 3$), lion-head goose (ST, $n = 3$), Huoyan goose (HY, $n = 3$), Lianhua white goose (LH, $n = 3$), Linxian white goose (LX, $n = 3$), Taihu goose (TH, $n = 3$), Yunnan goose (YN, $n = 3$), and Sichuan white goose (SC, $n = 3$). A Chinese domestic goose breed, Yili goose (YL, $n = 10$), from Northwest China, Xinjiang, and three European domestic goose breeds, Roman goose (RM, $n = 10$), Hortobagy goose (HT, $n = 10$), and Landaise goose (LD, $n = 9$), are regarded as originating from the greylag (Additional file 1: Table S1). The data of Roman and Hortobagy geese were downloaded from the National Center for Biotechnology Information (NCBI, BioProject ID: PRJNA722049).

Sequencing and variant calling

Genomic DNA was extracted using the standard phenol/chloroform extraction method [29]. The quality and integrity of the extracted DNA were verified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Paired-end (150 bp) libraries were sequenced and built using the Illumina Novaseq 6000 platform, according to the manufacturer's instructions. Quality control of raw sequencing data was performed using the NGS QC Toolkit (v2.3.3) with the default parameters [30].

Paired-end reads were mapped to the goose (*Anser cygnoides domesticus*) genome version 1.0, using BWA (v 0.7.17) [31, 32] with default parameters. The alignments were sorted and duplicate sequences were removed using Picard [33]. The reads were realigned using RealignerTargetCreator and IndelRealigner in GATK (v3.8) [34] to reduce the error rate of alignment. Variant calling was performed using

UnifiedGenotyper in GATK (v3.6) with a minimum base quality of 20. Variant filtering was performed using VariantFiltration in GATK (v3.6) with the recommended parameters. After filtering, 24,891,622 SNPs and 2,885,072 indels were obtained and used in the subsequent analyses.

Population structure

VCFtools software (v0.1.16) [35] and PLINK (v1.90) [36] was used to convert the VCF file to the PLINK format. Stringent filtering criteria were applied with the following parameters: `--geno 0.1`, `--maf 0.1`, and `--indep-pairwise 20 10 0.5`. After filtering, 11,410,883 variants were obtained for principal component analysis (PCA) with PLINK (v1.90). This analysis was performed using default parameters to extract the top 20 principal components of the variance-standardized relationship matrix.

The filtered dataset of 11,410,883 variants was used to analyze the admixture of each individual and population clustering using ADMIXTURE (v1.3.0) [37]. The VCF file was converted to the ADMIXTURE input file format using PLINK (v1.90). The analysis was performed by setting the assumed number of ancestral populations, K , from 2 to 16. Cross-validation (CV) was performed with a five-fold value to verify the optimal ancestral populations. When the CV error value was the lowest, the corresponding K value represented the optimal number of ancestral populations. The final population clustering and individual admixture results were visualized using Pophelper package [38].

Phylogenetic tree reconstruction was performed according to the SNPhylo protocol [39]. The VCF files were converted to HapMap format using custom Perl (v5.32.0) scripts. Maximum likelihood (ML) trees were constructed using DNAML programs in PHYLIP (v3.697), with bootstrapping analysis performed with Phangorn. The phylogenetic trees were visualized using iTOL [40].

PopLDdecay software was used to calculate the linkage disequilibrium (LD) decay of the 16 geese breeds [41]. The SNPs were filtered using the following parameters: `-MAF 0.05` and `-Miss 0.25`. Then, the LD measurement r^2 was calculated using the filtered dataset and a final LD decay plot was generated.

Demographic analysis

Population demographic history analysis of domestic geese and their wild ancestors was performed using the SMC++ [42]. The parameters for this analysis were set based on the goose genome and studies from other birds to an average mutation rate (μ) per generation of 1×10^{-9} and generation time (g) of 2 years [43].

Demographic inferences were made using DaDi [44]. To avoid the effect of selection on the demographic analysis, we focused on non-coding regions and extracted non-coding SNPs. The final dataset contained 23,080,074 SNPs, spanning 1,041,100,496 bp. To avoid the linkage of SNPs, we thinned the non-coding SNPs to 1% and obtained a dataset of 230,801 SNPs. Owing to the unknown ancestral state of each SNP, we folded the frequency spectrum. The projection value was selected based on the strategy of maximizing the number of segregated sites.

We tested several models to estimate the demographic history of Chinese and European domestic geese and the gene flow which occurred during domestication. As the Yili and European domestic geese share the same origin (see results), we added the Yili goose samples to the European domestic geese group. For each model, multiple runs were performed to ensure that the parameters converged to a similar log-likelihood. The starting parameter values for the first round were randomly assigned, and the best parameters obtained after completion of a round were used as the starting parameters for the subsequent rounds. After the convergence of the parameters, we retained the model with the highest log-likelihood as the final simulation result. The parameters of the optimal model were converted into absolute units using the average mutation rate per generation and generation interval. Confidence intervals for the parameters were generated using the Godambe information matrix (GIM) with 100 bootstraps [45].

Introgression analysis

Gene migration events between populations were inferred using TreeMix [46]. TreeMix uses genome-wide allele frequency data to infer patterns of differentiation and mixing across multiple populations. The software inputs data as allele frequencies for multiple populations to infer the population migration events. The VCF file was converted to the plink format using VCFtools (v 0.1.16) and PLINK (v1.90) was used to calculate the allele frequencies. A ML tree was constructed using TreeMix, and the parameters were set at the migration events ($m=1$).

The ABBA-BABA test, also known as the D -statistic, was used to infer the existence of gene flow between the populations. This analysis of D -statistic values was performed using Dsuite software [47], which can calculate D -statistics at the genome scale across all combinations of populations with VCF input files. According to the principle of ABBA-BABA test calculation, (P1, P2, P3, O) represent four different groups: P1 is a domestic goose group, P2 is another domestic goose group and sister group with P1, P3 is a closely related species of P1 and P2, and O is the outgroup. Gene flow between domestic

geese and their wild ancestors was inferred with the maldar as an outgroup. Dtrios was used to calculate the D and f_4 -ratio statistics for all trios of populations in the dataset, with a default value of 20,000 for the Jackknife block size. Then, using the Dinvestigate program to calculate the D -value for windows containing useable-size SNPs, the sliding window consisted of 2500 SNPs and a step of 500 SNPs. The locations of the windows with the top 5% D values were obtained and the genes in these windows were regarded as candidate genes for introgression. Functional annotation of the obtained gene dataset was performed using DAVID (v6.8) [48]. Gene ontology (GO) categories were visualized using R (v3.6) [49].

The degree of introgression in Yili goose individuals and the direction of introgression between Yili and other Chinese domestic geese were detected by calculating the F_{ST} between Yili individuals and European domestic geese and the F_{ST} between Yili individuals and other Chinese domestic geese, respectively, with a sliding window of 100kb and a step of 50kb. The ML tree was constructed for the candidate fragments using MEGAX software [50].

The potential mechanisms affecting body size in geese influenced by introgression were evaluated using the gene and 5' regulatory region sequences of *IGF-1* (insulin-like growth factor-1) in geese from NCBI and a study by Wang et al. [51]. The 5' *IGF-1* regulatory region sequence was obtained from the reference genome using BLAST [52], the 5' *IGF-1* regulatory-region SNP dataset of 74 samples were extracted, and JASPAR was used to predict the transcription factor-binding sites [53].

Results

Sequencing and variations

Whole-genome resequencing of the swan geese, the greylag geese, and 14 domestic goose breeds of two origins (Fig. 1B) was performed, and sequences were mapped to the swan goose genome (v1.0). After quality control and filtering, 923.6 Gb of high-quality sequences were obtained, with an average of 12.5 Gb per individual. Across all samples, a total of 5938 million mapped reads was obtained, with an average depth of 11.2X and an average coverage of 95.8% per individual. After variant calling, a total of 27.78 million variants was obtained, including 24.89 million SNPs and 2.885 million indels. The genomic SNPs of the five populations (swan geese, Chinese domestic geese, greylag goose, European domestic geese, Yili geese) are shown in Fig. 1C. Yili geese had 72% of its SNPs in common with that of the Chinese domestic geese and 81% in common with that of the European domestic geese.

Population structure

The PCA indicated clustering between domestic geese and their wild ancestors (Fig. 1D). In the first principal component, most Chinese domestic geese (DZ, FC, YK, ST, HY, LH, LX, TH, YN, and SC) and swan geese clustered, while Yili geese, European domestic geese (RM, HT, and LD), and greylag goose clustered. The PCA also showed divergence within the clustering; the Yili geese deviated towards other Chinese domestic geese in the first principal component and the wild ancestors separated from domestic geese in the second principal component.

Further, the population structure was inferred using ADMIXTURE to deduce the population admixture proportions by assigning ancestral populations K from 2 to 16. At $K=2$ (CV error=0.47774), we observed two separate clusters: Chinese domestic geese (DZ, FC, YK, ST, HY, LH, LX, TH, YN, and SC) and the swan geese formed one cluster, while Yili, European domestic geese (RM, HT, and LD), and the greylag goose formed the second cluster. At $K=3$ (CV error=0.47738), CV error was the lowest, three clusters were observed, the greylag goose clustered alone, European domestic and Yili geese clustered, and the swan geese and other Chinese domestic geese (DZ, FC, YK, ST, HY, LH, LX, TH, YN, and SC) clustered. At $K=4$ (CV error=0.50603), four clusters were observed: greylag goose, swan geese, and Chinese domestic geese (DZ, FC, YK, ST, HY, LH, LX, TH, YN, SC) clustered separately, European domestic geese and Yili clustered while showing different degrees of mixing from the greylag goose and Chinese domestic geese (Fig. 1E).

Genome-wide LD statistics were plotted for the 16 populations (Fig. 1F). The LD decay rate of domestic breeds subjected to long-term selection was slower than that of their wild ancestors, the swan geese and greylag goose showed the fastest LD decay and the smallest r^2 , indicating that they had higher diversity as wild ancestors.

The phylogenetic relationships between individual samples were inferred based on the ML trees constructed, with 1 ML tree constructed using whole genome SNPs and the other using mitochondrial SNPs (Fig. 2). In the genome-wide tree, Yili, European domestic geese, and greylag goose clustered, and the Chinese domestic geese and the swan geese clustered. The clustering observed in the mitochondrial tree supports the above results.

In conclusion, we obtained genomic evidence that most Chinese domestic goose breeds originated from the swan goose, while Yili and European domestic geese originated from the greylag goose. The Yili geese was mixed with other Chinese domestic geese.

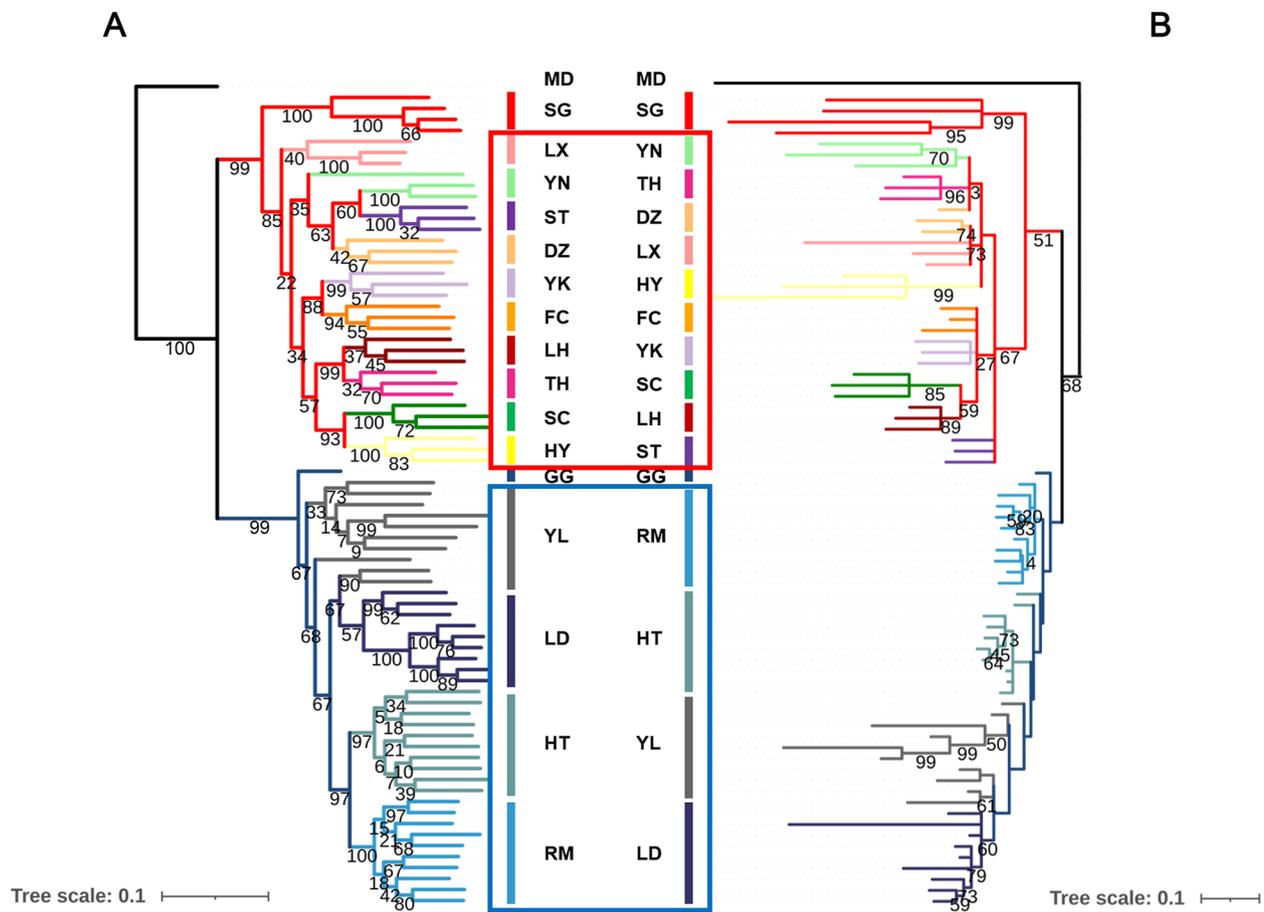


Fig. 2 Maximum likelihood phylogenetic tree of the goose population, showing the two origins of domesticated geese. **A** Genome-wide phylogenetic tree. **B** Mitochondrial phylogenetic tree. The color of the branches indicated the different breeds of geese. The value on the branches indicated the bootstrap values. MD represents the mallard (*Anas platyrhynchos*), the outgroup used in the phylogenetic analysis. SG represents the swan goose, GG represents the greylag geese, the blue box indicates European domestic and Yili geese, and red box indicates Chinese domestic geese

Demographic analysis

The SMC++ method was used to estimate changes in the historical effective population size (N_e) of the goose populations (Additional file 2: Fig. S1). The effective population size of both wild and domestic geese showed contraction and expansion with time, and domestic geese of two origins showed concordant demographic trajectories with their wild ancestors, respectively.

The demographic history of Chinese and European domestic geese was inferred by selecting four models (models 1–4). Model 1: split into two populations with symmetric migration; model 2: split into two populations with different migration; model 3: split with symmetric migration followed by isolation; and model 4: split with asymmetric migration followed by isolation (Fig. 3A).

For parameter estimation of the demographic history of Chinese domestic geese (Fig. 3B, Additional

file 2: Fig. S2), model 1 had the highest log-likelihood and the lowest Akaike Information Criteria (AIC) value, indicating the best fit of the model to the data (log-likelihood = -3291.24 ; AIC = 6590.48). The demographic parameters estimated using model 1 indicated that Chinese domestic geese and swan geese diverged 3499 years ago with a 95% confidence interval (CI) of ± 120 years (95% CI ± 120). The N_e values of swan geese and Chinese domestic geese were 3057.302 (95% CI ± 51.399) and 1866.709 (95% CI ± 46.379), respectively. Gene flow estimation between swan geese and Chinese domestic geese revealed 2.085×10^{-4} migrations per generation (95% CI $\pm 2.152 \times 10^{-7}$).

For parameter estimation of the demographic history of European domestic geese (Fig. 3C, Additional file 2: Fig. S3), model 1 was the best fit to the data (log-likelihood = -2549.66 ; AIC = 5107.32). Model 1 indicated that symmetric gene flow existed between European domestic

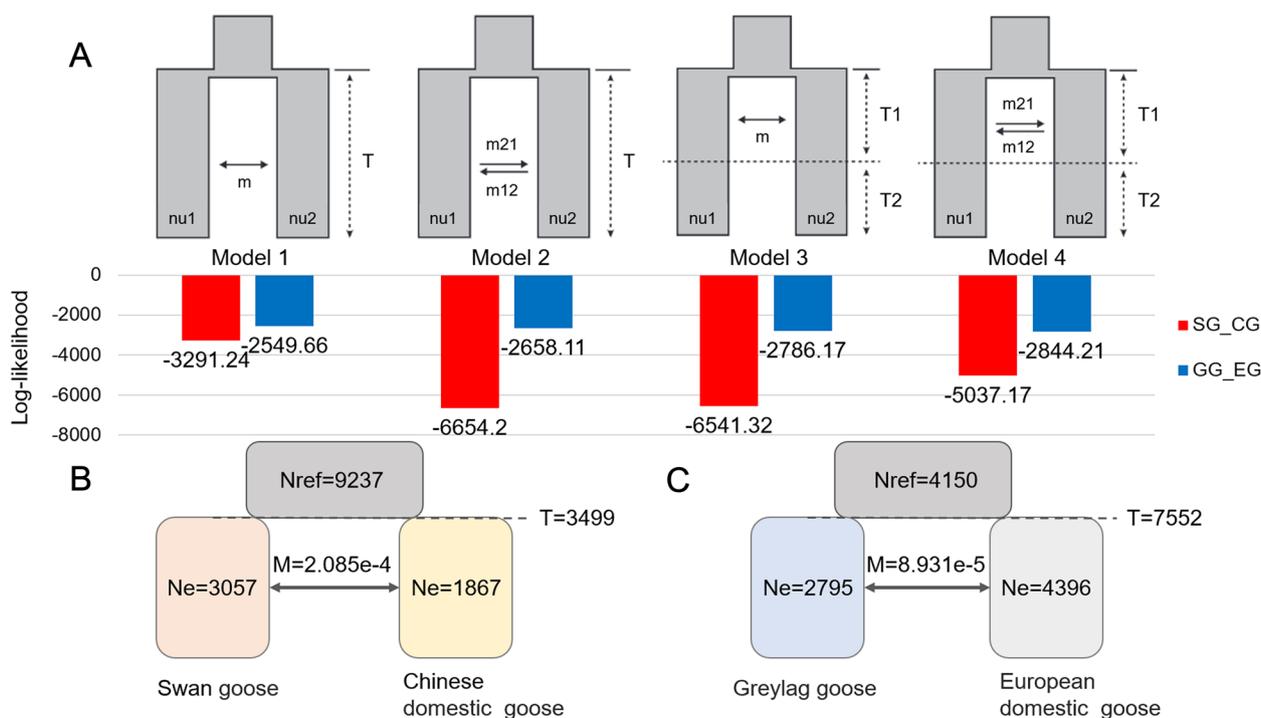


Fig. 3 Estimation of demographic parameters to show the time of goose domestication. **A** Four different demographic models and the corresponding log-likelihood values. SG_CG indicates the demographic inference of the swan geese and the Chinese domestic geese. GG_EG indicates the demographic inference of the greylag goose and European domestic geese. **B** Best-fit estimates of parameters for the Chinese domestic goose domestication model. **C** Best-fit estimates of parameters for the European domestic goose domestication model. The time units are years and migration event units are number of migrants per generation

geese and greylag goose after domestication. Demographic parameters estimated that European domestic geese and greylag goose diverged 7552 years ago (95% CI ± 10.574). The N_e values of greylag goose and European domestic geese were 2795.154 (95% CI ± 11.206) and 4395.884 (95% CI ± 4.696), respectively. Gene flow estimation showed 8.9309×10^{-5} migrations per generation between greylag goose and European domestic geese (95% CI $\pm 4.295 \times 10^{-7}$).

Introgression analysis

Migration events between populations were estimated using TreeMix and constructing ML phylogenetic trees. The results showed that gene flow from Chinese domestic geese to Yili geese occurred when the migration event was set to 1 ($M=1$). This result corroborated the PCA and ADMIXTURE results that demonstrated the existence of admixture between Yili and other Chinese domestic geese (Fig. 4A).

To further analyze introgression events, we performed D statistics for all individuals based on SNP frequency differences. To detect introgression of Chinese domestic geese to Yili geese, we set (P1, P2, P3, O) as (European domestic geese, YL, Chinese domestic geese,

mallard) (Fig. 4B, Additional file 1: Table S4). The D statistics showed significant introgression events ($Z > 3$, $P < 0.001$) from all 10 Chinese domestic goose breeds to Yili geese, with (LD, YL, TH) having the largest D value of 0.318905, indicating the strongest introgression from the Taihu geese to Yili geese. This was followed by (LD, YL, SC) with a D value of 0.284785, that shows, introgression from Sichuan geese to Yili geese.

The introgression region was further defined by identifying 1265 genes in the top 5% D -value window positions and presuming these to be candidates for introgression (Additional file 1: Table S5). GO analysis was performed on candidate genes, resulting in 227 GO enrichment terms (Additional file 1: Table S6), and the top 20 significant GO entries are shown in Fig. 4C. GO analysis revealed that genes involved in introgression were enriched in 191 biological processes, including positive regulation of cell proliferation, positive regulation of osteoblast differentiation, skeletal system development, muscle organ development, and hindlimb morphogenesis. These results suggest that the change in body size of Yili geese may be attributed to the introgression of Chinese domestic geese into Yili geese. Moreover, we identified genes associated with body size

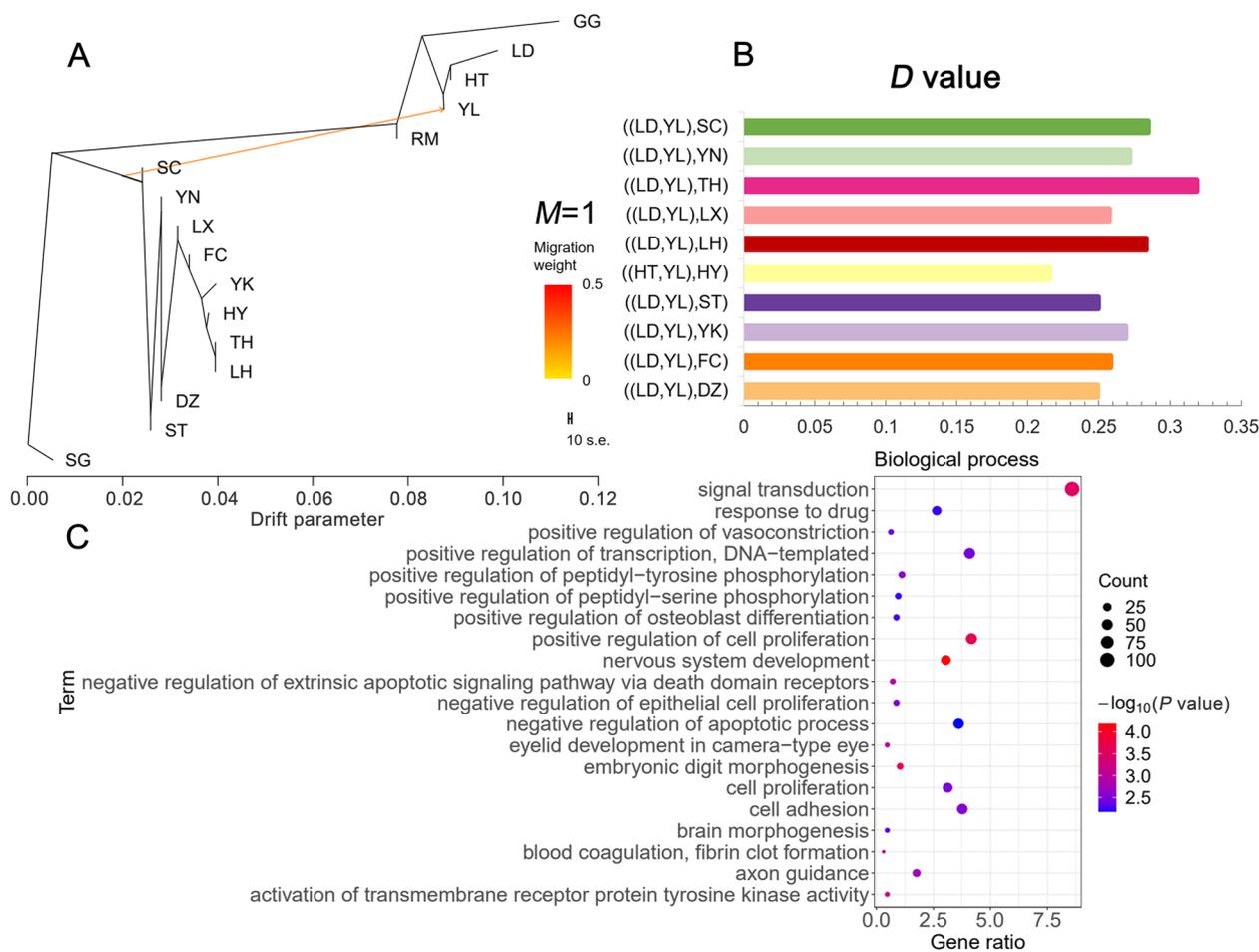


Fig. 4 Introgession analysis of goose population. **A** Genetic migration inferred using TreeMix ($M=1$). **B** D -statistics for introgression between Yili goose breed and Chinese domestic goose populations. More details in Additional file 1: Table S4. **C** GO enrichment analysis of candidate genes. The plot is shown with the top 20 significant biological processes

alterations, including *IGF-1*, *SIRT1* (sirtuin 1), *SALL3* (spalt like transcription factor 3), *GJA5* (gap junction protein alpha 5), and *SOX4* (SRX-Box transcription factor 4).

The degree and direction of introgression in the Yili goose population was further detected by calculating the F_{ST} between Yili individuals and Chinese domestic and European domestic geese, respectively, based on the results of the D -statistic analysis (Fig. 5A). The Yili goose samples were divided into three groups based on F_{ST} values: YL pop1 includes YL3; YL pop2 includes YL2, YL5, YL7, and YL8; and YL pop3 includes YL1, YL4, YL6, YL9, and YL10. We obtained a region located on scaffold NW_013185696.1 (100–350 kb), which contained the candidate gene *IGF-1*. In the region, the Chinese domestic geese and YL (pop1) were poorly differentiated ($F_{ST}=0$), European domestic geese and YL (pop1) were highly differentiated ($F_{ST}=0.89$), YL

(pop2) was moderately differentiated from both Chinese and European domestic geese, Chinese domestic geese and YL (pop3) were highly differentiated ($F_{ST}=0.78$), and European domestic geese and YL (pop3) were poorly differentiated ($F_{ST}=0.10$).

An ML tree was constructed using *IGF-1* gene sequences (Fig. 5B). In the ML tree, YL3 clustered with Chinese domestic geese; YL1, YL4, YL6, YL9, and YL10 clustered with European domestic geese; and YL2, YL5, YL7, and YL8 formed separate branches clustered with greylag goose and European domestic geese. Additionally, F_{ST} was calculated and phylogenetic trees constructed for regions where other candidate genes related to body size development were located, including *SIRT1*, *SALL3*, *GJA5*, and *SOX4*. The results of the analysis were consistent with those above, which also showed differences in Yili geese individuals (Additional file 2: Fig. S4–6). It is further suggested that there were

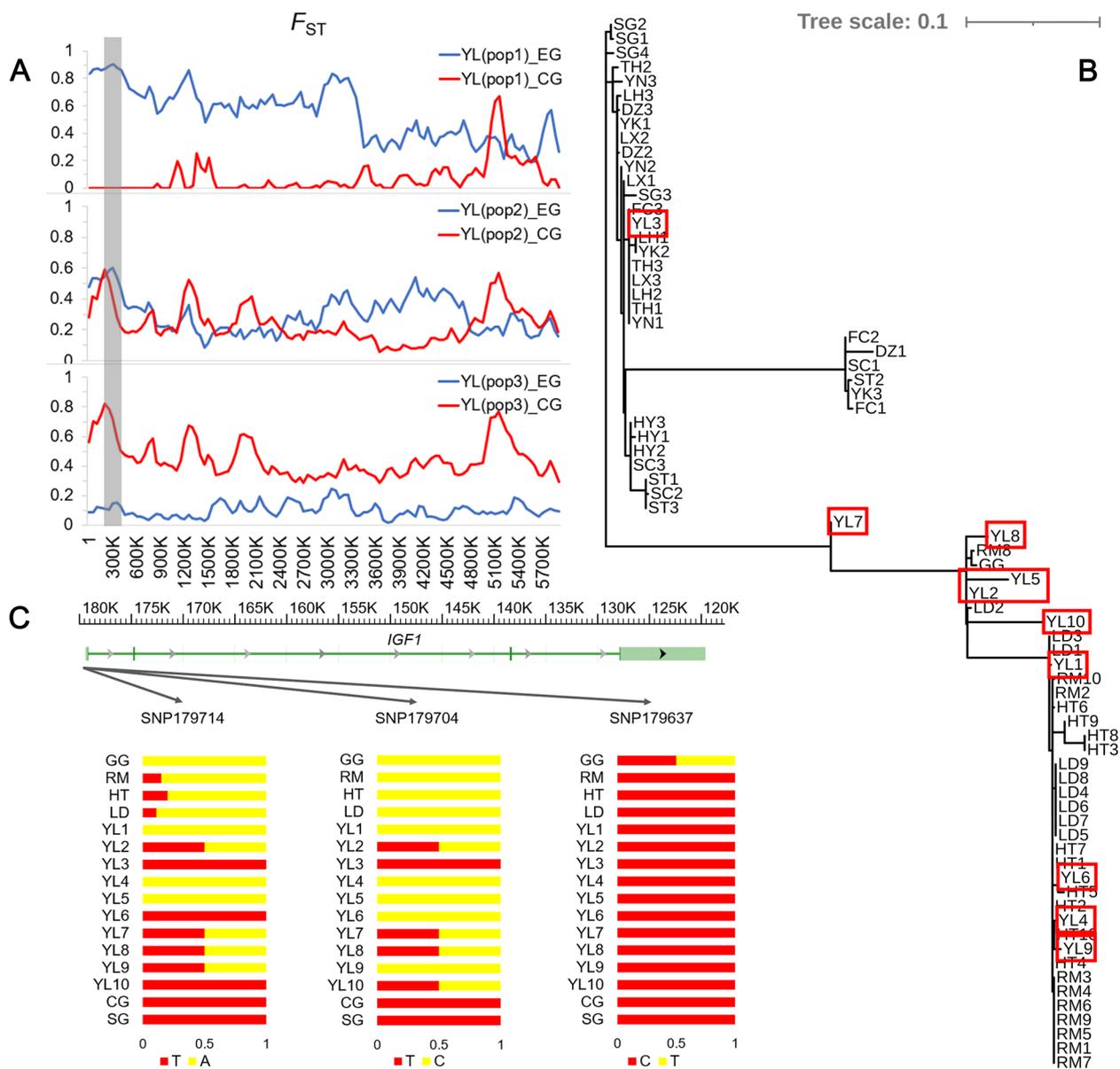


Fig. 5 Regions with introgression between Chinese domestic geese and Yili geese individuals. **A** F_{ST} between Yili goose individuals and the Chinese and European domestic goose populations across scaffold NW_013185696.1. The YL goose samples were divided into three groups based on F_{ST} values, YL (pop1) includes YL3; YL (pop2) includes YL2, YL5, YL7, and YL8; YL (pop3) includes YL1, YL4, YL6, YL9, and YL10. The blue line represents F_{ST} between Yili and European domestic geese, and red line represents F_{ST} between Yili and Chinese domestic geese. The gray region indicates the location of the *IGF-1* gene. **B** ML tree constructed using *IGF-1* sequences of scaffold region NW_013185696.1 (119,577–179,398 bp). The red frame highlights the branches of Yili goose individuals. **C** Allele frequencies of three SNPs in the 5' regulatory region of the *IGF-1* gene in goose populations, the SNP names represent their positions

introgression events in Yili geese from Chinese domestic geese, and that individual Yili geese were subjected to different degrees of introgression.

To explain the mechanism underlying the influence of introgression on body size, we studied variations in the *IGF-1* gene. The variation dataset of the coding sequence (CDS) region and *IGF-1* 5' regulatory region

was extracted; no mutations in the CDS region and three SNPs in the 5' regulatory region were found (Fig. 5C). A SNP (NW_013185696.1179704T>C) was found to be different in Chinese and European domestic goose populations. This SNP was not found in the swan geese and Chinese domestic goose populations, but was homozygous in greylag goose and European domestic geese, and

was different among individuals in Yili goose population. This SNP was homozygous in YL1, YL4, YL5, YL6, and YL9, heterozygous in YL2, YL7, YL8, and YL10, and not found in YL3. Using JASPAR to predict transcription factor binding sites for the 5' regulatory region sequence, the results showed that this SNP changed the transcription factor binding sites in the 5' regulatory region of *IGF-1*, including SRY, ISL2, EGR2, HES family, KLF family, and SP family (Additional file 1: Table S8). The mutation caused the occurrence of a binding site for transcription factor Sp1 (Specificity protein 1) in the 5' regulatory region of *IGF-1*, and the predicted sequence of the Sp1 binding site was 5'-gcagacacgcacaca-3' with a relative score higher than 0.8.

Discussion

In this study, we investigated the domestication history of Chinese and European domestic geese using whole-genome resequencing data from 14 domestic breeds and two wild populations. The results showed that the Chinese domestic geese were domesticated from the swan geese, and the European domestic geese were domesticated from the greylag geese. The Yili goose, which is distributed in Xinjiang, China, originates from the greylag geese. Gene flow was detected between Chinese domestic geese and Yili geese, and the direction of introgression and resulting effects were inferred. This study is the first to reveal the phylogenetic relationship and domestication history of the domestic geese with the swan geese and greylag at the genome-wide level and provides inferences on the causes of body size changes in the Yili geese.

Two clusters were obtained from the phylogenetic tree, the swan geese with the Chinese domestic geese cluster, and the greylag geese with the European domestic geese and the Yili geese cluster. This confirms the hypothesis that domestic geese have two origins. Our results are consistent with those of previous studies using mitochondrial data to determine the origins of Chinese and European domestic geese [20, 54]. In 2020, Heikkinen et al. used genomic data to show that the ancestor of European domestic geese was Greylag goose [25], which is consistent with our results. The topological structure of Yili geese in the genome-wide ML tree is also noteworthy, where Yili geese clustered as a group, but two individuals clustered with Landaise geese, possibly because Yili geese underwent hybridization events with Landaise geese. A similar phenomenon has been observed with the Tibetan chicken, which is a local Chinese chicken breed [55]. The two domestication clusters were also evident in the PCA and ADMIXTURE. These population structure analyses showed that the European domestic goose populations have genetic components of wild geese and Chinese

domestic geese, which is maybe due to the often observed interbreeding of geese [25, 54]. Additionally, they showed that Yili individuals contained genetic components from Chinese domestic geese, to varying degrees. This may be an introgression event due to increased geographic proximity causing crossbreeding between Yili geese and Chinese domestic geese.

DaDi was used to estimate the demographic history of Chinese and European domestic geese, and showed that the domestication of Chinese domestic geese occurred approximately 3499 years ago. This is consistent with ancient Chinese records, as there was an official position for breeding livestock and geese during the Zhou Dynasty (1046–256 BC) [13]. Domestication of poultry is generally considered to have occurred earliest in chickens and latest in ducks [56]. Domestic chickens were originated from red jungle fowl subspecies ~8000 years ago [56, 57], and ducks in East Asia were domesticated in China before 2200 BP [8]. However, recently published archaeological research has found goose bone fossils in a 7000-year-old cultivation village in the lower Yangtze River, China [14]. Unfortunately, the study did not extract ancient DNA for analysis. The effective population size of swan geese was larger than that of Chinese domestic geese, which is in agreement with the biological phenomenon that the effective population size of wild geese is larger than that of domesticated geese under long-term artificial selection [8]. The effective population size of greylag geese was smaller than that of European domestic geese, which differs from the results of Heikkinen et al. [25]. This may be because of the small number of greylag geese samples in the study, resulting in the biased estimation [58]. In our study, the domestication of European domestic geese occurred approximately 7552 years ago. In 2020, Heikkinen et al. concluded that the split between greylag geese and European domestic geese occurred at 14,000 BCE [25]. Heikkinen et al. noted that their inferred domestication time has large confidence intervals (2014.45–6503.75 generations). The domestication time inferred in our study was 3770.53–3781.13 generations. Therefore, our inferred time range for European geese domestication is consistent with that of Heikkinen et al. [25]. Additionally, the mutation rate and generation interval used were not the same as those used by Heikkinen et al., and the setting of the two parameters is important for explaining the results of the demographic analysis [43, 59]. Meanwhile, using the reference genome of *A. cygnoides* to obtain a variant dataset for European domestic geese may lead to the absence of specific variants in *A. anser*. Assembly genome of the greylag geese and expansion of the sampling range of European domestic geese may improve the accuracy of the estimated divergence time.

We found that gene flow existed between Chinese and European domestic geese and their wild ancestors. This finding supports the widespread occurrence of introgression in birds, especially among goose species [59], and provides a basis for subsequent introgression analyses. Our results showed introgression from Chinese domestic geese to the Yili geese, and 1265 candidate protein-coding genes were obtained, including gene related to neurodevelopment, cell signaling, transcription, translation, and skeletal development.

We identified *IGF-1* located in NW_013185696.1 (119,577–179,398 bp) by annotation of the candidate genes. *IGF-1* is a member of the IGF family, which plays an important role in cell differentiation, proliferation, individual growth, and development and is also the main mediator of growth hormone (GH) to initiate growth activity [60]. The gene structure and expression of *IGF-1* have a direct impact on animal growth and development, and have been reported in humans [61], pigs [62], mice [63], and birds [64].

The basic characteristics of the *IGF-1* gene in chickens are essentially the same as those in mammals, with a length of 50 kb; it is shorter than the human *IGF-1* gene, which is 70 kb in length [65]. Some studies have cloned the mRNA sequence, 5' regulatory region, and coding region of the *IGF-1* gene in geese and found that the homology of the *IGF-1* gene in geese, chicken, and duck is >95%, indicating that this gene is highly conserved in poultry [51, 66, 67]. Additionally, Mittanck et al. found that the 5' untranslated region (5'UTR) upstream of exon 1 of the *IGF-1* gene is of great significance for high-level basic gene transcription [68]. Therefore, research on the polymorphism of the 5' regulatory region plays an important role in understanding the transcription and expression levels of *IGF-1*. In the current study, we detected three SNP loci in the 5' regulatory region of the *IGF-1* gene, one of which was located 569 bp upstream of exon 1 present in greylag goose and European domestic geese, but absent in swan geese and Chinese domestic geese, and with different forms in individual Yili geese. The mutation was completely associated with the European domestic goose and also indicated the introgression of Chinese domestic geese into the Yili geese, suggesting that this was possibly a causative mutation located in the 5' regulatory region of *IGF-1*. The mutation causes the transcription factor binding site to be altered; the binding site for transcription factor Sp1 was present at the mutation site in greylag goose and European domestic geese, while it was absent in swan geese and Chinese domestic geese. Sp1 is an important member of the Sp family, which is a DNA-binding protein containing zinc-finger structures. Sp1 is a general transcription factor that recognizes and binds to GC boxes [69]. Our study indicated

that a polymorphic site c.-306 T>C in the 5' regulatory region of the *IGF-1* gene was associated with the Sp1 binding site. Sp1 plays an important regulatory role in many housekeeping genes [70, 71] and is a *trans*-activator with four domains [72], two of which are glutamine-rich and can act as strong activation domains [73]. Zhu et al. found that Sp1 is involved in the transcriptional regulation of *IGF-1* in rats and that *IGF-1* expression is decreased in cells with a mutation in the Sp1 binding site, indicating the importance of Sp1 in activating *IGF-1* expression [74]. Wang et al. found that *IGF-1* c.-366 A>C is associated with fat deposition in chickens and that this SNP affects *IGF-1* expression through differences in transcription factor binding [75]. Tang et al. found different levels of *IGF-1* expression in goose breeds of different sizes, with significantly higher levels in large geese compared with that in small geese [76]. In summary, we suggest that introgression of Chinese domestic geese into Yili geese affects *IGF-1* expression levels by increasing transcription factor Sp1 binding sites, which influence the body size of Yili geese.

The candidate genes associated with body size were *SOX4*, *SALL3*, and *SIRT1*. *SOX4* knockdown reduced brain and body size in *Xenopus* embryos [77]. *SALL3* is associated with small body size in Chinese local chicken breeds [78]. Polymorphisms in the *SIRT1* promoter region may modify fat mass and body size in cattle [79]. Therefore, we infer that the change in the body type of Yili geese is due to introgression from Chinese geese. Yili geese body size was similar to that of greylag goose, with an oval shape, short neck, thick and short legs, and adept at flying. Most Chinese domestic geese are smaller than Yili geese, with a boat-shaped body, slender neck, and longer legs than Yili geese. In recent years, there has been a trend that the body type of Yili geese has changed to Chinese domestic geese. Our study suggests that the cause of this phenomenon may be the introgression of several genes related to body size and skeletal development from Chinese domestic geese into Yili geese.

We found that the introgression events were of different proportions in Yili individuals by constructing the ML tree for the *IGF-1* gene sequences located on scaffold NW_013185696.1 and comparing this with the genome-wide tree. This result is consistent with the results of the ADMIXTURE and PCA analyses. This also indicated that the introgression event was incomplete in the Yili population; therefore, and further studies can be performed to elucidate gene introgression within Yili geese.

Conclusions

In this study, we describe the origins, timing and introgression of domestic geese. Our study was based on deep whole-genome sequencing of 74 individuals from

multiple wild and domestic populations. Using this dataset and a suite of cutting-edge population genomic and functional genetic analyses, we showed that the domestic geese have two origins: the ancestor of the Chinese domestic geese is the swan geese, and the ancestor of the Yili geese and European domestic geese is the greylag geese. We also identified the timing of domestication of domestic geese; the domestication of the Chinese geese occurred approximately 3499 years ago, and that of the European geese occurred approximately 7552 years ago. Furthermore, we found that gene flow occurs between domestic geese and their wild ancestors, and between domestic goose species of different origins. Introgression analysis showed that Yili geese had been introgressed by the Chinese domestic geese, and the body size of Yili geese could be influenced by introgression events of some growth-related genes, including *IGF-1*. Therefore, our results have important implications for broader questions on the evolution of complex phenotypes and inter-specific introgression.

Abbreviations

AIC	Akaike Information Criteria
CDS	Coding sequence
CI	Confidence interval
EPAS1	Endothelial PAS domain protein 1
GJA5	Gap junction protein alpha 5
GO	Gene ontology
IGF-1	Insulin-like growth factor-1
LD	Linkage disequilibrium
ML	Maximum likelihood
N_e	Effective population size
PCA	Principal component analysis
SALL3	Spalt like transcription factor 3
SIRT1	Sirtuin 1
SOX4	SRY-Box transcription factor 4
Sp1	Specificity protein 1
5'UTR	5' untranslated region

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40104-022-00826-9>.

Additional file 1: Table S1. Sample sources of domestic geese; **Table S2.** Summary of genome sequencing and mapping statistics; **Table S3.** Summary of SNPs and indels; **Table S4.** Results of *D* statistics; **Table S5.** Gene names in top 5% introgression regions; **Table S6.** Total GO terms of genes located in top 5% introgression regions; **Table S7.** The parameters estimated in the demographic model; **Table S8.** Prediction of transcription factor binding sites in the *IGF1* 5' regulatory region.

Additional file 2: Fig. S1. Demographic history of domestic geese and their wild ancestors; **Fig. S2.** Site frequency spectrum of swan and Chinese domestic geese populations; **Fig. S3.** Site frequency spectrum of populations of greylag and European domestic geese; **Fig. S4.** Regions of *SIRT1* with introgression between Chinese domestic and Yili geese individuals; **Fig. S5.** Regions of *SALL3* with introgression between Chinese domestic and Yili geese individuals; **Fig. S6.** Regions of *GJA5* with introgression between Chinese domestic and Yili geese individuals.

Acknowledgements

We gratefully acknowledge our colleagues in the Poultry Team at the National Engineering Laboratory for Animal Breeding of China Agricultural University for their assistance with sample collection and helpful comments on the manuscript.

Authors' contributions

Conceptualization, LQ; methodology, TZ, JZ, and XL; formal analysis, JW; investigation, JW; resources, HL, HW, JY, and ZN; writing-original draft preparation, JW; writing-review and editing, LQ and ZJ. All authors have read and approved the final manuscript.

Funding

This research was funded by the Open Project of Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, the National Nature Science Foundation of China (grant numbers: BRZD2104 and 31960648), and the Joint Plan of Liaoning Province in the field of Livelihood Science and Technology (Rural Revitalization Science and Technology Support).

Availability of data and materials

Whole-genome resequencing analysis of 18 Chinese domestic geese available from NCBI (BioProject ID: PRJNA662026). The data for 35 other Chinese domestic geese, Landaise geese, Yili geese, and Swan geese are available from NCBI (BioProject IDs: PRJNA767757 and PRJNA817006). The greylag geese data were obtained from a previous study [28]. Roman and Hortobagy geese data are available from NCBI (BioProject ID: PRJNA722049).

Declarations

Ethics approval and consent to participate

This study conformed to protocols approved by the Laboratory Animal Welfare and Animal Experimentation Ethics Review Committee of China Agricultural University (No. XK622).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 5 August 2022 Accepted: 14 December 2022

Published online: 13 February 2023

References

- MacHugh DE, Larson G, Orlando L. Taming the past: ancient DNA and the study of animal domestication. *Annu Rev Anim Biosci*. 2017;5:329–51. <https://doi.org/10.1146/annurev-animal-022516-022747>.
- Darwin CR. *The variation of animals and plants under domestication*. 1st ed. London: John Murray; 1868.
- Wilkins AS, Wrangham RW, Fitch WT. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197(3):795–808. <https://doi.org/10.1534/genetics.114.165423>.
- Larson G, Karlsson EK, Perri A, Webster MT, Ho SY, Peters J, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A*. 2012;109(23):8878–83. <https://doi.org/10.1073/pnas.1203005109>.
- Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science*. 2000;287(5461):2254–7. <https://doi.org/10.1126/science.287.5461.2254>.
- Zhang H, Pajmans JL, Chang F, Wu X, Chen G, Lei C, et al. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nat Commun*. 2013;4:2755. <https://doi.org/10.1038/ncomms3755>.

7. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010;464(7288):587–91. <https://doi.org/10.1038/nature08832>.
8. Zhang Z, Jia Y, Almeida P, Mank JE, van Tuinen M, Wang Q, et al. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *Gigascience*. 2018;7(4). <https://doi.org/10.1093/gigascience/gjy027>.
9. Crawford RD. Poultry genetics and breeding. New York: Elsevier; 1990.
10. Romanov MN. Evolution of domestic geese. Centres of domestication and dispersion. In: Gudina, Alexander N, editors. New investigations on Palearctic geese. Zaporizhya Branch of the Ukrainian Ornithological Society, Zaporizhya State University. 1995. p. 7.
11. Zeuner FE. A history of domesticated animals. London: Hutchinson & Co. Ltd.; 1963.
12. Mannermaa K. Goose: domestication. In: Smith C, editor. Encyclopedia of global archaeology. New York: Springer New York; 2014. p. 3096–8.
13. Bo WC. On the origin of Chinese goose. *Agric Archaeol*. 1996;16(3):268–72.
14. Eda M, Itahashi Y, Kikuchi H, Sun G, Hsu KH, Gakuhari T, et al. Multiple lines of evidence of early goose domestication in a 7,000-year-old rice cultivation village in the lower Yangtze River, China. *Proc Natl Acad Sci U S A*. 2022;119(12):e2117064119. <https://doi.org/10.1073/pnas.2117064119>.
15. Ministry of Agriculture and Rural Affairs of China. National list of breeds of livestock genetic resources. http://www.moa.gov.cn/gk/tzgg_1/tz/202005/t20200529_6345586.htm. Accessed 29 May 2020.
16. Delacour JM, Mayr E. The family Anatidae. *Wilson Bull*. 1945;57:3–54.
17. China national commission of animal genetic resources. Animal genetic resources of China - poultry. Beijing: China Agricultural Press; 2011.
18. Buckland R, Guy G. Goose production. Rome: Food and Agriculture Organization of the United Nations; 2002.
19. Mead D. Domesticated geese and ducks—and allied species. Indonesia: Sulawesi Language Alliance; 2013.
20. Shi XW, Wang JW, Zeng FT, Qiu XP. Mitochondrial DNA cleavage patterns distinguish independent origin of Chinese domestic geese and Western domestic geese. *Biochem Genet*. 2006;44(5–6):237–45. <https://doi.org/10.1007/s10528-006-9028-z>.
21. Bhatnagar MK. Mitotic chromosomes of white Chinese geese. *J Hered*. 1968;59(3):191–5. <https://doi.org/10.1093/oxfordjournals.jhered.a107685>.
22. Shi XW, Zeng FT, Qiu XP, Zhang YP. Mitochondrial DNA polymorphism and origin differentiation in major Chinese goose breeds. *J Genet*. 1998;25(6):499–507.
23. Li HF, Zhu WQ, Chen KWHY, Xu WJ, Song W. Two maternal origins of Chinese domestic goose. *Poult Sci*. 2011;90(12):2705–10. <https://doi.org/10.3382/ps.2011-01425>.
24. Gao G, Zhao X, Li Q, He C, Zhao W, Liu S, et al. Genome and metagenome analyses reveal adaptive evolution of the host and interaction with the gut microbiota in the goose. *Sci Rep*. 2016;6:32961. <https://doi.org/10.1038/srep32961>.
25. Heikkinen ME, Ruokonen M, White TA, Alexander MM, Gunduz I, Dobney KM, et al. Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history. *G3*. 2020;10(9):3061–70. <https://doi.org/10.1534/g3.120.400886>.
26. Cao YH, Xu SS, Shen M, Chen ZH, Gao L, Lv FH, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol*. 2021;38(3):838–55. <https://doi.org/10.1093/molbev/msaa236>.
27. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun*. 2018;9(1):2337. <https://doi.org/10.1038/s41467-018-04737-0>.
28. Ottenburghs J, Megens HJ, Kraus RHS, Madsen O, van Hooft P, van Wieren SE, et al. A tree of geese: a phylogenomic perspective on the evolutionary history of true geese. *Mol Phylogenet Evol*. 2016;101:303–13. <https://doi.org/10.1016/j.ympev.2016.05.021>.
29. Herrmann BG, Frischauf A-M. Isolation of genomic DNA. *Methods Enzymol*. 1987;152:180–3. [https://doi.org/10.1016/0076-6879\(87\)52018-3](https://doi.org/10.1016/0076-6879(87)52018-3).
30. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619. <https://doi.org/10.1371/journal.pone.0030619>.
31. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
32. Lu L. Anser cygnoides domesticus genome version 1.0. https://www.ncbi.nlm.nih.gov/assembly/GCA_000971095.1/. Accessed 1 May 2021.
33. Broad Institute. Picard Toolkit. <https://broadinstitute.github.io/picard/>. Accessed 10 May 2021.
34. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
35. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
36. 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(3):559–75. <https://doi.org/10.1086/519795>.
37. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64. <https://doi.org/10.1101/gr.094052.109>.
38. Francis RM. Pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017;17(1):27–32. <https://doi.org/10.1111/1755-0998.12509>.
39. Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*. 2014;15:162. <https://doi.org/10.1186/1471-2164-15-162>.
40. Interactive Tree of Life. <https://itol.embl.de/>. Accessed 25 Oct. 2021.
41. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35(10):1786–8. <https://doi.org/10.1093/bioinformatics/bty875>.
42. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*. 2017;49(2):303–9. <https://doi.org/10.1038/ng.3748>.
43. Ottenburghs J, Honka J, Muskens G, Ellegren H. Recent introgression between taiga bean goose and tundra bean goose results in a largely homogeneous landscape of genetic differentiation. *Heredity*. 2020;125(1–2):73–84. <https://doi.org/10.1038/s41437-020-0322-z>.
44. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009;5(10):e1000695. <https://doi.org/10.1371/journal.pgen.1000695>.
45. Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. Computationally efficient composite likelihood statistics for demographic inference. *Mol Biol Evol*. 2016;33(2):591–3. <https://doi.org/10.1093/molbev/msv255>.
46. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8(11):e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
47. Malinsky M, Matschiner M, Svardal H. Dsuite - fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*. 2021;21(2):584–95. <https://doi.org/10.1111/1755-0998.13265>.
48. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57. <https://doi.org/10.1038/nprot.2008.211>.
49. R Core Team. R: a language and environment for statistical computing. <https://www.R-project.org/>. Accessed 10 Jan. 2021.
50. Kumar S, Stecher G, Li M, Nkayaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547–9. <https://doi.org/10.1093/molbev/msy096>.
51. Wang L, Wang B, Jia X, Yang Z, Liu Q, Zhang M, et al. Cloning and analysis of the 5' regulatory region sequence of goose IGF-I gene. *China Anim Husbandry Vet Med*. 2007;(3):3.
52. Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Accessed 15 Dec. 2021.
53. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48(D1):D87–92. <https://doi.org/10.1093/nar/gkz1001>.
54. Heikkinen ME, Ruokonen M, Alexander M, Aspi J, Pyhäjärvi T, Searle JB. Relationship between wild Greylag and European domestic geese

- based on mitochondrial DNA. *Anim Genet.* 2015;46(5):485–97. <https://doi.org/10.1111/age.12319>.
55. Nie C, Almeida P, Jia Y, Bao H, Ning Z, Qu L. Genome-wide single-nucleotide polymorphism data unveil admixture of Chinese indigenous chicken breeds with commercial breeds. *Genome Biol Evol.* 2019;11(7):1847–56. <https://doi.org/10.1093/gbe/evz128>.
 56. Larson G, Fuller DQ. The evolution of animal domestication. *Annu Rev Ecol Syst.* 2014;45(1):115–36. <https://doi.org/10.1146/annurev-ecolsys-110512-135813>.
 57. Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, et al. 863 genomes reveal the origin and domestication of chicken. *Cell Res.* 2020;30(8):693–701. <https://doi.org/10.1038/s41422-020-0349-y>.
 58. Wang J, Santiago E, Caballero A. Prediction and estimation of effective population size. *Heredity.* 2016;117(4):193–206. <https://doi.org/10.1038/hdy.2016.43>.
 59. Ottenburghs J, Megens HJ, Kraus RHS, van Hooft P, van Wieren SE, Crooijmans R, et al. A history of hybrids? Genomic patterns of introgression in the true geese. *BMC Evol Biol.* 2017;17(1):14. <https://doi.org/10.1186/s12862-017-1048-2>.
 60. Laron Z. Insulin-like growth factor 1 (IGF-1): a growth hormone. *Mol Pathol.* 2001;54(5):311–6. <https://doi.org/10.1136/mp.54.5.311>.
 61. Tavakkol A, Elder JT, Griffiths CE, Cooper KD, Talwar H, Fisher GJ, et al. Expression of growth hormone receptor, insulin-like growth factor 1 (IGF-1) and IGF-1 receptor mRNA and proteins in human skin. *J Invest Dermatol.* 1992;99(3):343–9. <https://doi.org/10.1111/1523-1747.ep12616668>.
 62. Etherton TD, Wiggins JP, Chung CS, Evock CM, Rebhun JF, Walton PE. Stimulation of pig growth performance by porcine growth hormone and growth hormone-releasing factor. *J Anim Sci.* 1986;63(5):1389–99. <https://doi.org/10.2527/jas1986.6351389x>.
 63. Rui L, Aguirre V, Kim JK, Shulman GI, Lee A, Corbould A, et al. Insulin/IGF-1 and TNF- α stimulate phosphorylation of IRS-1 at inhibitory Ser307 via distinct pathways. *J Clin Invest.* 2001;107(2):181–9. <https://doi.org/10.1172/JCI10934>.
 64. Duclos MJ. Insulin-like growth factor-I (IGF-1) mRNA levels and chicken muscle growth. *J Physiol Pharmacol.* 2005;56(Suppl 3):25–35.
 65. Kajimoto Y, Rotwein P. Structure of the chicken insulin-like growth factor I gene reveals conserved promoter elements. *J Biol Chem.* 1991;266(15):9724–31.
 66. Wang L, Wang B, Jia X, Yang Z, Liu G. Cloning, analysis and prokaryotic expression of IGF-I gene cDNA in goose. *Chin J Anim Sci.* 2008;35(1):4–7.
 67. Xiao C, Li X, Tong C, Yan G, Liu G. Cloning of goose insulin-like growth factor-I gene and its mRNA expression changes in tissues. *Chin Vet Sci.* 2006;36(3):216–9.
 68. Mittanck DW, Kim SW, Rotwein P. Essential promoter elements are located within the 5' untranslated region of human insulin-like growth factor-I exon I. *Mol Cell Endocrinol.* 1997;126(2):153–63. [https://doi.org/10.1016/S0303-7207\(96\)03979-2](https://doi.org/10.1016/S0303-7207(96)03979-2).
 69. Suske G. The Sp-family of transcription factors. *Gene.* 1999;238(2):291–300. [https://doi.org/10.1016/S0378-1119\(99\)00357-1](https://doi.org/10.1016/S0378-1119(99)00357-1).
 70. Zaid A, Li R, Luciaková K, Barath P, Nery S, Nelson BD. On the role of the general transcription factor Sp1 in the activation and repression of diverse mammalian oxidative phosphorylation genes. *J Bioenerg Biomembr.* 1999;31(2):129–35. <https://doi.org/10.1023/a:1005499727732>.
 71. Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. *Eur J Cancer.* 2005;41(16):2438–48. <https://doi.org/10.1016/j.ejca.2005.08.006>.
 72. Kadonaga JT, Courey AJ, Ladika J, Tjian R. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science.* 1988;242(4885):1566–70. <https://doi.org/10.1126/science.3059495>.
 73. Courey AJ, Tjian R. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell.* 1988;55(5):887–98. [https://doi.org/10.1016/0092-8674\(88\)90144-4](https://doi.org/10.1016/0092-8674(88)90144-4).
 74. Zhu JL, Kaytor EN, Pao CI, Meng XP, Phillips LS. Involvement of Sp1 in the transcriptional regulation of the rat insulin-like growth factor-1 gene. *Mol Cell Endocrinol.* 2000;164(1–2):205–18. [https://doi.org/10.1016/S0303-7207\(00\)00207-0](https://doi.org/10.1016/S0303-7207(00)00207-0).
 75. Wang WJ, Guo YQ, Xie KJ, Li YD, Li ZW, Wang N, et al. A functional variant in the promoter region of IGF1 gene is associated with chicken abdominal fat deposition. *Domest Anim Endocrinol.* 2021;75:106584. <https://doi.org/10.1016/j.domaniend.2020.106584>.
 76. Tang J, Guo M, Fu J, Ouyang H, Tian Y, Shen X, et al. Polymorphism analysis and expression patterns of the IGF1 gene in the Shitou goose. *Arch Anim Breed.* 2021;64(2):315–23. <https://doi.org/10.5194/aab-64-315-2021>.
 77. Zawerton A, Yao B, Yeager JP, Pippucci T, Haseeb A, Smith JD, et al. De novo SOX4 variants cause a neurodevelopmental disease associated with mild dysmorphism. *Am J Hum Genet.* 2019;104(4):777. <https://doi.org/10.1016/j.ajhg.2019.01.014>.
 78. Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. *J Mol Cell Biol.* 2016;8(6):542–52. <https://doi.org/10.1093/jmcb/mjw044>.
 79. Li M, Sun X, Zhang L, Wang J, Huang Y, Sun Y, et al. A novel c.-274C>G polymorphism in bovine SIRT 1 gene contributes to diminished promoter activity and is associated with increased body size. *Anim Genet.* 2013;44(5):584–7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

