HAYATI
Journal of Biosciences

Vol. 31 No. 6, November 2024 1130-1143
DOI:10.4308/hjb.31.6.1130-1143
ISSN: 1978-3019
EISSN: 2086-4094

Characterization of CDS Region of Exons 1 and 2 of *SOX9* Gene as Potential Gene in Construction of Syrinx Structure in Junglefowl (*Gallus* sp.)

Achmad Alfiyan[1], Achmad Farajallah[1*], Maria Ulfah[2], Dyah Perwitasari-Farajallah[1,3], Muladno Muladno[2]

[1]*Department of Biology, Faculty of Mathematics and Natural Sciences, IPB University, Kampus IPB Dramaga, Bogor 16680, Indonesia*
[2]*Department of Animal Production and Technology, Faculty of Animal Science, IPB University, Kampus IPB Dramaga, Bogor 16680, Indonesia*
[3]*Primate Research Center, IPB University, Bogor 16151, Indonesia*

ARTICLE INFO

Article history:
Received March 2, 2024
Received in revised form June 6, 2024
Accepted June 12, 2024

KEYWORDS:
Gallus gallus,
Gallus lafayetii,
Gallus sonneratii,
Gallus varius
Sequence Read Archive

ABSTRACT

The crowing of male *Gallus* exhibits diverse sound patterns. This is believed to be related to the phenotypic diversity of vocal organs, one of which is influenced by the nucleotide diversity of the associated genes. The *SOX9* gene, involved in cartilaginous tissue growth and development, is reported to contribute e in the development of larynx and syrinx. This study aimed to characterize the CDS regions of exons 1 and 2 of the *SOX9* gene in junglefowl to assess its diversity. Genomic DNA was extracted from ten individuals of *G. varius* from Lombok and Sumbawa. The CDS regions of *SOX9* gene exons 1 and 2 were amplified using two primer pairs. Additionally, the CDS regions of *SOX9* gene exons 1 and 2 from 54 junglefowl SRA data in an online repository were mapped and analyzed. The study identified all nucleotide sequences as CDS regions of *SOX9* gene exons 1 and 2. Six shared, and 24 unique haplotypes were constructed. A putative amino acid sequence common to all *Gallus* species was also identified. The diversity observed in the CDS regions of *SOX9* gene exons 1 and 2 nucleotide sequence showed a different level with the diversity observed in its amino acid sequence.

1. Introduction

The crowing sounds of chickens, a form of vocalization typically produced by male *Gallus* birds, exhibit diverse patterns, as observed in domestic chickens and junglefowl. It is hypothesized that phenotypic diversity in the organ responsible for sound production (syrinx) is associated with the observed variation in rooster crowing patterns. The syrinx, comprises various tissues, including cartilage. The cartilage in the syrinx serves as its structural framework and the configuration of this framework exhibits considerable diversity across bird taxa (Morejohn 1966; Gaban-Lima and Hofling 2006). It is postulated that the varied architecture of this syrinx framework plays a pivotal role in the sound production process (Morejohn 1966; Gaban-Lima and Hofling 2006).

* Corresponding Author
E-mail Address: achamad@apps.ipb.ac.id

As structural frameworks, the cartilage in syrinx also acts as the attachment site of muscle and other tissues. Consequently, the cartilaginous structure provides support and facilitates the activities of other tissues during the vocalization process. The variations in syrinx structure, reported by Morejohn in 1966, between junglefowl and domestic chickens, specifically in the shape of the first bronchial half-ring attached to the pessulus and the number of syringeal bars, suggest the potential for different degrees of support provided by the framework during vocalization. In the study conducted by During *et al.* (2013), which aimed to construct a three-dimensional morphological map of the bird's syrinx, it is suggested that the MVC cartilage (medial ventral cartilage) within the syrinx plays a crucial role in regulating the frequency of vocalization sounds.

Tissue phenotype, including configuration and other inherent characteristics, is influenced by many factors, with genetics being critical. Genetic

variability, shown as differences in nucleotide sequences within a gene, results in diverse properties in the encoded protein or RNA, thereby contributing to a range of tissue phenotypes. Among the group of genes associated with the processes of cartilage formation and maintenance, the *SOX9* gene is of particular importance. The Sox9 protein, encoded by the *SOX9* gene, serves as a transcription factor that regulates gene expression linked to the processes of cartilage formation and maintenance (Wright *et al.* 1995; Ng *et al.* 1997; Zhao *et al.* 1997; Oh *et al.* 2014; Liu and Lefebvre 2015). Gokhman *et al.* (2020) assert the involvement of the *SOX9* gene in determining the anatomy of the human vocal tract, as evidenced by their investigation employing methylation mapping of bone-related genes across modern humans, archaic humans (Neanderthals and Denisovans), and chimpanzees. The findings of Longtine *et al.* (2024) showed that *SOX9* gene involved in the development of avian syrinx.

In the *Gallus* species, RNA expression analysis employing domesticated chickens as the experimental model indicates, the *SOX9* gene is involved not only in cartilage development but also plays roles in limb development, feather growth and development, sex determination, gonadal sex differentiation, and neural crest development (Kent *et al.* 1996; Healy *et al.* 1999; Cheung and Briscoe 2003; Scheider *et al.* 2014; Ayers *et al.* 2015; Montero *et al.* 2017; Su *et al.* 2019). The information concerning the *SOX9* gene in *Gallus* is currently constrained to domestic chickens and red junglefowl, discernible through *SOX9* gene nucleotide sequence data in online repositories. Hence, the primary objective of this investigation is to provide a characterization of the CDS region of exons 1 and 2 of the *SOX9* gene in junglefowl to assess the diversity of this region.

2. Materials and Methods

2.1. Sample Collection

Gallus varius specimens were obtained from two breeders, a hunters and a pet seller at the local market. Before the sampling process, brief interviews were conducted to validate the geographical origin of each specimen. Feather samples were procured from 10 *G. varius* individuals, with 8 originating from Lombok and 2 from Sumbawa (Table 1). Adhering to the DNA Diagnostic Center (DDC) protocol (c2019), feathers were collected from the chest region, and following the collection process, feathers from each individual were organized into clean paper containers, each labeled accordingly. The samples were then stored in the laboratory freezer at -20°C.

2.2. Total Genomic DNA Extraction

The genomic DNA was extracted from the feather base, precisely the tip of the calamus section, where DNA-containing subcutaneous tissue is located (DDC c2019). Utilizing a sterile surgical scissors, the calamus tip was excised and placed into a 1.5 ml tube. Then, GT Buffer solution, one of the buffer components of the DNA extraction kit, was added. The calamus tip was fragmented into smaller portions using the same sterilized surgical scissors. The subsequent steps were performed according to the protocol provided in the GENEAID DNA Extraction Kit "Tissue Genomic DNA Mini Kit" (Geneaid, Canada).

2.3. Amplification and Sequencing of the Exons 1 and 2 Regions of the *SOX9* Gene in *G. vairus*

Amplification of exons 1 and 2 regions of the *SOX9* gene in *G. varius* samples was performed using the Touchdown Polymerase Chain Reaction (TD-PCR) technique on the Biometra Thermo Cycle instrument. The amplification employed two manually designed primer pairs based on the *G. gallus* sequence (NC_052549). The primer pair AF635 (forward) 5'-TTTTCTCTCCGTTTTCTCCTC-3' and AF636 (reverse) 5'-ACAGAGCTGATGCAATCTAGG-3' was utilized for exon 1 amplification, while the primer pair AF637 (forward) 5'-CTCTCGTTTGGTCATTGAAAC-3' and AF638 (reverse) 5'-AAGAGAGAGTGTGAGCGTGAT-3' was used for exon 2 amplification (Figure 1). The GoTaq® Green Mastermix reagent was used for the amplification process. The PCR conditions comprised an initial pre-denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 1 minute, annealing at 62-57.5°C for the first 10 cycles (with a temperature decrease of 0.5°C per cycle) and 57°C for the subsequent 25 cycles for 1 minute, and extension at 72°C for 1 minute. This was followed by a post-elongation step at 72°C for 2 minutes and a finalization step at 15°C for 10 minutes. Subsequently, the amplicons were migrated in a 1% agarose gel supplemented with Florosafe staining dye (1st Base, MY) at 80 V for 55 minutes. The DNA bands were visualized under UV light using Geldoc. Amplicons with high-quality DNA bands were sent to 1st Base, a sequencing service company, for sequencing using the Sanger method (Sanger *et al.* 1977).

Table 1. CDS region of exons 1 and 2 of *SOX9* gene of *Gallus* species used in data analysis

Species	Origin of location	N	Sample name	Accession number	Source	Ref
*G. varius***	Lombok, Indonesia	8	Gv_L3,4, 6 Gv_L7, 10-13	- -	Pet seller Breeder	*
	Sumbawa, Indonesia	2	Gv_Sw1 Gv_Sw2	- -	Hunter Breeder	
	Madura, Indonesia	7	-	DRX083687-93	GenBank	A
	East Java, Indonesia	1	-	DRX083685		
	Central Java, Indonesia	1	-	ERX4842472		B
	Bali, Indonesia	1	-	ERX4842466		
	Indonesia	3	-	SRX5173445-47		C
	Zoological Park, Japan	1	-	SRX9334260		
	Zoological Park, Taiwan	1	-	SRX9334259		D
	-	1	-	SRX7909216		-
*G. lafayetii***	Sri Lanka	2	-	ERX4842448-49	GenBank	b
		2	-	SRX5173433; SRX5173439		c
	Zoological Park, France	2	-	SRX9334251-52		d
*G. sonneratii***	Zoological Park, France	2	-	SRX9334254; SRX9334256	GenBank	d
	Andhra Pradesh, India	2	-	SRX9334257-58		
	-	1	-	SRX7909214		-
	Indian subcontinent	1	-	SRX5173441		c
*G. gallus****	Fayetteville, Arkansas (USA)	2	-	NC_052549; NC_052590	GenBank	-
*G. gallus gallus***	Forest, Chiang-Mai, Thailand	3	-	SRX9334255, SRX9334262-63	GenBank	d
	Aceh, Indonesia	2	-	ERX4842911; ERX4842913		b
*G. gallus bankiva***	Zoological Park, France	2	-	SRX9334243-44	GenBank	d
	East Java, Indonesia	1	-	ERX4842909		b
	Central Java, Indonesia	1	-	ERX4842907		
*G. gallus spadiceus***	Ruili, Yunnan, China	2	-	ERX4843040-41	GenBank	d
	Mangshi, Yunnan, China	2	-	ERX4843025; ERX4843032		
	Forest, Chiang-Mai, Thailand	2	-	SRX9334245; SRX9334249		
*G. gallus jabouillei***	Baise, Guangxi, China	3	-	ERX4842930; ERX4842935-36	GenBank	b
	Luchuan, Guangxi, China	2	-	ERX4842928-29		
*G. gallus murghi***	Jammu & Kashmir, India	1	-	SRX9334265	GenBank	d
	Dehadrun, India	1	-	SRX9334266		
	Uttar Pradesh, India	1	-	SRX9334267		
	Haryana, India	2	-	ERX4842982; ERX4842987		b
	Bihar, India	1	-	ERX4842997		

N: Number of DNA sequence; Ref: Reference; "-": data unavailable. "*": this study; '**': junglefowl; '***': domesticated chicken; a: Ulfah *et al.* 2016; b: Wang *et al.* 2020; c: Lawal *et al.* 2020; d: Mariadassou *et al.* 2021.

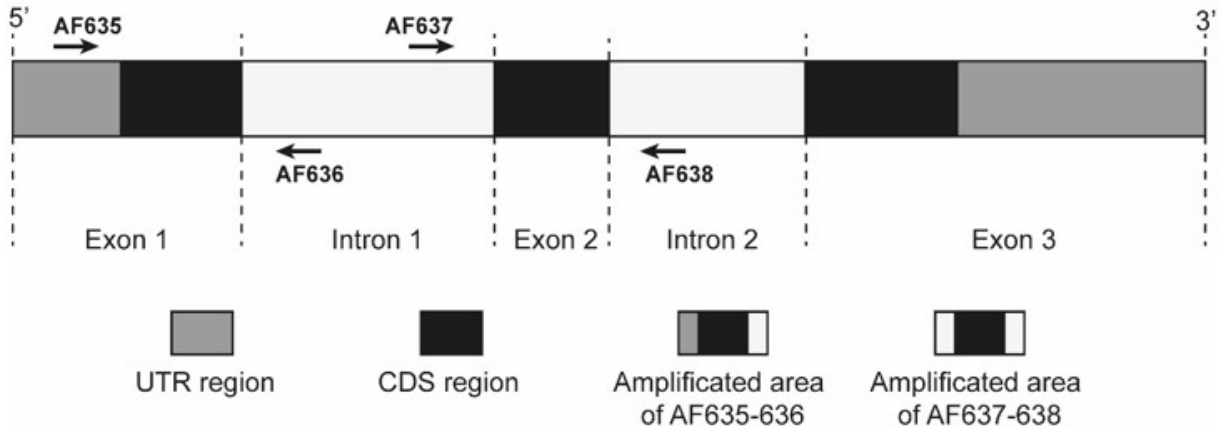

Figure 1. Schematic structure of *SOX9* gene in domestic chicken and amplification target region of primers, constructed based on nucleotide sequence of *G. gallus* (NC_052549)

2.4. Assembling of *SOX9* Gene Exons 1 and 2 of *Gallus* from SRA Database

The nucleotide sequences of protein- coding regions (CDS) for the *SOX9* gene's exons 1 and 2 in four *Gallus* species, stored in SRA data format in the GenBank (Table 1), were mapped using BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with maximum number of aligned sequences to display is 100 (Altschul *et al.* 1990; Johnson *et al.* 2008). The reference sequence selected for this mapping encompassed the CDS region of exon 1 of the *SOX9* gene, along with its flanking region (116 bp before and 153 bp after), and exon 2, along with its flanking region (145 bp before and 162 bp after), from *G. gallus* (NC_052549). Following the mapping process, the assembly of the mapped CDS region of exons 1 and 2 sequences of *Gallus* was carried out using MEGA7 (Kumar *et al.* 2016). The assembly process only used a full length sequence of nucleotide sequence fragment to construct CDS region of exons 1 and 2 of *SOX9* gene. After the assembling process, the CDS region of exons 1 and 2 was combined using MEGA7.

2.5. Bioinformatic Data Analysis

The sequencing products from two primer pairs were edited using MEGA7. Subsequently, each sequence was compared to reference sequences stored in GenBank using BLASTN. After the comparative analysis between the sample data and the database, 10 sequences of *SOX9* gene CDS region exons 1 and 2 from each *G. varius* sample were concatenated. These sequences were then aligned with *SOX9* gene sequences assembled from SRA data format and with reference sequences of domestic chicken (NC_052549 and NC_052590) (Table 1) using the ClustalW tool integrated within MEGA7. After alignment, a series of analyses, encompassing assessments of nucleotide composition, nucleotide variation, haplotype construction (nucleotide and amino acid sequence), and putative amino acid composition and variation, were undertaken. Nucleotide composition assessment, putative amino acid construction, and amino acid variation analyses were conducted within MEGA7. In addition to MEGA7, putative amino acid construction was executed using web-based software, BLASTX (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.* 1990; Johnson *et al.* 2008) and EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/) (Rice *et al.* 2000). For the analysis of nucleotide variation and haplotype construction (nucleotide sequence), DnaSP6 software (Rozas *et al.* 2017) was utilized. As for haplotype construction based on amino acid sequence were conducted manually.

3. Results

3.1. Comparison and Nucleotide Composition of Exons 1 and 2 *SOX9* Gene

The *G. varius* specimens from Lombok and Sumbawa Islands yielded a set of 20 *SOX9* gene sequences, comprising 10 sequences for each CDS region of exons 1 and 2, complemented by partial flanking regions like UTR regions and introns. The length of these sequences ranged from 653 to 796 bp for exons 1 and its flanking region and from 461 to 864 bp for exon 2 and its flanking region. Sequence database comparison using BLASTN revealed a significant degree of similarity

with the *SOX9* gene sequences of *G. gallus,* including the full-length gene, mRNA region, and CDS region (Supplementary Table 1). The confirmation of these findings is supported by the significant values observed for the E-value, Identity and Query Cover parameters, elucidated in Table 2.

The nucleotide composition analysis of 64 CDS regions spanning exons 1 and 2 of the *SOX9* gene obtained from four *Gallus* species, 54 assembled SRA data and 10 sample data, reveals a consistent distribution pattern closely resembling that of the CDS regions in exons 1 and 2 of the *SOX9* gene of *G. gallus* (NC_052549 and NC_052590) (Supplementary Table 2). The outcomes derived from BLASTN analysis, coupled with examining nucleotide distribution patterns, collectively authenticate the identity of the CDS regions within exons 1 and 2 of the *SOX9* gene in the sampled sequence and the assembled sequence from 54 SRA data formats of *Gallus*.

3.2. Heterozygosity, Nucleotide Variation and Haplotype of Exons 1 and 2 *SOX9* Gene

The CDS regions of exons 1 and 2 within the *SOX9* gene sequence, spanning 658 bp across 66 *Gallus* species, exhibited a total of 34 nucleotide variations (Supplementary Table 3) and a nucleotide diversity value of 0.00214. Of these variations, 20 were identified within the CDS of exon 1 (nucleotide positions 16-416), while 14 were founded in the CDS of exon 2 (nucleotide positions 433-680) (Supplementary Table 3). Heterozygosity, identified by IUPAC ambiguity nucleotide symbols within nucleotide sequences, was observed in 20 *Gallus* sequences (Table 2 and Supplementary Table 3). Heterozygotes were recognized when the frequency of each of the two-nucleotide type at a position fell within 40 to 60% of the depth of coverage. Heterozygous positions were detected in every junglefowl species, including sequences obtained through both the Sanger method (our sample, *G. varius*) and the NGS method (*Gallus* SRA data). These heterozygous positions contribute significantly to nucleotide variation (>80%). The total heterozygous sites (each species) in these 20 individuals ranged from 1 to 13 bp, with the lowest occurrence observed in *G. sonneratii, G. gallus gallus, G. gallus bankiva*, and *G. gallus jabouillei*, and the highest in *G. gallus murghi* (Table 2). A haplotype diversity of 0.711 was observed in the 66 nucleotide sequences of *Gallus* species, where 30 distinct haplotype groups were formed from the 34 identified nucleotide variations (Supplementary Table 4). Among these haplotypes, six were shared either among different species (Hap_1) or subspecies (Hap_16) or within the same species (Hap_13, Hap_14 and Hap_15) or subspecies (Hap_22), while the remaining 24 were unique (Supplementary Material, Table S4). Based on haplotype number and diversity, the genetic diversity of the CDS region of *SOX9* exons 1 and 2 in *Gallus* was found to be moderately high.

3.3. Amino Acid Variation of Exons 1 and 2 *SOX9* Gene

The CDS of exons 1 and 2 within the *SOX9* gene sequence from 64 *Gallus* species (junglefowl), spanning 658 bp, encoded 228 putative amino acids. The amino acid sequences derived from MEGA7, Transeq, and BLASTX for the same nucleotide sequence were identical. BLASTX comparison analysis across all junglefowl indicated a high similarity with the amino acid sequence of the *SOX9* gene in *G. gallus*, as evidenced by substantial values for E-value, Query Cover, and Identity parameters (Table 3). The previously mentioned 34 bp nucleotide variations impacted 31 amino acid sequences; however, not all of these 31 amino acids exhibited variations. Amino acid variations were observed in only 19 positions

Table 2. Summarize of heterozygous position in CDS region of exons 1 and 2 of *SOX9* gene in *Gallus*

Species	N	n	Nucleotide position which heterozygous	Number of heterozygous positions in each sequence
G. varius	5	5	170; 433; 487; 513; 654	1-3
G. lafayetii	2	2	384; 447	1
G. sonneratii	1	1	633	1
G. gallus gallus	1	1	102	1
G. gallus bankiva	1	1	411	1
G. gallus spadiceus	4	8	22; 102; 146; 250; 416; 447; 627; 680	1-3
G. gallus jabouillei	1	1	102	1
G. gallus murghi	5	13	16; 102; 181; 183; 204; 303; 374; 395; 401; 434; 435; 452; 602	1-6

N: number of sequences with heterozygosity, n: number of heterozygous

(non-synonymous), while the remaining 12 were synonymous (Supplementary Table 5). From those, 19 non-synonymous variations formed 12 amino acid haplotypes (Supplementary Table 6), with haplotype diversity value at 0.241. Notably, there are certain putative amino acid sequences that was shared among all *Gallus* species (Hap_aa_1) (Supplementary Table 6). Amino acid variations were observed solely in *G. varius* and *G. gallus* species with heterozygous sites. Due to that, some of *G. varius* (sample and SRA data), *G. gallus spadiceus*, and *G. gallus murghi* sequences which showed heterozygosity formed unique haplotypes (Hap_aa_2-11) (Table 4). Based on this finding, it was shown that the amino acid encoded by exons 1 and 2 of the *SOX9* gene was relatively conserved.

4. Discussion

Significant value in E-value, Identity, and Query Cover parameters obtained from BLASTN or BLASTX comparisons of sample nucleotide or amino acid sequences against reference sequences indicate a strong relationship or sequence similarity to the reference. The E-value gauges the likelihood of chance-based encounters with alignments whose scores are at least as high as the observed scores when query sequence comparison. A lower E-value (near 0) suggest a non-random similarity, implying potential functional or evolutionary significance (shared ancestry, similar function or structural). The Query Coverage (Query Cover) signifies the proportion of the query sequence that aligns with the database sequence. A heightened Query Cover value indicates that more nucleotides or amino acids from the query aligning or are being covered by the database. The Percentage of Identity indicates the proportion of matching nucleotides between the query and subject sequences. A greater percentage of Identity signifies a higher resemblance between the nucleotides or amino acids in the aligned query and those in the database.

Based on the haplotype diversity value (0.711), it suggested that the CDS region of *SOX9* gene exons 1 and 2 nucleotide sequence had a relatively high level of diversity. The main factor contributing to this result was the presence of heterozygous positions (28 out of 34 nucleotide variations) (Supplementary Table 3). The red junglefowl species, specifically *G. gallus spadiceus* and *G. gallus murghi*, exhibited the highest number of heterozygous positions, totalling 8 and 13 bp, respectively (Table 2 and Supplementary Table 3). Moreover, the red junglefowl species showed the highest number of nucleotide variants, comprising

Table 3. BLASTX result of CDS region of exons 1 dan 2 of *SOX9* gene of *Gallus*

Description	Species	Accession number	Value Range		
			Query cover (%)	E-value	Identity (%)
SOX9 protein	*G. gallus*	NP_989612	99	7e-133 - 2e-128	97.81-100
SOX9 protein isoform X1		XP_046785201	99	1e-132 - 3e-128	97.81-100
SOX9 protein		AAB09663	99	1e-127 - 4e-123	95.63-97.82
SOX10 protein		NP_990123	52	1e-78 - 6e-75	88.43-91.74

Table 4. Putative amino acid sequence variation in CDS region of exons 1 and 2 of *SOX9* gene between *Gallus* species

Haplotype group	N	Amino acid position																			
		6	8	11	49	57	61	68	49	84	102	125	132	134	139	145	151	163	171	201	227
Hap_aa_1	75	P	M	T	Q	D	K	Y	Q	T	K	D	N	E	L	L	K	V	D	N	H
Hap_aa_2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	M	E	.	.
Hap_aa_3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.
Hap_aa_4	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	E	.	.
Hap_aa_5	1	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Hap_aa_6	1	.	.	P	.	.	.	.	.	P	.	.	.	.	.	.	.	.	.	.	.
Hap_aa_7	1	.	L	P	.	.	.	.	.	P	.	.	.	.	R	.	.	.	.	.	.
Hap_aa_8	1	.	.	.	P	.	.	D	P	.	.	.	.	.	.	.	.	.	.	.	R
Hap_aa_9	1	T	.	.	.	.	.	.	.	.	.	A	T	A	.	.	T	.	.	.	.
Hap_aa_10	1	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.
Hap_aa_11	1	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	T	.
Hap_aa_12	1	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	R	.	.	.	.	.

N: number of sequences; Red-coloured number: heterozygous position

21 heterozygous and 5 non-heterozygous positions (Supplementary Table 3). This finding was intriguing and in need of exploration.

The depth of coverage in NGS sequencing refers to the number of read covering each base in a sequenced genome or target region. The depth of coverage implicates the quality and reliability of the obtained data, the higher the sequencing depth, the more accurate the base calling, reducing sequencing error (Bentley *et al.* 2008; Kim *et al.* 2015; Borisevich *et al.* 2017). Each nucleotide position in this study had a diverse depth of coverage in all *Gallus* species (SRA data). A different methodology in conducting NGS sequencing was demonstrated, affecting the depth of coverage (Borges *et al.* 2020). Considerable debate revolves around determining the minimum depth of coverage necessary for generating sufficiently accurate results and minimizing the likelihood of false positives and negatives (Desai *et al.* 2013; Kim *et al.* 2015; Borisevich *et al.* 2017; Petrackova *et al.* 2019). The least minimum number of depth of coverage was reported by Borisevich *et al.* (2017). Their research demonstrated that to achieve accurate base calling of heterozygotes and single nucleotide variants (SNVs), a depth of coverage of 12-fold is required.). In *G. gallus spadiceus* and *G. gallus murghi*, the majority of depth of coverage in the nucleotide variant position is lower than 10-fold, which made the accuracy in those positions was low. It might be due to this the total nucleotide variant that found in these 2 species was considerably higher than the others. Furthermore, while mapping *G. gallus murghi* SRA data, we identified nucleotide variants at numerous positions in the majority of *G. gallus murghi* sequence. Therefore, we speculated about the potential occurrence of interbreeding between red junglefowl and domestic chicken within the red junglefowl dataset, specifically within *G. gallus murghi* and *G. gallus spadiceus*. The study by Ulfah *et al.* (2016) delved into the genetic and phylogenetic relationships among red and green junglefowl along with domestic chicken breeds, detecting indications of interbreeding between red junglefowl and domestic chicken breeds in the red junglefowl data which stored in the online repository.

From the viewpoint of amino acid sequences of the CDS region of exons 1 and 2 of the *SOX9* gene, there is no clear difference among 4 junglefowl species. This uniformity is underscored by the predominant clustering of putative amino acid sequences within the Hap_aa_1 (>87%) and its small haplotype diversity value (0.241). This result implies that the amino acid encoded by exons 1 and 2 of the *SOX9* gene is considerably conserved and had low diversity level. Codon redundancy and selective constraint on protein function is 2 of the factors that could influence the differences in diversity level between nucleotide and amino acid sequences. The majority of amino acids are encoded by multiple codons, resulting in redundancy in the genetic code. This redundancy permits synonymous mutations. Selective pressures on amino acid sequences maintain protein function and structure making non-synonymous mutations often deleterious and subject to negative selection. Synonymous mutations, however, are usually neutral, contributing to higher nucleotide diversity over time.

Although nucleotide variations within coding regions (CDS) can influence phenotype via changes in encoded amino acids, it is acknowledged that other genomic segments, such as untranslated regions (UTRs) and introns, also contribute to phenotypic outcomes. Regulatory elements located within the untranslated regions (UTRs) and introns of genes modulate the gene expression, thereby impacting phenotypic traits (Touriol *et al.* 2003; Fablet *et al.* 2009; Fejes-Toth *et al.* 2009; Kühn *et al.* 2009; McClelland *et al.* 2009; Raveh-Amit *et al.* 2009; Beaudoin and Perreault 2010; Smith *et al.* 2010). Throughout cartilage tissue formation and development, the *SOX9* gene plays critical roles, including chondrogenic mesenchymal condensation, maintenance of chondrocyte viability, facilitation of chondrocyte differentiation and proliferation, regulation of chondrocyte hypertrophy, extracellular matrix (ECM) component regulation, and modulation of cartilage-specific transcription factors (TFs) (Bi *et al.* 2001; Akiyama *et al.* 2002; Ikegami *et al.* 2011; Liu and Lefebvre 2015; He *et al.* 2016). Given its importance in cartilage development and has been hypothesized as a genetic factor involved in larynx formation and development, it is appropriate to continue to study the *SOX9* gene, including analysis of regions such as CDS exon 3, UTRs, introns, and other regulatory elements. Longtine *et al.* (2024) suggest that changes in syrinx developmental pathway signals correlate with syrinx diversification.

Within the coding sequence (CDS) regions of exons 1 and 2 of the *SOX9* gene, the nucleotide sequence demonstrates high diversity among the four junglefowl species. This diversity is evident in both nucleotide variation and its haplotype diversity value. Notably, distinct nucleotide characteristics are observed exclusively in *G. sonneratii* and *G. lafayetii* within the CDS regions of exons 1 and 2 of the *SOX9* gene, delineated by the formation of unique haplotype groups. Of the species studied, *Gallus gallus* (red junglefowl) displayed the highest nucleotide variations. However, validation is necessary due to the low depth of coverage for individual nucleotide variants and the possibility of interbreeding between red junglefowl and domestic chicken lineages. From an amino acid perspective, there is no clear differentiation among the four junglefowl species within the CDS region of exons 1 and 2 of the *SOX9* gene underscores a considerable level of conservation in amino acid sequences, and based on its haplotype diversity value, the amino acid sequence had low diversity level. To attain a more comprehensive insight of the genetic features of junglefowl *SOX9* gene, it is important to integrate additional data from other regions, such as the CDS region of exon 3, untranslated regions (UTRs), and/or introns.

Acknowledgements

We would like to thank to Ministry of Research and Higher Education for financial support through the grant of PMDSU no. 1/E1/KP.PTNBH/2021.

References

[DDC] DNA Diagnostic Center c2019. Avian sample collection procedure, feather collection. Available at: https://www.vetdnacenter.com/files/collecting-feather-samples.pdf. [date accessed: 25 August 2019]

Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A., de Crombrugghe, B., 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 16, 2813-2828. https://doi.org/10.1101/gad.1017802

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ayers, K.L., Lambeth, L.S., Davidson, N.M., Sinclair, A.H., Oshlack, A., Smith, C.A., 2015. Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. *BMC Genomics.* 16, 1-19. https://doi.org/10.1186/s12864-015-1886-5

Beaudoin, J.D., Perreault, J.P., 2010. 5′-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 38, 7022-7036. https://doi.org/10.1093/nar/gkq557

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Cheetham, R.K., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Catenazzi, M.C.E., Chang, S., Cooley, R.N., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fajardo, K.V.F., Furey, W.S., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Jones, T.A.H., Kang, G.D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ng, B.L., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Pinkard, D.C., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Rodriguez, A.C., Roe, P.M., Rogers, J., Bacigalupo, M.C.R., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Sohna, J.E.S., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 456, 53-59. https://doi.org/10.1038/nature07517

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., de Crombrugghe, B., 2001. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc. Natl. Acad. Sci. U S A.* 98, 6698-6703. https://doi.org/10.1073/pnas.111092198

Borges, M.G., Rocha, C.S., Carvalho, B.S., Lopes-Cendes, I., 2020. Methodological differences can affect sequencing depth with a possible impact on the accuracy of genetic diagnosis. *Genet. Mol. Biol.* 43, e20190270. https://doi.org/10.1590/1678-4685-gmb-2019-0270

Borisevich, D.I., Krasnenko, A.Y., Stetsenko, I.F., Plakhina, D.A., Ilinsky, V.V., 2017. The impact of sequencing depth on accuracy of single nucleotide variant calls. *Bull. Russ. State Medical Univ.* 3, 48-52. https://doi.org/10.24075/brsmu.2017-03-06

Cheung, M., Briscoe, J. 2003. Neural crest development is regulated by the transcription factor Sox9. *Development.* 130, 5681–5693. https://doi.org/10.1242/dev.00808

Desai, A., Marwah, V.S., Yadav, A., Jha, V., Dhaygude, K., Bangar, U., Kulkarni, V., Jere, A., 2013. Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data. *PLoS ONE.* 8, e60204. https://doi.org/10.1371/journal.pone.0060204

Düring, D.N., Ziegler, A., Thompson, C.K., Ziegler, A., Faber, C., Müller, J., Scharff, C., Elemans, C.P., 2013. The songbird syrinx morphome: a three-dimensional, high-resolution, interactive morphological map of the zebra finch vocal organ. *BMC Biol.* 11, 1-27. https://doi.org/10.1186/1741-7007-11-1

Fablet, M., Bueno, M., Potrzebowski, L., Kaessmann, H., 2009. Evolutionary origin and functions of retrogene introns. *Mol. Biol. Evol.* 26, 2147-2156. https://doi.org/10.1093/molbev/msp125

Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G.J., Kapranov, P., Foissac, S., Willingham, A.T., Duttagupta, R., Dumais, E., Gingeras, T.R., 2009. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature.* 457, 1028–1032. https://doi.org/10.1038%2Fnature07759

Gaban-Lima, R., Hofling, E., 2006. Comparative anatomy of the syrinx in the tribe Arini (Aves: Psittacidae). *Braz. J. Morphol. Sci.* 23, 501-512.

Gokhman, D., Nissim-Rafinia, M., Agranat-Tamir, L., Housman, G., García-Pérez, R., Lizano, E., Cheronet, O., Mallick, S., Nieves-Colón, M. A., Li, H., Alpaslan-Roodenberg, S., Novak, M., Gu, H., Osinski, J. M., Ferrando-Bernal, M., Gelabert, P., Lipende, I., Mjungu, D., Kondova, I., Bontrop, R., Kullmer, O., Weber, G., Shahar, T., Dvir-Ginzberg, M., Faerman, M., Quillen, E.E., Meissner, A., Lahav, Y., Kandel, L., Liebergall, M., Prada, M.E., Vidal, J.M., Gronostajski, R.M., Stone, A.C., Yakir, B., Lalueza-Fox, C., Pinhasi, R., Reich, D., Marques-Bonet, T., Meshorer, E., Camel, L., 2020. Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* 11, 1-21. https://doi.org/10.1038/s41467-020-15020-6

He, X., Ohba, S., Hojo, H., McMahon, A.P., 2016. AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development.* 143, 3012-3023. https://doi.org/10.1242/dev.134502

Healy, C., Uwanogho, D., Sharpe, P.T., 1999. Regulation and role of SOX9 in cartilage formation. *Dev. Dynam.* 215, 69–78. https://doi.org/10.1002/(sici)1097-0177(199905)215:1<69::aid-dvdy8>3.0.co;2-n

Ikegami, D., Akiyama, H., Suzuki, A., Nakamura, T., Nakano, T., Yoshikawa, H., Tsumaki, N., 2011. Sox9 sustains chondrocyte survival and hypertrophy in part through Pik3ca-Akt pathways. *Development.* 138, 1507-1519. https://doi.org/10.1242/dev.057802

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, 5-9. https://doi.org/10.1093/nar/gkn201

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., Koopman, P., 1996. A male-specific role for SOX9 in vertebrate sex determination. *Development.* 122, 2813-2822. https://doi.org/10.1242/dev.122.9.2813

Kim, K., Seong, M.-W., Chung, W.-H., Park, S.S., Leem, S., Park, W., Kim, J., Lee, K., Park, R.W., Kim, N., 2015. Effect of next-generation exome sequencing depth for discovery of diagnostic variants. *Genomics Inform.* 13, 31-39. https://doi.org/10.5808/gi.2015.13.2.31

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., Wahle, E., 2009. Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J. Biol. Chem.* 284, 22803-22814. https://doi.org/10.1074/jbc.m109.018226

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. DOI:10.1093/molbev/msw054

Lawal, R.A., Martin, S.H., Vanmechelen, K., Vereijken, A., Silva, P., Al-Atiyat, R.M., Aljumaah, R.S., Mwacharo, J. M., Wu, D.D., Zhang, Y.P., Hocking, P.M., Smith, J., Wragg, D., Hanotte, O., 2020. The wild species genome ancestry of domestic chickens. *BMC Biol.* 18, 1-18. https://doi.org/10.1186/s12915-020-0738-1

Liu, C.F., Lefebvre, V., 2015. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res.* 43, 8183-8203. https://doi.org/10.1093/nar/gkv688

Longtine, C., Eliason, C.M., Mishkind, D., Lee, C., Chiappone, M., Goller, F., Love, J., Kingsley, E.P., Clarke, J.A., Tabin, C.J., 2024. Homology and the evolution of vocal folds in the novel avian voice box. *Curr Biol.* 34, 461-472. DOI:10.1016/j.cub.2023.12.013

Mariadassou, M., Suez, M., Sathyakumar, S., Vignal, A., Arca, M., Nicolas, P., Faraut, T., Esquerré, D., Nishibori, M., Vieaud, A., Chen, C.-F., Manh Pham, H., Roman, Y., Hospital, F., Zerjal, T., Rognon, X., Tixier-Boichard, M., 2021. Unraveling the history of the genus *Gallus* through whole genome sequencing. *Mol. Phylogenet. Evol.* 158, 107044. https://doi.org/10.1016/j.ympev.2020.107044

McClelland, S., Shrivastava, R., Medh, J.D., 2009. Regulation of translational efficiency by disparate 5' UTRs of PPARγ splice variants. *PPAR Res.* 2009, 1-8. https://doi.org/10.1155/2009/193413

Montero, J.A., Lorda-Diez, C.I., Francisco-Morcillo, J., Chimal-Monroy, J., Garcia-Porrero, J.A., Hurle, J.M., 2017. SOX9 expression in amniotes: species-specific differences in the formation of digits. *Front. Cell Dev. Biol.* 5, 23. https://doi.org/10.3389/fcell.2017.00023

Morejohn, G.V., 1966. Variation of the syrinx of the fowl. *Poult. Sci.* 45, 33-39. https://doi.org/10.3382/ps.0450033

Ng, L.J., Wheatley, S., Muscat, G.E.O., Conway-Campbell, J., Bowles, J., Wright, E., Bell, D.M., Tam, P.P.L., Cheah, K.S.E., Koopman, P., 1997. SOX9 binds DNA, activates transcription and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev. Biol.* 183, 108-121. https://doi.org/10.1006/dbio.1996.8487

Oh, C., Lu, Y., Liang, S., Mori-Akiyama, Y., Chen, D., de Crombrugghe, B., Yasuda, H., 2014. SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. *PLOS ONE.* 9, e107577. https://doi.org/10.1371/journal.pone.0107577

Petrackova, A., Vasinek, M., Sedlarikova, L., Dyskova, T., Schneiderova, P., Novosad, T., Papajik, T., Kriegova, E., 2019. Standardization of sequencing coverage depth in NGS: recommendation for detection of clonal and subclonal mutations in cancer diagnostics. *Front. Oncol.* 9, 851. https://doi.org/10.3389/fonc.2019.00851

Raveh-Amit, H., Maissel, A., Poller, J., Marom, L., Elroy-Stein, O., Shapira, M., Livneh, E., 2009. Translational control of protein kinase Cs by two upstream open reading frames. *Mol. Cell. Biol.* 29, 6140-6148. https://doi.org/10.1128/mcb.01044-09

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the european molecular biology open software suite. *Trends Genet.* 16, 276-277. https://doi.org/10.1016/s0168-9525(00)02024-2

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299-3302. https://doi.org/10.1093/molbev/msx248

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463-5467. https://doi.org/10.1073/pnas.74.12.5463

Scheider, J., Afonso-Grunz, F., Hoffmeier, K., Horres, R., Groher, F., Rycak, L., Oehlmann, J., Winter, P., 2014. Gene expression of chicken gonads is sex-and side-specific. *Sex. Dev.* 8, 178-191. https://doi.org/10.1159/000362259

Smith, L., Brannan, R.A., Hanby, A.M., Shaaban, A.M., Verghese, E.T., Peter, M.B., Pollock, S., Satheesha, S., Szynkiewicz, M., Speirs, V., Hughes, T.A., 2010. Differential regulation of oestrogen receptor β isoforms by 5' untranslated regions in cancer. *J. Cell. Mol. Med.* 14, 2172-2184. https://doi.org/10.1111/j.1582-4934.2009.00867.x

Su, L.N., Li, H., Tan, S.W., Fang, G.J., Yu, H., Yang, Y.L., 2019. Mechanisms of early- and late-feathering in Qingyuan partridge chickens. *Biotechnol. Biotechnol. Equip.* 33, 1172–1181. https://doi.org/10.1080/13102818.2019.1645619

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C., Vagner, S., 2003. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell.* 95, 169-178. https://doi.org/10.1016/s0248-4900(03)00033-9

Ulfah, M., Kawahara-Miki, R., Farajalllah, A., Muladno, M., Dorshorst, B., Martin, A., Kono, T., 2016. Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. *BMC Genomic.* 17, 1-9. https://doi.org/10.1186/s12864-016-2652-z

Wang, M.S., Thakur, M., Peng, M.S., Jiang, Y., Frantz, L.A.F., Li, M., Zhang, J.J., Wang, S., Peters, J., Otecko, N.O., Suwannapoom, C., Guo, X., Zheng, Z.Q., Esmailizadeh, A., Hirimuthugoda, N.Y., Ashari, H., Suladari, S., Zein, M.S.A., Kusza, S., Sohrabi, S., Kharrati-Koopaee, H., Shen, Q.K., Zeng, L., Yang, M.M., Wu, Y.J., Yang, X.Y., Lu, X.M., Jia, X.Z., Nie, Q.H., Lamont, S.J., Lasagna, E., Ceccobelli, S., Gunwardana, H.G.T.N., Senasige, T.M., Feng, S.H., Si, J.F., Zhang, H., Jin, J.Q., Li, M.L., Liu, Y.H., Chen, H.M., Ma, C., Dai, S.S., Bhuiyan, A.K.F.H., Khan, M.S., Silva, G.L.L.P., Le, T.T., Mwai, O.A., Ibrahim, M.N.M., Supple, M., Shapiro, B., Hanotte, O., Zhang, G., Larson, G., Han, J.L., Wu, D.D., Zhang, Y.P., 2020. 863 genomes reveal the origin and domestication of chicken. *Cell Res.* 30, 693-701. https://doi.org/10.1038/s41422-020-0349-y

Wright, E., Hargrave, M. R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A., Koopman, P., 1995. The SRY-related gene SOX9 is expressed during chondrogenesis in mouse embryos. *Nat. Genet.* 9, 15-20. https://doi.org/10.1038/ng0195-15

Zhao, Q., Eberspaecher, H., Lefebvre, V., de Crombrugghe, B., 1997. Parallel expression of SOX9 and Col2a1 in cells undergoing chondrogenesis. *Dev. Dynam.* 209, 377–386. https://doi.org/10.1002/(sici)1097-0177(199708)209:4<377::aid-aja5>3.0.co;2-f

Supplementary Materials

Supplementary Table 1. BLASTN result of exons 1 and 2 of *SOX9* gene of *G. varius* sample

Species	Accession number	Exon	Value Range		
			Query cover (%)	E-value	Identity (%)
G. gallus	CP100572	1	100	0.0	99.24-99.69
		2	99-100	0.0	98.10-99.35
	XM_046929245	1	58-66	0.0	99.79-100
		2	30-56	8e-142-4e-130	96.23-100
	NM_204281	1	58-66	0.0	99.79-100
		2	29-55	2e-133-2e-128	98.46-100
	AB012236	1	58-66	0.0	99.79-100
		2	29-55	4e-132-2e-127	98.44-100

Supplementary Table 2. Percentage mean of nucleotide composition of the CDS region of exons 1 and 2 of the *Gallus SOX9* gene

Species group	Mean length (bp)	Percentage mean of each nucleotide type			
		T	C	A	G
G. gallus (domesticated chicken)	685	12.6	12.6	24.5	29.6
G. varius	685	12.6	12.6	24.5	29.6
G. lafayetii	685	12.8	12.8	24.5	29.7
G. sonneratii	685	12.8	12.8	24.5	29.6
G. gallus gallus	685	12.6	12.6	24.5	29.7
G. gallus bankiva	685	12.6	12.6	24.6	29.6
G. gallus spadiceus	685	12.6	12.6	24.5	29.7
G. gallus murghi	685	12.6	12.6	24.4	29.7

Supplementary Table 3. Nucleotide variation and heterozygous position in CDS region of exons 1 and 2 of *SOX9* gene between *Gallus* species

Sequence name	Nucleotide position																
	16	22	31	99	102	146	156	170	181	183	204	250	303	306	374	384	395
*G_gallus*_NC_052549	C	A	A	C	C	A	C	A	A	G	A	T	C	A	A	G	A
*G_gallus*_NC_052590	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Sumbawa_2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Sumbawa_1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX9334260	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX9334259	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX7909216	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173447	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173446	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173445	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_ERX4842472	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_ERX4842466	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083693	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083692	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083691	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083690	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083689	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Red-coloured number: heterozygous position

Supplementary Table 3. Continued

Sequence name	Nucleotide position																
	16	22	31	99	102	146	156	170	181	183	204	250	303	306	374	384	395
*G_varius*_DRX083688	C	A	A	C	C	A	C	A	A	G	A	T	C	A	A	G	A
*G_varius*_DRX083687	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083685	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX9334252	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX9334251	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX5173439	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.
*G_lafayetii*_SRX5173433	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.
*G_lafayetii*_ERX4842449	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	K	.
*G_lafayetii*_ERX4842448	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.
*G_sonneratii*_SRX9334258	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_sonneratii*_SRX9334257	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_sonneratii*_SRX9334256	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_sonneratii*_SRX9334254	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_sonneratii*_SRX7909214	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_sonneratii*_SRX5173441	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_gallus*_SRX9334263	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_gallus*_SRX9334262	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_gallus*_SRX9334255	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_gallus*_ERX4842913	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_gallus*_ERX4842911	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_bankiva*_SRX9334244	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_bankiva*_SRX9334243	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_bankiva*_ERX4842909	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_bankiva*_ERX4842907	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.	G	.
*G_gallus_spadiceus*_SRX9334249	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_spadiceus*_SRX9334245	.	.	.	.	S	.	.	.	.	.	.	.	T	.	.	G	.
*G_gallus_spadiceus*_ERX4843041	.	M	C	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_spadiceus*_ERX4843040	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	G	.
*G_gallus_spadiceus*_ERX4843032	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_spadiceus*_ERX4843025	.	.	.	.	C	M	.	.	.	.	.	K	.	.	.	G	.
*G_gallus_jabouillei*_ERX4842936	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_jabouillei*_ERX4842935	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_jabouillei*_ERX4842930	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_jabouillei*_ERX4842929	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_jabouillei*_ERX4842928	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_murghi*_SRX9334265	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_murghi*_ERX4842982	M	.	.	.	C	.	.	.	.	.	.	.	.	.	M	G	M
*G_gallus_murghi*_SRX9334266	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_murghi*_SRX9334267	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.
*G_gallus_murghi*_ERX4842987	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.
*G_gallus_murghi*_ERX4842997	.	.	.	.	C	.	.	.	M	R	M	.	Y	.	.	G	.

Red-coloured number: heterozygous position

Supplementary Table 3. Continued

Sequence name	Nucleotide position																
	401	411	416	433	434	435	447	452	474	487	513	602	627	633	654	672	680
*G_gallus*_NC_052549	A	G	T	C	T	G	C	A	C	G	C	A	G	C	C	C	A
*G_gallus*_NC_052590	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Sumbawa_2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Sumbawa_1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_11	.	.	.	M	.	.	.	.	.	.	S	.	.	.	.	.	.
*G_varius*_Lombok_10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_7	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.
*G_varius*_Lombok_6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_Lombok_4	.	.	.	M	.	.	.	.	.	R	S	.	.	.	.	.	.
*G_varius*_Lombok_3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX9334260	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX9334259	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX7909216	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173447	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173446	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_SRX5173445	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_ERX4842472	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_ERX4842466	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083693	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083692	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083691	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083690	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083689	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083688	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_varius*_DRX083687	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.
*G_varius*_DRX083685	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX9334252	.	.	.	.	.	.	Y	.	T	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX9334251	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
*G_lafayetii*_SRX5173439	.	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.
*G_lafayetii*_SRX5173433	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
*G_lafayetii*_ERX4842449	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
*G_lafayetii*_ERX4842448	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
*G_sonneratii*_SRX9334258	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.
*G_sonneratii*_SRX9334257	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.
*G_sonneratii*_SRX9334256	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.
*G_sonneratii*_SRX9334254	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
*G_sonneratii*_SRX7909214	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
*G_sonneratii*_SRX5173441	.	.	.	.	.	.	T	.	.	.	.	.	.	Y	.	.	.
*G_gallus_gallus*_SRX9334263	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_gallus*_SRX9334262	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_gallus*_SRX9334255	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_gallus*_ERX4842913	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_gallus*_ERX4842911	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_bankiva*_SRX9334244	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_bankiva*_SRX9334243	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_bankiva*_ERX4842909	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_bankiva*_ERX4842907	.	R	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.

Red-coloured number: heterozygous position

Supplementary Table 3. Continued

Sequence name	Nucleotide position																
	401	411	416	433	434	435	447	452	474	487	513	602	627	633	654	672	680
*G_gallus_spadiceus*_SRX9334249	A	G	T	C	T	G	C	A	C	G	C	A	R	C	C	C	A
*G_gallus_spadiceus*_SRX9334245	.	.	.	.	.	.	Y	.	.	.	.	.	.	C	.	.	.
*G_gallus_spadiceus*_ERX4843041	.	.	K	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_spadiceus*_ERX4843040	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_spadiceus*_ERX4843032	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_spadiceus*_ERX4843025	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	R
*G_gallus_jabouillei*_ERX4842936	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_jabouillei*_ERX4842935	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_jabouillei*_ERX4842930	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_jabouillei*_ERX4842929	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_jabouillei*_ERX4842928	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_murghi*_SRX9334265	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_murghi*_ERX4842982	M	.	.	.	.	.	.	M	.	.	.	.	.	C	.	.	.
*G_gallus_murghi*_SRX9334266	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_murghi*_SRX9334267	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
*G_gallus_murghi*_ERX4842987	.	.	.	.	.	.	.	.	.	.	.	M	.	C	.	.	.
*G_gallus_murghi*_ERX4842997	.	.	.	.	K	K	.	.	.	.	.	.	.	C	.	.	.

Red-coloured number: heterozygous position

Supplementary Table 4. Haplotype of nucleotide sequences of the CDS region of exons 1 and 2 of the *SOX9* gene

Haplotype group	N	Number of nucleotide sequence for each *Gallus* species								
		a	b	c	d	e	f	g	h	i
Hap_1	46	26	-	-	2	5	2	2	5	4
Hap_2	1	1	-	-	-	-	-	-	-	-
Hap_3	1	1	-	-	-	-	-	-	-	-
Hap_4	1	1	-	-	-	-	-	-	-	-
Hap_5	1	1	-	-	-	-	-	-	-	-
Hap_6	1	1	-	-	-	-	-	-	-	-
Hap_7	1	-	1	-	-	-	-	-	-	-
Hap_8	1	-	1	-	-	-	-	-	-	-
Hap_9	1	-	1	-	-	-	-	-	-	-
Hap_10	1	-	1	-	-	-	-	-	-	-
Hap_11	1	-	1	-	-	-	-	-	-	-
Hap_12	1	-	1	-	-	-	-	-	-	-
Hap_13	2	-	2	-	-	-	-	-	-	-
Hap_14	4	-	-	4	-	-	-	-	-	-
Hap_15	3	-	-	3	-	-	-	-	-	-
Hap_16	5	-	-	-	-	1	-	-	2	2
Hap_17	1	-	-	-	-	-	1	-	-	-
Hap_18	1	-	-	-	-	-	1	-	-	-
Hap_19	1	-	-	-	-	-	1	-	-	-
Hap_20	1	-	-	-	-	-	-	1	-	-
Hap_21	1	-	-	-	-	-	-	1	-	-
Hap_22	2	-	-	-	-	-	-	2	-	-
Hap_23	1	-	-	-	-	-	-	1	-	-
Hap_24	1	-	-	-	-	-	-	1	-	-
Hap_25	1	-	-	-	-	-	-	1	-	-
Hap_26	1	-	-	-	-	-	-	1	-	-
Hap_27	1	-	-	-	-	-	-	-	1	-
Hap_28	1	-	-	-	-	-	-	-	1	-
Hap_29	1	-	-	-	-	-	-	-	1	-
Hap_30	1	-	-	-	-	-	-	-	1	-

N: number of sequences, -: 0, a: *G. varius*, b: *G. lafayetii*, c: *G. sonneratii*, d: *G. gallus*, e: *G. gallus gallus*, f: *G. gallus bankiva*, g: *G. gallus spadiceus*, h: *G. gallus murghi*, i: *G. gallus jabouillei*

Supplementary Table 5. Number of synonymous and non-synonymous mutations in the putative amino acid sequence of the CDS region of exons 1 and 2 of the *SOX9* gene in *Gallus*

Type of mutation	Number of amino acid containing mutation type	Number of amino acid containing	
		Heterozygous site	Non-heterozygous site
Synonymous	12	8	4
Non-synonymous	19	17	2

Supplementary Table 6. Haplotype of putative amino acid sequences of the CDS region of exons 1 and 2 of the *SOX9* gene in *Gallus*

Haplotype group	N	Number of nucleotide sequence for each *Gallus* species								
		a	b	c	d	e	f	g	h	i
Hap_aa_1	75	27	8	7	2	6	5	7	7	6
Hap_aa_2	1	1	-	-	-	-	-	-	-	-
Hap_aa_3	1	1	-	-	-	-	-	-	-	-
Hap_aa_4	1	1	-	-	-	-	-	-	-	-
Hap_aa_5	1	1	-	-	-	-	-	-	-	-
Hap_aa_6	1	-	-	-	-	-	-	1	-	-
Hap_aa_7	1	-	-	-	-	-	-	1	-	-
Hap_aa_8	1	-	-	-	-	-	-	1	-	-
Hap_aa_9	1	-	-	-	-	-	-	-	1	-
Hap_aa_10	1	-	-	-	-	-	-	-	1	-
Hap_aa_11	1	-	-	-	-	-	-	-	1	-
Hap_aa_12	1	-	-	-	-	-	-	-	1	-

N: number of sequences, -: 0, a: *G. varius*, b: *G. lafayetii*, c: *G. sonneratii*, d: *G. gallus*, e: *G. gallus gallus*, f: *G. gallus bankiva*, g: *G. gallus spadiceus*, h: *G. gallus murghi*, i: *G. gallus jabouillei*