

RESEARCH

Open Access



# Molecular characterization and phylogenetic analyses of the mitogenome of Wan-Xi white goose, a native goose breed in China

Lunbin Xia<sup>1,2,3\*</sup>, Shaoshuai Bi<sup>1,2</sup>, Yafei Zhang<sup>1,2</sup>, Cunwu Chen<sup>1</sup> and Naidong Chen<sup>1,3\*</sup>

## Abstract

**Background** The Wan-Xi white goose (WXG), an indigenous Chinese waterfowl (*Anserini: Anserinae*), is crucial for goose germplasm conservation. This study aimed to sequence and analyze the complete mitochondrial DNA (mtDNA) of WXG using the BGISEQ-500 platform. The mtDNA's structure and function were investigated to gain insights into its genetic diversity and population structure.

**Results** The mtDNA was found to be 16,743 bp long and comprised 22 transfer RNA (tRNA) genes, 2 ribosomal RNA genes, a complement of 13 protein-coding genes (PCGs), as well as a single noncoding control region known as the D-loop. Notably, all tRNA genes, except for *trnS1-tRNA* which lacked the dihydrouridine stem, were predicted to adopt the typical cloverleaf structure. Given the genetic variability across the mtDNA of *Anser* spp. and the intergenic gaps identified by codon analysis, the codon usage patterns were comprehensively examined via comparative analysis of the mtDNAs of WXG and 24 other *Anser* spp. The relative synonymous codon usage (RSCU) values of the 13 mitochondrial PCGs of WXG were consistent with those of the mitochondrial PCGs of the 24 other *Anser* spp. Analysis of the neutrality (GC3-GC12), the effective number of codons (ENCs)-GC3, and parity rule 2-bias plots further revealed that natural selection emerged as the primary factor influencing codon bias in *Anser* sp. High nucleotide diversity ( $\pi > 0.02$ ) was observed in several regions, including the *D-loop*, *ATP6*, *12S rRNA*, *ND1*, *16S rRNA\_ND1*, *COX2*, and *ND5*. Furthermore, the results of nonsynonymous ( $K_a$ )/synonymous ( $K_s$ ) analysis of the 13 mitochondrial PCGs of the 25 species under *Anser* revealed that the genes were subject to strong purifying selection. The findings of phylogenetic analysis further revealed that WXG and 10 other members of *Anser cygnoides* clustered into a single branch to form a monophyletic group.

**Conclusion** This research provides valuable insights into the mtDNA of WXG, highlighting its genetic diversity and population structure. The identified mutation hotspots and purifying selection on mitochondrial PCGs suggest potential areas for future research on *Anser cygnoides*. The findings contribute to our understanding of this rare species and its conservation efforts.

**Keywords** *Anser cygnoides*, Mitogenome, Wan-Xi white goose, Phylogenetic analysis

\*Correspondence:

Lunbin Xia  
xlwxc@126.com  
Naidong Chen  
2004cnd@163.com

<sup>1</sup>College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an 237012, P.R. China

<sup>2</sup>Small Animal Diagnosis and Treatment Experimental Training Center, West Anhui University, Lu'an 237012, P.R. China

<sup>3</sup>Anhui Province Key Laboratory for Quality Evaluation and Improvement of Traditional Chinese Medicine, Lu'an 237012, P.R. China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

## Introduction

The Wan-Xi white goose (WXG), an indigenous Chinese breed with over 2,000 years of domestication history, is primarily distributed in the hill regions of western Anhui Province, China. Recognized as a national genetic resource under China's Livestock and Poultry Conservation List, this breed exhibits exceptional traits including robust disease resistance, environmental adaptability, and unique feather development patterns that contribute to its globally renowned down quality. Despite its economic significance—with annual production exceeding 8 million birds valued for nutrient-rich meat (adult males averaging 6.8 kg and females 5–6 kg) and premium down accounting for 15% of global high-grade feather exports—the WXG faces population growth constraints due to limited reproductive efficiency, producing only 25 eggs per 150-day laying cycle compared to 55 eggs in Italian breeds [1, 2]. Mitochondrial haplotype analyses reveal its divergence from European goose lineages [3], yet critical gaps persist in characterizing its complete mitogenome structure and codon usage patterns, which are essential for elucidating evolutionary adaptations and informing conservation strategies.

Mitochondria are semiautonomous organelles that serve as the primary sites of aerobic metabolism in eukaryotes and can generate ATP in cells via oxidative phosphorylation. In addition, they can simultaneously break down sugars, fats, and amino acids via oxygenation to provide energy for cellular growth and metabolism [4], which remains a significant area of interest in molecular biology. The rapid advancements in genome sequencing technologies over the past decade have facilitated the extensive application of mitogenomics in systematic studies on animals, including livestock and poultry [5–7]. This is primarily attributed to the numerous advantages of mitochondrial DNA (mtDNA), including the presence of relatively fewer genes, simple structures and organizations, presence of conserved gene sequences, high rates of evolution [8, 9], and high copy numbers per cell, compared to those of nuclear DNA, which facilitate the easy isolation, sequencing, and assembly of mtDNA [10]. Additionally, mtDNA exhibits maternal inheritance, which makes it preferable to nuclear DNA for certain applications aimed at elucidating the population genetic diversity and molecular phylogenetic relationships of various species [9, 11]. The identification and management of genetic diversity in endemic species are crucial for implementing strategies aimed at expanding breeding populations and ensuring the effective utilization of germplasm resources [12, 13]. Consequently, mtDNA has emerged as an extensively used source of molecular data in studies focusing on population genetics, evolutionary patterns, and phylogenetic analysis of livestock species [14, 15].

Recent studies on vertebrate mtDNAs have demonstrated that the mtDNAs of animals are typically 15–24 kb-long, with a closed circular double-stranded structure [16], and comprise a heavy strand (H-strand) and a light strand (L-strand) containing a total of 37 genes, including 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, 13 protein-coding genes (PCGs) [17, 18], and one or two A + T-rich noncoding control region (D-loop) [19]. Compared to nuclear DNA, mtDNAs are abundant in cells, contain a higher number of conserved genes, are devoid of introns, and encode more significant phylogenetic information [20, 21]. These advantages make mtDNAs highly valuable for the biological identification of different species, as well as for studies on their taxonomy, phylogeny, and genetic structures [17, 22].

According to statistics, China possesses the richest goose breed resources worldwide, with 30 native goose breeds [23]. The WXG is globally the most renowned goose breed for its down feathers and is one of the varieties native to China. Therefore, the city of Lu'an is designated as the world's "Down Capital" due to the origin of the WXG in this area. Previous studies have mainly focused on the reproductive physiology [1, 24, 25], nutritional regulation, production performance [26, 27] and disease prevention [28] of the WXG. The obtained findings promoted the development of the goose industry in China, which has led to the extensive exploitation and utilization of geese. However, the preservation of the purebred population, which is vital for the future breeding of WXG, has been overlooked to date. Understanding the phylogeny and evolution of WXG is essential for ensuring the sustainability of its breeding. Extensive phylogenetic studies have established that Chinese domestic geese predominantly originated from *Anser cygnoides*, primarily through mitochondrial DNA (mtDNA) analyses [3, 29]. However, critical gaps persist in molecular investigations of the WXG—notably, its complete mitogenome remains absent from GenBank, and no studies have integrated codon usage bias with positive selection analysis in *Anatidae*. Therefore, obtaining the complete sequence of the mtDNA of WXG would have immense scientific value. This study aimed to sequence, assemble, and annotate the complete mtDNA sequence of WXG using a high-throughput sequencing technology, characterizing its structural organization, base composition, and tRNA secondary structures. Through comparative mtDNA of 25 species geese, we systematically analyzed: lineage-specific codon usage patterns driven by mutation-selection dynamics, positively selected genes under adaptive evolution, and phylogenetic relationships reconstructed via maximum likelihood (ML) method. The resultant high-confidence phylogeny clarifies WXG's taxonomic position within *Anser* and reveals mitogenomic adaptations underlying its unique traits.

By establishing the first comprehensive mitochondrial profile for WXG, this work provides an essential molecular foundation for conservation prioritization, breed improvement strategies, and future studies on waterfowl evolutionary genomics.

## Materials and methods

### Specimen collection and DNA extraction

A female WXG were purchased in April 2024 from an purebred conservation center of WXG in Lu'an, Anhui Province, China (longitude: 116° 33' 23'' E, and latitude: 31° 50' 29'' N) (Fig. 1). The geese were euthanized under anesthesia (involved administering Pentobarbital Sodium at 120 mg/kg, once unconscious, 10% Potassium Chloride at 0.5 mL/kg, both intravenously) in the veterinary pathology laboratory of West Anhui University, following which samples of leg muscles were collected, and stored at -80 °C for the next stage of the experiment. The total DNA was extracted using a DNA Extraction Kit (Roche). After quantification, the genomic DNA concentration was determined to be 95.4 ng/μl, with a total yield of 5.2 μg. The extracted DNA sample was fragmented by ultrasonication and subsequently subjected to fragment

purification, end-repair, and 3'-end adenylation, following which the sequencing adapters were ligated. The size of the DNA fragments was ascertained by using agarose gel electrophoresis technique, and the sequence libraries were subsequently prepared by polymerase chain reaction (PCR) amplification. The libraries thus constructed were initially inspected, and the validated libraries were sequenced using a BGISEQ-500 sequencing platform (Bio & Data Biotechnology Co., Ltd., China).

### Sequencing, assembly, annotation, and analysis of mtDNA

The total DNA of WXG thus isolated was subjected to high-throughput sequencing on a BGISEQ-500 sequencing platform, which generated 12,343,778,400 bp of raw sequencing data (Supplementary Table S1). The raw sequencing data underwent filtering utilizing the fastp tool (-5 -3-n 0 -f 5 -F 5 -t 5 -T 5 -q 20), specifically version 0.23.2 [30], to obtain 11,234,296,800 bp of clean data. The clean data were subsequently aligned to the sequence of the vertebrate mitochondrial core gene that served as a reference, using Minimap2 (< target. fa >[query.fa]), version 2.1 [31]. The coding gene with the highest sequence coverage was subsequently used



**Fig. 1** Physical characteristics and geographical distribution of WXG

as the seed sequence for the de novo assembly of WXG mtDNA using NOVOPlasty (Type = mito, Genome range = 15,500–17000, K-mer = 39/49/59/69, Seed Input = seed. fa) [32]. To assess the completeness and accuracy of the genome assembly, all sequence reads were remapped to the candidate mtDNA using Geneious Prime (version 2024.0, with default parameters). This step confirmed the assembly of the complete mitochondrial genome sequence of WXG [9]. The circular mtDNA of WXG was finally obtained.

The rRNA genes, tRNA genes, PCGs, and D-loop region in the mtDNA of WXG were annotated using the online MITOS2 tool (<https://usegalaxy.org/>). The results of sequence annotation were refined by comparing the sequences with those of other species under order Anseriformes in the NCBI database, and manually curated to enhance the annotation accuracy. A circular map of the mtDNA of WXG was finally generated using the online OGDRAW tool (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>). The secondary structures of the tRNAs were projected utilizing the tRNAscan-SE webserver (<https://lowelab.ucsc.edu/tRNAscan-SE/>) [33]. The skew values of AT and GC in the complete sequence of WXG mtDNA, as well in the PCGs, tRNA genes, rRNA genes, and D-loop regions, were additionally determined via the formulae: GC-skew =  $(G-C)/(G+C)$ , AT-skew =  $(A-T)/(A+T)$  [34].

### Comparative analyses of mtDNAs

The mtDNAs of closely related *Anser* spp. were retrieved from the NCBI database (Supplementary Table S2). The length and base characteristics of the mtDNAs were statistically analyzed using the Seqkit tool, version 0.16.1 [35]. Codon bias was determined by relative synonymous codon usage (RSCU) analysis of the individual codons in the mitogenomes of *Anser* spp. [36]. The ENC value and G + C (GC) content of the first (GC1), second (GC2), and third (GC3) codon positions, and the overall GC content of the protein-coding regions, were determined using EMBOSS [37, 38]. The balance between mutation and selection in generating codon bias can be evaluated by GC3-GC12 analysis, which is also known as neutrality plot analysis. In this study, GC12 represented the average GC content in GC1 and GC2 codon positions, respectively. The statistical correlation between GC12 and GC3 indicates whether natural selection or mutational pressure is the primary driving force underlying the codon bias of a species. The base content of GC3 was additionally determined to analyze the effective number of codons (ENCs)-GC3 (ENC vs. GC3) and parity rule 2 (PR2)-bias plots. The ENC-GC3 plot is generally used to analyze whether the codon usage of a specific gene is solely affected by mutation or other factors, such as natural selection [39]. PR2-bias plots are analyzed based on

the A3/(A3 + U3) vs. G3/(G3 + C3) ratio, and they are used to determine the magnitude and direction of gene bias [40]. In this study, the ENC-GC3 and PR2-bias plots were constructed using the ggplot2 package of R, version 4.3.2 [38].

### Analyses of nucleotide variability (Pi) and selection pressure

The diversity of the mtDNA sequences and the selection pressure on the PCGs in the mtDNAs of *Anser* spp. was estimated by calculating the Pi values of 25 species under the *Anser* genus using a sliding window approach in the DnaSP software, version 6.12.03 [41]. The nonsynonymous (Ka) and synonymous (Ks) substitution rates and the Ka/Ks ratios of each PCG were determined with the KaKs\_Calculator 2.0 toolkit [42]. The density distributions of Ka, Ks, and Ka/Ks were graphically represented using the ggplot2 package of R, version 4.3.2.

### Phylogenetic analyses

Subsequently, an investigation was conducted into the phylogenetic relationships of WXG with other species within the *Anatidae*, for which 49 mtDNA sequences from *Anatidae* were retrieved from NCBI, including 2 species under *Phasianinae* that served as the outgroup, 19 species under *Anatinae*, 4 species under *Cygnus*, and 25 species under *Anser*, and subjected to phylogenetic analyses. The sequence data for these 50 species are provided in Table S3. The mitogenomes were aligned using the MAFFT software [43], and the maximum likelihood (ML) tree was constructed using IQtree software [44]. The GTR + F + I + G4 model was selected based on the Bayesian Information Criterion (BIC).

## Results

### Composition of WXG mtDNA

The complete mtDNA of WXG (GenBank accession ID: PQ154620) spans a length of 16,743 bp (Fig. 2) and has a circular double-stranded structure typical of vertebrate mtDNAs. It contains 37 typical genes, including 22 tRNA genes, 2 rRNA genes, 13 PCGs, and one D-loop region, similar to the mtDNAs of most vertebrates. The H-strand of WXG mtDNA comprises a total of 28 genes, of which 12 are PCGs, 2 are rRNA genes, and the remaining 14 are involved in tRNA synthesis. On the other hand, the L-strand contains 9 genes, including 1 PCG and 8 tRNA genes. The mtDNA of WXG contains a 1,178 bp-long control region that is located between the *trnE* (UUC) and *trnF* (GAA) genes, and plays a crucial role in regulating gene transcription [45].

The molecular makeup of the mtDNA of WXG is summarized in Table 1, which shows that the contents of A, T, C, and G were 30.204%, 22.690%, 32.031%, and 15.075%, respectively. The content of A + T (52.894%) in the



**Fig. 2** Gene map of WXG mtDNA. The figure utilizes colors to denote various types of regions and genes. Detailed information is provided in the lower left corner of Figure. The arrows indicate the positive and negative strands, with genes encoded by the positive strand (heavy strand) located on the outer side of the circle, and genes encoded by the negative strand (light strand) located on the inner side. The inner gray circle represents the GC content

**Table 1** Nucleotide composition of WXG's mtDNA

WXG	Length(bp)	A(%)	T(%)	C(%)	G(%)	A + T(%)	C + G(%)	AT skew	GC skew
mtDNA	16,743	30.204	22.690	32.031	15.075	52.894	47.106	0.142	-0.360
PCGs	11,402	29.337	22.610	34.240	13.813	51.947	48.053	0.129	-0.425
tRNAs	1,541	33.420	24.594	25.049	16.937	58.014	41.986	0.152	-0.193
rRNAs	2,601	33.218	20.338	26.375	20.069	53.556	46.444	0.240	-0.136
D-loop	1,178	28.523	25.891	31.070	14.516	54.414	45.586	0.048	-0.363

complete mitogenome of WXG exceeded the C + G content (47.106%). Similarly, the content of A + T exceeded the content of C + G in the tRNA and rRNA genes, PCGs, and D-loop regions, indicating a pronounced A + T bias. Analysis of the nucleotide skewness in the mitogenome of WXG revealed that it has a positive AT skew value of 0.142 and a negative GC skew value of -0.360.

### PCGs and codon usage

The 13 PCGs in the mtDNA of WXG were found to have a total length of 11,402 bp, which accounts for 68.10% of the entire mitogenome (Table 1). Of the 13 PCGs, only

1 gene, *ND6*, was located on the L-strand, whereas the remaining 12 genes were located on the H-strand. Further analysis revealed that *ND5* is the longest PCG with a length of 1818 bp, while *ATP8* is the shortest PCG with a length of 168 bp (Fig. 2, Table 2). The A + T content (51.947%) of the PCGs was greater than the G + C content (48.053%), which was similar to the base makeup of the complete mitogenome of WXG. Analysis of the overall base composition of the 13 PCGs revealed that the contents of A, T, C, and G were 29.337%, 22.610%, 34.240%, and 13.813%, respectively. The PCGs exhibited a positive AT skew value of 0.129, and a negative GC

**Table 2** Annotation of the whole mtDNA of WXG

Gene	Strand	Location		Size (bp)	Intergenic length	Anticodon	codon	
		Start	End				Start	Stop
<i>trnF(GAA) tRNA</i>	H	1	68	68	-1	GAA	-	-
<i>12S rRNA</i>	H	68	1057	990	-1		-	-
<i>trnV(uac) tRNA</i>	H	1057	1127	71	0	TAC	-	-
<i>16S rRNA</i>	H	1128	2738	1611	0		-	-
<i>TrnL2(UAA) tRNA</i>	H	2739	2812	74	6	TAA	-	-
<i>ND1 CDS</i>	H	2819	3796	978	-2		ATG	AGG
<i>trnI(GAU) tRNA</i>	H	3795	3867	73	7	GAT	-	-
<i>trnQ(UUG) tRNA</i>	L	3875	3945	71	-1	TTG	-	-
<i>trnM(CAU) tRNA</i>	H	3945	4013	69	0	CAT	-	-
<i>ND2 CDS</i>	H	4014	5054	1041	-2		ATG	TAG
<i>trnW(UCA) tRNA</i>	H	5053	5125	73	5	TCA	-	-
<i>trnA(UGC) tRNA</i>	L	5131	5199	69	1	TGC	-	-
<i>trnN(GUU) tRNA</i>	L	5201	5273	73	3	GTT	-	-
<i>trnC(GCA) tRNA</i>	L	5277	5342	66	-1	GCA	-	-
<i>trnY(GUA) tRNA</i>	L	5342	5412	71	1	GTA	-	-
<i>COX1 CDS</i>	H	5414	6964	1551	-9		GTG	AGG
<i>TrnS2(UGA) tRNA</i>	L	6956	7028	73	2	TGA	-	-
<i>trnD(GUC) tRNA</i>	H	7031	7099	69	1	GTC	-	-
<i>COX2 CDS</i>	H	7101	7787	687	1		GTG	TAA
<i>trnK(UUU) tRNA</i>	H	7789	7857	69	1	TTT	-	-
<i>ATP8 CDS</i>	H	7859	8026	168	-10		ATG	TAA
<i>ATP6 CDS</i>	H	8017	8700	684	-1		ATG	TAA
<i>COX3 CDS</i>	H	8700	9483	784	0		ATG	T--
<i>trnG(UCC) tRNA</i>	H	9484	9552	69	0	TCC	-	-
<i>ND3 CDS</i>	H	9553	9904	351	1		ATG	TAG
<i>trnR(UCG) tRNA</i>	H	9906	9976	71	0	TCG	-	-
<i>ND4L CDS</i>	H	9977	10,273	297	-7		ATG	TAA
<i>ND4 CDS</i>	H	10,267	11,644	1378	0		ATG	T--
<i>trnH(GUG) tRNA</i>	H	11,645	11,713	69	0	GTG	-	-
<i>TrnS1(GCU) tRNA</i>	H	11,714	11,780	67	-1	GCT	-	-
<i>TrnL1(UAG) tRNA</i>	H	11,780	11,850	71	0	TAG	-	-
<i>ND5 CDS</i>	H	11,851	13,668	1818	7		GTG	AGA
<i>CYTB CDS</i>	H	13,676	14,818	1143	2		ATG	TAA
<i>trnT(UGU) tRNA</i>	H	14,821	14,888	68	8	TGT	-	-
<i>trnP(UGG) tRNA</i>	L	14,897	14,965	69	10	TGG	-	-
<i>ND6 CDS</i>	L	14,976	15,497	522	0		ATG	TAG
<i>trnE(UUC) tRNA</i>	L	15,498	15,565	68	0	TTC	-	-
<i>D-loop</i>	H	15,566	16,743	1178	0		-	-

T-- denotes an incomplete termination codon. Negative values indicate nucleotide overlap

skew value of  $-0.425$ , which reflected the preference for A and C bases over the T and G.

Of the 13 PCGs, *COX1*, *COX2*, and *ND5* utilized the GTG start codon, while the remaining 10 PCGs utilized the ATG start codon. The PCGs utilized the AGG, TAG, TAA, and AGA stop codons, of which the TAA stop codon was the most common (Table 2). *COX2*, *ATP8*, *ATP6*, *ND4L*, and *CYTB* utilized the TAA termination codon, *ND2*, *ND3*, and *ND6* employed the TAG stop codon, *ND1*, and *COX1* utilized the AGG termination codon, and *ND5* utilized the AGA stop codon. *ND4* and *COX3* employed the incomplete “T-” termination codon, where the “T-” sequence was situated at the 5' terminus of the adjacent gene [46]. With the exception of eight tRNA genes, namely, *trnQ* (UUG), *trnA* (UGC), *trnN* (GUU), *trnC* (GCA), *trnY* (GUA), *trnS* (UGA), *trnP* (UGG), and *trnE* (UUC), and one PCG (*ND6*), all the mtDNA genes were encoded on the H-strand. Gene overlapping and intergenic spacing were evident in the mtDNA of WXG, and a total of 11 gene overlaps and 15 intergenic spaces of length 1–10 bp were identified. Specifically, *ATP8* and *ATP6* shared 10 nucleotides, *COX1* and *trnS* (UGA) shared 9 nucleotides, *ND4L* and *ND4* shared 7 nucleotides, and *ND2* and *trnI* (GAU) shared 2 nucleotides.

#### Analysis of rRNA and tRNA genes and noncoding regions

The findings revealed that the mtDNA of WXG contains two rRNA genes, namely, 16S rRNA and 12S rRNA genes, of length 1611 bp and 990 bp, respectively. These genes are located on the H-strand between the *trnF* (GAA) and *trnL* (UAA) tRNA genes, and are separated by the *trnV* (UAC) gene (Fig. 2). These findings are consistent with previous observations reported in the majority of vertebrate species. The contents of A + T and C + G in the rRNA genes were 53.556% and 46.444%, respectively, while the AT and GC skew values of the rRNA genes were positive (0.240) and negative ( $-0.136$ ), respectively (Table 1).

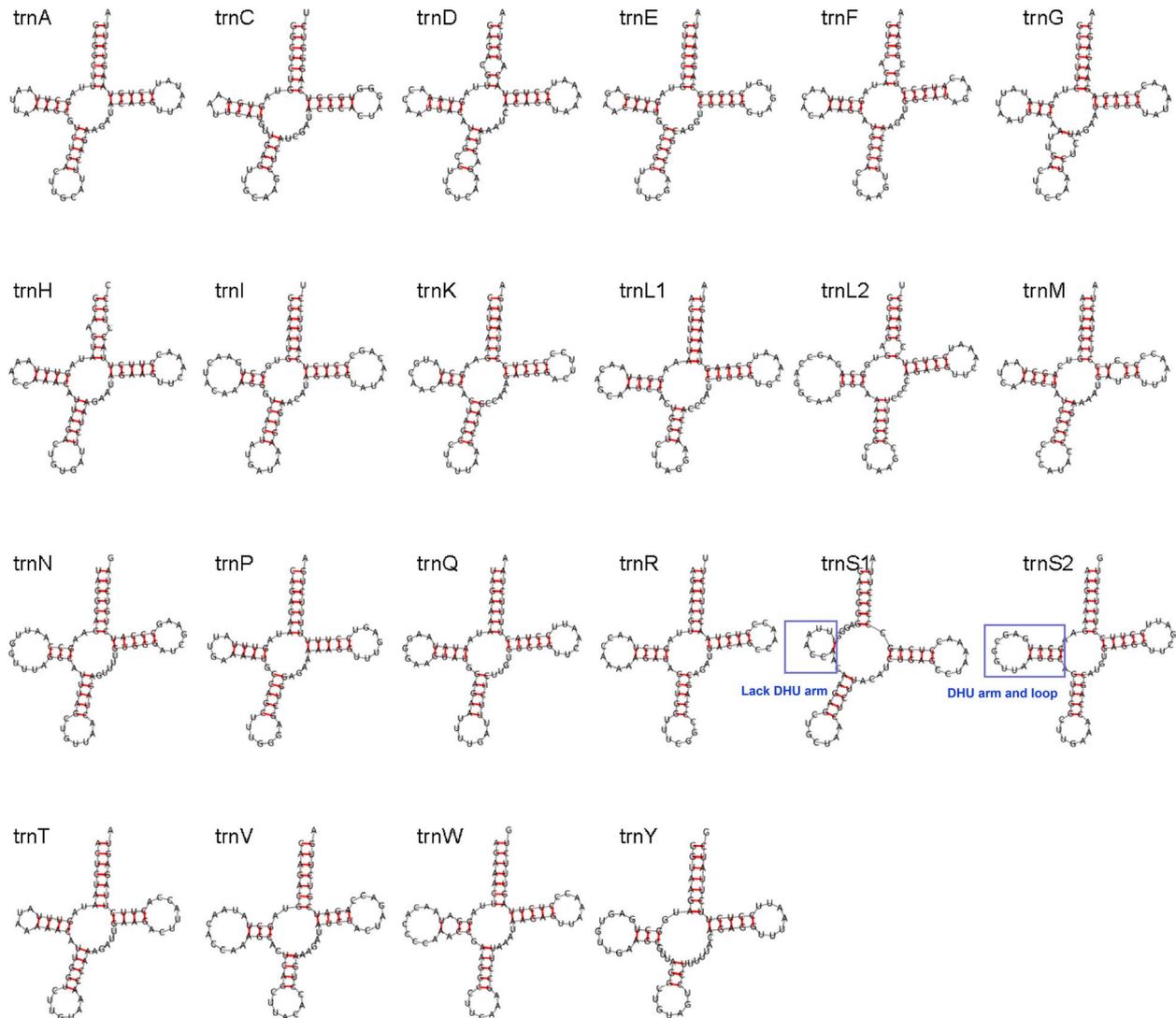
The mtDNA of WXG contains 22 typical tRNA genes (Fig. 2, Table 2), encompassing a total length of 1541 bp, and ranging from 66 bp (*trnC* (GCA)) to 74 bp (*trnL* (UAA)) in length. Of these 22 tRNA genes, 8 were located on the L-strand, while 14 genes were situated on the H-strand. Similar to the rRNA genes, the A + T content (58.014%) of the tRNA genes was significantly higher than the G + C content (41.986%), and the AT and GC skew values of the tRNA genes were positive (0.152) and negative ( $-0.193$ ), respectively (Table 1). The secondary structures of the 22 tRNA genes were further predicted, as illustrated in Fig. 3, which revealed that nearly all the tRNA genes displayed the classic cloverleaf structure. However, the *trnSI*-tRNA gene deviated from this

pattern due to the absence of the dihydrouridine (DHU) arm, resulting from unmatched base pairs.

The noncoding regions of vertebrate mtDNAs typically comprise a D-loop region along with several intergenic spacers. Specifically, in the mtDNA of WXG, the D-loop region spans 1,178 bp and is situated between the *trnE* (UUC) and *trnF* (GAA) tRNA genes on the H-strand (Fig. 2, Table 2). Analysis of the base composition of the D-loop region revealed that the contents of A, T, G, and C were 28.523%, 25.891%, 14.516%, and 31.070%, respectively. Similar to that of the rRNA and tRNA genes, the A + T content (54.414%) of the D-loop region was significantly higher than the G + C content (45.586%). The AT and GC skew values of the D-loop region were additionally calculated, which revealed that A was more predominant than T (AT skew = 0.048), while the frequency of G was lower than that of C (GC skew =  $-0.363$ ).

#### Analysis of codon bias

The analysis of codon bias revealed that mtDNAs had the highest content of Leu, with Thr and Ala following closely behind (Supplementary Table S4). The utilization of synonymous codons in the coding regions was further evaluated using the RSCU tool, with higher RSCU values indicating a stronger bias. As depicted in Fig. 4, all the amino acids were encoded by two or more synonymous codons. Specifically, Leu and Ser have six synonymous codons each, Ala, Arg, Gly, Pro, Ter, Thr, and Val have four each, Met has three, and the others have two. A total of 29–31 codons with RSCU values  $>1$  were classified as high-frequency codons, and the high RSCU values indicated a strong bias in the utilization of these codons in the mtDNAs of the 25 *Anser* spp. selected herein. With the exception of the UAG termination codon, all codons with RSCU  $>1$  preferentially ended in A (13–14 codons) or C (16 codons). In contrast, the majority of codons with RSCU  $<1$  terminated with U or G, with the exception of a few codons that ended in A or C (Supplementary Table S4). These findings are consistent with the results of previous studies and suggest that codons ending with A/C are preferentially used in vertebrates [9]. As depicted in Supplementary Table S4, the codon preferences were highly conserved across the different *Anser* spp. The CUA (Leu) and GUG (Met) codons exhibited the highest and lowest RSCU values, respectively, which represented the maximal and minimal values across the 25 *Anser* spp. Although the mtDNAs of the 25 *Anser* spp. selected herein all contained 61 different codons along with 4 termination codons, significant interspecies differences were observed, especially in terms of the respective codon usage frequencies.



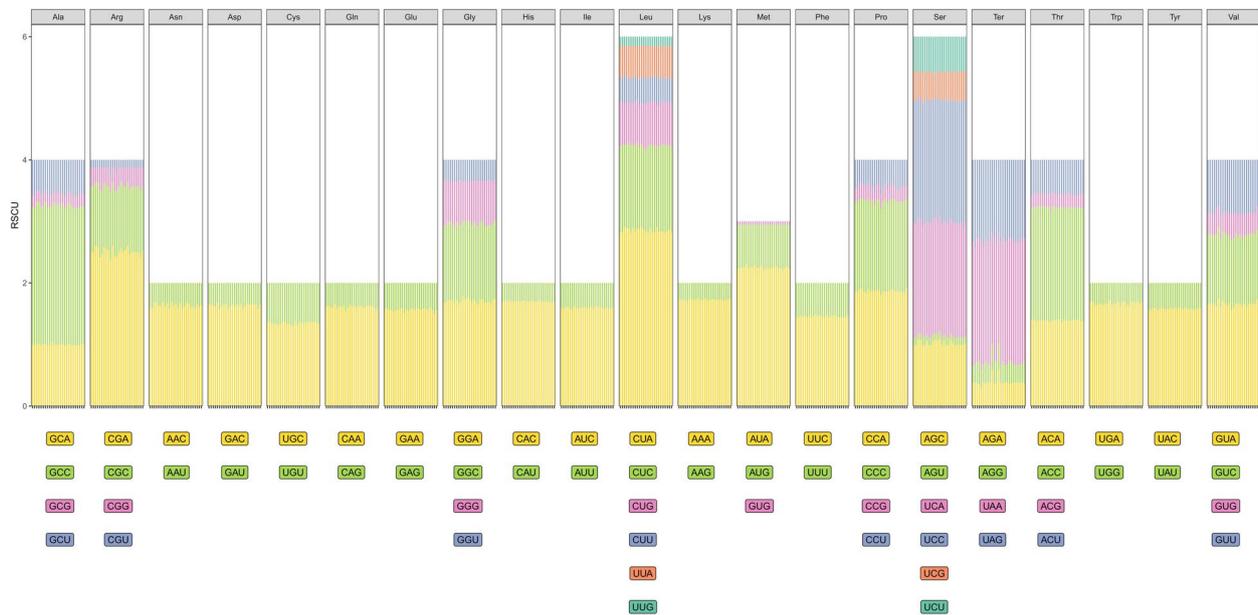
**Fig. 3** Secondary structural configuration of 22 tRNAs in WXG mtDNA

### Analyses of GC3-GC12, ENC-GC3, and PR2-bias plots

The effects of natural selection and mutational pressure on the codon usage bias in the mtDNAs of 25 closely related *Anser* spp. were assessed by analyzing the correlation between GC12 and GC3 via neutrality plot analysis (Fig. 5A). The findings revealed that the content of GC3 ranged from 41.070 to 57.580%, while the content of GC12 varied from 44.075 to 54.050% (Supplementary Table S5). The slopes of the regression lines (regression coefficients) ranged from  $-0.388$  for *Anser fabalis* (GenBank accession ID: HQ890328) to 0 for *Anser anser* (GenBank accession ID: OQ134125), which suggested a slight correlation between GC3 and GC12 in the mitochondrial codons of *Anser* spp. Additionally, the  $R^2$  value of the standard curve ranged from 0 to 0.290, indicating that the data points deviated markedly from the trend line. Statistical analysis indicated no significant correlation between the GC3

and GC12 contents ( $p > 0.05$ ), suggesting that natural selection played a pivotal role in shaping the codon bias observed in the mtDNA of *Anser* spp.

ENC-GC3 plot analysis was utilized to evaluate the impacts of natural selection and mutational pressure on codon usage bias. As depicted in Fig. 5B, the ENC values of the mtDNAs of the 25 *Anser* spp. ranged from 28.842 to 57.630 (Supplementary Table S6), which confirmed a pronounced codon usage preference in *Anser* spp. Analysis of the ENC-GC3 plot indicated that all observed values fell below the anticipated ENC-GC3 curve, with only a minor portion of the values approaching the expected curve. These findings indicated that the codon usage preference in the mtDNAs of these 25 *Anser* spp. was primarily determined by natural selection and other factors, while mutational pressure contributed only partially to the selection pressure [38].



**Fig. 4** RSCU analysis in mitochondrial PCGs of 25 *Anser* spp

The randomness of mutation ensures that there is an equal probability of GC3 being A/T or C/G when under the sole influence of mutation pressure. However, the frequency of usage of A/T or G/C bases becomes unequal under natural selection pressure [47]. As depicted in Fig. 5C, the PR2 plot was centered at 0.5 and divided into 4 regions. The PCGs in the mtDNAs of the 25 *Anser* spp. selected herein were unevenly distributed in the 4 regions, and nearly all the genes were distributed far from the center. This distribution pattern suggests a bias in the utilization of bases in the third codon, and that natural selection was the primary factor that influenced the codon bias. Furthermore, the majority of the genes were distributed in the lower right corner, while only a few genes were distributed in the upper left corner. This indicated that the frequency of the third base is: A > T and C > G, which suggests a preference for A/C bases in GC3.

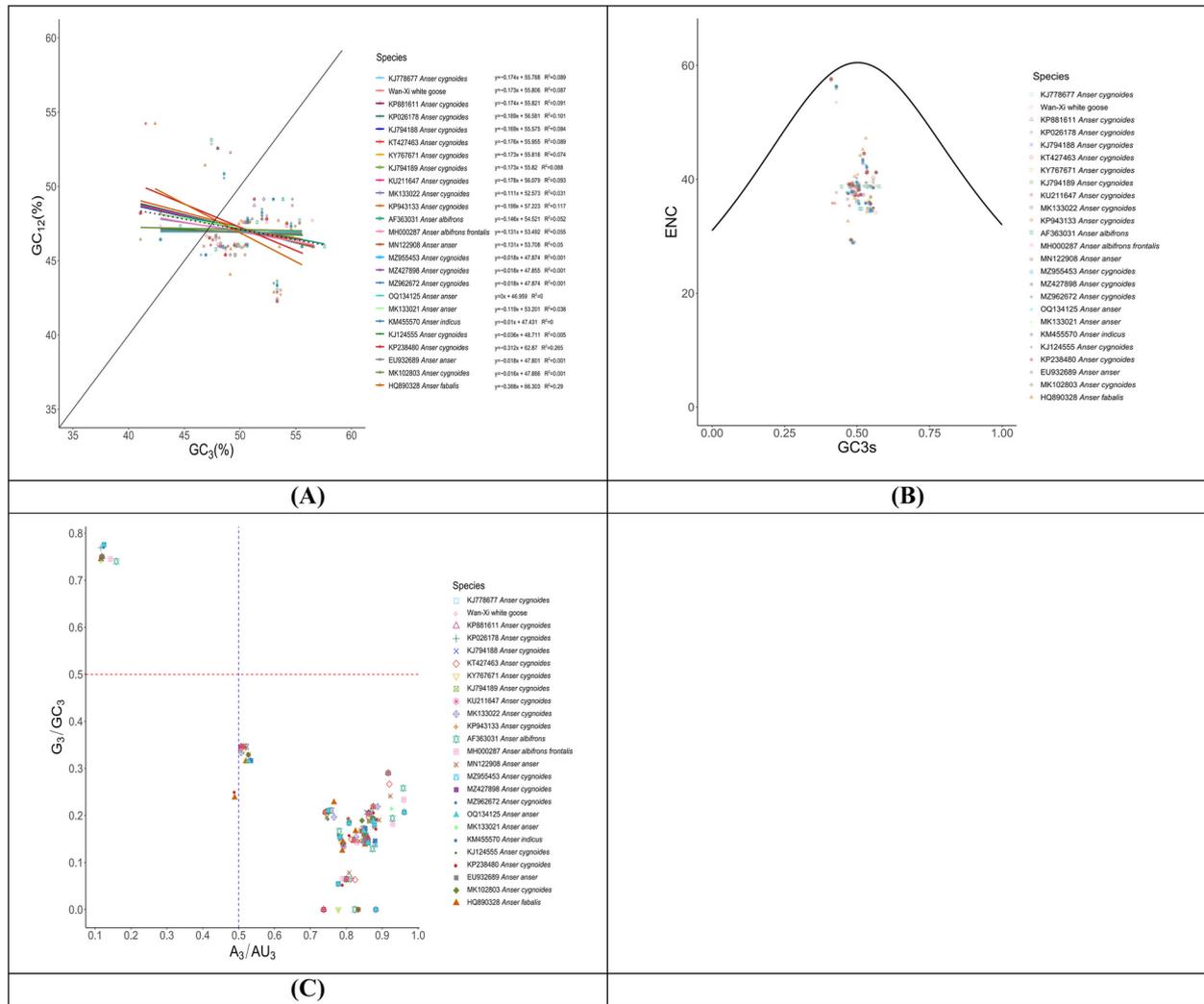
#### Analysis of Pi values

The mutation hotspots in the mtDNA of the 25 *Anser* spp. were identified by calculating the values of Pi using the DnaSP software, version 6.12.03. The top 10 regions with relatively high Pi values were labeled according to their locations within the mitogenomes and the positions of the genes (Supplementary Table S7). As depicted in Fig. 6A, the Pi values ranged from 0 to 0.04603, and 10 regions with Pi values greater than 0.017 were identified, including *12S rRNA*, *16S rRNA*, *16S rRNA\_ND1*, *ND1*, *ND2*, *COX2*, *ATP6*, *ND5*, *CYTb*, and *D-loop*. The results of the Pi analysis indicated that the Pi values varied markedly across the different regions of the mitogenomes. The *D-loop* region (Pi = 0.04603) exhibited the

highest level of polymorphism, followed by *ATP6* (Pi = 0.02668), *12S rRNA* (Pi = 0.02429), *ND1* (Pi = 0.02382), *16S rRNA\_ND1* (Pi = 0.02169), *COX2* (Pi = 0.0206), and *ND5* (Pi = 0.02032), while the *16S rRNA*, *ATP8*, *COX1*, *COX3*, *ND2*, *ND4*, *ND6*, *ND4L*, and *CYTb* regions exhibited low sequence variability (Pi < 0.02).

#### Effects of selection pressure on mitochondrial PCGs

The ratio of Ka/Ks is a reliable indicator of selection pressure. To investigate the effects of selection pressure on the mitochondrial PCGs, a total of 13 shared PCGs were used to evaluate the Ka/Ks ratios of the mtDNAs of the 25 *Anser* spp. selected herein. The density distribution diagrams of Ka, Ks, and Ka/Ks are provided in Fig. 6B. Analysis of the Ka, Ks, and Ka/Ks values of the homologous mitochondrial PCG pairs across the 25 *Anser* spp. revealed that the values of Ka and Ks were less than 0.35 and 1.0, respectively, and that 96.54% of the Ka/Ks values were distributed in the Ka/Ks < 1 region (Supplementary Table S8). These findings suggested that nearly all the PCGs underwent purifying selection during the evolution of the 25 *Anser* spp. and within the mtDNAs of the respective species, while only a few genes experienced neutral or positive selection, as indicated by Ka/Ks ratios (Ka/Ks < 1, purifying selection; Ka/Ks = 1, neutral evolution; Ka/Ks > 1, positive selection [48]). The findings further revealed that 103 of the homologous PCG pairs were under positive selection, and included 13 gene pairs with 1.5 > Ka/Ks > 1 and 90 gene pairs with 50.5 > Ka/Ks > 49 (Supplementary Table S8). The results of the functional prediction of these genes encoded proteins were subjected to statistical analysis, with the most frequently identified genes being *ND5*, *ND4*, *ND1*, *COX3*, *COX2*,



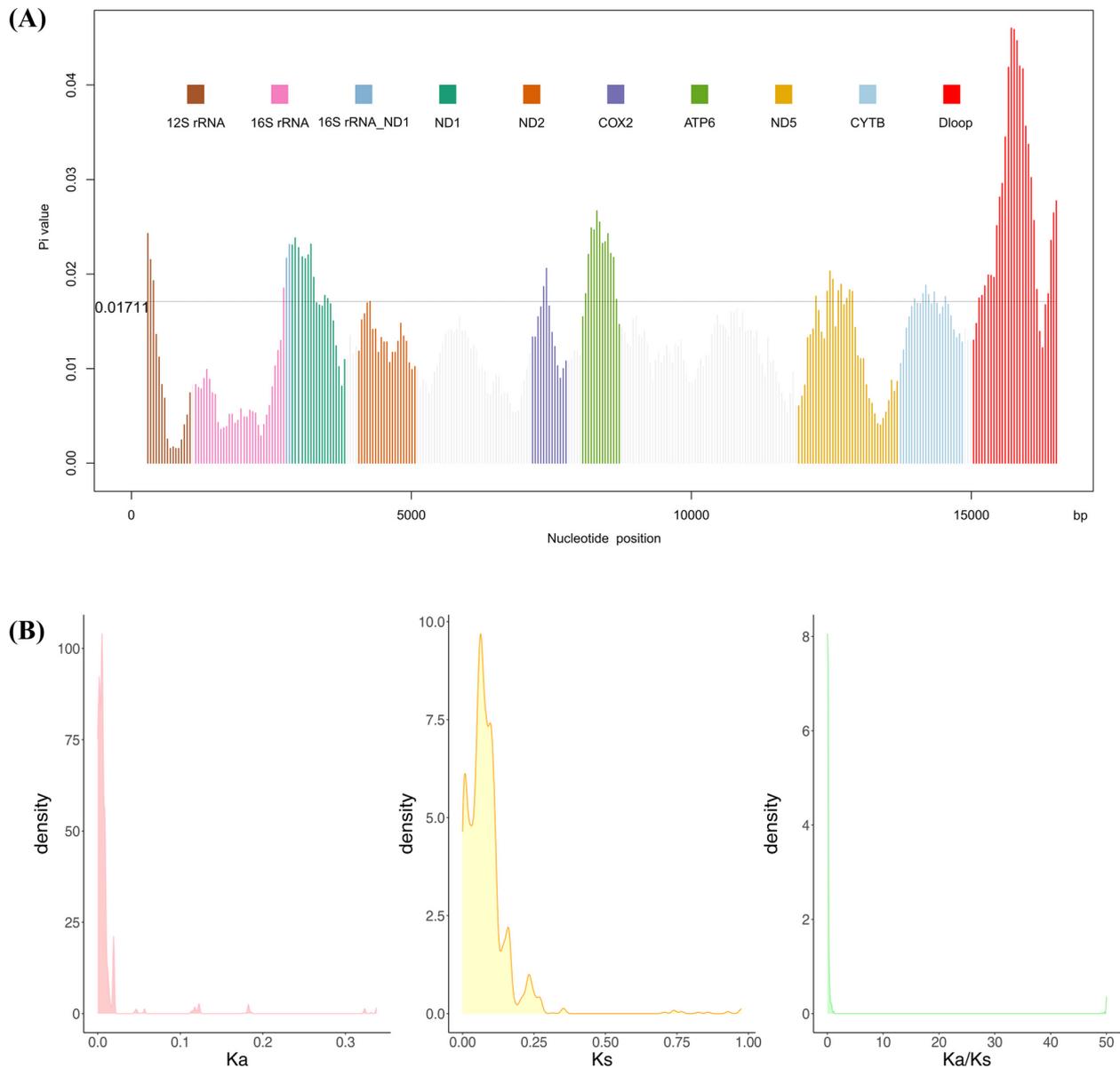
**Fig. 5** A The neutrality plot analysis in mtDNA of 25 *Anser* spp. B ENC-GC<sub>3</sub>s plot analysis in mtDNA of 25 *Anser* spp.. The solid line depicts the expected curve influenced solely by mutation pressure on codon usage bias. C PR2 plot analysis of the PCGs in mtDNA of 25 *Anser* spp

*ND6*, *ATP6*, *ATP8*, and *ND3* (Table 3). These results demonstrated that these genes are pivotal in the energy metabolism and environmental adaptation processes of *Anser* sp..

**Analysis of the phylogenetic relationships of WXG**

The ML-based phylogenetic tree constructed using mtDNA sequences is depicted in Fig. 7. The results of phylogenetic analyses revealed that the 50 species clustered into three major branches, namely, the *Anserinae*, *Anatinae*, and *Phasianinae* branches. The *Phasianinae* subfamily included 2 species that served as an outgroup. The results of phylogenetic analysis led to the identification of some close evolutionary relationships; however, 19 species under *Anatinae* and 29 species from *Anserinae* formed two distinct clusters, which indicated significant differences between their mtDNA sequences, and these findings were consistent

with the reports of previous research [4, 8, 16]. The 29 species under *Anserinae* could be further divided into two genera, namely, *Anser* and *Cygnus*, of which the species under *Anser* were found to be closely related to *Cygnus cygnus*. These taxa exhibited a sister relationship with the *Cygnus* species assemblage, and this finding was consistent with the results of previous morphology-based analyses [49]. Our target species, WXG, belongs to the *Anser* genus under the *Anserinae* subfamily, and is phylogenetically close to *Anser cygnoides*. In this study, WXG and the clade comprising KJ778677, KP881611, KP026178, KJ794188, KT427463, KY767671, KJ794189, KU211647, MK133022, and KP943133, all of which belong to *Anser cygnoides*, clustered into a single branch with well-supported values, suggesting that the 11 species formed a monophyletic group. Notably, other six *Anser cygnoides* formed a distinct branch phylogenetically distant from WXG,



**Fig. 6** **A**  $P_i$  of the mtDNA of 25 *Anser* spp.. A sliding window test with 500-bp windows and 50-bp steps annotated 10 mutation hotspots. The X-axis depicts genomic regions, while the Y-axis shows  $P_i$  values. The grey bars represent the threshold lines for the top 10 regions with relatively high  $P_i$  values. **B** Analysis of selection pressure. Density distribution of  $K_a$ ,  $K_s$ , and  $K_a/K_s$  for homologous PCG pairs of the mtDNA of 25 *Anser* spp

suggesting complex evolutionary relationships within the *Anser* genus that warrant further investigation.

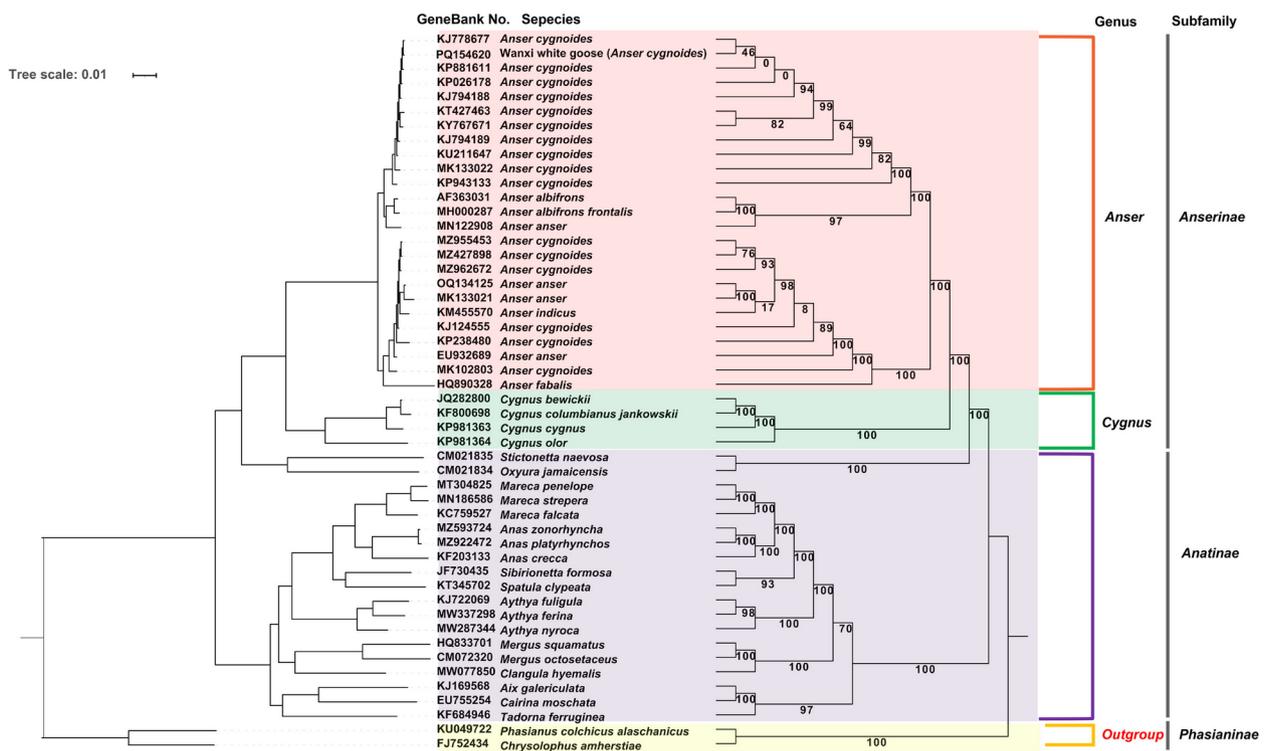
## Discussion

Mitochondria serve as the primary energy producers in eukaryotic cells, which enable them to perform essential activities. Furthermore, mtDNA sequences are extensively used as molecular markers for inferring the phylogenetic relationships among animals [50]. In our research, we examined and characterized the mtDNA composition of WXG. The findings revealed that the mtDNA of WXG totaled 16,743 bp in length, and the contents of the A, T, C, and G bases were 30.204%,

22.690%, 32.031%, and 15.075%, respectively, which followed the order: C > A > T > G. The mtDNA of WXG has a typical circular double-stranded structure and contains 22 tRNA genes, 2 rRNA genes, 13 PCGs, and one D-loop region. The base composition and structural features of the mtDNA of WXG were found to be similar to those of other breeds of geese [9]. The start and termination codons of the 13 mitochondrial PCGs were identical to those of the other *Anser* spp. selected in this study. The findings revealed that the majority of the PCGs utilize ATG as the start codon, while only a few codons employ GTG as the start codon. The PCGs utilized AGG, TAA, TAG, AGA, and T– as the termination codons, of which

**Table 3** The function of the homologous gene with positive selection effect in mtDNA PGCs of 25 *Anser* spp

Gene	Function	Number gene pairs
<i>ND5</i>	Involved in electron transport in the cellular respiratory chain and the assembly of a mitochondrial respiratory chain complex	22
<i>ND4</i>		20
<i>ND1</i>		17
<i>ND6</i>		7
<i>ND3</i>		1
<i>COX3</i>	Involved in electron transport in the cellular respiratory chain and REDOX-driven transmembrane transporter activity	17
<i>COX2</i>		12
<i>ATP6</i>	Involved in the energy production of mitochondria	6
<i>ATP8</i>		1



**Fig. 7** Phylogenetic tree based on mtDNA for 50 vertebrate spp.. Numbers beside the nodes represent bootstrap values

TAA was the most predominant. Additionally, the lengths of the mtDNAs of the majority of reported *Anser cygnoides* breeds range between 16,688 and 16,743 bp, with D-loop lengths varying from 1,173 bp to 1,182 bp and GC content between 47.08% and 47.33% (Supplementary Table S2). These observations demonstrate that there is a high degree of conservation among the mtDNAs of the majority of *Anser cygnoides* breeds.

The present study revealed that the mtDNA of WXG contained 22 typical tRNA genes, of which the majority exhibited typical cloverleaf structures, except for *trnS1*, which was devoid the DHU stem. This structural anomaly in *trnS1* has been detected in several other members of the *Anatidae* family, including *Aythya baeri* [8] and

Chaohu duck [51], and is a common feature of vertebrate tRNA genes [52]. It is considered that structural abnormalities in the encoded tRNAs and their functions are typically restored during post-transcriptional RNA editing [53].

By employing comparative genomic analysis, previous studies have demonstrated that the usage of synonymous codons varies across species, and that certain codons are more frequently utilized than others [40]. Additionally, analyses of the codon bias of various species can provide key insights into their genetic structures and evolutionary patterns [54]. In this study, comparison of the mtDNAs of 25 *Anser* spp. revealed that the RSCU values exhibited similar patterns across different *Anser*

spp. (Supplementary Table S4). The frequency of the Leu codon was highest among the mitochondrial PCGs of the 25 *Anser* spp., and there was a preference for codons ending with A/C bases. These findings are consistent with those of previous studies on several other animal species [16, 51, 55]. There were nearly no differences among the codon preferences of the 25 *Anser* spp., which suggests that these species are closely related, and further highlights the high degree of conservation in the mtDNA of *Anser* spp.

Mutation pressure and natural selection are two primary factors influencing codon usage bias [56]. In this study, analyses of the neutrality, ENC-GC3, and PR2-bias plots supported the role of natural selection in influencing the codon bias of the PCGs in the mtDNAs of 25 *Anser* spp. (Fig. 5). Altogether, these findings revealed that there were certain discrepancies in the codon usage patterns between and within *Anser* spp. However, a common trend observed herein was that the mitochondrial PCGs of all members of *Anserini* were subject to strong natural selection.

The results of Pi analysis led to the identification of ten highly variable regions, namely, *12S rRNA*, *16S rRNA*, *16S rRNA\_NDI*, *ND1*, *ND2*, *COX2*, *ATP6*, *ND5*, *CYTB*, and *D-loop*, of which the Pi values of the *D-loop*, *ATP6*, *12S rRNA*, *ND1*, *16S rRNA\_NDI*, *COX2*, and *ND5* regions were higher than 0.02. Although the five PCGs, *ATP6*, *12S rRNA*, *ND1*, *COX2*, and *ND5* had relatively high Pi values, they also exhibited a high degree of conservation. Therefore, the aforementioned *D-loop* and *16S rRNA\_NDI* regions are more suitable as candidate barcode labels than PCGs for the identification of *Anser* spp. [57].

The Ka/Ks ratio is an effective indicator of the evolutionary relationships and selection pressures between homogenous and heterogeneous species [58]. The present study demonstrated that 96.54% of the Ka/Ks values were distributed in the Ka/Ks < 1 region (Fig. 6B). This indicated that nearly all the PCGs underwent purifying selection during the evolution of these 25 *Anser* spp. and within the mtDNA of individual species. Notably, a total of 103 homologous PCG pairs in the mtDNA of the 25 *Anser* spp. selected herein had Ka/Ks values > 1 (Supplementary Table S8). The findings indicated that these genes exhibited a relatively high rate of non-synonymous mutations, with their dominant evolutionary patterns being strongly influenced by positive selection [59]. These genes were enumerated (Table 3), and the findings revealed that the majority, including *ND5*, *ND1*, and *COX2*, were located within the mutation hotspots identified by Pi analysis. These findings indicated that the results of Pi and Ka/Ks analyses were essentially consistent. The observed mutation hotspots in these genes may be associated with environmental adaptation and

genetic diversity in *Anser* spp. [60]. The genus *Anser* demonstrates a broad geographical distribution, inhabiting diverse ecological environments ranging from Arctic cold regions to subtropical warm zones. This wide distribution may have intensified evolutionary pressures on the mitochondrial genome. The mutation hotspots could provide these species with enhanced capacity for rapid environmental adaptation. For instance, certain mutations in cold environments may improve mitochondrial energy metabolism efficiency, thereby enhancing cold resistance [61]. Furthermore, frequent hybridization events may occur among *Anser* spp.. These mutation hotspots could facilitate genetic diversity in hybrid offspring, potentially conferring survival advantages in complex environments. For example, hybrid individuals inheriting mutation hotspot regions from both parents might acquire improved population-level disease resistance and environmental adaptability [62]. The observed differences in this study could be attributed to adaptive evolution to the environment, which conferred a survival or reproductive advantage to *Anser* spp. and enabled the nonsynonymous substitutions to be retained and fixed in this group.

The present study also performed phylogenetic analysis based on the mtDNA sequences of *Anatidae*, which encompassed 48 species across two subfamilies, and an outgroup comprising 2 species from the *Phasianinae* subfamily. The phylogenetic relationships among the taxa were evident from the resulting phylogenetic tree (Fig. 7). The clustering patterns of the three subfamilies were notably consistent with those reported in previous studies [5, 9, 17, 63], which further highlighted the utility of mtDNA in phylogenetic analyses. Analysis of the phylogenetic tree revealed that WXG clustered with 10 other goose breeds belonging to the *Anser cygnoides* clade. These 10 goose breeds are indigenous Chinese breeds, except for KU211647, which is native to South Korea (Supplementary Table S3). The findings strongly supported that the domestic geese exhibited a sister relationship with the *Cygnus* species assemblage. However, it remains to be determined whether the evolutionary differences among the indigenous breeds of Chinese geese are attributable to factors such as habitat selection and geographical isolation. Interestingly, we observed that six other *Anser cygnoides* individuals formed a distinct phylogenetic branch distant from WXG. We hypothesize that potential gene flow between *Anser cygnoides* and other *Anser* spp. may have contributed to this complex phylogenetic distribution pattern. Previous studies have demonstrated significant genetic diversity within *Anser* spp., with documented cases of natural interspecific hybridization (both intra-genus and with other *Anatidae* species). Such hybridization

events may have influenced phylogenetic relationships, resulting in anomalous clustering patterns of certain individuals in the evolutionary tree. Research suggests that reproductive isolation mechanisms in waterfowl evolve relatively slowly [62, 64, 65]. Therefore, the complex phylogenetic distribution observed in *Anser cygnoides* may represent genetic legacies of these historical hybridization events. Future studies incorporating multi-omics approaches and experimental validations are warranted to elucidate the temporal dynamics of these introgressive hybridization events and their evolutionary consequences.

## Conclusions

In this study, the complete 16,743 bp-long mtDNA of WXG was sequenced, assembled, and annotated (GenBank accession number: PQ154620). The findings revealed that the mtDNA of WXG has a typical circular, double-stranded structure, and comprises 13 PCGs, 22 tRNA genes, 2 rRNA genes, and one D-loop region. Additionally, the A + T content in the PCGs, tRNA and rRNA genes, D-loop region, and the complete mtDNA was found to be markedly higher than the C + G content. The codon usage bias, nucleotide polymorphisms, and evolutionary selection pressure of the PCGs in the mitogenomes of WXG and 24 other *Anser* spp. were explored by comparative genomic analysis. The evolutionary status of WXG was further verified by phylogenetic analysis based on the comparative assessment of the mtDNA sequences of WXG and 49 other species. The results of phylogenetic analysis revealed that WXG belongs to the *Anser cygnoides* clade, and formed a monophyletic group with 10 other members of *Anser cygnoides*. The findings obtained herein enhance our understanding of the phylogenetic status of WXG and provide valuable insights into the composition of its mtDNA. The study serves as a key reference for the preservation of this unique goose species and offers crucial insights for future studies on the phylogenetic relationships and evolutionary patterns of the *Anserinae* family.

## Abbreviations

ATP6	ATP synthase subunit 6
COX2	Cytochrome C oxidase subunit 2
CYTB	Cytochrome b
DHU	Dihydrouridine
D-loops	Noncoding control regions
mtDNA	Mitochondrial DNA
PCGs	Protein-coding genes
ND1	NADH dehydrogenase subunit 1
Pi	Nucleotide Variability
PR2	Parity rule 2
rRNA	Ribosomal RNA
RSCU	Relative synonymous codon usage
tRNA	Transfer RNA
WXG	Wan-Xi white goose

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-025-01326-1>.

Supplementary Material 1: Table S1. Sequencing data statistics using BGISEQ-500 platform; Table S2. Accessions used for comparative analysis in this study; Table S3. Accessions used for phylogenetic analysis in this study; Table S4. RSCU result in the mtDNA of 25 *Anser* spp.; Table S5. The GC contents of three positions of codons in the mtDNA of 25 *Anser* spp.; Table S6. The ENC and GC3 s values of the PCGs in the mtDNA of 25 *Anser* spp.; Table S7. The nucleotide diversity of the mtDNA of 25 *Anser* spp.; Table S8. The values of Ka, Ks, and Ka/Ks for homologous gene pairs in the mitochondrial PCGs of 25 *Anser* spp.

## Acknowledgements

We are deeply grateful to the Anhui Provincial Department of Education and West Anhui University for funding the research project. In addition, we are grateful to the Wan-Xi white goose farm owners for giving access to the samples. We also thank the editor and anonymous reviewers for their constructive comments that help to improve the quality of this paper.

## Authors' contributions

Conceptualization, investigation, writing—original draft preparation and funding acquisition, L.X.; software, S.B.; validation, Y.Z.; formal analysis, N.C.; writing—review and editing, N.C., S.B., Y.Z., C.C. and L.X.; visualization, Y.Z. and S.B. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by Natural Science Research Key Project of Universities in Anhui Province (grant number No. KJ2021 A0956), Anhui Scientific Research and Innovation Team of Quality Evaluation and Improvement of Traditional Chinese Medicine (grant number No. 2022 AH010090), Anhui Provincial Quality Engineering Project of Higher Education (grant number No. 2022ZYBJ108), Platform Collaborative Innovation Project of Anhui Province (grant number No. 0041122066) and Teaching Research Quality Project of West Anhui University (grant number No. WXXY2022007).

## Data availability

All data generated or analyzed during this study are included in this published article. The nucleotide sequence of the mitochondrial genome of the Wan-Xi white goose has been deposited in GenBank under accession ID: PQ154620.

## Declarations

### Ethics approval and consent to participate

Ethical approval was granted from the Ethics Committee of Western Anhui University for studies involving animals. All animal experimentation procedures adhere to the Basel Declaration (<https://animalresearchtomorrow.org/en>) and in compliance with the Guide for the Care and Use of Laboratory Animals published by the Chinese Academy of Sciences, and written consent was obtained from the farm handlers prior to sampling.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 12 December 2024 / Accepted: 5 May 2025

Published online: 13 May 2025

## References

- Du Y, Chen X, Yang H, Sun L, Wei C, Yang W, Zhao Y, Liu Z, Geng Z. Expression of oocyte vitellogenesis receptor was regulated by C/EBP $\alpha$  in developing follicle of Wanxi White Goose. *Animals (Basel)*. 2022;12(7):874. <https://doi.org/10.3390/ani12070874>.

2. Kuzniacka J, Biesek J, Banaszak M, Adamski M. Evaluation of egg production in Italian white geese in their first year of reproduction. *Europ Poult Sci.* 2019;83:279. <https://doi.org/10.1399/eps.2019.279>.
3. 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>.
4. Gray MW. Mitochondrial evolution. *Cold Spring Harb Perspect Biol.* 2012;4:a011403. <https://doi.org/10.1101/cshperspect.a011403>.
5. Ren T, Liang S, Zhao A, He K. Analysis of the complete mitochondrial genome of the Zhedong White goose and characterization of NUMTs: Reveal domestication history of goose in China and Euro. *Gene.* 2016;577(1):75–81. <https://doi.org/10.1016/j.gene.2015.11.018>.
6. Wang D, Teng J, Ning C, Wang W, Liu S, Zhang Q, Tang H. Mitogenome-wide association study on body measurement traits of Wenshang Barred chickens. *Anim Biotechnol.* 2023;34(7):3154–61. <https://doi.org/10.1080/10495398.2022.2137035>.
7. Kanakachari M, Chatterjee R.N., Reddy M.R., Dange M, Bhattacharya T.K. Indian Red Jungle fowl reveals a genetic relationship with South East Asian Red Jungle fowl and Indian native chicken breeds as evidenced through whole mitochondrial genome sequences. *Front Genet.* 2023; 14:1083976. <https://doi.org/10.3389/fgene.2023.1083976>.
8. Subramanian S, Denver DR, Millar CD, Heupink T, Aschrafi A, Emslie SD, Baroni C, Lambert DM. High mitogenomic evolutionary rates and time dependency. *Trends Genet.* 2009;25(11):482–6. <https://doi.org/10.1016/j.tig.2009.09.005>.
9. Liu D, Zhou Y, Fei Y, Xie C, Hou S. Mitochondrial genome of the critically endangered Baer's pochard, *Aythya baeri*, and its phylogenetic relationship with other *Anatidae* species. *Sci Rep.* 2021;11(1):24302. <https://doi.org/10.1038/s41598-021-03868-7>.
10. Clayton DA. Mitochondrial DNA replication: what we know. *IUBMB Life.* 2003;55:213–7. <https://doi.org/10.1080/1521654031000134824>.
11. Schmidt, Chloé, Garraway C.J. The conservation utility of mitochondrial genetic diversity in macrogenetic research. *Conservation Genetics.* 2021;22:323–327. doi: 10.1007/s10592-021-01333-6.
12. Li C, Wang J, Chen J, Schneider K, Veetil RK, Elmer KR, Zhao J. Native bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix* populations in the Pearl River are threatened by Yangtze River introductions as revealed by mitochondrial DNA. *J Fish Biol.* 2020;96:651–62. <https://doi.org/10.1111/jfb.14253>.
13. Jang JE, Kim JH, Kang JH, Kang JH, Baek SY, Wang JH, Lee HG, Choi JK, Choi JS, Lee HJ. Genetic diversity and genetic structure of the endangered Manchurian trout, *Brachymystax lenok tsinlingensis*, at its southern range margin: conservation implications for future restoration. *Conserv Genet.* 2017;18:1023–36. <https://doi.org/10.1007/s10592-017-0953-7>.
14. Jain K, Panigrahi M, Nayak SS, Rajawat D, Sharma A, Sahoo SP, Bhushan B, Dutt T. The evolution of contemporary livestock species: Insights from mitochondrial genome. *Gene.* 2024;927: 148728. <https://doi.org/10.1016/j.gene.2024.148728>.
15. Wang X, Pei J, Xiong L, Bao P, Chu M, Ma X, La Y, Liang C, Yan P, Guo X. Genetic diversity, phylogeography, and maternal origin of *yak* (*Bos grunniens*). *BMC Genomics.* 2024;25(1):481. <https://doi.org/10.1186/s12864-024-10378-z>.
16. Mo R, Zhu D, Sun J, Yuan Q, Guo F, Duan Y. Molecular identification and phylogenetic analysis of the mitogenome in endangered giant nuthatch *Sitta magna* (*Passeriformes, Sittidae*). *Heliyon.* 2024;10(9): e30513. <https://doi.org/10.1016/j.heliyon.2024.e30513>.
17. Liu G, Zhou L, Zhang L, Luo Z, Xu W. The complete mitochondrial genome of bean goose (*Anser fabalis*) and implications for *Anseriformes* taxonomy. *PLoS ONE.* 2013;8:e63334. <https://doi.org/10.1371/journal.pone.0063334>.
18. Lin Q, Jiang GT, Cao R, Yun L, Li GJ, Dai QZ, Zhang SR, Hou DX, He X. Determination and analysis of the complete mitochondrial genome sequence of Wugangtong grey goose. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2016;27(2):1008–9. <https://doi.org/10.3109/19401736.2014.926527>.
19. Huang YX, Ren FJ, Bartlett CR, Wei YS, Qin DZ. Contribution to the mitogenome diversity in Delphacinae: Phylogenetic and ecological implications. *Genomics.* 2020;112(2):1363–70. <https://doi.org/10.1016/j.ygeno.2019.08.005>.
20. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res.* 1999;27:1767–80. <https://doi.org/10.1093/nar/27.8.1767>.
21. Kumazawa Y, Endo H. Mitochondrial genome of the Komodo dragon: efficient sequencing method with reptile-oriented primers and novel gene rearrangements. *DNA Res.* 2004;11:115–25. <https://doi.org/10.1093/dnares/11.2.115>.
22. Castro JA, Picornell A, Ramon M. Mitochondrial DNA: a tool for populational genetics studies. *Int. Microbiol.* 1998;1:327–332. doi: 10.2436/im.v1i4.36.
23. China Committee on Animal Genetic Resources. Animal genetic resources in China-poultry. China Agriculture Press: Beijing, China, 2011; pp. 15–16.
24. Wei C, Chen X, Peng J, Yu S, Chang P, Jin K, Geng Z. BMP4/SMAD8 signaling pathway regulated granular cell proliferation to promote follicle development in Wanxi white goose. *Poult Sci.* 2023;102(1): 102282. <https://doi.org/10.1016/j.psj.2022.102282>.
25. Chen X, Bai H, Li L, Zhang W, Jiang R, Geng Z. Follicle characteristics and follicle developmental related Wnt6 polymorphism in Chinese indigenous Wanxi-white goose. *Mol Biol Rep.* 2012;39(11):9843–8. <https://doi.org/10.1007/s11033-012-1850-2>.
26. Wang B, Liu Z, Chen X, Zhang C, Geng Z. Green cabbage supplementation influences the gene expression and fatty acid levels of adipose tissue in Chinese Wanxi White geese. *Anim Biosci.* 2023;36(10):1558–67. <https://doi.org/10.5713/ab.22.0345>.
27. Chen X, Liu X, Du Y, Wang B, Zhao N, Geng Z. Green forage and fattening duration differentially modulate cecal microbiome of Wanxi white geese. *PLoS ONE.* 2018;13(9): e0204210. <https://doi.org/10.1371/journal.pone.0204210>.
28. Zhai Q, Zhao L, Jia C.L., Zhou X.R., Luo S.J., Wen X.H., Lv D.H. Duplex SYBR Green I real-time RT-PCR for Simultaneous Detection of Goose Astrovirus Genotypes 1 and 2. *Pak Vet J.* 2023;43(4): 828-830. <https://doi.org/10.29261/pakvetj/2023.067>.
29. Sun J, Zhang S, He DQ, Chen SY, Duan ZY, Yao YG, Liu YP. Matrilineal genetic structure of domestic geese. *J Poult Sci.* 2014;51:130–7. <https://doi.org/10.2141/JPSA.0120152>.
30. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
31. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
32. Dierckxens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2017;45(4): e18. <https://doi.org/10.1093/nar/gkw955>.
33. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol.* 2019;1962:1–14. [https://doi.org/10.1007/978-1-4939-9173-0\\_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
34. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol.* 1995;41(3):353–8. <https://doi.org/10.1007/BF00186547>.
35. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE.* 2016;11(10): e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
36. Meade JC, Shah PH, Lushbaugh WB. *Trichomonas vaginalis*: analysis of codon usage. *Exp Parasitol.* 1997;87:73–4. <https://doi.org/10.1006/expr.1997.4185>.
37. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000;16:276–7. [https://doi.org/10.1016/s0168-9525\(00\)02024-2](https://doi.org/10.1016/s0168-9525(00)02024-2).
38. Chang H, Guo J, Li M, Gao Y, Wang S, Wang X, Liu Y. Comparative genome and phylogenetic analysis revealed the complex mitochondrial genome and phylogenetic position of *Conopomorpha sinensis* Bradley. *Sci Rep.* 2023;13(1):4989. <https://doi.org/10.1038/s41598-023-30570-7>.
39. Li Q, Luo Y, Sha A, Xiao W, Xiong Z, Chen X, He J, Peng L, Zou L. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. *Front Microbiol.* 2023;14:1134228. <https://doi.org/10.3389/fmicb.2023.1134228>.
40. Wu P, Xiao W, Luo Y, Xiong Z, Chen X, He J, Sha A, Gui M, Li Q. Comprehensive analysis of codon bias in 13 Ganoderma mitochondrial genomes. *Front Microbiol.* 2023;14:1170790. <https://doi.org/10.3389/fmicb.2023.1170790>.
41. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol.* 2017;34(12):3299–302. <https://doi.org/10.1093/molbev/msx248>.
42. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics.* 2010;8(1):77–80. doi: 10.1016/S1672-0229(10)60008-3.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.

44. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol.* 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015>.
45. Nicholls TJ, Minczuk M. In D-loop: 40 years of mitochondrial 7S DNA. *Exp Gerontol.* 2014;56:175–81. <https://doi.org/10.1016/j.exger.2014.03.027>.
46. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature.* 1981;290(5806):457–65. <https://doi.org/10.1038/290457a0>.
47. Sueoka N. Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J Mol Evol.* 1999;49:49–62. <https://doi.org/10.1007/pl00006534>.
48. Zhu H, Lv J, Zhao L, Tong X, Zhou B, Zhang T, Guo W. Molecular evolution and phylogenetic analysis of genes related to cotton fibers development from wild and domesticated cotton species in *Gossypium*. *Mol Phylogenet Evol.* 2012;63(3):589–97. <https://doi.org/10.1016/j.ympev.2012.01.025>.
49. Livezey BC. A phylogenetic analysis of geese and swans (*Anseriformes: Anserinae*), including selected fossil species. *Syst Biol.* 1996;45:415–50. <https://doi.org/10.1093/sysbio/45.4.415>.
50. Zardoya R, Meyer A. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol Biol Evol.* 1996;13(7):933–42. <https://doi.org/10.1093/oxfordjournals.molbev.a025661>.
51. Jia Y, Qiu G, Cao C, Wang X, Jiang L, Zhang T, Geng Z, Jin S. Mitochondrial genome and phylogenetic analysis of Chaohu duck. *Gene.* 2023;851: 147018. <https://doi.org/10.1016/j.gene.2022.147018>.
52. Sérgio Luiz Pereira. Mitochondrial genome organization and vertebrate phylogenetics. *Genet Mol Biol.* 2000;23:745–52. <https://doi.org/10.1590/S1415-47572000000400008>.
53. Masta SE, Boore JL. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. *Mol Biol Evol.* 2004;21(5):893–902. <https://doi.org/10.1093/molbev/ms0096>.
54. Xiao G, Zhou J, Huo Z, Wu T, Li Y, Li Y, Wang Y, Wang M. The Shift in Synonymous Codon Usage Reveals Similar Genomic Variation during Domestication of Asian and African Rice. *Int J Mol Sci.* 2022;23(21):12860. <https://doi.org/10.3390/ijms232112860>.
55. Han S, Ding H, Peng H, Dai C, Zhang S, Yang J, Gao J, Kan X. *Sturnidae sensu lato Mitogenomics: Novel Insights into Codon Aversion, Selection, and Phylogeny.* *Animals (Basel).* 2024;14(19):2777. <https://doi.org/10.3390/ani14192777>.
56. Akashi H, Kliman RM, Eyre-Walker A. Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. *Genetica.* 1998;102–103(1–6):49–60. <https://doi.org/10.1023/A:1017078607465>.
57. Zhao Z, Zeng MY, Wu YW, Li JW, Zhou Z, Liu ZJ, Li MH. Characterization and Comparative Analysis of the Complete Plastomes of Five *Epidendrum (Epidendreae, Orchidaceae)* Species. *Int J Mol Sci.* 2023;24(19):14437. <https://doi.org/10.3390/ijms241914437>.
58. Li X, Huang Y, Lei F. Comparative mitochondrial genomics and phylogenetic relationships of the Crossoptilon species (*Phasianidae, Galliformes*). *BMC Genomics.* 2015;16(1):42. <https://doi.org/10.1186/s12864-015-1234-9>.
59. Ji L, Jia Z, Bai X. Comparative Analysis of the Mitochondrial Genomes of Three Species of *Yangiella (Hemiptera: Aradidae)* and the Phylogenetic Implications of *Aradidae*. *Insects.* 2024;15(7):533. <https://doi.org/10.3390/insects15070533>.
60. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics.* 2008;9:119. <https://doi.org/10.1186/1471-2164-9-119>.
61. Mahalingam S, Cheviron ZA, Storz JF, McClelland GB, Scott GR. Chronic cold exposure induces mitochondrial plasticity in deer mice native to high altitudes. *J Physiol.* 2020;598(23):5411–26. <https://doi.org/10.1113/JP280298>.
62. Ottenburghs J. Multispecies hybridization in birds. *Avian Res.* 2019;10:20. <https://doi.org/10.1186/s40657-019-0159-4>.
63. Lee MY, Jeon HS, Choi YS, Joo S, An J. Sequencing and analyzing complete mitochondrial genome of *Anser cygnoides (Anserini: Anserinae)*. *Mitochondrial DNA B Resour.* 2017;2(1):228–9. <https://doi.org/10.1080/23802359.2017.1310607>.
64. Ottenburghs J, van Hooft P, van Wieren SE, Ydenberg RC, Prins HH. Hybridization in geese: a review. *Front Zool.* 2016May;12(13):20. <https://doi.org/10.1186/s12983-016-0153-1>.
65. Ottenburghs J, Megens HJ, Kraus RHS, van Hooft P, van Wieren SE, Crooijmans RPMA, Ydenberg RC, Groenen MAM, Prins HHT. A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evol Biol.* 2017;17(1):201. <https://doi.org/10.1186/s12862-017-1048-2>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.