GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Complete mitochondrial genome of the Chinese Hwamei *Garrulax canorus* (Aves: Passeriformes): the first representative of the Leiothrichidae family with a duplicated control region

D.S. Chen[1]*, C.J. Qian[1]*, Q.Q. Ren[1,2], P. Wang[1], J. Yuan[1], L. Jiang[1], D. Bi[1], Q. Zhang[1], Y. Wang[1] and X.Z. Kan[1,2]

[1]Provincial Key Laboratory of the Conservation and Exploitation Research of Biological Resources in Anhui, College of Life Sciences, Anhui Normal University, Wuhu, China
[2]The Institute of Bioinformatics, College of Life Sciences, Anhui Normal University, Wuhu, China

*These authors contributed equally to this study.
Corresponding author: X.Z. Kan
E-mail: xianzhao@ahnu.edu.cn

Genet. Mol. Res. 14 (3): 8964-8976 (2015)
Received January 11, 2015
Accepted May 8, 2015
Published August 7, 2015
DOI http://dx.doi.org/10.4238/2015.August.7.5

ABSTRACT. The Chinese Hwamei *Garrulax canorus*, a member of the family Leiothrichidae, is commonly found in central and southern China, northern Indochina, and on Hainan Island. In this study, we sequenced the complete mitochondrial genome of *G. canorus*. The circular mitochondrial genome is 17,785 bp in length and includes 13 protein-coding genes, 22 transfer RNA (tRNA) genes, and two ribosomal RNA genes. In addition, two copies of highly

similar putative control regions were observed in the mitochondrial genome. As found in other vertebrates, most of the genes are coded on the H-strand, except for one protein-coding gene (*nad6*; NADH dehydrogenase subunit 6) and eight tRNA genes ($tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Ser(UCN)}$, $tRNA^{Pro}$, and $tRNA^{Glu}$). All the protein-coding genes start with ATG, with the exception of *cox1* (cytochrome oxidase subunit 1), which starts with GTG. All tRNA genes have the potential to fold into the typical clover-leaf structure. Conserved sequences in three domains were observed in the two putative control regions. These results provide basic information for future phylogenetic analyses among species of the order Passeriformes.

Key words: Chinese Hwamei; *Garrulax canorus*; Passeriformes; Mitochondrial genome; Leiothrichidae

INTRODUCTION

The order Passeriformes is the largest order or birds and is divided into 124 families, occupying approximately 60% of all existent avian species (Clements et al., 2011). The monophyly of Passeriformes is strongly supported by morphological and molecular data (Raikow, 1982; Sibley and Ahlquist, 1990; Johansson et al., 2001; Cracraft et al., 2004). According to DNA-DNA hybridizations (Sibley and Ahlquist, 1990), the order Passeriformes is divided into two groups, the Oscines and the Suboscines. Recent molecular data suggest that the New Zealand wrens constitute the sister-group to all other passerines (Barker et al., 2002; Ericson et al., 2002; Barker, 2004). The oscine passerines consist of two groups, named Corvida and Passerida, and the latter is divided into three superfamilies: Muscicapoidea, Sylvioidea, and Passeroidea (Sibley and Ahlquist, 1990), with Muscicapoidea basal relative to the other two clades (Nabholz et al., 2010). *Garrulax canorus*, commonly known as the Chinese Hwamei, is a member of the Sylvioidea and is located in the family Leiothrichidae (Clements et al., 2011; Gill and Donsker, 2012). It is widely distributed in central and southern China, northern Indochina, and on Hainan Island (MacKinnon and Phillipps, 2000). The name "Hwamei" comes from Chinese and means "painted eyebrow", referring to the distinctive marking around the bird's eyes. The species is a popular cage-bird because of its attractive song (Li et al., 2006).

A typical mitochondrial genome contains 13 protein-coding genes (PCGs), 2 ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes, ranging from 15 to 20 kb in size, and is double stranded (Kan et al., 2010a). The superiority of the mitochondrial genome, such as its small size and fast evolutionary rate, has led to it becoming ever more popular in phylogenetic, phylogeography, and evolutionary studies among animals (Ballard and Whitlock, 2004; Chesser et al., 2010; Kan et al., 2010b,c; Zhang et al., 2012). However, no complete mitochondrial sequence has been reported from members of the family Leiothrichidae. Here, we present the first complete mitochondrial genome of *G. canorus*, which provides further information for the future study of phylogenetic analyses among passerines, especially species belonging to Leiothrichidae.

MATERIAL AND METHODS

Sample collection and DNA extraction

Adult *G. canorus* [voucher code AHNU (Anhui Normal University), No. A0040] was collected from Wuhu (31°21'N, 118°22'E), southeast China, in 2009. A voucher specimen was deposited at the College of Life Sciences, Anhui Normal University, China. Total genomic DNA was extracted from muscle tissue using the standard phenol/chloroform method (Sambrook and Russell, 2001).

Polymerase chain reaction (PCR) amplification and sequencing

First, four long overlapping fragments were amplified using the long and accurate-PCR (LA-PCR) kit (Takara, Dalian, China) to minimize the possibility of obtaining nuclear copies of mitochondrial genes. The resulting amplification fragments (approximately 6 kb long) were used as templates for nested PCR amplification using 18 primer pairs (Table 1), which were designed based on the available mitochondrial genome sequences of passerines, and overlapped by at least 50 bp at both ends of the sequence. The LA-PCR and nested PCR were conducted as described by Kan et al. (2010a). The band with the expected size was cut from the gel and purified using the TIANgel MiDi Purification Kit (Tiangen Biotech Co., Ltd., Beijing, China) and then cloned using a pUCm-T vector kit (Bio Basic Inc., Canada). All expected clones were sequenced on an ABI-PRISM 3730 sequencer using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with the corresponding primers (Table 2).

Sequence analysis

DNA sequences were analyzed using the BioEdit 7.1.3 software (Hall, 1999). The ContigExpress program (a component of Vector NTI Suite 6.0) was used to assemble the contig. The boundaries of rRNA genes and PCGs were identified using DOGMA (Wyman et al., 2004), initially with the default settings, and then refined by alignment with mitochondrial genomes of other species of Passeriformes (Table 2). Most tRNA genes were identified using tRNAscan-SE 1.21 (Schattner et al., 2005) using the 'cove only' search mode, with the vertebrate mitochondrial genetic code and 'mito/chloroplast' source. A number of tRNA genes not identified using tRNAscan-SE 1.21 were identified by proposed secondary structures and anti-codons (Zhang et al., 2012). The gene map of the complete mitochondrial genome of *G. canorus* was initially generated by OGDRAW (Lohse et al., 2007) and then modified manually.

RESULTS AND DISCUSSION

Genome organization and base composition

The complete mitochondrial genome of *G. canorus* (GenBank accession No. JQ348398) was sequenced and was found to be 17,785 bp in length, and included 13 PCGs,

two rRNA genes (*srRNA* and *lrRNA*), 22 tRNA genes, and 2 putative control regions (D-loop) (Figure 1 and Table 3). The size of the mitochondrial genome of other passerine species ranges from 16,809 bp (*Pseudopodoces humilis*) to 18,154 bp (*Tachycineta cyaneoviridis*) (Table 4). Gene distributions within the mitogenome of *G. canorus* were similar to the gene distributions observed in the mitogenomes of 23 other passerines, with all the genes encoded on the H-strand, except for one protein-coding gene (*nad6*; NADH dehydrogenase subunit 6) and eight tRNA genes ($tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Ser(UCN)}$, $tRNA^{Pro}$, and $tRNA^{Glu}$).

Table 1. Primers used in the analysis of the *Garrulax canorus* mitochondrial genome.

No. of primer pair	Name	Sequences (5'-3')	Size (bp)
1	KLPASMTF14	CCCACACCATCAAACATCTC	20
	KLPASMTR3	ACTCTTTGTTGATGGCTGCT	20
2	KLPASMTF4	GAGGTGAAAAGCCAATCGAGC	21
	KLPASMTR7	ATGGATAGGACGTAGTGGAA	20
3	KLPASMTF8	GAATGGACGTAGACACCCG	20
	KLPASMTR10	GTTTTGCCTGGGTAGTATG	19
4	KLPASMTF11	CCTCCTACAATGCTAAAAAT	20
	KLPASMTR13	GAAGGCAGTTGCTATGAGG	19
5	QPASHMMTF1	ACGCTCAATGACTTTCCGC	19
	QPASHMMTR1	AGGAACGGTTGATAATGCTG	20
6	KLPASMTF2	ATCTCCAACTCCCAAAGCT	19
	QPASHMMTF2(2)	TCTCCGGGATCTAAAGCCT	19
	QPASHMMTR2(2)	AGGGTGGATGGATGGTGAG	19
	KLPASMTR2	GGTATCTAATCCCAGTTTG	19
7	KLPASMTF3	CCCACGGGTATTCAGCAGT	19
	KLPASMTR3	ACTCTTTGTTGATGGCTGCT	20
8	KLPASMTF4	GAGGTGAAAAGCCAATCGAGC	21
	KLPASMTR4	GCTAGGGAGAGGATTTGAACC	21
9	KLPASMTF5	AGTCCTACGTGATCTGAGTT	20
	KLPASMTR5	GGCCCGATAGCTTGTTTAG	19
10	KLPASMTF6	GATAAAGTGAACATAGAGGT	20
	KLPASMTR6	ATCGAAGCCCATCTGCCTA	19
11	KLPASMTF7	GCCTTCAAAGCCTTAAACAA	20
	KLPASMTR7	ATGGATAGGACGTAGTGGAA	20
12	KLPASMTF8	GAATGGACGTAGACACCCG	19
	QPASHMMTR8	GTGGCTGTAGAGATAAGTTG	20
13	PASMTF9	GGACGCCTAAACCAAACCTC	20
	PASMTR9	GCTTCTGTAATACTGTGGTG	20
14	KLPASMTF10	AGAACTAGGAGGACAATGAC	20
	KLPASMTR10	GTTTTGCCTGGGTAGTATG	19
15	KLPASMTF11	CCTCCTACAATGCTAAAAAT	20
	KLPASMTR11	CTTTCACTTGGATTTGCACC	20
16	KLPASMTF12	AAAACCTTCTTACCTGCCGA	20
	QPASHMMTR12	TGTGTAGACTGCGGTGAATG	20
17	KLPASMTF13	TCCTACACATCTCAACGCAC	20
	KLPASMTR13	GAAGGCAGTTGCTATGAGG	19
18	KLPASMTF14	CCCACACCATCAAACATCTC	20
	KLPASMTR14	GGCTTACAAGACCAATG	17

The overall A+T content of the H-strand is 52.1% (A = 29.96%, T = 23.15%, G = 14.81%, C = 33.08%), which is the lowest among known passerine mitogenomes (range from 52.1-57.7%; Table 4). The AT and GC skews, which can be calculated as (A-T) / (A+T) or (G-T) / (G+T) (Perna and Kocher, 1995), for the complete mitochondrial genome of *G. canorus* are 0.11 and -0.38, respectively, showing a strong bias against guanine. This phenomenon has also been reported in other birds (San Mauro et al., 2004; He et al., 2009; Kan et al., 2010a,b; Zhang et al., 2012).

Table 2. Avian species used for alignment of the mitochondrial genome in *Garrulax canorus*.

Family	Species	Accession No.	Reference
Acrocephalidae	*Acrocephalus scirpaceus*	NC-010227	(Singh et al., 2008)
Calyptomenidae	*Smithornis sharpei*	NC-000879	(Mindell et al., 1999)
Corvidae	*Corvus frugilegus*	NC-002069	(Härlid and Arnason, 1999)
Corvidae	*Podoces hendersoni*	NC-014879	(Ke et al., 2010)
Estrildidae	*Taeniopygia guttata*	NC-71619	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta bicolor*	JQ071614	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta cyaneoviridis*	JQ071617	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta euchrysea*	JQ071616	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta leucorrhoa*	JQ071621	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta meyeni*	JQ071622	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta stolzmanni*	JQ071618	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta thalassina*	JQ071615	(Cerasale et al., 2012)
Hirundinidae	*Progne chalybea*	JQ071623	(Cerasale et al., 2012)
Hirundinidae	*Tachycineta albiventer*	JQ071620	(Cerasale et al., 2012)
Leiothrichidae	*Garrulax canorus*	JQ348398	This study
Menuridae	*Menura novaehollandiae*	NC-007883	(Slack et al., 2007)
Paridae	*Pseudopodoces humilis*	NC-014341	(Yang et al., 2010)
Pycnonotidae	*Pycnonotus sinensis*	NC-013838	Unpublished[a]
Pycnonotidae	*Pycnonotus taivanus*	NC-013483	Unpublished[b]
Sylviidae	*Sylvia atricapilla*	NC-010228	(Singh et al., 2008)
Sylviidae	*Sylvia crassirostris*	NC-010229	(Singh et al., 2008)
Tyrannidae	*Cnemotriccus fuscatus*	NC-007975	(Slack et al., 2007)
Viduidae	*Vidua chalybeata*	NC-000880	(Mindell et al., 1999)

Unpublished data: [a]Chang HW, Lin ZH, Su YF, et al.; [b]Chang HW, Su YF, Yao CT, et al.

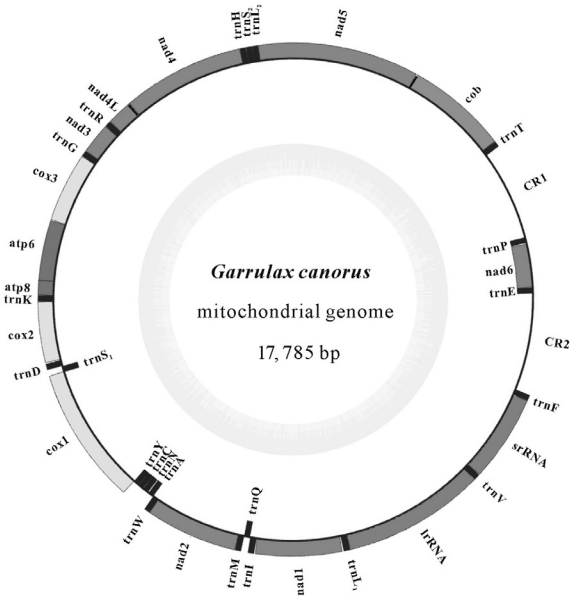

Figure 1. Gene map of the mitochondrial genome of *Garrulax canorus*. Genes encoded on the heavy or light strands are shown outside or inside the circular gene map, respectively. The inner ring displays the GC content. Twenty-two transfer RNA (tRNA) genes are designated by single-letter amino acid codes. Light gray boxes represent the protein-coding genes (PCGs) cytochrome oxidase subunits 1-3 (*cox1*, *cox2* and *cox3*), while dark gray boxes indicate the remaining PCGs. The figure was initially generated using OGDRAW and modified manually.

Table 3. Localization and features of genes in the mitochondrial genome of *Garrulax canorus*.

Gene/region	Strand	Position		Size (bp)		Codon		Anticodon	Intergenic
		From	To	Nucleotide	Amino acid	Start	Stop[a]		nucleotides[b]
tRNA^{Phe}	H	1	68	68				GAA	0
srRNA	H	69	1,047	979					0
tRNA^{Val}	H	1,048	1,117	70				TAC	0
lrRNA	H	1,118	2,718	1601					0
tRNA$^{Leu(UUR)}$	H	2,719	2,793	75				TAA	12
nad1	H	2,806	3,783	978	325	ATG	AGG		8
tRNA^{Ile}	H	3,792	3,864	73				GAT	5
tRNA^{Gln}	L	3,870	3,940	71				TTG	-1
tRNA^{Met}	H	3,940	4,008	69				CAT	0
nad2	H	4,009	5,048	1040	346	ATG	TA-		0
tRNA^{Trp}	H	5,049	5,118	70				TCA	1
tRNA^{Ala}	L	5,120	5,188	69				TGC	10
tRNA^{Asn}	L	5,199	5,271	73				GTT	1
tRNA^{Cys}	L	5,273	5,338	66				GCA	-1
tRNA^{Tyr}	L	5,338	5,408	71				GTA	1
cox1	H	5,410	6,960	1551	516	GTG	AGG		-9
tRNA$^{Ser(UCN)}$	L	6,952	7,024	73				TGA	4
tRNA^{Asp}	H	7,029	7,098	70				GTC	10
cox2	H	7,109	7,792	684	227	ATG	TAA		0
tRNA^{Lys}	H	7,793	7,862	70				TTT	1
atp8	H	7,864	8,031	168	55	ATG	TAA		-10
atp6	H	8,022	8,705	684	227	ATG	TAA		5
cox3	H	8,711	9,494	784	261	ATG	T-		0
tRNA^{Gly}	H	9,495	9,563	69				TCC	0
nad3	H	9,564	9,914	351	116	ATG	TAA		-1
tRNA^{Arg}	H	9,914	9,983	70				TCG	1
nad4L	H	9,985	10,281	297	98	ATG	TAA		-7
nad4	H	10,275	11,652	1378	459	ATG	T-		0
tRNA^{His}	H	11,653	11,722	70				GTG	0
tRNA$^{Ser(AGY)}$	H	11,723	11,788	66				GCT	-1
tRNA$^{Leu(CUN)}$	H	11,788	11,858	71				TAG	0
nad5	H	11,859	13,676	1818	605	ATG	AGA		8
cob	H	13,685	14,827	1143	380	ATG	TAA		1
tRNA^{Thr}	H	14,829	14,897	69				TGT	0
CR1	H	14,898	15,965	1068					0
tRNA^{Pro}	L	15,966	16,034	69				TGG	9
nad6	L	16,044	16,562	519	172	ATG	TAG		1
tRNA^{Glu}	L	16,564	16,635	72				TTC	0
CR2	H	16,636	17,785	1150					0

tRNA = transfer RNA; *rRNA* = ribosomal RNA; *lrRNA* = large ribosomal RNA; *srRNA* = small ribosomal RNA; *nad1-6* = NADH dehydrogenase subunits 1-6; *cox1-3* = subunits 1-3 of mitochondrial cytochrome oxidase; *cob* = mitochondrial cytochrome b; *atp6* and *atp8* = subunits 6 and 8 of ATP synthase; CR = control region; [a]"-" Indicates termination codons completed via polyadenylation; [b]Negative values represent overlapping nucleotides.

Protein-coding genes

The total length of the PCGs in the mitogenome of *G. canorus* is 11,400 bp, which is similar to most other species of Passeriformes (range from 11,391-11,403 bp). The longest PCG of *G. canorus* is *nad5* (1818 bp) and the shortest is ATP synthasesubunit 8 (*atp8*; 168 bp; Table 5). The AT composition for the mitogenome of *G. canorus* at the first, second, and third codon positions are 47.4, 58.4, and 47.3%, respectively (Table 4). All 13 PCGs initiate with ATG, with the exception of *cox1* (cytochrome oxidase subunit 1), which begins with GTG. Six types of stop codons are used by the coding genes, including AGG for *nad1* and *cox1*; AGA for *nad5*; TAG for *nad6*; TAA for *cox2*, *atp8*, *atp6*, *nad3*, *nad4L*, and *cob* (cytochrome oxi-

Table 4. Genomic characteristics of the mitochondrial DNA of species belonging to the order Passeriformes.

Species	Heavy-strand		Protein-coding					lrRNA gene		srRNA gene		tRNA genes		Control region	
	Length (bp)	AT%	Length (bp)	AT% (all)	AT% (1st)	AT% (2nd)	AT% (3rd)	Length (bp)	AT% (bp)	Length	AT% (bp)	Length	AT% (bp)	Length	AT%
G. canorus	17,788	52.1	11,400	51.0	47.4	58.3	47.3	1601	54.1	979	51.7	1547	56.4	2218	53.7
C. fuscatus	17,171	57.7	11,403	57.5	51.6	58.9	62.1	1580	55.7	981	53.9	1543	60.0	1618	61.3
C. frugilegus	16,931	55.7	11,406	55.1	47.4	59.2	58.8	1601	56.5	975	51.7	1546	59.5	1339	58.6
M. novaehollandiae	17,839	55.2	11,394	53.4	48.0	58.4	53.5	1602	55.7	973	51.7	1548	57.6	2280	63.7
P. sinensis	16,923	54.0	11,400	53.4	48.8	58.3	53.1	1600	55.1	978	51.9	1541	58.1	1341	54.7
P. taivanus	16,923	54.0	11,400	53.5	48.9	58.3	53.2	1601	55.0	977	51.9	1533	58.1	1341	54.5
A. scirpaceus	17,903	52.1	11,403	51.3	47.8	58.5	47.8	1603	54.8	977	50.9	1544	56.7	2328	51.8
S. atricapilla	17,937	55.5	11,400	55.3	49.6	58.8	57.5	1599	57.0	976	51.0	1546	58.0	2367	56.0
S. crassirostris	17,207	53.8	11,397	52.9	47.8	58.3	52.3	1598	55.8	975	51.2	1544	58.6	1640	55.9
T. guttata	16,853	54.1	11,400	53.4	47.7	58.8	53.8	1594	56.2	979	50.4	1545	57.9	1275	56.3
V. chalybeata	16,895	54.2	11,400	53.4	46.8	58.5	54.8	1600	56.3	978	51.7	1542	58.2	1295	57.0
S. sharpei	17,344	54.8	11,391	53.9	50.5	58.3	53.0	1601	53.7	976	52.7	1546	58.5	2055	59.2
P. humilis	16,809	52.9	11,400	52.1	48.0	58.3	50.1	1597	55.0	976	51.1	1548	56.5	1240	54.6
P. hendersoni	16,867	54.5	11,400	53.8	48.5	58.8	54.2	1603	55.8	978	52.2	1545	58.2	1290	56.3
T. albilinea	17,923	53.4	11,400	52.4	48.5	58.1	50.5	1607	54.5	974	50.8	1537	56.6	2346	56.6
T. bicolor	17,945	53.6	11,396	52.8	48.7	58.1	51.6	1603	54.2	973	51.0	1538	56.1	2377	56.3
T. cyaneoviridis	18,154	54,3	11,400	53.2	48.4	58.1	53.1	1604	54.0	973	51.5	1540	56,2	2580	59.4
T. euchrysea	17,929	53.8	11,400	52.8	48.4	58.3	51.7	1607	54.3	973	51.1	1542	57.2	2352	57.7
T. leucorrhoa	17,965	53.5	11,400	52.4	48.5	58.1	50.6	1603	54.6	975	51.0	1540	56.8	285	57.1
T. meyeni	18,012	53.4	11,400	52.4	48.4	58.1	50.7	1603	54.7	975	51.0	1541	56.7	2431	56.4
T. stolzmanni	17,932	53.4	11,400	52.4	48.3	58.1	50.9	1605	54.5	974	51.0	1538	56.9	2359	56.5
T. thalassina	18,118	53.8	11,400	52.8	48.4	58.0	51.8	1605	54.1	973	51.5	1540	56.8	2540	57.5
P. chalybea	18,030	53.2	11,400	52.1	48.3	58.1	49.8	1606	54.7	971	50.4	1539	56.0	2451	57.0
T. albiventer	17,923	53.4	11,400	52.4	48.4	58.1	51.0	1607	54.5	974	50.8	1537	56.6	2346	56.6

lrRNA = large ribosomal RNA; srRNA = small ribosomal RNA; tRNA = transfer RNA.

dase b); and the incomplete stop codon TA- or T- for *nad2*, *cox3*, and *nad4*. The *nad6* gene of the *G. canorus* mitogenome shows strong skews of T versus A (-0.57) and G versus C (0.58), while *nad4L* has a slight skew of T versus A. All other PCGs have a slight skew of A versus T (AT skew = 0.04-0.19), and a strong skew of C versus G (GC skew = -0.58-0.30) (Table 5). A common phenomenon found in other avian mitogenomes is gene overlapping, which was also observed in the *G. canorus* mitogenome (Table 3). Specifically, *cox1* shares 9 nucleotides with $tRNA^{Ser(UCN)}$, *atp8* and *atp6* share 10 nucleotides, *nad3* and $tRNA^{Arg}$ share 1 nucleotide, and *nad4L* and *nad4* share 7 nucleotides.

Table 5. Base composition for protein-coding genes found in the mitochondrial genome of *Garrulax canorus*.

Gene	Length (bp)	Proportion of nucleotides (%)					AT skew	GC skew
		A	C	G	T	A+T		
nad1	978	26.07	35.48	14.52	23.93	50.00	0.04	-0.42
nad2	1041	28.63	37.08	12.39	21.90	50.53	0.13	-0.50
cox1	1551	27.27	32.24	17.28	23.21	50.48	0.08	-0.30
cox2	684	28.95	33.33	15.50	22.22	51.17	0.13	-0.37
atp8	168	32.14	36.31	9.52	22.02	54.17	0.19	-0.58
atp6	684	28.22	38.01	10.09	23.68	51.90	0.09	-0.58
cox3	786	26.46	34.99	16.16	22.39	48.85	0.08	-0.37
nad3	351	28.49	35.04	12.54	23.93	52.42	0.09	-0.47
nad4L	297	23.57	36.03	14.48	25.93	49.49	-0.05	-0.43
nad4	1380	28.91	36.96	11.81	22.32	51.23	0.13	-0.52
nad5	1818	29.65	35.31	12.38	22.66	52.31	0.13	-0.48
cob	1143	27.91	35.00	13.04	24.06	51.97	0.07	-0.46
nad6	519	10.60	10.60	40.08	38.73	49.33	-0.57	0.58
Average		26.68	33.57	15.37	24.38	51.07	0.04	-0.38

nad1-6 = NADH dehydrogenase subunits 1-6; *cox1-3* = subunits 1-3 of mitochondrial cytochrome oxidase; *cob* = mitochondrial cytochrome b; *atp6* and *atp8* = subunits 6 and 8 of ATP synthase.

In this study, the pattern of codon usage in the *G. canorus* mitogenome was also investigated (Table 6). There are 3800 codons for all the 13 PCGs after the exclusion of stop codons. Of these, the amino acids that are used most frequently are Leu (17.63%), Ala (8.52%), Thr (8.42%), Ile (7.52%), and Ser (7.30%).

Table 6. Codon usage of 13 protein-coding genes in the mitochondrial genome of *Garrulax canorus*.

Amino acid	Codon	Number	Frequency (%)	Amino acid	Codon	Number	Frequency (%)	Amino acid	Codon	Number	Frequency (%)	Amino acid	Codon	Number	Frequency (%)
Phe	TTT	46	1.21	Ser	TCT	37	0.97	Tyr	TAT	24	0.63	Cys	TGT	4	0.11
	TTC	175	4.61		TCC	93	2.45		TAC	88	2.32		TGC	27	0.71
Leu	TTA	46	1.21		TCA	74	1.95	Stop	TAA	9	0.24	Trp	TGA	91	2.39
	TTG	20	0.53		TCG	11	0.29		TAG	1	0.03		TGG	16	0.42
	CTT	49	1.29	Pro	CCT	32	0.84	His	CAT	6	0.16	Arg	CGT	5	0.13
	CTC	191	5.03		CCC	76	2.00		CAC	100	2.63		CGC	16	0.42
	CTA	300	7.89		CCA	106	2.79	Gln	CAA	84	2.21		CGA	44	1.16
	CTG	64	1.68		CCG	12	0.32		CAG	13	0.34		CGG	9	0.24
Ile	ATT	56	1.47	Thr	ACT	43	1.13	Asn	AAT	11	0.29	Ser	AGT	4	0.11
	ATC	230	6.05		ACC	149	3.92		AAC	115	3.03		AGC	58	1.53
Met	ATA	110	2.89		ACA	122	3.21	Lys	AAA	79	2.08	Stop	AGA	1	0.03
	ATG	44	1.16		ACG	6	0.16		AAG	8	0.21		AGG	2	0.05
Val	GTT	24	0.63	Ala	GCT	50	1.32	Asp	GAT	9	0.24	Gly	GGT	19	0.50
	GTC	68	1.79		GCC	162	4.26		GAC	61	1.61		GGC	93	2.45
	GTA	74	1.95		GCA	102	2.68	Glu	GAA	66	1.74		GGA	72	1.89
	GTG	28	0.74		GCG	10	0.26		GAG	19	0.50		GGG	36	0.95

Ribosomal and transfer RNA genes

Two rRNA genes, small subunit rRNA (*srRNA*) and large subunit rRNA (*lrRNA*), were found in the *G. canorus* mitogenome, located between $tRNA^{Phe}$ and $tRNA^{Leu(UUR)}$ and separated by $tRNA^{Val}$. The lengths of *srRNA* and *lrRNA* are 979 and 1601 bp, respectively. The A+T content of *srRNA* and *lrRNA* are 51.7 and 54.1%, respectively, both of which are similar to the other known species of Passeriformes.

The complete mitochondrial sequence contains 22 tRNA genes, which are interspersed in the genome and range in size from 66 ($tRNA^{Cys}$ and $tRNA^{Ser(AGY)}$) to 75 ($tRNA^{Leu(UUR)}$) nt (Table 3). All the tRNA gene sequences have the potential to fold into typical cloverleaf secondary structures (Figure 2). $tRNA^{Cys}$ and $tRNA^{Ser(AGY)}$, which were not found by the tRNAscan-SE as reported in other species (Kan et al., 2010a; Zhang et al., 2012), were identified by comparing the sequence with other passerine species counterparts.

Non-coding regions

Two putative control regions (CR) were found in the *G. canorus* mitogenome [1068 bp (CR1) and 1150 bp (CR2) in length], located between the $tRNA^{Thr}$ and $tRNA^{Phe}$ genes and separated by $tRNA^{Pro}$, *nad6*, and $tRNA^{Glu}$. The total length of CRs observed in other passerine species varies between 1240 bp (*Pseudopodoces humilis*) and 2580 bp (*Tachycineta cyaneoviridis*), with the AT content ranging from 51.8% (*Acrocephalus scirpaceus*) to 63.7% (*Menura novaehollandiae*) (Table 4). The nucleotide composition of the *G. canorus* CR is A = 24.0%, T = 29.5%, C = 33.4%, and G = 13.2%, with a distinct bias against G. Based on the distribution of the conserved motifs in other avian CRs (Brown et al., 1986; Saccone et al., 1991; Randi and Lucchini, 1998; Zhang et al., 2012), the CR1 and CR2 of *G. canorus* can be separated into three domains: ETAS (extended termination-associated sequences) Domain I, Central Conserved Domain II, and CSB (conserved sequence block) Domain III (Figure 3). Domain I consists of parts A and B. In part A of CR1, ETAS1 and ETAS2 are found from position 57-121 and 116-163 nt, respectively, and overlap by 6 nt, with 65.2 and 56.3% similarity to the consensus mammalian ETAS1 and ETAS2 (Sbisa et al., 1997). Furthermore, a CSB1-like block in part B of CR1 has 55.6% similarity to the CSB1 in domain III (Figure 3). In CR1, four conserved sequence boxes were observed in Domain II, which were named boxes F, E, D, and C (Figure 3), and Poly (T) sequences were located just a few nucleotides downstream from the putative CSB1 in Domain III (Figure 3). In *G. canorus*, the CR1 and CR2 sequences were identical for over 1000 bp (only the first ~70 bp of both CRs and the last ~90 bp of CR2 were not identical). The high level of similarity of CR1 and CR2 sequences suggests either a recent and independent duplication of the CR or concerted evolution (Singh et al., 2008). In comparison with CR1, CR2 has bidirectional LSP/HSP-like promoters, a sequence similar to the mammalian LSP/HSP (light- and heavy-strand transcription promoters) (Randi and Lucchini, 1998) (Figure 3).

This is the first study to present the complete mitochondrial genome of the Chinese Hwamei *G. canorus* distributed in the Wuhu, Anhui Province, China. We report the genome organization, codon usage, and duplicated control region of the *G. canorus* mitochondrial DNA. These results provide basic information for future phylogenetic analyses among species of the order Passeriformes.

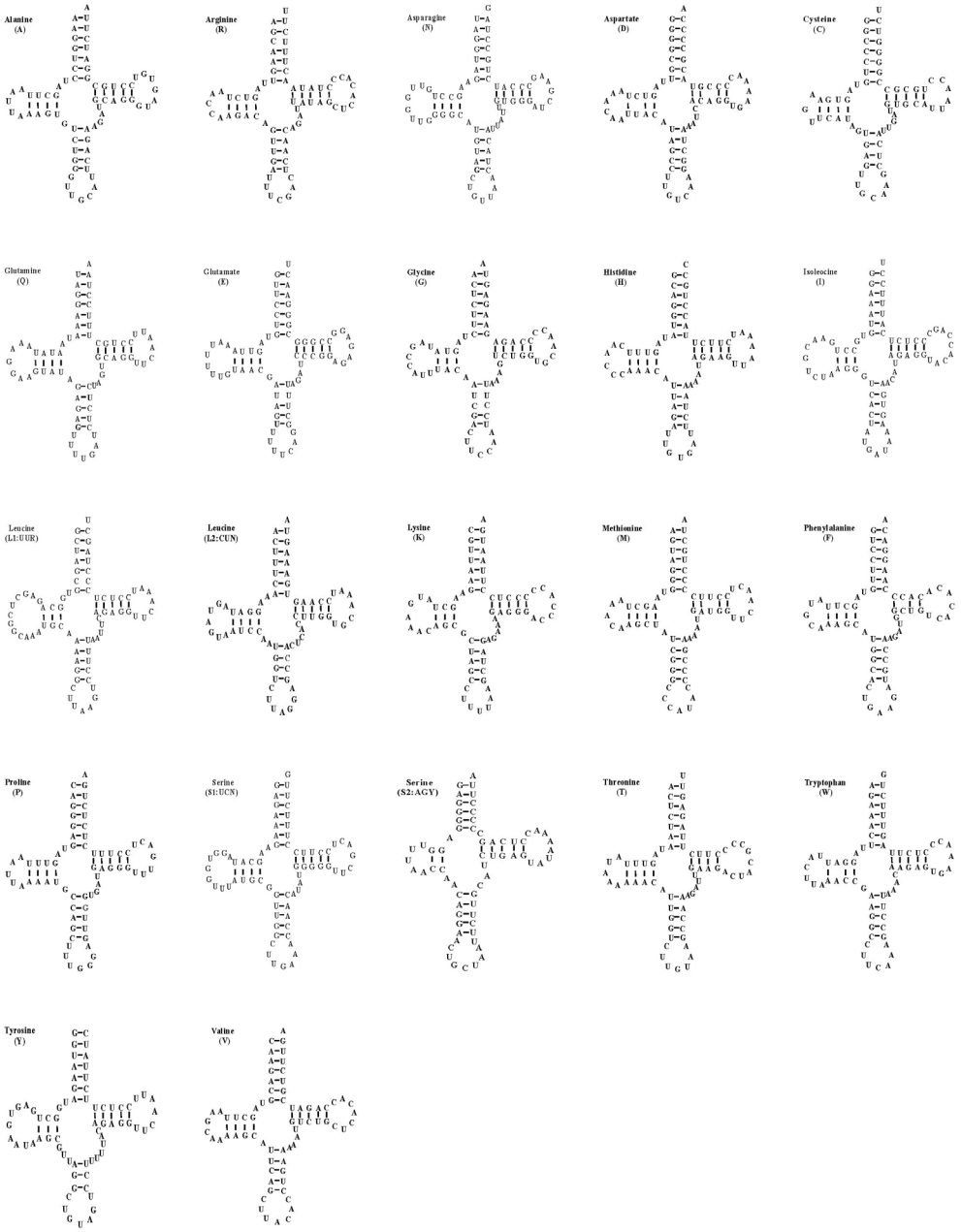

Figure 2. Inferred secondary structures of 22 transfer RNAs (tRNAs) found in the *Garrulax canorus* mitochondrial genome.

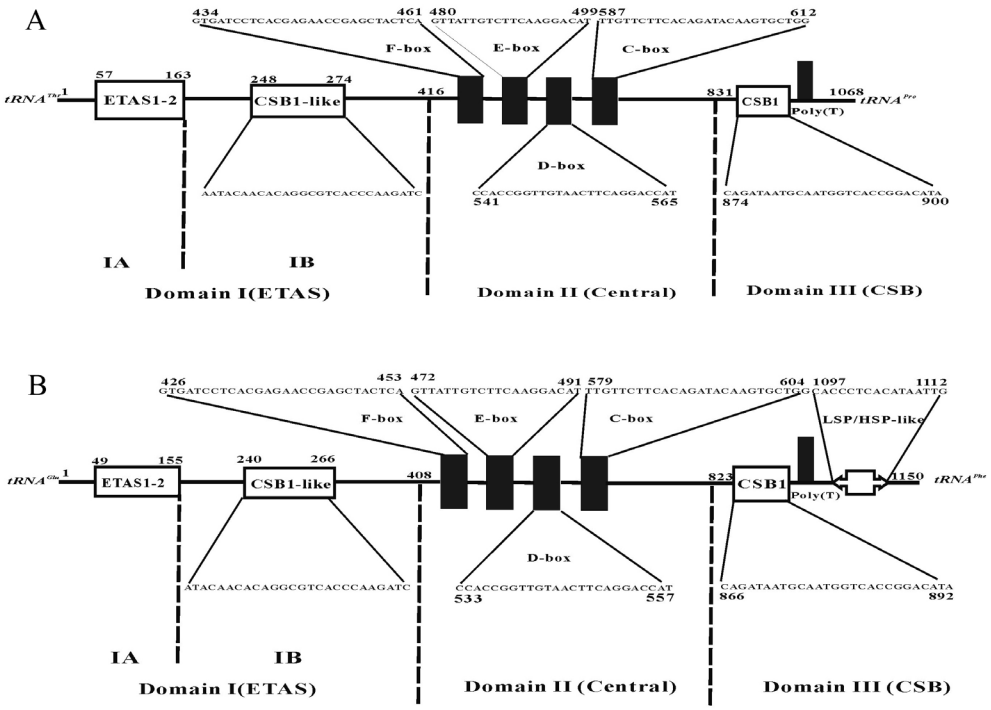

Figure 3. Schematic representation of the organization of the *Garrulax canorus* control region CR1 (**A**) and CR2 (**B**) ETAS, extended termination-associated sequences; F through C boxes, conserved sequence boxes in the central domain; *CSB*, conserved sequence block; CSB-like, a sequence similar to the CSB; LSP, light-strand transcription promoter; HSP, heavy-strand transcription promoter; LSP/HSP-like, a sequence similar to the LSP/HSP.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to thank Binghua Li, Jianke Yang, Zhichun Guo, and Xifeng Li for assistance, suggestions, and comments. Research supported by the National Natural Science Foundation of China (#31071266), the Key Laboratory of Biotic Environment and Ecology Safety in Anhui Province, the Key Program of Natural Science Foundation of the Anhui Higher Education Institutions (#KJ2013A127), and the Anhui Provincial Natural Science Foundation (#1308085MC38).

REFERENCES

Ballard JW and Whitlock MC (2004). The incomplete natural history of mitochondria. *Mol. Ecol.* 13: 729-744.

Barker FK (2004). Monophyly and relationships of wrens (Aves: Troglodytidae): a congruence analysis of heterogeneous mitochondrial and nuclear DNA sequence data. *Mol. Phylogenet. Evol.* 31: 486-504.

Barker FK, Barrowclough GF and Groth JG (2002). A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. *Proc. R. Soc. B-Biol. Sci.* 269: 295-308.

Brown GG, Gadaleta G, Pepe G, Saccone C, et al. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.* 192: 503-511.

Cerasale DJ, Dor R, Winkler DW and Lovette IJ (2012). Phylogeny of the *Tachycineta* genus of New World swallows: insights from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 63: 64-71.

Chesser RT, Yeung CKL, Yao CT, Tian XH, et al. (2010). Molecular phylogeny of the spoonbills (Aves: Threskiornithidae) based on mitochondrial DNA. *Zootaxa* 2603: 53-60.

Clements JF, Schulenberg TS, Lliff MJ, Sullivan BL, et al. (2011). The Clements checklist of birds of the world. Version 6.6. Cornell University Press, New York.

Cracraft J, Barker FK, Braun MJ, Harshman J, et al. (2004). Phylogenetic relationships among modern birds (Neornithes): towards an avian tree of life. In: Assembling the tree of life (Cracraft J and Donoghue MJ, eds.). Oxford University Press, New York, 468-489.

Ericson PGP, Christidis L, Cooper A, Irestedt M, et al. (2002). A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proc. R. Soc. B-Biol. Sci.* 269: 235-241.

Gill F and Donsker D (2012). IOC World Bird Names (v3.1). Available at [http://www.worldbirdnames.org].

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Härlid A and Arnason U (1999). Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters. *Proc. R. Soc. Lond. B-Biol. Sci.* 266: 305-309.

He L, Dai B, Zeng B, Zhang X, et al. (2009). The complete mitochondrial genome of the Sichuan Hill Partridge (*Arborophila rufipectus*) and a phylogenetic analysis with related species. *Gene* 435: 23-28.

Johansson US, Parsons TJ, Irestedt M and Ericson PGP (2001). Clades within the 'higher land birds', evaluated by nuclear DNA sequences. *J. Zool. Syst. Evol. Res.* 39: 37-52.

Kan XZ, Li XF, Lei ZP, Wang M, et al. (2010a). Complete mitochondrial genome of Cabot's tragopan, *Tragopan caboti* (Galliformes: Phasianidae). *Genet. Mol. Res.* 9: 1204-1216.

Kan XZ, Li XF, Zhang LQ, Chen L, et al. (2010b). Characterization of the complete mitochondrial genome of the Rock pigeon, *Columba livia* (Columbiformes: Columbidae). *Genet. Mol. Res.* 9: 1234-1249.

Kan XZ, Yang JK, Li XF, Chen L, et al. (2010c). Phylogeny of major lineages of galliform birds (Aves: Galliformes) based on complete mitochondrial genomes. *Genet. Mol. Res.* 9: 1625-1633.

Ke Y, Huang Y and Lei FM (2010). Sequencing and analysis of the complete mitochondrial genome of *Podoces hendersoni* (Ave, Corvidae). *Yi Chuan* 32: 951-960.

Li SH, Li JW, Han LX, Yao CT, et al. (2006). Species delimitation in the Hwamei *Garrulax canorus*. *Ibis* 148: 698-706.

Lohse M, Drechsel O and Bock R (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52: 267-274.

MacKinnon J and Phillipps K (2000). A Field Guide to the Birds of China: ornithology. Oxford University Press, New York.

Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, et al. (1999). Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol.* 48: 138-152.

Nabholz B, Jarvis ED and Ellegren H (2010). Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. *Mol. Phylogenet. Evol.* 57: 466-470.

Perna NT and Kocher TD (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* 41: 353-358.

Raikow RJ (1982). Monophyly of the Passeriformes: test of a phylogenetic hypothesis. *Auk* 99: 431-455.

Randi E and Lucchini V (1998). Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *J. Mol. Evol.* 47: 449-462.

Saccone C, Pesole G and Sbisa E (1991). The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J. Mol. Evol.* 33: 83-91.

Sambrook J and Russell DW (2001). Molecular Cloning: A Laboratory Manual. 3rd edn. Cold Spring Harbor Laboratory Press, New York.

San Mauro D, Garcis-Paris M and Zardoya R (2004). Phylogenetic relationships of discoglossid frogs (Amphibia:Anura:Discoglossidae) based on complete mitochondrial genomes and nuclear genes. *Gene* 343: 357-366.

Sbisà E, Tanzariello F, Reyes A, Pesole G, et al. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205: 125-140.

Schattner P, Brooks AN and Lowe TM (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33: W686-W689.

Sibley CG and Ahlquist JE (1990). Phylogeny and Classification of Birds: A study in molecular evolution. Yale University Press, New Haven.

Singh TR, Shneor O and Huchon D (2008). Bird mitochondrial gene order: insight from 3 warbler mitochondrial genomes. *Mol. Biol. Evol.* 25: 475-477.

Slack KE, Delsuc F, McLenachan PA, Arnason U, et al. (2007). Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol. Phylogenet. Evol.* 42: 1-13.

Wyman SK, Jansen RK and Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.

Yang C, Lei FM and Huang Y (2010). Sequencing and analysis of the complete mitochondrial genome of *Pseudopodoces humilis* (Aves, Paridae). *Dongwuxue Yanjiu* 31: 333-344.

Zhang L, Wang L, Gowda V, Wang M, et al. (2012). The mitochondrial genome of the Cinnamon Bittern, *Ixobrychuscinnamomeus* (Pelecaniformes: Ardeidae): sequence, structure and phylogenetic analysis. *Mol. Biol. Rep.* 39: 8315-8326.