

Nucleotide Sequence and Evolution of Coding and Noncoding Regions of a Quail Mitochondrial Genome

Paul Desjardins* and Réjean Morais

Département de biochimie, Faculté de médecine, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec H3C 3J7, Canada

Summary. Segments of the Japanese quail mitochondrial genome encompassing many tRNA and protein genes, the small and part of the large rRNA genes, and the control region have been cloned and sequenced. Analysis of the relative position of these genes confirmed that the tRNA^{Glu} and ND6 genes in galliform mitochondrial DNA are located immediately adjacent to the control region of the molecule instead of between the cytochrome b and ND5 genes as in other vertebrates. Japanese quail and chicken display another distinctive characteristic, that is, they both lack an equivalent to the light-strand replication origin found between the tRNA^{Cys} and tRNA^{Asn} genes in all vertebrate mitochondrial genomes sequenced thus far. Comparison of the protein-encoding genes revealed that a great proportion of the substitutions are silent and involve mainly transitions. This bias toward transitions also occurs in the tRNA and rRNA genes but is not observed in the control region where transversions account for many of the substitutions. Sequence alignment indicated that the two avian control regions evolve mainly through base substitutions but are also characterized by the occurrence of a 57-bp deletion/addition event at their 5' end. The overall sequence divergence between the two gallinaceous birds suggests that avian mitochondrial genomes evolve at a similar rate to other vertebrate mitochondrial DNAs.

Key words: *Coturnix japonica* — Mitochondrial DNA — Mitochondrial genes — Gene rearrangements — Nucleotide substitutions — Evolution

Introduction

Studies of the mitochondrial genome in closely related vertebrate species allow the inference of phylogenetic relationships, and also can reveal some important features of mitochondrial (mt) DNA evolution that are not observed when more distant species are compared. Vertebrate mtDNA has a high mutation rate (Brown et al. 1979) and, over evolutionary time, becomes rapidly saturated by multiple substitutions at the same nucleotide site (Brown et al. 1982).

In the course of studying relationships among avian mitochondrial genomes, we have cloned and sequenced the chicken (*Gallus gallus domesticus*) mitochondrial genome (Desjardins et al. 1989; Desjardins and Morais 1990). Sequence analysis revealed that chicken mtDNA displays two characteristics not seen in other vertebrate mtDNAs sequenced thus far, that is, a novel gene order and the absence of an equivalent to the light (L)-strand replication origin. To determine if the peculiar features observed in chicken mtDNA represent a more general characteristic of the avian mitochondrial genome, we have undertaken the sequence determination and analysis of selected portions of the mtDNA molecule in closely related birds of the order Galliformes. A previous study on mtDNA evolution in gallinaceous birds (Glaus et al. 1980) using restriction polymorphism data and the mathematical model of Nei and Li (1979) revealed 9–15% sequence difference between chicken, turkey, pheasant, Guinea fowl, and Japanese quail mtDNAs. The present study reports the nucleotide sequence of segments of the Japanese quail (*Coturnix japonica*) mitochondrial genome comprising many tRNA and

* Present address: Institute for Molecular Biology and Biotechnology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Offprint requests to: R. Morais

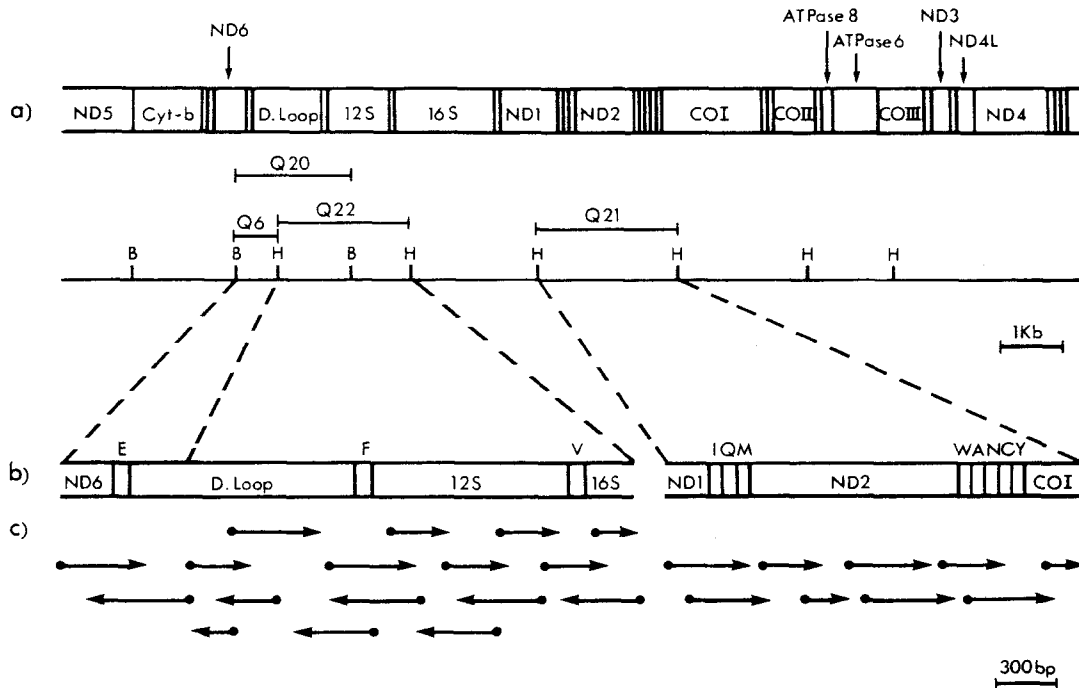


Fig. 1. a A linear representation of the chicken mtDNA molecule showing the localization of the genes for cytochrome oxidase subunits I, II, and III (COI, COII, and COIII), ATPase subunits 6 and 8 (ATPase 6 and 8), apocytochrome b (Cyt-b), NADH dehydrogenase subunits 1–6 (ND1, 2, 3, 4, 4L, 5, 6), the small (12S) and large (16S) ribosomal RNA subunits, 22 tRNAs (narrow boxes), and the control region (D-loop). Alignment is with the physical map of Japanese quail mtDNA and relative

position of clones pMtQ 6, 20, 21, and 22. B and H are restriction sites for BamHI and HindIII, respectively. b Japanese quail mtDNA sequenced regions and their gene content. The tRNA genes are indicated by the one-letter amino acid code. c Sequencing strategy. Arrows indicate the start, length, and direction of sequencing for the individual phage M13 subclones. Clone pMtQ 20 was used to sequence the nucleotides at the junction between pMtQ 6 and 22 using a synthetic oligonucleotide.

protein genes, the small and part of the large rRNA genes, as well as the complete control region.

Materials and Methods

Extraction and Cloning of mtDNA. Mitochondrial DNA was extracted from Japanese quail liver as described by Morais et al. (1988). Three mtDNA fragments totaling nearly 5.0 kb were generated by endonuclease digestion and cloned into M13 mp18/mp19 phages. These clones designated pMtQ 6, 21, and 22 (Fig. 1a) were further subcloned after exonuclease digestion of the 3' end with phage T4 DNA polymerase (IBI) according to the procedure of Dale et al. (1985).

DNA Sequencing. Sequencing reactions were performed by the dideoxynucleotide chain termination method (Sanger et al. 1977) using 35 S-dATP (1000 Ci/mmol; Amersham), the modified phage T7 DNA polymerase (Sequenase, USB) and the universal M13 forward primer. The sequencing strategy is depicted in Fig. 1c. Clones pMtQ 6 and 22, which encode the complete small rRNA subunit and the D-loop region, were sequenced in both orientations. Clone pMtQ 21 consists mainly of open reading frames and was sequenced at least twice using numerous overlapping subclones having the same orientation. The nucleotide sequence at the junction between pMtQ 6 and 22 was confirmed by priming a 2.0-kb BamHI fragment (pMtQ 20 in Fig. 1a) with a synthetic oligonucleotide. Computer programs used for sequence data treatment have been described (Desjardins and Morais 1990).

Results and Discussion

Sequence and Genomic Organization

Desjardins and Morais (1990) reported that chicken mtDNA has a novel gene order among vertebrates (Desjardins and Morais 1990), the contiguous tRNA^{Glu} and ND6 genes being located immediately upstream from the control region (Fig. 1a), instead of the contiguous tRNA^{Pro}, tRNA^{Thr}, and cytochrome b genes as in other vertebrate mitochondrial genomes (Anderson et al. 1981, 1982; Bibb et al. 1981; Roe et al. 1985; Gadaleta et al. 1989). Cross-hybridization of Japanese quail, turkey, pheasant, and Guinea fowl mtDNAs with selected portions of the chicken mitochondrial genome as probes has further suggested that this unusual gene order is a feature common to all gallinaceous birds (Desjardins and Morais 1990). This is confirmed here in Japanese quail by sequence analysis of clone pMtQ 6 (Fig. 2a), which encodes the N-terminal portion of the ND6 polypeptide, the tRNA^{Glu} gene, and the 5' end of the control region (Fig. 1b). The gene order in clones pMtQ 21 and 22 is identical to that found in homologous regions in chicken and other vertebrate mtDNAs. The three cloned mtDNA fragments show, respectively, 81.2%, 85.9%, and 83.0% over-

all sequence conservation with their chicken counterparts. Clone pMtQ 21 contains eight tRNA genes and three large open reading frames encoding the N-terminus of cytochrome oxidase subunit 1, the C-terminus of subunit 1 and the complete subunit 2 of the NADH dehydrogenase complex. Clone pMtQ 22 encodes most of the control region, the complete 12S rRNA, the 5' end of 16S rRNA, and three tRNAs. As in other vertebrate mtDNAs, all the structural genes, except ND6, are H-strand encoded and punctuated by one or more tRNA genes (Fig. 1b).

Further analysis of the sequences (Fig. 2a and b) revealed that most of the Japanese quail mitochondrial genes are separated by short noncoding sequences of 1–5 nucleotides in length. Although these intergenic sequences are poorly conserved in primary sequence, their respective size and position are quite similar to those occurring between the homologous chicken genes. Other coding sequences, such as the ND1 and tRNA^{Ile} genes or the ND2 and tRNA^{Met} genes are butt-joined. There are also three instances of overlapping genes. The first event is a 1-nucleotide overlap between the tRNA^{Gln} and tRNA^{Met} genes, which are encoded on opposite strands. The two other overlapping pairs involve genes encoded by the same strand: a 1-nucleotide overlap between the tRNA^{Cys} and tRNA^{Tyr} genes and a 2-nucleotide overlap between the tRNA^{Trp} and the 3' end of the ND2 genes. In the latter case, it is highly likely that cleavage of the primary transcript occurs immediately after the T residue found upstream from the 5' end of the tRNA^{Trp} gene, the termination codon of the ND2 mRNA being subsequently restored by polyadenylation as previously demonstrated in human mitochondria (Ojala et al. 1981). The case of tRNA^{Cys} and tRNA^{Tyr} is more puzzling because it involves two tRNA genes. The tRNAs could be produced by differential processing giving rise to only half concentrations of these tRNAs. Another possibility would involve a post-transcriptional maturation mechanism, i.e., the addition of a G residue at the 3' end of tRNA^{Tyr} or at the 5' end of tRNA^{Cys}, depending upon localization of the RNA cleavage. Addition of a G residue at the 5' end of the chicken mitochondrial tRNA^{His} is but one example of such a maturation process (L'Abbé et al. 1990).

Protein Genes and Codon Usage

The base composition of the coding strand for the ND1, ND2, and COI polypeptides is 32.6% C, 10.7% G, 26.2% T, and 30.5% A. The overall codon usage for these protein-encoding genes has a strong bias against the use of codons ending in G (Table 1), which account for less than 3% of all triplets. This

Table 1. Average base composition at silent positions in protein gene sequences

Gene	Average base composition % of total			
	Guanine	Adenine	Thymine	Cytosine
ND1 (3' end)	1.4	37.5	23.6	37.5
ND2	1.4	42.8	15.9	39.9
COI (5' end)	3.0	38.2	23.5	35.3
ND6 (5' end)	44.8	10.6	43.4	1.4

extremely low incidence of guanine at the third position has also been observed in chicken for similar genes (Desjardins and Morais 1990) and in a few distantly related birds for a part of the cytochrome b gene (Kocher et al. 1989). The same observation has been made for similar genes in mammals (Anderson et al. 1981, 1982; Bibb et al. 1981; Gadaleta et al. 1989) and amphibia (Roe et al. 1985). More than 75% of all codons end in A or C and about 20% in T. The moderate use of T at the third position resembles the situation in human and rat (Anderson et al. 1981; Gadaleta et al. 1989). The codon usage bias observed in Japanese quail reflects the base composition of the L-strand, which has a very low guanine content. Conversely, the ND6 gene, which is L-strand encoded, uses a greater proportion of codons ending in G or T (Table 1).

Comparison of the Japanese quail protein genes with their chicken counterparts reveals that they encode very similar products (Fig. 3 and Table 2). Nucleotide substitutions in these coding sequences are found mainly at the third position of codons and occur very infrequently at the second position (Table 3). Consequently, a high proportion (66–100%) of base substitutions are silent and do not result in amino acid replacement (Table 2). This functional constraint against replacement is especially high for the 5' end of the COI gene, which, over short evolutionary time, has accumulated base substitutions at approximately the same rate as other protein genes (Table 3), even though it still encodes an identical amino acid sequence in both avian genomes (Fig. 3).

The average sequence difference in protein genes (Tables 2 and 3) is 14%, a value slightly superior to the previous estimate made from restriction mapping (Glaus et al. 1980). These results suggest that the evolutionary rate (0.0152 substitution/bp/10⁶ years) proposed for galliform mtDNAs (Glaus 1980), on the basis of restriction polymorphisms, might have been underestimated. The rate appears to be closer to the value of 2% per million years reported for mammals (Brown et al. 1979; Brown and Simpson 1982).

As shown in Table 4, the transition events found in protein genes greatly outnumber transversions.

```

      D A A L S V S Y V F V V L M G G L Y I
1  GGATCCGCAGCTAAAGAAACAGAAATACACAAAACCCCAACATCCCCCTAAATAAATC
      M F L V L S V F S V G L N V L W G C G V
61  ATAAACAACACCAAAGAAACAAAAGAAACCCCTAAATTCACCAACCCCAACCCCTACC
      V S A L V L G V V G Y Y P S P N S A V A
121 ACAGACGCTAACCAACCCCAACCCCAATAATACGGCGAAGGATTAGACGCTACAGCC
      <---ND6
      L T G L M F C V G L F I V F Y T M *-----
181 AAAGTCCCAACATGAAACAACTCCTAAAAAATCACAAAATAAGTCATATATCCCGC
      ---tRNA-Glu-----**
241 TTGGATAGACCCCAAGGACTACGGCTTGAAGGCCATTGTTGTTCTCAACTACGGGAACA
      --- Control region -----
301 ACACTTTTTTAACTAACTCCCTACTTAGTGACCCCCCTTCCCCAGGGGGGTAT
      -----
361 ACTATGCATAATCGTGCATACATTTATATCCACATATACTATGGTACCGGTAATATATA
      -----
421 TTATATACGCTAAACCCATTATATGTATACGGGCATTACATATTGTCGCCATTTCTCC
      -----
481 CCATGTACATTAGTGCATGCTCCAAGACATAAACCATACGTTACCTAGTAATAGACTTT
      -----
541 CCACTAACAGGACACCATAACTATGAATGGTTGCAGGACATAAGCTTACTAAATACTTAG
      29 bp sequence          HindIII
601 CTCCCACTTTGGTTATGCTAGACGTACCAGATGGATTTATTGATCGTACACCTCACGAGA
      -----
661 GATCACCAACCCCTGCTGTAATGTATTCGGTACTAGCTTACGGCCCATTTCTTCCCC
      -----
721 CTACACCCCTCGCCCTCTTGTCTTTTGGCCTCTGGTTCCTCGGTACGGCACATCCCC
      -----
781 CGCATAACTCCTGAACCTTCTCACTTTTCACGAAGTCACTGTGCGTATCCTGCCCTC
      -----
841 TTCAGTCCGTGATCGCGGCATTTCCCTTCTCTGCTGTGGTTCTTCTCTCTTTGGG
      -----
901 GCTTCTCACAGGTTGCCCTTACAGTGCCTGCGCCAGAGTCTACTCAAGTGAAGCCTG
      -----
961 GACTACTCCTCGCTGCGTCTATCCTAGACCTCTAGTGTCCCTCGATGATACGGTTTGC
      -----
1021 GTGTATTGGGTATCATTTTGACACTGATGCACTTTGGATCGCATTTGGTTATGGCTCTC
      -----
1081 CACCCCCCGGATAATGGTGTATATAGTGAATGCTTGCAGGACATATTTTTACCAATTT
      CSB-1
1141 TCACTTCTCTATTTTTCTAAACAACTAGGGAGTTTCCAAATTTTTTTGTTGTTTT
      -----
1201 GTTTAAATTTTTTTTTGTTTTTAAAAACATTTTTTAAAAAACTAAATACATACAAACT
      -----
1261 CACCGCACAAATACCCTCAAACCAAAAAAGCTTTTACGTTTAGTATATATACATTGTTA
      -----
1321 CGTTTATATAATTATAGAGAACTCCACTACCGAAACAAACATAAAAAACAACAAACT
      -----
1381 ATTCGGCATAAACTTCCCAACAATATATATTTATATGTTAATTACAATCAATTTTA
      -----
1441 TCCCCCTAGAAAGCCCATAGCTTAAACACAAAGCATGGCACTGAGATGCCAAGACGGT
      -----
1501 ACACAGATTACCTGTGGACAAAAGACTCAGTCTAACCCTACTATTGGTTTTGCTAGAC
      -----
1561 ATATACATGCAAGTATCCGCATCCCACTGAGAAATGCCCCACAAACCTTTATCTTCAAAG
      -----
1621 CAAAAGGAGCAGGTATCAGGCGCACTTTAATAGTTAGCCCCAAAACGCTTGTTCAGCCA
      -----
1681 CACCCCCACGGGTATTACAGCAGTAATTAACATTAAAGCAATAAGTGTAACTTGACTTAGC
      -----
1741 CATAGCAAGACCACAGGGTTGGTAAATCTTGTGCCAGCCACCGCGTCAACAAGAAGCC
      -----
1801 CAAATCAATAGCCATCCGGCGTAAAGAGTGGCCACATGTTATCTTCAATTAACATAGATCA
      -----
1861 AAATGGAACCAAGCTGTCATAAGCCAGGATCCACTAAGCCCAACATAAAAAACCATCTT
      -----
1921 AGCCACATGATCAATTTCAATCCACGAAAGCCAGGGCACAACCTGGGATTAGATACCCCA
      -----
1981 CTATGCCTAGCCCTAAATCTAGATACCCACATCTTATGTATCCGCTGAGAACTAGCAG
      -----
2041 CACAACCGCTTAAACTCTAAGGACTTGGCGGTGCCCAAAACCCACCTAGAGGAGCCTGT
      -----
2101 TCTATAACCGATAATCCAGCATCTACCCAACCCCTTGCACAACAGCCTACATACCG
      -----
2161 CCGTCGCCAGCCACTTAATGAAAGAACAACAGTGAGCTCAATAGCCGCCACTAATAAGA
      -----
2221 CAGGTCAAGGTATAGCCTATGGGATGGAAGAAATGGGCTACATTTTCTAAAAATAGAACAA
      -----
2281 ACGAAAAAGGACATGAAACCTGGTCTTGGGAGGAGGATTAGCAGTAAAAATGGGATCAC
      -----
2341 TTTGCCCACTTTAAGATGGCCCTGAGGACAGTACATACCGCCCGTACCCCTCTTCAAAG
      -----
2401 CTACTAATACCGATAAATAACACCCCAACCATTAAAGCCAAAGACGAGGTAAGTCGTAAACA
      -----
2461 GGTAAGTGTACCCGGAAGTCACTTAGACACCAAGGCGTAGCTATAAGTCCAAAGCATT
      -----
2521 CAGCTTACACCTGAAAGATGCCCCCAATAAGGGTGCCTTGTATTGCTACCTCTAGCC
      -----
2581 CAACCAACAAATCTTCAAAAAATCCAAATACCTCACACAGTTTAATTAACATTTAA
      -----
2641 TCTTATCCTAGTATAGGCGATAGAACCAGCCGAGGCGCAATAGAGACCAACCGTACCGT
      -----
2701 AAGGGAAAGATGAAATAACAATGAAAAACACAAGCAAAAAGCAGTAAAGACAAACCTTG
      -----
2761 TACCTCTTGCATCATGATTTAGCAAGAACAACCAAGCAAAAGCGGACTAAAGCTT

```

Fig. 2. Nucleotide sequence of clone pMtQ 6 and 22 (A) and 21 (B). Sequences shown are those of the L-strand and are numbered commencing from the first nucleotide at the 5' end of pMtQ 6 and 21, respectively. ND6 and six tRNAs (Glu, Gln, Ala, Asn, Cys, Tyr) are L-strand encoded. Other protein and tRNA genes are H-strand encoded. The control region, tRNA, and rRNA genes are indicated by broken lines above the nucleotide sequence and are delimited by asterisks to indicate the putative 5'- and 3'-encoded nucleotide. Anticodons are underlined. Translation of the four mitochondrial proteins is indicated above the sequence using the one-letter amino acid code with polarity given by the arrows. Stop codons are designated by asterisks. The 29-bp sequence and the HindIII restriction site between pMtQ 6 and 22 are indicated and the nucleotide recognition sequences underlined. Also shown in the control region is the position for CSB-1. The overlap between the tRNA^{Trp} and the 3' end of the ND2 gene involves nucleotides 1474 and 1475. Continued on next page.

ND1 --->
 1 AAGCTTCTGGAGCCGCCCTCAGAACTATTCTCTATTACACTCGCCACTAAAGTCTACT 1081
 L S S S F L W I R A S Y P R F R Y D Q L L
 61 TCTCTCATCTTCTTATGAATCCGAGCTTCATACCCACGATTCGGCTATGATCAACT 1141
 M H L L W K N F L P L T L A M C L W H T
 121 AATACATCTTCTATGAAAAATTTCTACCCCTAACCCCTAGCCATATGTCTGTGACATA 1201
 S M P I S V A G L P P A * *---tRNA-Ile-----
 181 CAGCATACCTATGATGTCGGCCTTCCCCAGCTTAAGGAAGCGTGCCTGAACAAAA 1261
 -----* *---
 241 GGGCCACTATGATAAAGTGGACATAGAGGTAAAACAACCCCTCTCGCCTCTAAACCTTAG 1321
 ---tRNA-Gln-----
 301 AAAAGTAGGAATTGAACCTACACGAAAGAGATCAAAACTCTTCACTATCCCTTATATTA 1381
 -----tRNA-Met-----
 361 TTTTCTAGTAAGGTCACTAATCAAGTATCGGGCCATACCCCGAAAATGATGGTTTTAA 1441
 ND2--->
 421 ATCTTCCCTCACTAATGAATCCCCATGCAAACTAATCTCTATCAATAAGCTAGCCCTA 1501
 * *---* *---
 481 G T S I T I S S N H W I L A W T G L E I
 GGGACCATGATCACAATCTCTAGCAACCATGAACTCTAGCCTGAACAGGTCTAGAAATC 1561
 N T L A I I P L I S K S H H P R A I E A
 541 AATACCCCTAGCCATCCCACTAATCTCAAAATCACACCACCCCTCGAGCAATCGAAGCA 1621
 A I K Y F L T Q S T A S A L I L F S S M
 601 GCAATCAAATACTTCTAACCATCAACCGCATCCGCCCTAATCTCTTCTCAAGCATA 1681
 N N A W S T G Q W D I T Q L N H P T S C
 661 AACAAAGCCTGATCTACTGGTCAATGAGACATTACACAACCTTAACCCACCCATCCTGC 1741
 L I L T M A I A I K L G L V P P H F W F
 721 CTAATCCTAACAATAGCCATGCAATCAAACTAGGTCTAGTTCATTCACCTTCTGACTA 1801
 P E V L Q G S S L I T A L L L S T L M K
 781 CCTAAGTCTTCTCAAGGCTCCCTAATCACAGCCCTTCTACTCTCTACCCCTAATAAAA 1861
 L P P M T L L L M T S Q S L N P A L L T
 841 CTCCCCCAATAACCCCTTCTCTCATAACATCCCAATCCCTTAACCCCTGCCTCTCACT 1921
 L L A V S S A L V G G W M G L N Q T Q T
 901 CTACTAGCAGTTTCTCCGATTAGTTGGAGGCTGAATAGGCCATAAACCAACACAAAA 1981
 R K I L A F S S I S H L G W M I V I I I
 961 CGAAAACTCTAGCCTTCTCATCCATCTCCACCTAGGATGAATAATCGTGATTATTATT 2041
 Y N P K L T I L T F I I Y S L M T S T V
 1021 TACAACCCAAAACCTACCATCTTAACCTTCACTCATCTACTATTATAAATCACTCACTGA 2101

Fig. 2. Continued

Table 2. Comparison of Japanese quail mitochondrial genes with their chicken counterparts

tRNA gene	Nucleotide conservation				rRNA gene	Control region	%
	%	Protein gene	%				
Val	87.8	ND1 (3' end)	83.6	12S	From tRNA ^{Glu}		
Phe	89.1	ND2	84.5	16S	To tRNA ^{Phe}	80.0	
Glu	98.5	COI (5' end)	88.3				
Ile	90.3	ND6 (5' end)	88.7				
Gln	90.1						
Met	91.3						
Trp	89.5						
Ala	88.4						
Asn	86.3						
Cys	81.8						
Tyr	94.4						

The percent amino acid conservation for the protein genes ND1, ND2, COI, and ND6 is 88.9, 88.7, 100.0, and 89.5, respectively

This bias is observed at both the silent and replacement sites and is caused mainly by numerous C → T changes occurring in the L-strand (Table 5).

The possible utilization of a GTG codon as translational initiator of the COI gene has been suggested in chicken (Desjardins and Morais 1990) and has also been suggested to initiate the ND1 gene in rat (Gadaleta et al. 1989), the ND5 gene in *Drosophila yakuba* (Clary and Wolstenholme 1985), and a few

prokaryotic genes (Stormo et al. 1982). This unusual start codon is also found at the 5' end of the Japanese quail COI gene (Fig. 2b).

Control Region

The region of vertebrate mtDNA encompassing the heavy (H)-strand replication origin is a noncoding sequence of variable length bordered by the genes

Table 3. Sequence differences in four galliform genes

Gene	Size (codons)	Number of differences		
		Position 1	Position 2	Position 3
ND1	72	8	5	23
ND2	346	39	14	108
COI	102	4	0	30
ND6	76	8	3	15
Supergene	596	59	22	176

ND1 (C-terminal)

Q: SFLSPPELFSITLATKLVLLSSFLWIRASVPRFRYDQLMHLWKNFLPLTLAMCLWHT
 C: ***NL*P***P*A***T*****L*****I*****

Q: SMPISYAGLPPA
 C: *****I

ND2

Q: MNPHAKLISIMSLALGTSITISSNHWLAWTGLEINTLAIIPLISKSHHPRAIEAAIKYF
 C: *****CTV*IM*****T*****

Q: LTQSTASALILFSSMNNANSTGQWDITQLNHPTSCLLITMAIAIKLGLVPPHFVPEVLQ
 C: *****T*****M*****

Q: GSSLITALLSTLMKLPMTLLMSTQSLNPALLTLLAVSSALVGGWMLNQTRKILA
 C: *****T*****I*T*I*****

Q: FSSISHLGWMIVIIYINPKLITLTFIIYSLMSTVFLSLSQIKVLKLSLMLISWTKTPML
 C: *****M*S**Q*****L*TI*****A*****L*****

Q: NSTIMVTLTLLSAGLPLPTGFMPKWLIIQELTKQEMTPVATTIAMLSELLFFYLRLAYHS
 C: *A*V*L*****M**I*T*****S*****

Q: TITLPPNSSNHMKLWRINTPTNTPTAILTVLSISLLPLSPLITTLV
 C: *****T*K*L*****A*TT*****I*ML

COI (N-terminal)

Q: MTFINRWLFSTNHKDIGTLYLIFGTWAGMAGTALSLLIRAEIGQPGTLLGGDDQIYNVIVT
 C: *****

Q: AHAFVMIFFMVMPIMIGGFGNWLPLMIGAPDMAFPRMNMMS
 C: *****

ND6 (N-terminal)

Q: MTYFVIFLGVCFMLGLTAVASNPSPYVGVGLVLSVVGCGWLVNLGVSFVSLVLFMIYL
 C: *****I*****V*****V**M*****S*****A**LV**

Q: GGMLVVFVYSVSLAAD
 C: *****

Fig. 3. Comparison of the Japanese quail (Q) mitochondrial-encoded proteins with those of chicken (C). Amino acids are depicted using the one-letter code. Asterisks indicate identical residues. In each case, the Japanese quail sequence is shown in full. The GTG start codon of the COI gene (Fig. 2) is assumed to represent methionine.

for tRNA^{Phe} and tRNA^{Pro}. Despite its high functional importance, the control region is the most rapidly evolving part of vertebrate mtDNA molecules (Bibb et al. 1981; Anderson et al. 1982) even when closely related species are compared (Upholt and Dawid 1977). In addition to its high substitution rate, this region occasionally undergoes large deletion or addition events (Ferris et al. 1981) as well as direct tandem duplications (Moritz and Brown 1986, 1987; Poulton et al. 1989).

In contrast to other vertebrates, the control region of the Japanese quail and chicken mitochondrial genomes, and probably that of all gallinaceous birds (Desjardins and Morais 1990), is flanked on its 5' end

Table 4. Substitution events in Japanese quail mitochondrial DNA

Gene	% transition	% transversion
ND1		
Silent	50.0	13.9
Replacement	33.3	2.8
ND2		
Silent	39.8	23.6
Replacement	18.6	18.0
COI		
Silent	72.2	27.8
Replacement	—	—
ND6		
Silent	42.3	19.2
Replacement	29.6	11.6
Control region	40.0	60.0
12S	73.2	26.8
16S	47.9	52.1
tRNAs (11)	78.7	21.3

Table 5. Frequencies of specific base substitutions between aligned Japanese quail and chicken protein gene sequences

Type of changes	COI	ND1	ND6	ND2
Transition				
A-G	5	7	6	23
C-T	21	23	12	71
Total	26	30	18	94
Transversion				
A-C	6	3	0	36
A-T	3	3	3	22
G-C	0	0	0	6
G-T	1	0	5	3
Total	10	6	8	67
	36	36	26	161

by the tRNA^{Glu} gene (Fig. 1a and b). The Japanese quail and chicken control region sequences are highly similar (80%) and apparently evolve at a similar rate to most coding regions (Table 2). A comparison of the two avian sequences with more distantly related mtDNAs from mammals and amphibia reveals, in contrast, a poor degree of conservation (data not shown). These results suggest that the genetic constraints needed to preserve the function of the control region differ from those acting upon the coding sequences. The regions of sequence similarity detected in interclass comparisons include one of the three conserved sequence blocks (CSB-1) described in other vertebrates (Walberg and Clayton 1981; Wong et al. 1983) and known to lie in the vicinity of the H-strand replication origin. As shown in Fig. 4A, both avian putative CSB-1 are highly homologous to their human counterpart. However,

no sequence of convincing similarity with CSB-2 and -3 found in other vertebrates has been detected in either avian control region.

Comparison of the two avian control regions reveals a uniform pattern of base substitutions. As observed in interspecific comparison of the control region of closely related rat species (Brown and Simpson 1982), most of the base substitutions (60%) occurring in Japanese quail and chicken are transversions (Table 4). Furthermore, their distribution across the control region is similarly uneven, being mostly concentrated at the 5' and 3' ends of the sequence where the number of substitutions per base pair is the highest. In contrast, in the conserved central part of the control region, the percent transversion (40%) is similar to that found in protein genes. These observations raise the possibility that the 5' and 3' ends of the control region evolve at a more rapid rate (DeSalle et al. 1987) than that of the conserved central part and the protein genes. Functional constraints or selection at the level of gene products could contribute to the higher transition/transversion ratio determined in the latter two regions.

The avian control region shows a high frequency of length mutations. The size difference observed in the control region of chicken (1228 bp) and Japanese quail (1153 bp) is due to many small (1–3-bp) addition/deletion events and to a large 57-bp deletion that occurs at the 5' end of the Japanese quail control region (Fig. 4B). The deleted segment, which maps near the HindIII cloning site at the junction of pMtQ 6 and 22 (Figs. 1b and 2a), includes one of the two 29-bp repeats detected in the chicken sequence (Desjardins and Morais 1990) and most of the nucleotides that separate these repeats. The possibility that an additional HindIII fragment occurring in this region may have been lost during the cloning procedure was excluded by sequencing a cloned BamHI fragment (pMtQ 20 in Fig. 1a) containing the complete control region.

Short repeated sequences are known to occur in the control region of many vertebrates. These sequences (TAS) are involved in termination of D-loop strand synthesis (Doda et al. 1981) and occasionally share homology with CSB-1 (Walberg and Clayton 1981). Interestingly, the 29-bp monomer is well conserved in both avian sequences and differs at only a few positions from the putative CSB-1 (Fig. 4B), suggesting that these sequences are involved in arrest of DNA template synthesis.

Transfer RNA Genes

Eleven potential tRNA genes have been detected. Sequence comparison indicated that these genes are highly homologous to their chicken counterparts (Fig.



Fig. 4. A Comparison of the putative Japanese quail CSB-1 sequence (nucleotides 1102–1128) with those from chicken (Desjardins and Morais 1990; nucleotides 867–893) and human (Anderson et al. 1981; nucleotides 209–235). B Sequence comparison of the chicken 29-bp repeat to an equivalent sequence in Japanese quail starting at position 553 and to Japanese quail CSB-1. C Mitochondrial DNA sequence in the region encompassing the L-strand replication origin. Sequences shown are those of the L-strand. Transfer RNA regions flanking the replication origin sequence are overlined and delimited by asterisks. Gaps introduced to optimize alignment are indicated by dashes. Asterisks indicate identical residues. Species compared: Q, Japanese quail; C, chicken; X, *Xenopus laevis*; H, human.

5) with similarity ranging from 82% for tRNA^{Cys} up to 98% for tRNA^{Glu} (Table 2). Evolution of the avian mitochondrial tRNA genes involves mainly substitutions, although small insertion/deletion events of 1–3 nucleotides in length also occur in the DHU and T ψ C loops of the tRNAs specifying isoleucine, valine, and phenylalanine. The DHU and T ψ C loops are in fact the most rapidly evolving part of these avian tRNAs. The best conserved region besides the anticodon loop is the DHU stem. Substitutions in this region are observed only in a few tRNAs (Tyr, Ala, and Asn). The overall substitution pattern of the avian tRNA genes reveals a striking bias toward transitions (Table 4). Transversions, which are observed in only 5 of the 11 tRNA genes (Met, Ile, Phe, Trp, and Asn), occur infrequently and are mainly confined to the DHU and T ψ C loops.

Ribosomal RNA

Comparative analysis of the primary sequence of the mitochondrial 12S rRNA reveals a pattern of evolution similar to the protein and tRNA genes with both avian genomes encoding a similar product (Table 2). The two genes evolved mainly by substitutions but small insertion/deletion events varying from 1 to 6 bp in length also occur. The substitutions are uniformly distributed throughout the genes and involve mainly transition type mutations

	AA stem	D stem	D loop	D stem	AC stem	AC loop	AC stem	V loop	T stem	T loop	T stem	AA stem	
Met:	AGTAAGG *****	TC **	AGCT ****	AATCA **CT*	AGCT A ****	TCGGG *****	CCCATAC *****	CCCGA *****	AAAT *****	GATGG *****	TTAAAT *****CC	CCTTC *****	CCTCACTA **CT****
Glu:	GTTCCCG *****	TA **	GTTG ****	AGAA ****	CAAC A ****	ATGGC *****	TTTCAA *****	GCCGT *****	AGTC *****	CTTGG *****A	GGTCTAT *****A	CCAAG *****	CGGGAATA *****
Ile:	GGAAGCG *****	TG **	CCTG ****	AAC-AA **A**	AAGG G ****	CCACT T****	ATGATAA *****	AGTGG *****A	ACAT *****	AGAGG *****	TAAAACAAC **T*****	CCTCT *****	CGCCTCCT **A*T****
Gln:	TAGAAA *****	TA **	ATAT ****	AAGGGAA *GA**G*	GTAT G ****	AAGAG *****	TTTGAT *****	CTCTT *****	TCGT CT**	GTAGG *****	TTCAATT ***G***	CCTAC *****	TTTTCTAA *****G
Tyr:	GGTAAGA *****	TG **	GCCG ****	AGTGTCA ***T*G	AGGC G *A**	TTAGG *****	CTGTAGT *****	CCTTT *****	TTAC *****	AGAGG *****	TTCAATT *****	CCTCT *****	TCTTATCG *****
Val:	CAAGCG *****	TA **	GCTA ****	TAAGCTCCA ***-T**	AAGC A ****	TTCAG *****	CTTACAC *****	CTGAA *****	AGAT *****	GCCCC A***T	CAATA--- ***C*GAC	AGGGT *A***	CGCCTTGA *****
Phe:	GCCCACA ***C**	TA **	GCTT ****	AAC--ACA **C**	AAGC A ****	TGGCA *****	CTGAAGA *****	TGCCA *****	AGAC ***T	GGTAC *****	ACAGA CT*CT	TTACC A****	TGTGGACA *****G**
Trp:	AGAAACT *****	TA **	GGAT ****	TAATTGCACTTA **C**TCACC*	AACC A ****	AAGGC *****	CTTCAA *****	GCCTT *****	AAAT *****	AAGAG *****	TTGAACC **A***T	CTCTT *****	AGTTTCTG *****
Ala:	GAGGTCT ***C**	TA **	GCTT ****	AATTA *****	AAGT G ***C	TCTGG *****A	TTGCAT *****G	TCAGA *****	AGAT *****	ACAGG *****	TTAGTGC ***A**T	CCTGT *****	AGATCTTA T*G*****
Asn:	TAGACAG ***G**	AA **	GCTA ****	ATGGGTTA **C**T**GT	TGGC G ****	TCTAG *T***	CTGTTAA *****	CTAGA ***A*	ATGAT **TG*	ATGGG *****	ATCGAAG *****	CCCAT *****	CTGTCTAG *****
Cys:	GGCCCTG *A*T***	TG *A	GTGA ****	AGT ***	TCAT G ****	GTTGAG A****	TTGCAAG *****A	CTCAT ***G*	TGAT *****	GTATG ***CA	TTAAAG C****	CATGC TG***	CGGGTCT *****

Fig. 5. Sequence comparison of the Japanese quail tRNAs (sense strand) with those from chicken (Desjardins and Morais 1990). In each case, the Japanese quail sequence is shown in full. Residues in chicken homologous to those of Japanese quail are indicated by asterisks and missing nucleotides by dashes.

(Table 4). The putative overall secondary structure of the avian 12S rRNA (data not shown) is quite similar to those previously proposed for mammalian (Zwieb et al. 1981) and amphibian mitochondria (Dunon-Bluteau and Brun 1986).

Simultaneous comparison of both the primary and secondary structures reveals that most of the changes, base substitutions and length mutations occurring in 12S rRNA are found in nonpaired segments of the molecule. Substitutions in stem regions are infrequent and consist exclusively of transitions. Analysis of the 5' end of the two avian mitochondrial 16S rRNA genes reveals a higher substitution rate than in 12S rRNA, which may not be representative of the overall sequence.

Light-Strand Replication

The L-strand replication origin of mammalian and amphibian mtDNA consists of a short noncoding sequence located between the genes for tRNA^{Asn} and tRNA^{Cys} (Fig. 4C). This sequence can form a stable hairpin structure characterized by a GC-rich stem of variable length ranging from 9 bp in *Xenopus laevis* up to 12 bp in mouse mtDNA and a T-rich loop of 12–19 nucleotides. The bases of the stem overlap with the 5' end of the tRNA^{Cys} by a few nucleotides. The synthesis of the primer RNA has

been shown to begin in the T-rich stretch of the loop structure and the transition from RNA to DNA synthesis to occur near the base of the hairpin structure (Wong and Clayton 1985). Further experiments have shown that the pentanucleotide motif 5'-CGGCC-3' found at the 5' end of the tRNA^{Cys} is necessary for efficient replication (Hixson et al. 1986).

An L-strand replication origin equivalent to the conserved mammalian and amphibian sequences has not been found at this position in Japanese quail and chicken mtDNAs (Fig. 4C). However, the putative recognition sequence 5'-CGGCC-3' for RNase H activity and/or for transition from RNA to DNA synthesis is present in the amino acid acceptor stem of tRNA^{Cys} and is conserved in both avian mtDNAs. These observations raise the question as to whether L-strand replication in galliform mtDNAs is initiated in this genomic region through an equivalent stable stem-loop structure obtained by an alternate conformation of one of the tRNA genes present in this region (Fig. 2b).

Acknowledgments This work was supported by the Medical Research Council of Canada. P. Desjardins is a recipient of a fellowship from the Cancer Research Society, Inc., Montréal. The technical assistance of Thu Cue Buy was highly appreciated. We thank Danielle Beaudoin for her excellent secretarial assistance in preparing the typescript. We thank Robert Cedergren and Karen Zinkewich-Péotti for critical review of the manuscript.

References

- Anderson S, Bankier AT, Barrell BG, DeBruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlick DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-464
- Anderson S, DeBruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) The complete sequence of bovine mitochondrial DNA: conserved features of mammalian mitochondrial genome. *J Mol Biol* 156:683-717
- Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26:167-180
- Brown GG, Simpson MV (1982) Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc Natl Acad Sci USA* 79:3246-3250
- Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76:1967-1971
- Brown WM, Prager EM, Wong A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* 18:225-239
- Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 22:252-271
- Dale R, McClure BA, Houchins S (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. *Plasmid* 13:31-40
- DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J Mol Evol* 26:157-164
- Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *J Mol Biol* 212:599-634
- Desjardins P, L'Abbé D, Lang FB, Morais R (1989) Putative chicken muscle-specific 7S RNA is related to the mitochondrial ATPase 6 gene. *J Mol Biol* 207:625-629
- Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci USA* 78:6116-6120
- Dunon-Bluteau D, Brun G (1986) The secondary structures of the *Xenopus laevis* and human mitochondrial small ribosomal RNA subunit are similar. *FEBS Lett* 198:333-338
- Ferris SD, Wilson AC, Brown WM (1981) Evolutionary tree for apes and human based on cleavage maps of mitochondrial DNA. *Proc Natl Acad Sci USA* 78:2432-2436
- Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisà E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J Mol Evol* 28:497-516
- Glaus KR (1980) Structure, organization and evolution of avian mitochondrial DNA. PhD dissertation, Ohio State University, Columbus
- Glaus KR, Zassenhaus HP, Fechner NS, Perlman PS (1980) Avian mtDNA: structure, organization and evolution. In: Kroon AM, Saccone C (eds). Organization and expression of the mitochondrial genome, vol 2. Elsevier/North-Holland Biochemical Press, pp 131-135
- Hixson JE, Wong TN, Clayton DA (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J Biol Chem* 261:2384-2390
- Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86:6196-6200
- L'Abbé D, Lang BF, Desjardins P, Morais R (1990) Histidine tRNA from chicken mitochondria has an uncoded 5'-terminal guanylate residue. *J Biol Chem* 265:2988-2992
- Morais R, Desjardins P, Turmel C, Zinkewich-Péotti K (1988) Development and characterization of continuous avian cell lines depleted of mitochondrial DNA. *In Vitro Cell Dev Biol* 24:649-658
- Moritz C, Brown WM (1986) Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. *Science* 233:1425-1427
- Moritz C, Brown WM (1987) Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proc Natl Acad Sci USA* 84:7183-7187
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269-5273
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470-474
- Poulton J, Deadman HE, Gardiner RM (1989) Duplication of mitochondrial DNA in mitochondrial myopathy. *Lancet* i:236-239
- Roe BA, Ma DP, Wilson RK, Wong JFH (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J Biol Chem* 260:9759-9774
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain termination inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467
- Stormo GD, Schneider TD, Gold LM (1982) Characterization of translation initiator sites in *E. coli*. *Nucleic Acid Res* 10:2971-2996
- Upholt WB, Dawid IB (1977) Mapping of mitochondrial DNA of individual sheeps and goats: rapid evolution in the D-loop region. *Cell* 11:571-583
- Walberg MN, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L-cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res* 9:5411-5421
- Wong TW, Clayton DA (1985) In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell* 42:951-958
- Wong JFH, Ma DP, Wilson RK, Roe BA (1983) DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origins and flanking tRNA genes. *Nucleic Acids Res* 11:4977-4995
- Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. *Nucleic Acids Res* 9:3621-3640

Received March 12, 1990/Revised August 6, 1990