

The Complete Nucleotide Sequence and Gene Organization of Carp (*Cyprinus carpio*) Mitochondrial Genome

Yea-sha Chang,^{1,2} Fore-lien Huang,^{1,3} Tung-bin Lo^{1,2}

¹Institute of Biological Chemistry, Academia Sinica, P.O. Box 23-106, Taipei, Taiwan

²Graduate Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan

³Department of Zoology, National Taiwan University, Taipei, Taiwan

Received: 19 March 1992/Revised: 18 May 1992

Abstract. The complete sequence of the carp mitochondrial genome of 16,575 base pairs has been determined. The carp mitochondrial genome encodes the same set of genes (13 proteins, 2 rRNAs, and 22 tRNAs) as do other vertebrate mitochondrial DNAs. Comparison of this teleostean mitochondrial genome with those of other vertebrates reveals a similar gene order and compact genomic organization. The codon usage of proteins of carp mitochondrial genome is similar to that of other vertebrates. The phylogenetic relationship for mitochondrial protein genes is more apparent than that for the mitochondrial tRNA and rRNA genes.

Key words: Carp (*Cyprinus carpio*) — Mitochondrial DNA — Complete genome sequence — Homology analysis of gene sequences

Introduction

The complete mitochondrial DNA (mtDNA) sequences of several species of vertebrates have been determined including mammals (Anderson et al. 1981, 1982; Bibb et al. 1981; Gadaleta et al. 1989), chicken (Desjardins and Morais 1990), and *Xenopus laevis* (Roe et al. 1985; Dunon-Bluteau and Brun 1986). These mitochondrial genomes have a con-

served organization and gene content. They consist of two species of ribosomal RNA (rRNA) genes, 22 distinct transfer RNA (tRNA) genes, and 13 protein genes encoding components of the inner mitochondrial membrane (Cantatore et al. 1987). The data of partial sequences of mtDNAs of several species of fish (Araya et al. 1984; Gilbert et al. 1988; Davidson et al. 1988, 1989; Huang et al. 1990; Johansen et al. 1990) also suggest that a similar gene content and genomic organization exist in fish mtDNA. The complete sequence of a fish mtDNA has not been available. In order to obtain more information about the evolution of vertebrate mtDNA, we have sequenced the mtDNA of carp, *Cyprinus carpio*, and its complete sequence of 16,575 base pairs (bp) is compared with those of other vertebrates.

Materials and Methods

Isolation of Carp mtDNA and Cloning of mtDNA Fragments. Carp liver mtDNA was isolated as described by Tapper et al. (1983). The 1.2-, 3.4-, and 4.1-kilobase (kb) fragments of *EcoRI* digest and 0.6-, 0.7-, 1.1-, 2.4-, and 5.3-kb fragments of *HindIII* digest were cloned into pUC19. The 8.0-kb fragment of *EcoRI* digest and the 6.5-kb fragment of *HindIII* digest were further treated with other restriction endonucleases (RE) and then cloned into pUC19.

DNA Sequencing. The mtDNA fragments from the above clones were further treated with other REs to generate DNA fragments of appropriate size suitable for DNA sequencing. In this study, pUC/M13 sequencing primers (forward and reverse)

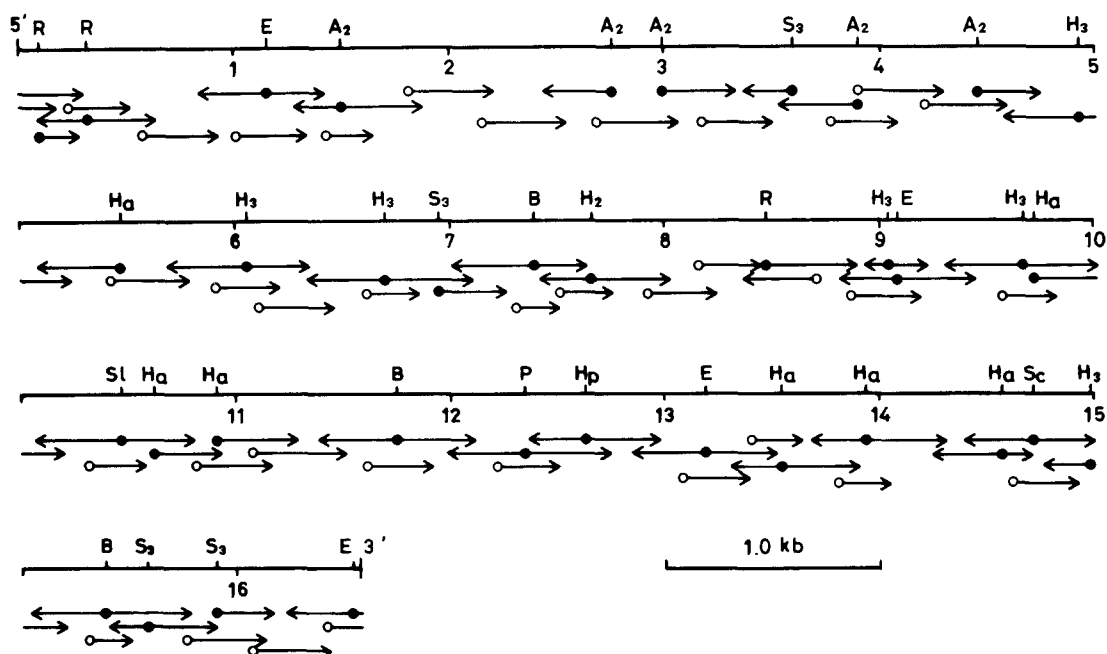


Fig. 1. The partial restriction endonuclease map and sequencing strategy of carp mtDNA. Only the restriction endonuclease sites used for subcloning are shown. The numbers represent the genomic length in kilobases. The arrows show the regions and direction of sequences determined. Solid and open circles represent the M13/pUC and synthetic primers used for sequencing, respectively. Abbreviations: A2, *AvaII*; B, *BamHI*; E, *EcoRI*; Ha, *HaeIII*; Hp, *HpaII*; H2, *HincII*; H3, *HindIII*; P, *PstI*; R, *RsaI*; Sc, *SacI*; Sl, *SalI*; S3, *Sau3AI*.

and synthetic oligonucleotide primers of specific sequences were used for DNA sequencing. The strategy of sequencing of carp mtDNA is shown in Fig. 1. Nucleotide sequencing was performed by the chain termination method of Sanger et al. (1977).

Results and Discussion

Sequence and Gene Content of mtDNA

The partial RE map of carp mtDNA is shown in Fig. 1, which is similar to that determined by Araya et al. (1984). The complete sequence of carp mtDNA has been recorded in the EMBL data library under the accession number X61010. The length of carp mtDNA is 16,575 bp, which is similar to that of the mtDNAs of human (16,569 bp), mouse (16,295 bp), and chicken (16,795 bp) but shorter than that of *X. laevis* mtDNA (17,553 bp).

As found in the mtDNAs of other vertebrates, carp mtDNA contains 2 rRNA genes, 22 distinct tRNA genes, and 13 protein genes. Most of these genes are encoded by the heavy (H)-strand, which includes two rRNA genes, 14 tRNA genes, and 12 protein genes. The light (L)-strand encodes the rest of genes. The same genes are also encoded, respectively, by the H- and L-strand of the mtDNAs of other vertebrates. The coding regions on both strands of carp mtDNA have about the same length as those found in other vertebrates. The organization of the carp mitochondrial genome is dia-

grammed in Fig. 2. The order of mtDNA genes in carp is identical to that of cod (Johansen et al. 1990), *X. laevis* (Roe et al. 1985), and mammals (Anderson et al. 1981, 1982; Bibb et al. 1981; Gadaleta et al. 1989) but different from that of chicken (Desjardin and Morais 1990). The D-loop is bordered by the tRNA^{Phe} and tRNA^{Pro} gene in the mtDNAs of carp, cod, *X. laevis*, and mammals but is bordered by the tRNA^{Phe} and tRNA^{Glu} gene in chicken mtDNA.

The majority of the protein genes and the two rRNA genes of the carp mitochondrial genome are interposed by at least one tRNA gene. These tRNA genes are thought to act as a recognition signal for mitochondrial RNA processing (Ojala et al. 1981). In carp mtDNA, some genes are punctuated by spacers of variable length from 1 to 13 nt, while some genes join end to end and some other genes overlap. Such a pattern of gene organization is also found in the mtDNAs of other vertebrates, although the length of intergenic spacer and the extent of overlapping may differ from those of carp.

Protein Genes

Alignment of the amino acid sequences of the 13 proteins encoded by carp mtDNA with those of other vertebrates (Fig. 3) reveals that COI, COII, COIII, Cyt b, and ND1 are conserved (identity from 64 to 90%) while ATPase 8, ND2, and ND6 (identity

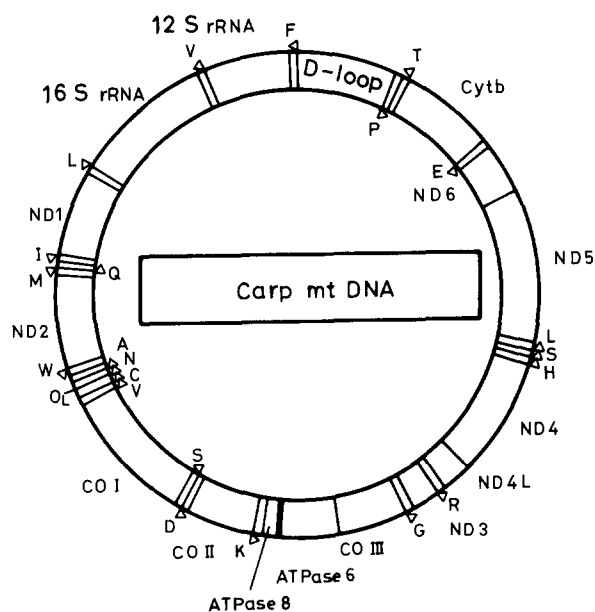


Fig. 2. The genomic organization of the carp mtDNA. Protein and tRNA genes are indicated by the one-letter amino acid code with polarity given by the open arrowheads. Protein and tRNA genes shown outside the circle are H-strand encoded with counter-clockwise polarity. The ND6 and 8 tRNA genes shown inside the circle are L-strand encoded with clockwise polarity. The origin of L-strand replication is indicated as O_L while the origin of H-strand replication has not been mapped.

from 33 to 63%) are diversified among vertebrates. The results of homology analysis of these 13 proteins are presented in Table 1, which clearly indicates the phylogenetic relationship of these protein between carp and other vertebrates. The proteins encoded by carp mtDNA have a greater similarity to those encoded by cod and amphibian mtDNAs than to those encoded by chicken and mammalian mtDNAs.

In carp, all but one protein encoded by mtDNA use the orthodox AUG codon as the translational initiator. The only exception is the COI gene, which begins with the sequence GTG. The presence of an in-phase GTG codon at the beginning of an mtDNA-encoded protein gene is also observed in the COI gene of cod (Johansen et al., 1990) and chicken (Desjardins and Morais 1990), in the ND1 gene of mouse (Bibb et al. 1981), and in the ND5 gene of *Drosophila yakuba* (Clary and Wolstenholme 1985). The GTG codon has also been reported to initiate prokaryotic genes (Stromo et al. 1982).

Among the 13 protein genes encoded by carp mtDNA, four genes end with TAA and one gene ends with TAG as the stop codon, whereas the other genes do not end with a complete stop codon but with T or TA. Therefore the stop codons of several species of carp mitochondrial mRNAs appear to be created by addition of adenylate residues to the primary transcript as proposed for other vertebrates (Ojala et al. 1981).

Table 1. Homology comparison of the sequences of proteins, transfer and ribosomal RNAs of vertebrate mitochondrial genomes

	Homology (%)			
	CP/CD*	CP/XE	CP/CH	CP/HU
Protein				
ND1		78	73	68
ND2		60	54	46
COI		90	88	85
COII		80	68	64
ATPase 6		80	74	53
ATPase 8	59	63	43	33
COIII		91	82	81
ND3	85	74	73	57
ND4L	84	51	67	49
ND4		66	61	61
ND5		64	59	46
ND6		49	51	32
Cyt b		80	75	70
tRNA				
Phe	75	87	77	66
Val		63	61	66
Leu(UUR)		73	78	62
Ile		74	74	71
Gln		75	75	68
Met		77	78	78
Trp	79	81	69	71
Ala	75	77	84	78
Asn	85	78	70	69
Cys	73	63	68	82
Tyr	79	81	70	69
Ser(UCN)		79	82	72
Asp		73	73	78
Lys	72	76	66	71
Gly	87	79	64	66
Arg	62	71	65	65
His	84	82	78	81
Ser(AGY)	60	54	61	64
Leu(CUN)	89	89	82	80
Glu		74	63	78
Thr		76	65	67
Pro		70	67	74
rRNA				
12 S		74	69	69
16 S		67	63	62

* CP, carp; CD, cod; XE, *Xenopus laevis*; CH, chicken; HU, human

The carp ND4L and ND4 genes overlap by 7 nt as found for the same pair of genes in other vertebrates. The ATPase 8 and ATPase 6 genes of carp mitochondria overlap by 7 nt as well. In other vertebrates, these latter two genes overlap by 10 nt in *X. laevis* and chicken and 40 to 46 nt in mammals. These four genes are translated in two different reading frames which differ by 1 nt. A single transcript is detected for each gene pair of ND4L/ND4 and ATPase 8/ATPase 6 in mammalian mitochondria (Ojala et al., 1980; Anderson et al. 1981) and for the gene pair of ATPase 8/ATPase 6 in chicken mitochondria (Desjardins et al. 1989). It appears that in chicken and mammalian mitochondria, a pair of

ND1
 CP:ML.NILMTHLINPLAYIVPVLLAVAFITLIERKVLGYMQLRKGNVVGYPYGLLQPIADGVKLFKEP
 XE:---.TII-----L-MI-I-----H-----I---T-I-----
 CH:-TLPT-TNL--MT-S--L-I-I-----V---I-S--A-----I---F-----V-----
 HU:-PM.....A-L-LL---I-I-M---M-T---I-----F--AM---T---

CP:VRPSTSSPFLFLAAPVLAITLMTLWAPIPMHPVTDNLGILFIALLSSLAIVSILSGWALNSKY
 XE:-----QTM--I--TM--A---SI---L---FSLA-----T-----SS----
 CH:I-----IIT-I---L--L-I-V-L-L-F-LA-----L--L-M---T---L-W---S---
 HU:LK-A--TIT-YIT--T---I-LL--T-L--N-LVN---L-----T-----W---S--N-

CP:ALIGALRAVAQTISYEVSLLGLILLSVLIIFSGGYTLQTFNITQESIWLIPAWPLAAMWYISTLAETN
 XE:-----T-----CM-MIA--F-YT-LMT--QM--I--G--M-----
 CH:-----T-AI---T-ML--N---S-LA---P-Y-IFS-----M-----
 HU:-----T-AI---TLLM--SFN-S-LIT---HL---L-S---M--F-----

CP:RAPFDLTEGESELVSGFNVEYAGGPFALFFLAEYANILLMNTLS.AVLFLGASHIPSVPELTINLM
 XE:-----S-----M-----YLI-----S-F.MNQ-----S--
 CH:-----A---M-----M-----T.T---NP-FLNLP---FP-ALA
 HU:-T---A-----I---A-----M---T---IM---TTTI---TTYDALS---Y-TYFV

CP:TKAALLSIIFLWVRASYPRFRYDQLMHLVWKNFLPLTLAFVLWHTALPIALAGLPQQL
 XE:I-SSI--M-----I---MT---IS---SML---S-T
 CH:--TL--SS---I-----L-----LC---SM--SY-----I
 HU:--TL--TSL---I-TA-----L-----LLM-YVSM--TISSI---T

ND2
 CP:MNPYRRATLLCSVGLGTTLTFASSHLLAWMGLEINTLAITPLMAQHHPRAVEATTKYFLTQATAA
 XE:---ITFSVV-T-LASEQF-AVS-----I---T--K---I--S-----A-S
 CH:---HAKLICTV-LIM--SI-IS-N--I---T-----I--ISKS-----I---I-----S--S
 HU:I---LAQPIVYSTIFA--LI-AL---FFT-V---M-M--FI-VLPKMN--ST--AI-----S

CP:AMILFASTTNAWMTGEWSINDLSNPIASTMFMAALALKIGLAPMHFWMPPEVLQGLDLLTGLILSTWQ
 XE:-LL--S-LN---L-----L--T--LSCATMTI-ICM-L---F---L-----S-T-----
 CH:-L---S-M---S--Q-D-TQ-NH-TSCL-LTM-I-I-L--V-F---F-----SS-I-A-L--LM
 HU:MIL-M-ILF-NMIS-Q-TMTNTT-QYS-L-I-M-M-M-L-M--F---V---T--TP-TS--L-L---

CP:KLAPLALIQTQATID.PLLLTLLGISSTLVGGWGGINQTLRKILAYSSIAHMGWMIIVIQYAPQL
 XE:---M-ILY-I-PMLNT-----L-T---I-----F-----L---SILPFS---
 CH:--P-IT-LL-L-S-SLNTT--.---A---I---M-----T---F---S-L---MI-S-N---
 HU:---ISIMY-ISPSLNVS---.S-L-IMA-S-----T---MA-LP-N-NM

CP:TLIALGTYIIMTSAAFLLTKMSLTTKISTLATWWSKSPILTSTTALVLLSLGGLPPLTGFMKWLIL
 XE:MILN-TI-L---TM--V--TISS---S---S---T-ST-ALSL-T-----S--V---F-I
 CH:-ILTFIL-T---TV--S-AQIKVL-L--LIS-T-T-M-NA-VM-T---A-----I
 HU:-ILN-TI---L-TT---L-NLNSS-TTLL-SR--N-LTW--PLIPST-----L---A-I

CP:QELTKQDLPIIATAMALTALISLYFYLRLLYAMTLTISPNNINSTTPWRTQTQTSLPLALFTTAAL
 XE:---S-NTT-L--TL--S--L--F---T-IV---S---TS-ASLT--HHSK-PT-L-SIALILSS
 CH:-----EMTPM--IITMLS-L--F---A-HS-I-LP--SS-HMKL---NK-LNTP.T-IL-ALST
 HU:E-F--NNSL--P-I--TIT-LN-----I-STSI-LL-MSN-VKMK-QFEH-KPTPF-PTLIALTT

CP:GLLPMAPIALMLTT
 XE:FII-IS-LT-T
 CH:T---LS-L-ITML
 HU:L---IS-FM--IL

COI
 CP:V.AITRWFFSTNHKDIGTLYLVFGAWAGMVGTALESLLIRAELSQPGSLLSDDQIYNVIVTAHAFVMI
 XE:M-----L-----L-----T-G-----I--
 CH:-TF-N--L-----I--T---A-----G---T--G-----
 HU:M.FAD--L-----L---VL-----G--N--GN-H-----

CP:FFMVMPILIGGFGNWLVPIMIGAPDMAFPRMNMSFWLLPPSFLLLASSGVEAGAGTGWSVYPPLA
 XE:-----M-----T-----
 CH:-----M-----T-----T-----
 HU:-----M-----L-----AM-----T-----

CP:GNLAHAGASVDLTIFSLHLAGVSSILGAINFITTTINMKPPAISQYQTPLFVWSVLVTAVLLLSLP
 XE:-----I-----M-----I-----
 CH:-----A--H.Y-----I-----L-----I--I-----
 HU:--YS-P-----I-----MT-----I-----

CP:VLAAGITHLLTDRNLNTTFDPAGGGDPILYQHLFWFFGHPEVYILILPGFGIISHVVAYYSKKEP
 XE:-----V-----M---I-T-----
 CH:-----M-----M-----A-----
 HU:-----M---I-T-----

CP:FGYMGVMWAMMAIGLLGFIVWAHMHFTVGMDDVDTRAYFTSATMIIAIPGKVKVSWLATLHGGSIKW
 XE:-----S-----DLN-----M---T---
 CH:-----LS-F-----R-N-----I-----T---
 HU:-----S--F-----SNM---

CP:ETPMLWALGFIFLFTVGGTLGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIMAAAFVHWFPLLT
 XE:DA-----A-----M-----GG-I---F-
 CH:DP-----I-----A-----A-----L-G-T---F-
 HU:SAAV-----A-----GG-I---F-

Fig. 3. Comparison of the amino acid sequences of the proteins encoded by carp mtDNA with those of other vertebrates. Species compared: CP, carp; CD, cod; XE, *X. laevis*; CH, chicken; HU, human. Amino acids are denoted by one-letter code. Residues identical to those of carp are indicated by dashes and gaps shown by dots are introduced to optimize alignment.

CP:GYTLHSTWTKIHFGVMFIGVNLTFPPQHFLGLSAMPRRYSYDPDAYALWNTVSSIGSLISLVAVIMF
 XE:-----E--A-----A-----T-----M
 CH:-F--PS--A--T-----AG-----T--L-----MT--L
 HU:----DQ-YA---TI-----G-----TT--IL--V--F--T--MLM

CP:LFILWEAFAAKREVLSVELTATNVEWLHGCCPPPYHTYEPAFVQIQSN
 XE:M--I-----TTY---S-ML---Q---T-----LKTSL---NHQMIKS
 CH:M--V---S---K--QP-----I--I-----F-----V-E
 HU:I-MI-----S--K--M--EPSM-L---Y-----F---VYMKS

COII
 CP:MAHPTQLGFKDAAMPVMEELLHFHDHALMIVLLISTLVLYIITAMVSTKLTNKVILDSQEIEIVWTI
 XE:----S---Q---S-I-----T--A-F-----I-MT-----TNLM-A---M---
 CH:--NHS---Q--SS-I---VE-----VA-A-CS---LL-L-LME--S.SNTV-A--V-LI---
 HU:--AA-V-LQ--TS-I---IT-----IF--CF---ALFLTLT-----TN-S-A--M-T---
 CP:LPAVILVLIALPSLRLLYLMDIEINDPHLTIKAMGHQWYWSYDYENLGFDSYVPTQDLAPGQFR
 XE:M--IS-IM-----V-----I-----N--D-S---I--N--T---
 CH:--IV--L---Q---M---DE-D--L--I-----T-----FKD-S---T--T--PL-H--
 HU:--I-----MT--V---S---SI-----T-----GG-I-N---L-PLF-E--DL-

CP:LLETDRHMVPMESPVRVLVSAEDVLHSAVPSLGVKMDAVPGRNLQAAFIASRPGVFYGCSEICG
 XE:--V-N-----T-L--T-----T--I---H-TS---T-----
 CH:--V---I-I---I--IIT-D-----A--T--I-----TS--TT-----
 HU:--DV-N-V-L-I-A-I-MMITSQ-----T--L-T--I-----TT-T-T---Y-----

CP:ANHSFMPIVVEAVPLEHFENWSSLMLEDA
 XE:-----TD-----S---AS
 CH:----Y---ST--K--A---LSS
 HU:-----L-LI--KI--MGPVFT-

ATPase 8
 CP:MPQLNPGPWFAILVFSWLIFLTIPTKILSHISPNEPTPVSAEKHK.TESWDWPW
 CD:-----A--M-FM-T-A-----L-P-VMA-TF---S-QGMTTP--AP-N---H
 XE:-----L--I---VL--F--P-V-K-KAF---TQTT--S-.PNP-N---T
 CH:-----N---S-MLLT-FT-SLL-QP-L--FTLT-N-ANKITT.T-.PTP-T---T
 HU:-----TTV-PTMITPML-TLFL-TQL-M-NTNYHLP-S-KPMKMKYNKP-EPK-TKICSLHSLPPQS

ATPase 6
 CP:MMVSFFDQFASPSYLGIPNIAVAIALPWVLYPTPPARWI.NNRLITIQGWFINRFTNQLMLPLNVGG
 XE:-NL-----M--VI---L--I-MLD-FT-ISW-IQSNGF-----L-S---LHN--TIFYQ.-TSP-
 CH:-NL-----S--CL---L-LPSLL--AL-L-S-GN---S---L--THLI-K---T---KA-
 HU:-NENL-AS-IA-TI--L-AAVLI-LF-PL-I..-TSKYLI-----T-Q-L-KLTSK-M-TMH-TK-

CP:HKWALLASLMIFLITINMLGLLPYTFTPTTQLSLNMGFAPVPLWLATVIIGMRNQPTIALGHLLPEG
 XE:-----T---LL-MSL-L-----L-----MA.SKPTNY-----
 CH:-----T---ILM-LS--L-----M--AL-L-----LLT-L---SAS-----
 HU:RT-S-M-V--I--IA-T-L---HS-----M-LAM-I--AG---M-F-SKIKN--A-F--Q-

CP:TPIPLIPVLIIETISLLIRPALGVRLTANLTAGHLLIQLIATAVFVLLPMMPTVAILTAAVLFL
 XE:--T-----F-----A---SI---SI---
 CH:--T---A-M---T-----S-TIA-----SISA---LI---
 HU:--T---M-V-----Q-M--A-----I-----MH--GS-TLAMSTINLPSTLIIFTI-I--

CP:TLLEVAVAMIQAYVFVLLSLYLQENV
 XE:----I-----
 CH:-I-----I
 HU:-I--I---L-----T-V---HD-T

COIII
 CP:MAHQAHAYHMVDPSPWPLTGAIAALLMTSGLAIWFHSTTLMTLGLILLLLTMYQWWRDIIREGTF
 XE:-----V---L---M---G-MI-L---TMV---I---V---
 CH:-----S-----IF--A---T---IM---YS---L-M--LSM--V-L---VV--S--
 HU:-T-S-----K-----LS-----M-----M--LM--LTNT-----VT--S-Y

CP:QGHHTPPVQKGLRYGMILFITSEVFFFLGFFWAFYHSSSLAPTPELGCCWPPTGITPLDPFEVPLLNT
 XE:-----I-----N-----Y---E-----N-----
 CH:-----T-----A-----F-----Q---VK--N-L-----
 HU:-----A-----N-----Q--H--R-----N-L-----

CP:AVLLASGVTVTAHHSIMEGERKQAIQSLALTILLGLYFTALQAMEYYEAPFTIADGVYGSTFFVAT
 XE:-----H-D--E---T-----
 CH:-I-----T--N---HA-T---F-----H--S-S--S-----
 HU:S-----SI-----L--NN-N-M--A-LI-----L--S--F-S---S--I-----

CP:GFHGLHVIIGSTFLAVCLLRQIQYHFTSEHHFGFEAAAWYWHFVDVVWLFVLSIYWWGS
 XE:-----L--S-----K-----
 CH:-----S--T---L-KF--PN-----II---M-M---
 HU:-----TI-FI--LMF---K-----

ND3
 CP:MNLIMITITVALSLILATVSWLPQMNPDAEKLSPYECGFDPLGSARLPFSLRFFLVAILFLFLD
 CD:--S-VIL-AS-----IL-----LS--Y-----I-----
 XE:-.TA--M-AMT--T--IL-----T--M-----M---M---I-----
 CH:--TLTFM-SLSFL--AA-T-MN--A--A--T-----I-----
 HU:--.FAL--M-NTL-A-L-MIIT-----L-GYM--ST-----MSP--V--MK-----T-----

CP:LEIALLLPLPWGDQLHNPTGTFWFATTVLILLTLGLIYEWTOGGLEWAE
 CD:-----S---L-M---S-A-----L-----
 XE:-----F-AA--NT-SIVIL--ALI-T-----L-----
 CH:-----AI--AH-MM-LT---IIA---F-----
 HU:-----AL-TT-LPLMVMSSLLLI-I-A-S-A---L-K--D-T-

ND4L

CP:MTPVHFSFSSAFILGLMGLAFHRTLLSALLCLEGMMLSLFIALALWALQFESPGFSTAPMLLLAFS
 CD:---T---TI---L-M-----A---A---S---S---LDAT-C-----M---
 XE:--LI---C---T---LN-SPI--I---L-MSMDGIV-TP-HLTIYLS-MMLYIM-P-A
 CH:-S-L---Y---TFSSL-----I-----S---M-P-SI-PVENQT-S-ALV-I-M---
 HU:-PLIYMNIML--TIS-L-MLVY-S--M-S-----MAT-MT-NTH-LLANIV-IAM-V-A

CP:ACEASTGLALLVATARTHTDRLQNLNLLQC
 CD:-----A-----HM-A-----
 XE:-P-A---S-NSDHYT-----K-FS---E-
 CH:---G---M---S---S-H-H-----
 HU:---AV-----SISN-Y-L-YVH-----

ND4

CP:MLKVLIPITIMLFPTIWLTSKWLWTTTTHASLLIASISLTLWKWTSETGWTSSNMYLAKGPLLTPLL
 XE:---I-L--L--I-ST---NK---PSL-SQ--I-SLL--M-FFNQ---T.HF--YLMTIDQIS---
 CH:---IIL---L--AL-SPA-SM--N--MY-----H--TPSYPT.KTTLTWGMDQIS---
 HU:---LIV---L-LT--SKKHHI-IN-T---I-SI-P-LFFNQINNNL.F-CSPTFSSD--T---

CP:VLTCWLLPLMILASQNHINPEPISRERLYITLLALLQTFLIMAFGATEIIMFYIMFEATLIPTLI
 XE:I-----L-----LSN-----Q-TF--M-VF--LS-----S---L-L-----I-----
 CH:--S--F---M--G-LQH--HK-K-MF-ST-III-P-I-L--S-L-LML--S-----L-
 HU:M--T---T-M--R-LSS--L-KK--LSM-IS--IS--T-T--L-----F--T-----A--

CP:TRWGNQTERLNAGTYFLFYTLAGSLPLLVALLLLQOSTGTLSILVLQYSQPLQLNSWGHMFVWAGCL
 XE:-----A-----S-YS-----LNL--LLPNHTPMT-ANYS--LA--
 CH:-----P---S---I-L-----IS-----SI-Y-HTN---HLPPIKLTH-NLPA--TSLSSLAL-
 HU:-----P-----V-----I--IYTHNTL-S-N--L-TLTAQELS---ANNLM-LAYT

CP:IAFLVKMPLYGVHLWLPKAHVEAPVAGSMVLAALLKLGYGMMRMVMDPLSKELAYPFILALW
 XE:L--M-----T-----I-----I-----II-ISIT-S-SM-----L--S--
 CH:M--M--A---L-----I---L--L-----I--VTLLME-V-NF-H--LT---
 HU:M--M-----L-----I-----LTLI-N--T-HM---LV-S--

CP:GIIMTGSICLRQTDLKSIIAYSSVSHMGLVAGGILIQTWPWGS.GAIILMIAHGLVSSALFCLANTA
 XE:-----S-----M-----ISAGNN--MKALT--M--NTSD--TH--C---KYQ
 CH:-AL--S-----IAASM--Q-S--.-M---S--T--L-----N
 HU:-M--S-----I---A--VTA-----S-T--V-----T--L-----SN

CP:.YERTHSRTMILARGLVIFPLTAVVWFIANLANLALPLPNLMGELMIITTLFNWSPWTILLTGLG
 XE:S-----ALL-S---ET-L--MGT--L-S---M---S--W---IT-M-A---S---I--D--
 CH:-----IL--T---PLL--MS---LL--T-M---TT--A--T-MVA-----SP--I--TA
 HU:-----I---SQ---TLL--M-F--LL-S-----TI--L---SVLV-T-S--NI-L-----N

CP:TLITAGYSLYMFMSQRGPTPNHIIIGLPFHTREHLLMTLHLIPVILLVTKPELMWGWGY
 XE:--L--S-----T---M--E-LNAIN-T---T---M---I-P-MM---I--LFF
 CH:--L--S-T---L-ST--TL-S--TTTPNSN-----I--MLT-IL---IS-TPL
 HU:M-V--L-----TTT-W-SLTH--NNMK-SF--NT--FM--S-IL--SLN-DIIT-FSS

ND5

CP:MTLIMHSSLLLIFFILVYPLLTTLNPNQQDSNIAG.ITKTAVSSAFFVSLLPLIIF..LNLKTEGII
 XE:-NFPLIFNSSMLIT-SILI-PILMSTFNMNIMNLHHLI--SVKT--LI-II-----..DQGL-S-T
 CH:-ETPLLLNT-TTLTL-TLLTPII-P-LLNLK-SPMS---I.KT--LI--I-TT--..IHSGA-S-A
 HU:--..-TMTTTLTSLI-PIL-TLV-PNKK-SYPHYV-SI-A-T-II--F-TTM-MC-DQEVI.-S

CP:TNWQWMNTQAFDVNISFKFDHYSILFVPIALYVTVSILEFALWYMHSDPNIDRFFKYLLTFLVAMII
 XE:--FH---INT--I-M-----I--S-I---F-----T---A---M-S-----V-
 CH:-H-E-QFIPN-KIP--L-M-M--MM-F---F-----T---A-E-F-TK--T-----I--LT
 HU:..-H-AT--TTQLSL--L-YF-MM-I-V--F---M--S---N---NQ-----I--IT-L-

CP:LVTANNIFQLFIGWEGVIMSFLFIGWWHGRADANTAALQAVIYNRVGDIGLIMTMAWLAMNLSWE
 XE:-----F--F-----YA--EP-----LS---V-----
 CH:-TI--M-L--V-----Q---E---M---I-----LS---SS--T--
 HU:-----L-----S---YA-----I--IL---I---F-LAL--FILHS---D

CP:IQQIFALSKNFDMTIPLMGLAATGKSAQFGLHPWLPVAMEGPTVPSALLHSSTMVAGIFLLIRL
 XE:M--V-M-NSD.NL-L--L--I-----A-----I-----
 CH:---THPNQT.P.-L--L--I-----A-----T
 HU:P--MAL-NA-.PSLT--L--L--A---L---S-----F

CP:HPRMENNQALALTCLCLGALTSFLTATCALTQNDIKKIVAFSTSSQLGLMMVTIGLNQPLAFHIC
 XE:S-M-N---T---I-----M-T---A-----IF----F---
 CH:--FLSS-KT-----ST--A-----I-----DL-----S
 HU:--LA--SP-IQ-LT---I-T--A-V-----I--H-----

CP:THAFFKAMFLFCSGSIIHSLNDEQDIRKMGGFLNIMPATSTYFTIGSLALTGTPTPLAGFFSKDAIIE
 XE:NN----VYY-F---QYSSC-----Q-SL-I-TSCL-----
 CH:-----L-----G-----C-QKTL-M-TSCL--N---M-----Y--L--
 HU:-----M-----N--N-----LKT--L--SL-----A-M---T--Y--H--

```

CP:ALNTSYLNAWALTTLTIATSFTAVYSFRLVYFVIMGTPRFLPLSPINENNPLVINTIKRLAWGSIIA
XE:-----QT-T-----I-----VIF-AS--H--SN-----KT---P-----V-
CH:N-----I-T---S---L-----T--L-MTLL-QT-HT-TPSNH-----T-PA-LP-M---L---M-
HU:TA-M--T-----SI-----L-SA--T-MILLTLT-Q---PT-TN-----TLL-P--G--A--LF-

CP:GLIITQNFPPIKTPIMTMSITLKMAALMVTIAGLLVAMELANLTSKQVKITPMISTHHSNMLGFYP
XE:--L-AS-ML-----NS-----PTLA-Q--II-SVT--II--D-SK--TYINQESKTNI.-S---L---F-
CH:--L-SSLIL-P---P---PTIT-T--II--TL-IIL-L--SS-SYSLTPPKHNPL.MN--SS--YFN
HU:--FL--N-IS-ASPFQT-IPLY--LT--A--FL---T-LD-NY--N-LKMKS-LCT.FY-----

CP:MIIHRLIPKLKLTGLQSAATQL.DKTWLETVGPKGLALTQAMAKITNDITRGMIKTYLTIFLLTLI
XE:T-----MM--TN-N-A-NI--H-I-LS-Y-KS--Q-MVNQ-LP-I-T-TN-QQ-L-----L--M-SA
CH:PLT--IS-SIL-HT--KI-SH-I-MA-YKKM--E---NLHLT-T--STTLHT-L--S--GS-A--IL
HU:S-T--T--Y-G-LTS-NLPLL-L-L---KLL--TISQH-ISTSI--ST.QK---L-FLS-FFP--

CP:LATIPVLL
XE:IIITLF
CH:TTILLIQK
HU:--TLLLTIT

ND6
CP:MTYFMFLLLMALVVGLVAVASNPTPYFAALGLVVAAGVCGVLVGHGGSFSLVLFLIYLGGMLVVF
XE:-I-MVSVSM-V--L-----S-FY-----L--A--L-I-SF-S---I-----
CH:---VIF-GICFML-VL-----S--YGVV-----SVM---W--SL-V--V--A---V-----
HU:-M-AL---SVC--M-F-GFS-K-S-IYGG-V-I-SGV---VIIINF--GYMG-MV-----M---

CP:AYAALAAE.PFPEAWGSRSVLGY.VLVYLLGVGLVAGIFWGGWYEGSWVVVDGLKEFSVLRGDISGV
XE:--S-AR-K.-Y-----W--VF-.-----I--.--WYL-L--VEVDGMNKSSE-GSYVM.--WV--
CH:V-SVSL-AD-Y-----DWR-V--.G-GFV-V-WMGVVLGGLVDFWKVG--TVDGGGVSA-L-F---
HU:G-TTAM-IEEY-----GVEVLVS---G-AMEVGFVLWVKEYDGVVVV-NFNSVGSWMIYE-EG--F

CP:AVMYSSGGGMLVICAWVLLTLLVLELTRGLSRGALRAV
XE:-L---CWWVIVVYVWVSIIINFVCGIWNKSMWESSC-
CH:--F--C-V-LFLVAG-G---A-F-----V-----I---
HU:IREDPi-A-A-YDYGRW-VVVTGWP-FVGVIIVIEIA-GN

Cyt b
CP:MA.SLRKTHPLIKIANDALVDLPTPSNISAWWNFGSLLGLCLITQILTGLFLAMHYTSDISTAFSSV
XE:--PNI--S-----I-NSFI-----SL-----V--A--I-----A-T-M-----
CH:--PNI--S---L-MI-NS-I--A-----AV--M-----L-----A-T-L-----
HU:--TP.M--IN--M-LI-HSFI-----A--L--T-----SP-A-----I

CP:THICRDVNYGWLIRNVHANGASFFFCIYMHIARGLYGYSYLYKETWNIGVVLVLLVMMTAFVGYVL
XE:A--F--L---L---L-----L--G-----F-----I--F--A-----
CH:A-T--N-Q-----L-----FL--G-----T--I--TL-A-----
HU:A--T-----I--YL-----M---LFL--G-----F--S-----II--AT-A--M---

CP:PWGQMSFWGATVITNLLSAVYPYMGDMLVQWIWGGFSDNATLTRFFAFHLLPFVIAAATIIHLLFL
XE:-----K--I-NV---SL-----I--G-S-L-----
CH:-----F--I--I-HT--E-A-----P-----L-----A-GI-----T---
HU:-----I--I-TD-----Y---SP-----T---I---I---LATL-----

CP:HETGSNNPIGLNSDADKVSFHPYFSYKDLLGFVIMLLALTLLALFSPNLLGDPENFTPANPLVTPPH
XE:-----T--T-----P-----L---T-----D-----I---
CH:--S-----L-IS--S--IP---Y-F--I--LTL--TPFLT-----
HU:-----L-IT-HS--IT---YTI--A--LLLF--S-MT-T---D-----D-Y-L---N---

CP:IKPEWYFLFAYAILRSIPNKLGGVALLFSILVLIVVPLLHTSKQRLTFRPITQFLFWTLVADMII
XE:-----M.-----VL---I-ALM-----S-M---F--IM--A---TL-
CH:-----AA-V-I-FLI-F--K---TM---LS-T---L---NLL-
HU:-----T---V-----L---I-AMI-I--M---QSM---LS-S-Y-L-A--LL-

CP:LTWIGGMPVEHPFIIIGQIASVLYFALFLIFMPLAGWLENKALKWA
XE:-----Q---D-YTM---L---I--SI-I-MF--M--V---L-N-
CH:-----SQ-----M---LS--TIL--LF-TI-T---M-NY
HU:-----Q--SY--T---V-----TTI--L--TISLI--M---

```

Fig. 3. Continued.

proteins would be translated from a single mRNA by using different reading frames and that the mRNA is not subjected to further processing to produce two different transcripts. Whether or not this is also the case in carp mitochondria remains to be investigated.

Genetic Codes and Codon Usage

Certain codons of animal mtDNAs deviate from the universal genetic code. In *X. laevis*, chicken, and

mammalian mitochondrial genomes. UGA codes for tryptophan instead of termination, AUA codes for methionine but not for isoleucine, and AGA and AGG code for termination rather than for arginine. The genetic code for amino acids of the proteins encoded by carp mtDNA is presently unknown. However, sequence comparison reveals that most of the 112 UGA and 126 AUA codons of the 13 protein genes encoded by carp mtDNA align, respectively, with tryptophan and methionine of the

Table 2. Comparison of codon usage in vertebrate mitochondrial genomes

	CP ^a	XE	CH	HU		CP	XE	CH	HU
TTT(Phe)	76	125	59	77	TCT(Ser)	29	66	35	32
TTC(Phe)	146	103	160	141	TCC(Ser)	57	44	101	99
TTA(Leu)	100	217	62	73	TCA(Ser)	82	123	98	83
TTG(Leu)	11	20	16	16	TCG(Ser)	9	11	4	7
TAT(Tyr)	50	60	26	46	TGT(Cys)	5	13	5	5
TAC(Tyr)	65	57	80	89	TGC(Cys)	20	18	23	17
TAA(Ter)	4	3	8	2	TGA(Trp)	112	101	96	93
TAG(Ter)	1	0	0	2	TGG(Trp)	8	16	12	11
CTT(Leu)	82	159	78	65	CCT(Pro)	33	52	21	41
CTC(Leu)	97	54	185	167	CCC(Pro)	52	20	108	119
CTA(Leu)	297	141	288	276	CCA(Pro)	105	121	103	52
CTG(Leu)	37	10	35	45	CCG(Pro)	12	11	3	7
CAT(His)	23	40	26	18	CGT(Arg)	9	10	1	7
CAC(His)	79	56	87	79	CGC(Arg)	15	5	16	25
CAA(Gln)	98	93	80	81	CGA(Arg)	45	50	45	29
CAG(Gln)	1	6	6	9	CGG(Arg)	10	1	10	2
ATT(Ile)	157	228	86	125	ACT(Thr)	41	83	55	51
ATC(Ile)	137	110	218	196	ACC(Thr)	104	71	164	155
ATA(Met)	126	161	135	167	ACA(Thr)	146	145	126	133
ATG(Met)	49	35	33	40	ACG(Thr)	8	8	4	10
AAT(Asn)	45	70	22	33	ACT(Ser)	9	15	3	14
AAC(Asn)	76	80	104	131	AGC(Ser)	46	42	54	39
AAA(Lys)	73	77	79	85	AGA(Ser)	0	1	0	0
AAG(Lys)	4	8	11	10	AGG(Ter)	0	0	1	0
GTT(Val)	43	62	31	30	GCT(Ala)	55	73	42	43
GTC(Val)	30	25	45	49	GCC(Ala)	144	83	144	124
GTA(Val)	118	75	57	70	GCA(Ala)	129	114	94	80
GTG(Val)	18	17	27	18	GCG(Ala)	10	6	9	8
GAT(Asp)	11	36	11	15	GGT(Gly)	32	44	20	24
GAC(Asp)	64	38	53	51	GGC(Gly)	44	40	79	88
GAA(Glu)	93	83	78	64	GGA(Gly)	137	113	87	67
GAG(Glu)	8	14	16	24	GGG(Gly)	35	23	29	34

^a CP, carp; XE, *X. laevis*; CH, chicken; HU, human.

homologous protein genes of other vertebrates. Furthermore, like the mtDNAs of other vertebrates, no internal AGA and AGG codons are found in carp mtDNA. Therefore the same codon assignment for the proteins encoded by the mtDNAs of other vertebrates is proposed for the proteins encoded by the mtDNA of carp as well.

The overall codon usage for all 13 protein genes encoded by carp mtDNA has a strong preference for A and/or C at the third position, 17% in T, 31% in C, 44% in A, and 6% in G (Table 2). A similar situation is also observed in other vertebrates. Such a biased distribution is reflected by the unequal frequency of codon usage in which some codons are used frequently but some other codons are seldomly used. The codon usage bias observed in carp and other vertebrates reflects the base composition of the L-strand of mtDNA, which is low in G (15.8% in carp and 12.3–13.5% in other vertebrates). On the other hand, the ND6 protein which is encoded in H-strand RNA sequence uses a high proportion of codons ending with G and/or T (A and C, 18.3%; T and G, 81.7%), a situation extremely different from that found in the genes encoded by the H-strand.

Transfer and Ribosomal RNA Genes

The sequences complementary to the anticodons of the 22 tRNAs encoded by carp mtDNA are identical to those of the homologous tRNAs encoded by the mtDNAs of other vertebrates (Fig. 4). As found in other vertebrates, all the 22 tRNAs encoded by carp mtDNA can be folded into cloverleaf secondary structure with 7 bp in the amino acid stem, 5 bp each in the T C and anticodon stems, and 4 bp in the DHU stem (Fig. 5). Like the tRNA genes in the mitochondrial genomes of other vertebrates, the tRNA genes of carp mitochondrial genome do not encode the 3' CCA terminus. Consequently, they must be added posttranscriptionally. Many mismatched base pairs are found in the stem regions of the tRNAs encoded by the mtDNAs of carp and other vertebrates. Among the 22 mtDNA-encoded tRNAs of vertebrates, the tRNA^{His} and tRNA^{Leu} (TAG) are well conserved (identity from 78 to 89%) while the tRNA^{Ser} (AGY) is divergent (identity from 54 to 64%) (Table 2).

Among the five contiguous tRNA genes encoded by the same strand of carp mtDNA, only the

Phe
tRNA
CP: GCT. AGCGTAGCTTAA... CACAAAGCATAGCACTGAAGATGCTAAGATGAGTCCTAAGAAA. CTCCGCATGCA
CD: ---T-----TT---AT-----AT---G-----GAC---.---GTC---A-A---
XE: ---T-----GT-----C-----A-----G-----C-----G-----T-A---
CH: ---CCC.-A-----C.C-----G-----C-----GTA---CT-T-C---GT..GG---
HU: -T-T-T.-----CCTC-T-----ATA-----A---T-T---C-G-CT-AC-TC-. -C---AT-AA---

Val
tRNA
CP: CAGGGTGTGGCTGAGTTAGTCAA. GCATCTCA. CTTACACCGAGAAGACATCCATGCAAATTGGA. TCGCCCTGA
XE: --AA--A-A---T-ACC-...--C-T--G.-----A-CA-T---TG-TA---CCC---T-.A-TT---
CH: --A-C--A---ATAACT.---A---.---G-----TGA---T-C---TCAAC-GACAAGG---T---
HU: --A---A---T-ACA...--A---C-CA-.-----TT--G---TT-CA-C-T--C---AC..--T---

Leu
tRNA (TAA)
CP: TACTGGGTGGCAGAGCATGGTAAATGCGAAAGGCCCTAAGCCCTTTTACCCAGAGGTTCAAATCCTCTCCAGTTT
XE: GCTA-C-----C---CT--.-----A-----T-----TAT---G-----C-.--G-TAAC
CH: GCTA-C-----TC-C---.---A---T-----AT.-----TACG-
HU: GTTAA-A-----CC---TC.-AT-AA-T---AA-T---AC-GT---C---T-----TT-ACA

Ile
tRNA
CP: AGGAACTGCGCTGAGCACCTAAGGACCCTTTGATAGAGTGAAT..TACAGGGGTTAAATCCCC.TCAGTTCCTA
XE: G-A--TG---C---AAG..-G---T-----A..-T-T---C---C---A---TC---
CH: G-A-G-G-----A..AA---T---A---A---C.A-G-.--TA---C-A---TC---C---
HU: --A--TA--T---T...AA---AGTT-----A---AA--.G-A-C---.C---.T.A--T---

Gln
tRNA
CP: TAGAAAGTGGTGTA. AGGAAGCACCAAGAGTTTGGATCTCTTGAGTATGGGTTCAATTCCCTTCTTTCTAG
XE: ---G---A--G-T--G--T--GG--G-----CAG--GCA-----TG-----A
CH: ---A--AA-A--G.-G--T-TG-----CT--G-A-----G-----TACT-----
HU: ---G-T-G---G-T--TG---GG--A---GAT--CAG-G-----G---T-A-AG-C---

Met
tRNA
CP: GGCGGGGTGCTAGCTAATAAAGCTTTGCGGCCCATACCCCGAACATGACGGTTAAAGTCCCT. CCTCCGCCA
XE: A-TAAA-----A-----T-----A-----TT-----CC---T---T-TA-T-
CH: A-TAA-----CT---A-----A---T---T---.C---T--C-TA-T-
HU: A-TAA-----AT---A-----TT-----T-.--T--CGTA-T-

Trp
tRNA
CP: AGGGATTAGGATAACATTAGACCAAAAG....CCTTCAAAGCTTTAAGTAGAAGTGAATACTTCTAATCCCTG
CD: ---GC-----T-----GG-....-----CC-C--CG--G-----C--C-GC-----
XE: --A---A--T---T---T---G-....-----CC---C--G--T-G-----C-----T---
CH: --AA-C-----T-AC-GTC-----CCAAAGG-----C---A--AG--T---C---CT--G-TT---
HU: --AA-----T--ATAC-----G-....-----CC-C---AG.T--C--TA---A..--TT---

Ala
tRNA
CP: AAGGCCTTAGCTTAATAAAGTGTCTGATTGCAATCAGA. GATGCAAGTTAATATCTTGTAGGCCTTA
CD: ---T-----T-----A-T--T-----TA---A---TGG-G--G-G--C--C--AT---
XE: ---T-----T-----T-T-G-----T---ATT---TTG-A---A---C--C-A-----
CH: G-----T-----T---C-----T---GA---A--G-----G--C--T--T---
HU: A--G-----T-----G-----GT---TT-----G-G-G--G-T--GC--T-----

Asn
tRNA
CP: TAGATGGATGCTCGCTGGCTTGGGCGTTTAGCTGTAACTAAAAGTTTGTAGGATCGAGGCCTTCCCATCTAG
CD: ---A-----T-----A-T-C-----G-----AA-----T-TG-----
XE: ---AT--A-----T---A--A-T..-----TG---CG-----CGT-TT-----
CH: ---GCA--A--CAAT--TG-T--A-----T-G-A-G-----A---CAT-TG-----
HU: ---T--A--CA-T--AT-A--T-C-----GT-----G--T-TA--T---CATGG-----

Cys
tRNA
CP: CAGCTCCGTGGTGAAGAACACATTGGATTGCAAAATCCGAAGACGAGATTAATACTCTGCCGGGGCTTT
CD: .--T--T-GA---TT.-C---G-CA-----T-----A-----C-T-C-G-T-----A---
XE: A---CT.-C---TT--C.-TGCCA-----TCG--A---A-C.G-AGG-T-----C
CH: G-.--T--A-----TT--T-A--AG-----CT--TT--T-T-C-C--AG.-G-----.-C-
HU: .-----A-----TTT--T---A-----T-----A---C--C-A--.-----

Tyr
tRNA
CP: GGTAGGGTAGCCGAGTGCTTAGGCGGTGGGTTGTAGCCC. GAAGACAGAGGTTTGAGTCCTCTTCTCTATCA
CD: ---A-G--T---AA....-----A---A---CAC-T-----C-----T-GC---

Fig. 4. Comparison of the nucleotide sequences of tRNA genes encoded by carp mtDNA with those of other vertebrates. The anticodons are *underlined*. Species compared are the same as those shown in Fig. 3. Residues identical to those of carp are indicated by *dashes* and gaps shown by *dots* are introduced to optimize alignment.

XE:---A---G-----AA-----C-A-----T-C-TGT-----CA-----T-----
 CH:---A-A-G---T-----T-G-AG---T-A---C-----T-TTTT-----CA-T-----T-----G
 HU:---AAA-G---T-----AA...-AT---AC---AAT-TA-----G-----..-G-----TT-C---

Ser

tRNA (TGA)
 CP:AAGGAAGTGACA..GAGTGGTT..ATGTGGCTGGCTTGAAACCAGCATATGGGGGTTC AATTCCTCCCTTCTCG
 XE:---A---A---G---..-----G---CAA---A-----T---AG---G-----G-----TT-----
 CH:---A---AAG---TTA-----TG---C---T-----A---G---A-----G-----T-----T-----
 HU:T---A---AA-T---..TG-A---CC..---G---T-----T-TG-----G-----T-----T-T-T

Asp

tRNA
 CP:AAGACATTAGTAAAAACGTAAATTACATCACCTTGTC AAGGTGAAATCGTAGGTAAATCCCTGCATGTCTTA
 XE:G---TG-----..-----GCA-G-----C-----A-CT-----G-CT---G---CA-C-C-
 CH:G---G-----C---..-----AGA-C-----ACT---AC---GC---A-----T-CA---C-
 HU:---GT-----A-----..C---T---A---T-----A---T---TA---C-----..AT---A-----

Lys

tRNA
 CP:CGCTAGGAAGCTAAATATTGGACAAAGCGTTGGCCTTTTAAGCCAAAGATTGGTGATCCCGACCACCTCT...AGTGA
 CD:-A---A-----..-G---TT---ACCA-----TGG-AGCA---C---A-----CT---...-A---
 XE:-A---A-----..-G---CATT---ACA-----TGT-----C---A-----CT---...-A---
 CH:-AT---A-----T.....-CACC---AC-A-----T-G---GA---G---CA---TC---..-CT---...-A---
 HU:-A---GTA-----..TT---A---AA-----TT-----AAGAGAA---..-A-----TTAC-----

Gly

tRNA
 CP:ATCTTTCTAGTATCAAAAGTTAGTACAAGTGA CTCCAATCACACAGTCTTGGTTAAACCCCAAG..GAAAGATA
 CD:..-----..-T---G---T---G-----C---G-----G-GT-----..-----
 XE:-CT---CT---T---CC..-----C-----A-----A---G-AT-T---A-----..-
 CH:GCTC-----ACTCA...-----C-----TTTA-AATC---ATC-A---..-A---..-GCA
 HU:-CTC---T---A---..-----CGT-A-----T-ACT---T---ACA-C..TT---AAA-G---..-

Arg

tRNA
 CP:AGGGAGTTAGTCCAAAACAAGACCTCTGATTTCCGGCTCAGAAAATCATGGTTAATTCATGACCCCTT
 CD:---TGA-----..T---GT---A-TACT-----AG-GTCTG---A---G---CA-TTG---A
 XE:GA-TT-----..T-----AGT-----AC---T---A---CC---A-TAA-TC-
 CH:-AA-----..T---CT---AG---G---A-C---C---T---A-ACCC-CCT..-A---TTT---
 HU:T---T-TA---T..T-----A---GAA-----A---TT---T---..A---CAT---TTA---AA

His

tRNA
 CP:GTAAGTATAGTTTAACTAAACATTAGATGTGTGATTCTAAAAACAGGGGTAAAGTCCCTTACTC.ACC
 CD:---GA-T-----C---G-----T---A-----C---TC---A---C---
 XE:---GA-----..-----C-----G-GT---A-----..C---TC-TA---A---
 CH:-C---AC-----CC-----T---A---T---CC-T---GTT-.G---
 HU:---A-----C-----C-----A---G-C---A---C-T-C-A-----T-T---

Ser

tRNA (AGY)
 CP:AAGGAAGAACAGAAAATAGTAAGCACTGCTAATCCTTACATTCATGGTTAAATCCGTGGCTTCCTTG
 CD:G---AG---GC-C---CGGC-A-G-AG-----CTA-C---CC..-T-----G-CC---AA---C---C-
 XE:G-ACTT---CTG---CCC...-A-----TAC-TACG..-TGT---C-T---AC---GT-C-
 CH:-G---G---GC-CA-GCC..-C---A-----T-C-G---CTG-GCT---..-TCA-TCC---
 HU:G---A---CT---..C---A-----CT-A-G-CCC..-TC---C-A-A..-T---CA

Leu

tRNA (TAG)
 CP:GCTTTCAAAGGATAACAGTTTCATCCGTTGGTCTTAGGAACCAAAAACCTCTTGGTGCAATCCAAGTGGAAGCT
 CD:---CCT-----T---C-----C-----C-----A-C---
 XE:---T-----A---CT---C-----G-----AA---
 CH:A---T-----G---CA-----CCTA-----AA---TA
 HU:A---T-----..CT---A-----CC-----..-T-----C---A-AA---TA

Glu

tRNA
 CP:GTTCTTGTAGTTGAATAACAACGGTGGTCTTCAAGTCATGGTCTCGGTTAAAGTCTGAGCAAGAATT
 XE:---C-A-----T---GA---CAA---TCT---G---CTG-TGG---
 CH:---CC---G---AA---C-T---G-A-TC---T---GGTCTAA-CA---GG---A
 HU:---AT---A---T---TA---GT---GT---C-T-G---A

Thr

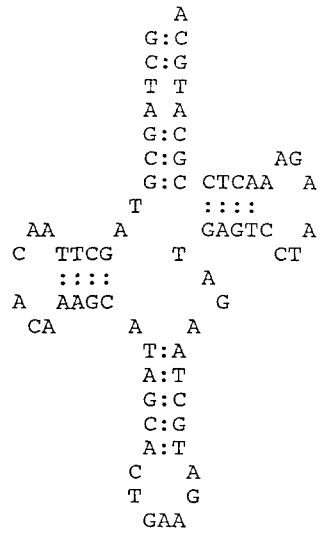
tRNA
 CP:GCCCTAGTAGCTTAGCCTAAAGCATCGGCTCTTGAATCCGAAGATCGGAGGTAAATTCCTCCCTAGCGCC
 XE:-T---GA---ATT---G-----T---C---AC---TC-AGA-T
 CH:A-T---A---T---TG..-A---T-----A---A---A-CT-A---AC-CC-.C---T-T---A-TA
 HU:-T---T---TA-AA---T..-C-A-----A---G---A-AACC...-TT---A---GA-A

Pro

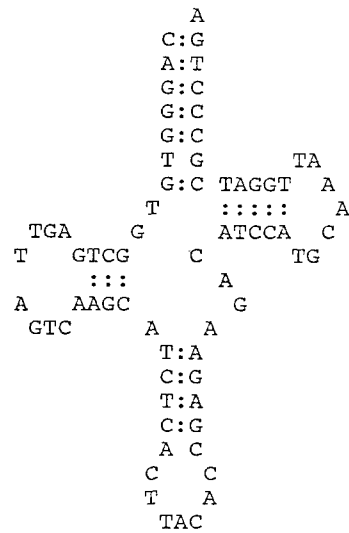
tRNA
 CP:CGAGAAAATAGTTTGTAGTATTCTGGCTTTGGTAGCCAGGGGTG.AGAGTTAAATCTCTCTTTTCTGGG
 XE:---G-G-G-A---AA---GT---GG-T---ATA---G---GT---G-G..-CT---..-C-A
 CH:---G-G---T---G-A---A-CA-----G---TG-A-A---G---GT---G-G.C-CTC-----A
 HU:---G-----AA-----CT-A-----GT---T-AT---..G.ACT-T---C---A

Fig. 4. Continued.

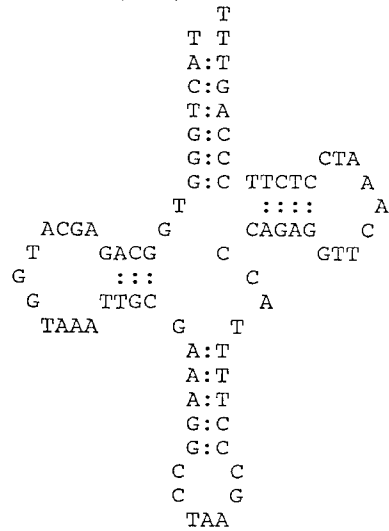
tDNA^{Phe}



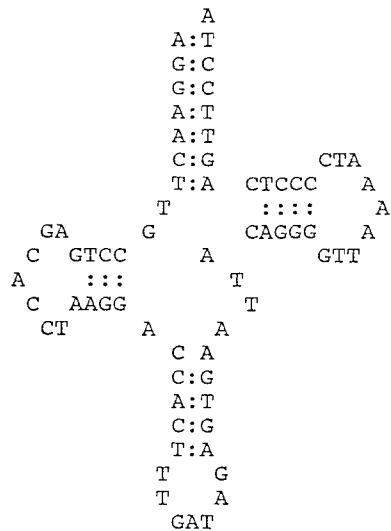
tDNA^{Val}



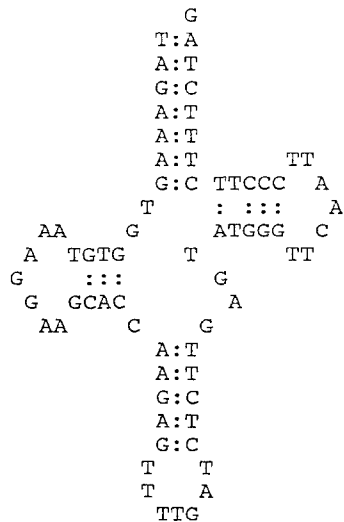
tDNA^{Leu} (TAA)



tDNA^{Ile}



tDNA^{Gln}



tDNA^{Met}

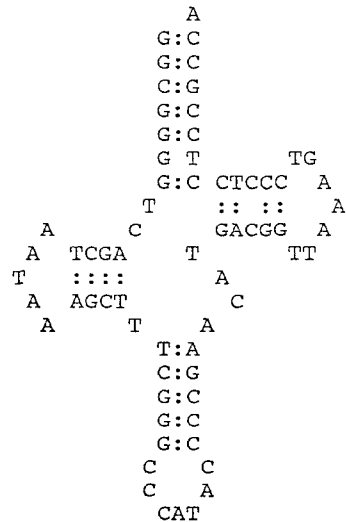


Fig. 5. The secondary structure of the 22 tRNA genes encoded by carp mtDNA represented in cloverleaf form. Standard base-pairings (G-C and A-T) are indicated by *colons* (:).

tDNA^{Trp}

```

      G
      A:T
      G:C
      G:C
      G:C
      A:T
      T:A      TA
      T:A TCTTC A
      T      : : : : A
      AA      A      AGAAG A
      C TAGG      T      TG
      A      : :      G
      T GACC      A
      TA      A      A
      A:T
      A:T
      A:T
      G:C
      C:G
      C      A
      T      A
      TCA

```

tDNA^{Ala}

```

      A
      A:T
      A:T
      G:C
      G:C
      C:G
      C:G      TA
      T:A TGTTC T
      T      : : : : A
      A      A      GCAAG A
      A TTCG      T      TT
      T      : : :      A
      A AAGT
      A      G      G
      T:A
      C:G
      T:A
      G:C
      A:T
      T      A
      T      A
      TGC

```

tDNA^{Asn}

```

      G
      T:A
      A:T
      G:C
      A:T
      T:A
      G:C      GG
      G:C CTCTCC A
      A      : : : : G
      TCG      T      T GTAGG C
      G      CTCG      T      AT
      G      : : :      T
      G GGGC      G
      CTT      G      A
      T:A
      T:A
      T:A
      A:T
      G:C
      C      A
      T      A
      GTT

```

tDNA^{Cys}

```

      T
      C T
      A:T
      G:C
      C:G
      T G
      C:G      TCA
      C:G CCGTC T
      G      : : : : A
      A      T      CGCAG A
      A GTGG      A      ATT
      G      :      G
      G ACAC
      A      A      A
      T:A
      T G
      G:C
      G:C
      A:T
      T      A
      T      A
      GCA

```

tDNA^{Tyr}

```

      A
      G:C
      G T
      T:A
      A:T
      G:C
      G:C
      G T      TG
      G T TCTCC A
      T      : : : : G
      GA      A      AGAGG T
      T      GCCG      C      TT
      G      : : :
      C AGGC      A
      TT      G      G
      G A
      T:A
      G G
      G:C
      G:C
      T      C
      T      G
      GTA

```

tDNA^{Ser}(TGA)

```

      G
      A C
      A:T
      G:C
      G T
      A:T
      A:T      TT
      G:C CCTCC A
      T      : : : : A
      GA      G      GGGGG C
      T GACA      T      TT
      G      : : : : A
      G ATGT      T
      TT      G      A
      G:C
      C:G
      T:A
      G:C
      G:C
      C      A
      T      A
      TGA

```

Fig. 5. Continued.

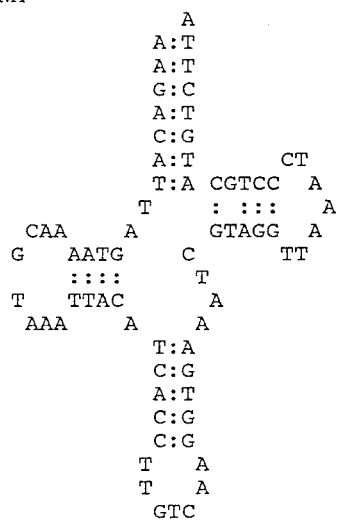
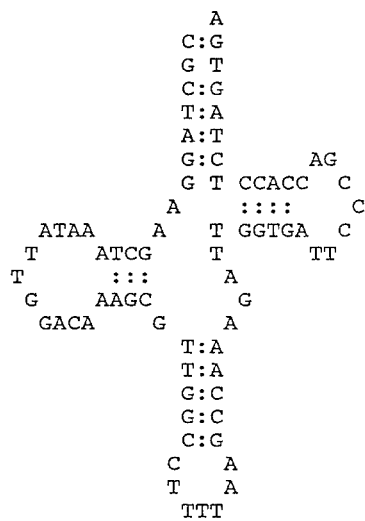
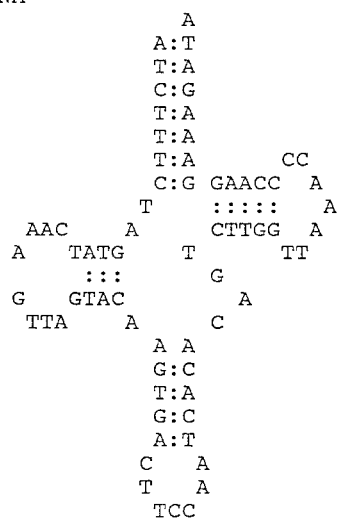
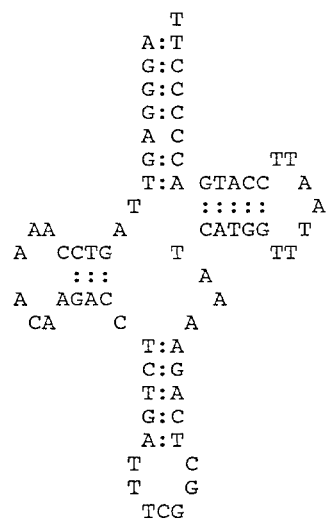
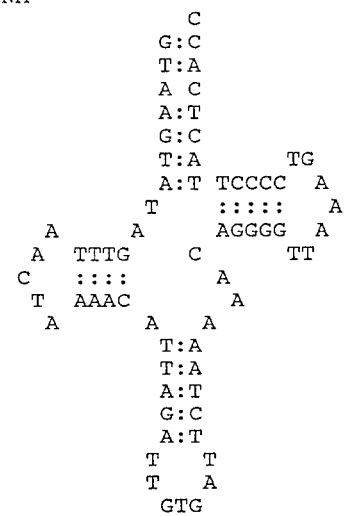
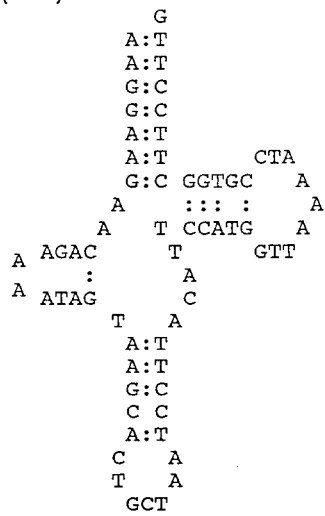
tDNA^{Asp}tDNA^{Lys}tDNA^{Gly}tDNA^{Arg}tDNA^{His}tDNA^{Ser} (AGY)

Fig. 5. Continued.

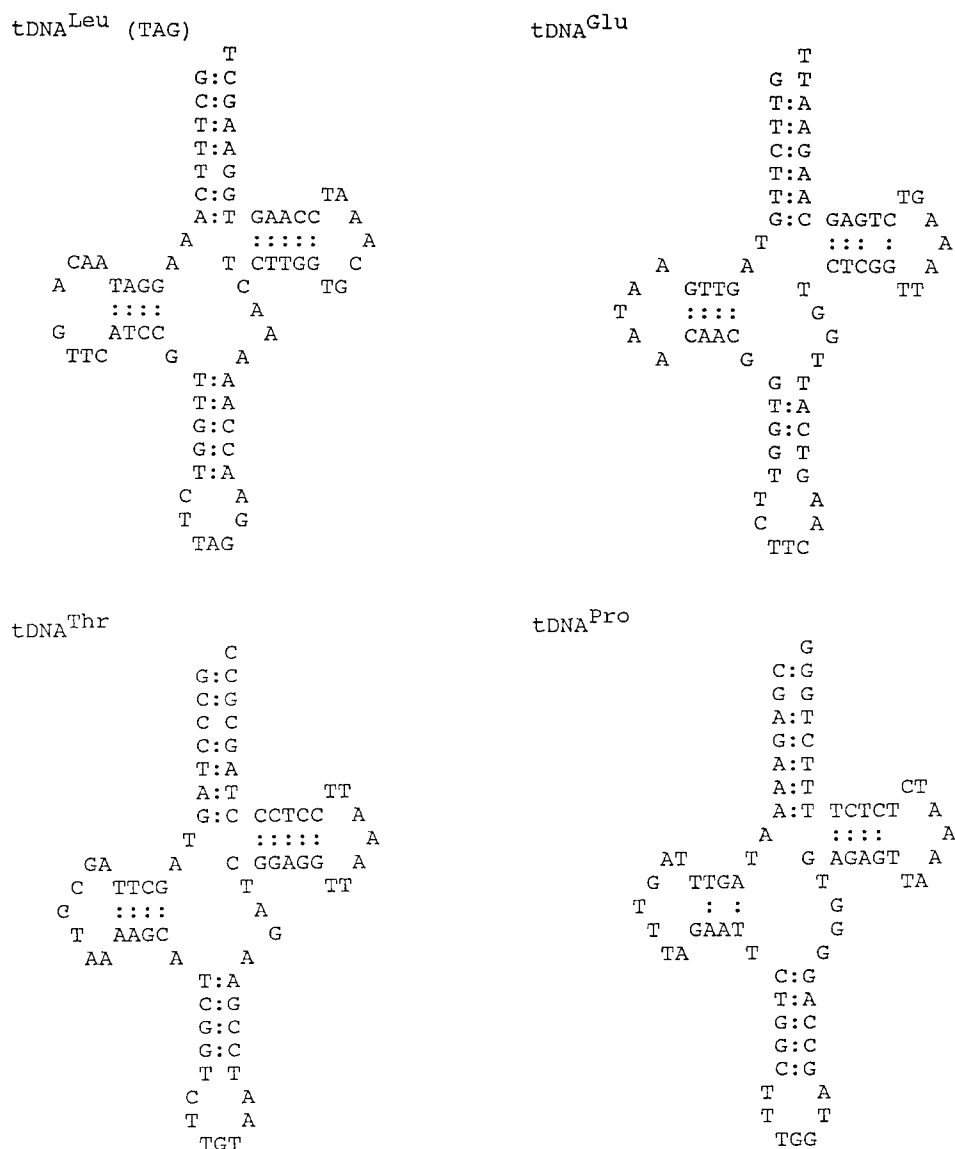


Fig. 5. Continued.

tRNA^{Cys} and tRNA^{Tyr} genes overlap each other by 2 nt, whereas other contiguous tRNA genes are either interposed by intergenic spacers (1–32 nt) or join end to end. Overlapping of these two tRNA genes is also observed in *X. laevis*, chicken, and human mtDNAs, but not observed in mouse and bovine mtDNAs. These observations suggest a possibility of posttranscriptional addition of a nucleotide at the 3' end of tRNA^{Tyr} or at the 5' end of tRNA^{Cys} . This process could occur by mechanisms similar to those involved in the addition of CCA terminus to the 3' end of tRNAs or the 5' end addition of a guanylate residue to tRNA^{His} of the chicken mitochondrial genome (L'Abbe et al. 1990).

In carp mtDNA, the 12S and 16S rRNA genes are, respectively, bordered by the tRNA^{Phe} and tRNA^{Val} gene at the 5' end and by the tRNA^{Val} and

tRNA^{Leu} (UUR) gene at the 3' end. The same organization of these two rRNA gene is also observed in other vertebrates. When these two rRNA genes of carp mtDNA are aligned with their homologous genes of *X. laevis*, chicken, and human, numerous highly conserved regions interposed with variable regions are observed (Fig. 6), suggesting that these constant regions are important for rRNA functions.

Homology analysis reveals that the sequences of the tRNA and rRNA genes encoded by vertebrate mtDNAs are relatively conserved (identity from 54 to 89% for tRNAs and 62 to 74% for rRNA genes) (Table 2). The phylogenetic relationships of these tRNAs and rRNA genes between carp and other vertebrates are not so apparent as that for the proteins encoded by the mtDNAs of carp and other vertebrates.

A

CP: CAAAGGCATGGTCCCGACCTTATTATCAGCTCTAACTCAACTTACACATGCAAGTCTCCGCAACCCA
 XE: T-----TT-----TAG-----A--T--TT--G-----CA-----CT--C
 CH: A- -A-T-A-----TA-----TC-----TG-T-T-TG--AG--A--T-----A-----T-----
 HU: A-T--TT-----TAG-----TC-----T-----T-G-A-GA-----CA--C-GTT--

CP: GTGAATATGCCCTCAATCCCCCAAGG. GGGACGAGGAGCGGCATCAGGCACAAACAT. .TAGC
 XE: -----A-----T--G--T--TT-AA-AG-----TA-----C--T-----CT-A. .--
 CH: -----A-----C--A--TTT-TT--CAA-C-AA-----A--T-----CT--GCAG--
 HU: ---G-TCA-----T-A...AT---ACGATC-AA-----A-AA-----A---GC-G-AATGC--

CP: CAA.GACGCCTAGCC. AAGCCACACCCCAAGG. AATTGAGCAGTGATAAACATTAAAGCCATAAGTG
 XE: -C-T--A--T--TCT. -----A-----G-----G-A. ---G--C-
 CH: -C-A-----T--TT-----C--GT-C-----A--T--C-----A-----
 HU: TC-AA-----T. -----C--G--A. -----T--C--T--A---AC-

CP: ...AAAACCTTGACTCAGTTAGTGTTA. .AGAGGGCCGGTAAACTCGTGACAGCCACCGCGGTTAG
 XE: ACAC---G--C--T-----CA--A-AT. .--A-TT--C--T-----C-----G-----T
 CH: ...T-----T--CC-TA-CA-CCC. ---TT-----T--T--C-----C-T
 HU: ...---GT--A--A--C--TAC-A-CCCC---TT--C--TTT---C-----C-C

CP: ACGAGAGGCCCTAGTTGATATTACAACGGCGTAAAGGGTGTT. AAGGATAAACA. .AAAATAAA
 XE: -----AA-T-A-----C--A--A--A--A. ---. .CCC--C--G-
 CH: --A--AA--A-A-CA--GCTACC-----A---CC. -CATG-T-T-TGCACC-GC--G
 HU: ---TTAA--A--CA--G. -AGC-----A---T--T-GATCACC-CTCCCC--. ---

CP: GTCAATGGCCCTTGGCCGTCATACGCTTCTAGGAGTCCGAAGCCCTAATACGAAAGTAACTTTAA
 XE: -----CTC-AA-CAA--T---GC. -----TCGTTG. . . .TTTGAAG--C--TC-C-A--G-AAC.
 CH: A-T--AT--AA-CAA--T-----A--C-AAGATCCA--T--AC--A-CCCAA-TCCATCT-AGCC
 HU: -CT--ACT-A--GA-TT--A-A-AA--C-AGTTG. . . .CA-AAA-TAGACT-C-A--G-GGCT

CP: TAAACCCACC. . . .TGACCCACGAAAGCTGAGAAACAACTGGGATTAGATACCCCACTATGC
 XE: -CT-----TATTACACT--A-T-----CC--AG-----
 CH: -C--G. -TTAAT. . .T-T-A-----AG--CC-----
 HU: -T--. .-TAT. . . .C--A-A--A-T---A--CC-----

CP: TCAGCCGTAAA. CTCAGACATCCAGCTACAATTAGATGTCCGCCAGGTACTACGAGCATT. .AGCT
 XE: CT-----A-----T-GACTACTT-CGC-A-AA-----A. .-----C-A. ---
 CH: CT-----C-----T--AGATACCT-CCA-CAC-CAT-TA. -----T-A-A-----CAAAC--
 HU: -T-----C-----C--ACAG-TA-ATC-AC-AA-CTGC--. ---AAC-----CAC. ---

CP: TAAACCCCAAAGGACCTGACGCTGTCTCAGACCCCTAGAGGAGCCTGTCTAGAACCGATAACCC
 XE: -----T--G-----CTC--A--A-----GT--T-----C--
 CH: -----T-T-----T--G-----C-C--A--A-----T--T-----T--
 HU: -----T-----G-----CT--T-T--T-----GT--T-----A--

CP: CCGTTCAACCTCACCACCTTCTAGCCACCCAGCCTATATACCGCCGTCGTCAGCTTACCCTG. .TGA
 XE: T--C-A-----T-----AA--C-----A-----C-----CC--TC. .--
 CH: A--A-TC--CA-----CC--T-----G-A-----C-----CC--TCTAA--
 HU: --A-----C--T--. . .T-----A--T-----AA-----A. ---

CP: AGGTAATAAAAGTAAGCAAAATGGGCACAACCCAAACGTCAGGTCGAGGTGTAGCGCATGAAGTGG
 XE: .-AG-T-CTT--G--TT--ATTTTTCAT--C-----A-----AT-----
 CH: .AAG-C--C--G--TC--A-C-C-TCG-T--T-A-A-----A--A-----CT--GG--
 HU: --CT-C--. ---GC--GTAC-. -CGTA--G--T-----A-----C-----G--

CP: GAAGAAATGGGCTACATTTTCTAATA. .TAGAATATTACG. AACATGCACCATGAACAATG. CTT.
 XE: -----ATGGGCTACATTTTCTAATA. .TAGAATATTACG. AACATGCACCATGAACAATG. CTT.
 CH: --. -----CA. -----C-A. ---A-A-G-A. .TG-----CCGCCC--A
 HU: C-----CCCC. ---A-C-----AT-GCCCT. .T-----TTAAGGG-C

CP: GAAGGAGGATTTAGTAGTAAAGGGAAGTAGAGTG. . . .TCCCTTTTGAACCCGGCTCTGAGACGC
 XE: -----C-----C-----GA-A--CA-----AGTT--C--T--AA-----C--GAG--
 CH: -----C-----GT-AG-TC-T-CCCCCTAAG-T-AC--T--GA-----G-A-
 HU: -----T-----C-----CTAAG-----G-TTAG-----AG--C--AG--

CP: GTACACACCGCCGCTCACTCTCCCTGTCAAACGACCAAAAATACATAATACAACAGCACTGAC.
 XE: -C-----C-----TT--A-----AT--A--TTT--. ---CA-T-A--A
 CH: -----T-----C--TT-ACAAGCC-. T--A--TC--A--TATACTTCC-CTCCCGG
 HU: -----C-----T-AAGT-T-CTT--AAGG-C--T. ---CTA--ACC-CTACG-A

CP: ...AAGGGGAGG. .CAAGTCGTAACCGGTAAGT. TACCGGAA. GTGCACCTGGATCAAACC
 XE: C.A--AA-----T-----CAC-----G--CTG-----AT
 CH: CTA--AC-----A-----G-----G--C-----A--CT-C
 HU: TTT-TATA--AGA-----T-----G--T--A-----CG--C

B

CP: GCCAAACAGCTAGCTTAACTACTTAATAACTAAACAATATAAATAAAACAAAATAAGCCTGAC. ACC
 XE: T-----A-C-----A-TC-A-T-AT-AT-AC--TA-CCTC-T--TCTCAT--TTCTAATTAA--
 CH: CTTGCC-CCCCTAGCCCGAC-AAC.TCGTACCCTT--C-----A--CTT-CCTCCCC-TCTT.A--
 HU: --T--. .-CTAGCCCCAA--CC--CTC-ACCTT-C.T-CC-G-CAA-CTT.AG--

CP: AAAAATAAACCATTTTCTTACCTGAGTATGGGAGACAGAAAAGGTTCCACAAAGCGATAGAAATAG
 XE: .-TT-----A.A--C--T-----C-ATCATA-T--. ---T-----A--
 CH: ---CA-T-TAA--G-CCC-GTAT--GCGATAGA-A--CT-CCCCGGCGC--TA--GGCT--CT-
 HU: ---CCA-T.TA-CCAAA-AA-GTAT--GCGATAGA-ATTG--CC-. GGCGC--TA--.

Fig. 6. Comparison of the nucleotide sequences of 12 S rRNA (A) and 16 S rRNA gene (B) of carp mtDNA with those of other vertebrates. Species compared are the same as those shown in Fig. 3. Residues identical to those of carp are indicated by *dashes* and gaps shown by *dots* are introduced to optimize alignment.

CP: TACCGCAAGGGAAAGCTGAAAGAGAAATGAAATAACCCATATAAGCACTAAAAAGCAAAGATTAAAC
 XE: -----T-----A-----T-----TT.-AC-----AC-----G---ACTT--
 CH: -----A-----T---C-----,-----AA---C-----CC--C-
 HU: -----A-----A-----TT-----TA-T-T-----G---C---C-

CP: CTCGTACCTTTTGCATCATGATTTAGCCAGTAC.ACCCAAGCAAAGAGACCTTTAGTTTGAACCCC
 XE: -----A-----G-C-----CAT-AT-----AC--AT--C-----CTA--
 CH: --T-----A--A--A-----T-G-AA-----CCTT--
 HU: -CTA-----C-----A---A---A-T--A-AT-A-TTT---G---G-CAA--C-AAG-C---

CP: GAAACCAGGT.GAGCTACCC.CGAGACAGCCTA.TTATGGCCACCCCGTCTCTGTGGCAAAGAGT
 XE: -----T-A-C.--T-----T-.-TA-AG-A-A-----
 CH: -----CAAGC-----TTG-----TA-AA-T--AG-G-A-----T-----C
 HU: -----AGAC-----TAA--A-----TA-AA...-AG-ACA-----A---A-----T---

CP: GGGAAGAGCTCCGGGTAGAAGTG.ACAGACCTACCGAACCTGGTGATAGCTGGTTGCCTAAGAAATG
 XE: -----T-----A---GG---T-----A-A---G---A-----TC-G-----
 CH: ---T---CT-G-CA-----G---A-AG-----G-TG-----A---GTC-----C-
 HU: -----TT-ATA-----G-C-.-A-----G-----T-C---T-.-

CP: AATAGAAGTTCAGCCTCGTGCACCTCAAATCACACAAACAAACAAGACTAAGAGAAACACATGAGA
 XE: -----T-----TA--CTAAAT-TAGATTT-TTA---TT---G.....-A-AGTCT--T-AG--
 CH: ---CT-----CC---TAAC-CA-C-CCTAA-G---CC--CCTTTGTCAACCTT--G-ACGT--G-G
 HU: ---CTT-----A-T-TAAATTGTC-C..A--G-.-CC-TC..T--AT-CCCTT-T--.-T-A-CT

CP:GTAGTTAAAGGGGTACAGCCCCTTT.AACAAAGGACACAACCTTTCCAGGAGGATAAAG
 XE:T---T-C--TCA-----TGA--G--AC-----T-----A-AATACT--G-----
 CH: TTAAGA-CA-T-CG-T-----T--A-CG--A-.-A-----CCT-CA-C-----TA
 HU:CC--A-A-A-A---T-T--GG-CACT---A-A-----GTAGA--AG---A

CP: ATCATAATACATAAAACATACTGTTCTAGTGGGCCTAAAAGCAGCCATCT.AAACAGAAAGCGTTAA
 XE: --T-----CTTC--GGA-AGT--AGTC-----C--GT--A---C-----C--
 CH: ---CCCCT-CCGC-----T-----TC-----C-AAC--A---GT---C-
 HU: -ATT---C--CC-.....-A-----C-A.-TTA-----C-

CP: AGCTCAGACAGAAAAAGTTTATTATCCCGA...TAAAAATCTTACTCCCCTAAATACTATAGGC
 XE: -----CT--ATC-TTTAACCC--TAATTAGTATA-CT--T---A-A.-C---C-A--C-GA--
 CH: -----CCT--T.....-A-AA---TATTTG---CCCT..
 HU: -----ACACCC-CT-CC.....CC-AA-ATA---C-GAAC-CCTCA

CP: CAACCCATGCCCCCATGGAAGAGATTATGCTAAATGAGTAACAAGAAGGCCCGCCCTTCTTCCAAG
 XE: T-TT-T--AAA-TATA-----CAC-----G--CT---TGT--TACA-GA...-C-T--A
 CH: -----A-A--AGTTGACCTAT--CA--AGA-G---C-A-GCT--A-T-AGTAAT-TGGAAC-T-TC
 HU: .C-----ATTGGA-CAATCTATC-CCC-ATAG--GAACCTA-TGTTAGTATAAGTAA-A-GAAAC-T

CP: AACAAAGTGAAGCCAAACCGGACCAACCATTTGGCACTAACGAACCTCAATCAAAGAGCAATGTGA
 XE: TGT-----AT--G-T--A-TA--T--C--AT--T-----TC--TCCCTG-GATCCTTGCAA-A-
 CH: CT-CTAC-GCGT...-A-TT--ATTAATACATT-TTAACA-----CTT-TACCCC--CAC..
 HU: TCTCCTCCGC-T...-G--T-CGTG-GA...TIA--AC-CT---G-CAATT-ACAGC-C-ATATC

CP: ATTACAAAAAACCCAGAAAAATCCCAACTAAGCTATCGTTACCCCCACACTGGAGTGCCCTAAA
 XE:CCATG--CT--TTACCGTTAAT-TAA---AGAACA-TTA-A...
 CH: ..A---C--T-CGT-TTC--C-ATCT,-----AG--A--C-A---C---AC..
 HU: ..-----TC-A---C--GTCATT-TT--CCTCACTGTC..-A---A---A--CA--T-AT--

CP: GGAAAGACTAAAAGAAAGGAAGGAACCTCGGCAAC.ACAAGCCTCGCCTGTTTACCAAAACAT.C
 XE: -----T-----CGCA-----T-TG-A--C-----T-
 CH: --T--T-----CCT-CA-----C-.-A-C-A-----C-----A
 HU: -----GT-----A-GTAA-----TCTT...-C-----

CP: GCCTCCTGCAACGAACCAA..GTATAGGAGGTCCAGCCTGCCCA.GTGACTAC..AAGTTCAACGGC
 XE: -----T---TTAA--AAC-TT-----A-----A-----GTAT--T-.-
 CH: -----T-A--TAAC-A-----T-A--GAT-----CC-CA-----
 HU: A-----TA--...ATCA-C...TA-----CA-C-----T---T-----

CP: CCGGCGGTTATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAGAGACCTGTATGAA
 XE: C.....ATC.....T-----AAT--G...-TG
 CH: C.....CC-A-----G-----A-----CCG-----T-----T-----
 HU: C.....CCC-A-----AT-----TCC-----G-----

CP: TGCTAAACGAGGGCTTAAGTGTCTCCCTTTCAAGTCAGTGAATGATCTACCCGTGCAGAAGCG
 XE: C---C-CGA,---T.-C-----TGCA--C-A--CA-T---C---C-.-G-----
 CH: -----T-----TG-AGGT-A-TA-----AG-A-T-----A--A-
 HU: -----CC-----T-C-G-----TTAC--TT-AC-----C-G-----A--G---

CP: GGTATAATACTACAAGA.CGAGAAGACCTTTGGAGCTT.AAGGTACAAAACCTCAACCA...CGTT
 XE: --G---GA--C-T-.-A-----A-----T--AC--A-G---GCCAAGTTGAACC-
 CH: A-A--GTG-AC-T-----G---A---T--AA-CACG-C-A-CTTA...CAACC-
 HU: --C---C--AG-----A-----T-TT--TT--TGCA--AG...TACC-

CP: AAGCAACTCAATAAAAAGCAAACCTTG.TGGATCATGAGATTTACCTTCGGTTGGGGCGACCAC
 XE: --C-C-TAAGGA--T--CA-TT--A-AA-CA-A-A-.-CC-AAAGTT-----
 CH: T-CAC-GC-CCACTGGGT-C-CCCA-ACATAAACCC-C--GTCGACATT-----TT
 HU: --CA--...CCAC-GGT.-CT---TA.CCAAACC...CAT-AAA-AT-----T-

CP: GGAGGAAAGAAAGCCCTCC...AGGTGGACTGGGAAACCTC.CTAAACCAA.GAGAGACAT...
 XE: -----A-T-A-----TTG-A-AAT-GG-CCT-CCA-C-TTTC-CAAAG-ACCACC-TTCTAAG
 CH: ---A---A---T-----AACC--A-ACC-C-A---T-C-CTAAG-.CCA-CT-C-....
 HU: ----C-G-ACCC-A-----GA-CA-T.ACATGCT-AGA-.T-C-CCA.GT.C-A-----

Fig. 6. Continued.


```

CP: ...CTCTAAGCACAGAACA.TCTGACCAAATATGATCCGGCTAACACATAGCCGATCAACGAACCAA
XE:TAA-AAA-TTT-T.--CT-TAA-TGATCC-.....GTCCTA-T-----
CH:...-AAAGTA-CA-C-GT-AC-A---C-----T---ATT-G---T-G-----
HU:...-GAAGTA-TA.T-CTCAAT---T-C-----ATA-CTT--C-----G-A---

CP:GTTACCCCTAGGGATAACAGCGCAATCCTCTCCAGAGTCCATATCGACGAGGGGGTTTACGACCTCG
XE:-----AT-T-A-A---T-C-----A-AT-----
CH:-C-----C-----TC---A---C-----A---A-----
HU:-----A-T-T-----A-A-TA-----

CP:ATGTTGGATCAGGACATCCTAATGGTGCAGCCGCTATTAAGGGTTCGTTTCAACGATTAA.AGT
XE:-----G---C-G-----C---A-----C---
CH:-----A-----A-----C---
HU:-----CG-----A-----

CP:CCTACGTGATCTGAGTTCAGACCGGAGCAATCCAGGTCAGTTTCTATCTG..TAACGCTACTTTTCC
XE:-----T-----A..G-A..-T---T---
CH:-----G-----A..GGAC..-CC---
HU:-----T-----G-----ACC-TCAA..-T-CC---

CP:TAGTACGAAAGGATCGGAAAGAGGGGCCCATACTTAA.AGCACGCCCCACCC..CTAATTTATGAA
XE:-----C--A-----AT-A---A--GT--.TA-TA-GC-T-TCT-TATA--.CA---T-
CH:-----C---G---T---T-A---CACTG-----A---A--TT---GCA-----
HU:CT-----CAA--G--ATAA---T-CTTCAC-A---G-CTT--C--GT.....AA---T

CP:AACAAATAAATAAAATAAGGAGAGCCAAATCCAGCTGGCCAAATAAGGACA
XE:GT---C---TG--A-T--AACT.....TA-T-T---G--T---G-TAG
CH:T-----C--CTGCCA-G-ACCCCTC--CC-CA---GAAGT..-T-G-A---TCCA
HU:-T--TC-C--CTT-G--TTATACC.....C-CA---C-CAAGA-C-GGGTTT

```

Fig. 6. Continued.

The D-loop Region and the Origin of L-Strand Replication

The D-loop region of carp mtDNA is 928 bp in length and its base composition is rich in A and T (65%). Its length is similar to that of other vertebrates but is shorter than that of *X. laevis*. Sequence analysis of vertebrate mtDNAs has revealed that the D-loop region is the most rapidly evolving part of the genome. However, the midportions of the D-loop region of the carp mtDNA (map position 310–530) and cod mtDNA (map position 450–674) are relatively conserved (60% identity), particularly the last one-third part. A relatively conserved sequence is also noted in the midportion of the D-loop region along the mtDNAs of carp (map position 326–382), mouse (map position 336–392), and human (map position 446–501). For the above sequences, the identity between carp and mouse is 67% while that between carp and human is 74%. No such homologous sequence can be identified in the corresponding region of *Xenopus* and chicken mtDNAs. In addition, the three conserved sequence blocks of *X. laevis* and mammals can be identified in the D-loop region of carp mtDNA (not shown). No repeated sequences of any length, as present in cod, *X. laevis*, and chicken, are found in the D-loop region of carp mtDNA. The origin of H-strand replication has not yet been mapped in the D-loop region of carp mtDNA.

A putative termination-associated sequence (TAS) of TACATATTAT is identified in the 3' end of the D-loop region of carp mtDNA (map position 28–36). This sequence is identical to that of pig (McKay et al. 1986) and homologous to those of other vertebrates (Foran et al. 1988). Furthermore, a putative promoter for H-strand transcription

(HSP) of the sequence of ACCAAAAATC-CCCAAAAAAGA is noted at 55 nt upstream of the 5' end of tRNA^{Phe} gene. This sequence is homologous to that of human (Chang and Clayton 1984) but different from that of *Xenopus* (Bogenhagen et al. 1986). The promoter of L-strand transcription has not yet been identified in the D-loop region of carp mtDNA.

A structure located between the tRNA^{Asn} and tRNA^{Cys} gene of carp mtDNA similar to that of the origin of L-strand replication of cod, *X. laevis*, and mammals can be identified. This sequence has the potential to form a stable stem-loop structure. In carp, this structure consists of a stem of 11 bp, a loop of 14 nt, and a 5' flanking sequence. This structure is involved in the transition from RNA to DNA synthesis during replication initiation (Wong and Clayton 1985). The sequence of the origin of L-strand replication is well conserved in the vertebrate mtDNAs, but is lacking in chicken mtDNA.

References

- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:475–485
- Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1982) Complete sequence of bovine mitochondrial DNA, conserved features of the mammalian genome. *J Mol Biol* 156:683–717
- Araya A, Amthauer R, Leon G, Krauskopf M (1984) Cloning, physical mapping and genome organization of mitochondrial DNA from *Cyprinus carpio* oocytes. *Mol Gen Genet* 196:43–52
- Bibb MH, van Etten RA, Wright CT, Walberg WM, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26:167–180

- Bogenhagen DF, Yoza BK, Cairns SS (1986) Identification of initiation sites for transcription of *Xenopus laevis* mitochondrial DNA. *J Biol Chem* 261:8488–8494
- Cantatore P, Gadaleta G, Roberti M, Saccone C, Wilson AC (1987) Duplication and remoulding of tRNA genes during evolutionary rearrangement of mitochondrial genomes. *Nature* 329:853–855
- Chang DD, Clayton DA (1984) Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell* 36:635–643
- Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecules of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J Mol Evol* 22:252–271
- Davidson WS, Bartlett SE, Birt TP, Green JM (1988) Cloning and sequence analysis of an Xba I fragment of rainbow trout mitochondrial DNA. *Curr Genet* 14:483–486
- Davidson WS, Birt TP, Green JM (1989) Organisation of the mitochondrial genome from Atlantic salmon (*Salmo salar*). *Genome* 32:340–342
- Desjardins P, L'Abbe D, Lang BF, Morais R (1989) Putative chicken "muscle-specific 7 S RNA" is related to the mitochondrial ATPase 6 gene. *J Mol Biol* 207:625–629
- 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
- Dunon-Bluteau D, Brun G (1986) The secondary structures of the *Xenopus laevis* and human mitochondrial small ribosomal subunit RNA are similar. *FEBS Letters* 198:333–338
- Foran DR, Hixson JE, Brown WM (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Res* 16:5841–5861
- Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisa 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
- Gilbert TL, Brown JR, O'Hara PJ, Buroker NE, Beckenbach AT, Smith MJ (1988) Sequence of tRNA^{Thr} and tRNA^{Pro} from white sturgeon (*Acipenser transmontanus*) mitochondria. *Nucleic Acids Res* 16:11825
- Huang CJ, Huang FL, Chang YS, Tsai YJ, Lo TB (1990) Nucleotide sequence of carp mitochondrial cytochrome C oxidase III. *Nucleic Acids Res* 18:1056
- Johansen S, Guddal PH, Johansen T (1990) Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. *Nucleic Acids Res* 18:411–419
- L'Abbe 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
- MacKay SLD, Olivo PD, Laipis PJ, Hauswirth WW (1986) Template-directed arrest of mammalian mitochondrial DNA synthesis. *Mol Cell Biol* 6:1261–1267
- Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondria DNA. *Cell* 22:393–403
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474
- Roe BA, Ma DP, Wilson RK, Wong FH (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J Biol Chem* 260:9759–9774
- Sanger F, Nicklen A, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Stormo GD, Schneider TD, Gold LM (1982) Characterization of translational sites in *E. coli*. *Nucleic Acids Res* 10:2971–2996
- Tapper DP, van Etten RA, Clayton DA (1983) Isolation of mammalian mitochondrial DNA and RNA and cloning of the mitochondrial genome. In Fleischer S, Fleischer B (ed) *Methods in enzymology*. Academic Press, New York 97:426–434
- 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