DNA sequence of the Xenopus laevis mitochondrial heavy and light strand replication origins and flanking tRNA genes

J.F.H.Wong, D.P.Ma, R.K.Wilson and B.A.Roe

Chemistry Department, University of Oklahoma, Norman, OK 73019, USA

Received 2 May 1983; Revised and Accepted 27 June 1983

ABSTRACT

We have determined the primary structure of the two regions of the <u>Xenopus laevis</u> mitochondrial genome which encompass the origins of heavy (H) and light (L) strand replication. The first segment, which consists of 2398 nucleotides, contains the displacement loop (D-loop), the tRNA genes for threonine, proline and phenylalanine, the origin of H-strand replication, and the promoters of H- and L-strand transcription. The second segment, which consists of 447 nucleotides, contains the L-strand replication origin flanked by the tRNA genes for tryptophan, alanine, aspragine, cysteine, and tyrosine. A comparison of the sequences of the <u>Xenopus laevis</u> mitochondrial L-strand replication origin region and the eight tRNA genes with their counterparts from the mammalian mitochondrial genomes reveals that these regions are quite homologous, while its D-loop region shows only slight homology with those of the mammalian mitochondrial genomes.

INTRODUCTION

All eucaryotes examined to date contain a double stranded mitochondrial specific DNA which codes for several mitochondrial components. In higher vertebrates, the mitochondrial DNA ranges in size from 14.5 to 19.5 kb (1) and codes for the 22 tRNAs and 2 ribosomal RNAs required for the mitochondrial protein synthesizing system. In addition, these genomes code for at least five known polypeptides (3 subunits of cytochrome C oxidase, ATPase subunit 6, and cytochrome b) and may code for an additional eight polypeptides in the other available unidentified open reading frames (URFs) (2-4).

The complete nucleotide sequences of the human, bovine, and mouse mitochondrial genomes (2-4), and the restriction maps and partial nucleotide sequences of several other higher eucaryote mitochondrial genomes have been reported (5-12). These studies reveal that higher eucaryote mitochondrial DNAs have common elements of overall gene organization but differ slightly in their nucleotide sequences. In contrast, the <u>Drosophila yakuba</u> mitochondrial ribosomal RNA genes and the origin of replication occupy similar relative positions (13-15), but the overall gene order differs from that observed for the mammalian mitochondrial genome (16).

Higher eucaryote mitochondrial DNAs replicate by a mechanism in which each strand contains its own, physically distinct replication origin (17). Heavy strand (H-strand) replication begins in the displacement loop (D-loop) region while light strand (L-strand) replication does not begin until approximately two-thirds of the H-strand has been replicated (18-19). This asynchronous mechanism of mitochondrial DNA replication requires the displacement of the H-strand by synthesis of a complementary DNA fragment to form a D-loop structure which is flanked by the genes for tRNA^{Phe} on one side and the genes for tRNA^{Thr} and tRNA^{Pro} on the other. The H-strand replication proceeds from this D-loop region without concomitant L-strand replication until it encounters a cluster of five tRNA genes with the first two separated from the last three by a G-C rich stem and T rich loop. It is here that the L-strand replication begins (20).

The <u>Xenopus</u> <u>laevis</u> mitochondrial genomic D-loop region contains a hybridized 14S single stranded DNA initiation segment, while the mammalian mitochondrial genomic D-loop regions contain a 7S single stranded DNA initiation segment. The amphibian 14S DNA consists of at least two species, 1350 and 1510 nucleotides in length (21). The mammalian 7S DNAs are smaller than their amphibian counterparts and are in the size range of between 500 and 630 nucleotides (21-22). Sequence analysis (2-4,9), electron microscopy (23), and direct isolation of expanded D-loop structures (20) have shown that the mammalian mitochondrial D-loop regions map in the segment flanked by the genes for tRNA^{Pro} and tRNA^{Phe}, and span 1122 nucleotides in the human, 910 in the bovine, 879 nucleotides in the mouse, and at least 717 nucleotides in the rat mitochondrial genome (2-4,9). In the <u>Xenopus laevis</u> mitochondrial genome the origin of H-strand replication has been located by electron microscopy studies (24) and restriction enzyme mapping (5).

In this report we present the complete nucleotide sequence of the region of the <u>Xenopus</u> <u>laevis</u> mitochondrial genome which begins at the tRNA^{Thr} and tRNA^{Pro} genes, continues through the 2134 nucleotide D-loop region and extends into the 12S ribosomal RNA gene immediately following the tRNA^{Phe} gene. In addition, we present the complete nucleotide sequence of the region which begins in URF2 and extends into the cytochrome C oxidase subunit I (CoI) gene. This region also contains the tRNA^{Trp} and tRNA^{Ala} genes, the origin of L-strand replication, and the tRNA^{Asn}, tRNA^{Cys} and tRNA^{Tyr} genes. Although there is little homology among the D-loop regions of the <u>Xenopus</u> <u>laevis</u> mitochondrial genome and those of other higher eucaryotes, the few regions of sequence similarities could represent segments which perform similar functions. In contrast, the regions encoding the eight tRNA genes, a portion of the 12S ribosomal RNA gene, a portion of the cytochrome C oxidase subunit I gene and the origin of L-strand replication show significant homology with their counterparts from other higher eucaryotes.

MATERIALS AND METHODS

A clone containing the entire <u>Xenopus</u> <u>laevis</u> mitochondrial genome inserted into the BamHl site of pBR-322 (pXlm-31) and transvected into <u>E</u>. <u>coli</u> strain Hb-101 was obtained from Dr. I. Dawid (NIH, Bethesda, Md.). The complete mitochondrial genomic insert was excised by restriction endonuclease cleavage with BamHl, purified by preparative electrophoresis on low melting 0.7% agarose gels, eluted by a modified freeze-thaw method (25), and concentrated by ethanol precipitation. After further digestion with selected restriction endonucleases, fragments of the <u>Xenopus</u> <u>laevis</u> mitochondrial genome were ligated into either M13-mp8, -mp9, -mp10 or -mp11 (26). In some instances the linearized genome was treated with EcoRl to cleave the DNA into three fragments, one of which encompassed the D-loop region (5,24). This fragment was gel purified as described above, treated with nuclease Bal-31 for short time periods, and blunt end ligated into the SmaI site of M13-mp9. After transvection into <u>E</u>. <u>coli</u> strain JM-101, the single stranded recombinant phage DNAs containing fragments of the <u>Xenopus</u> <u>laevis</u>



Figure 1. The restriction sites and sub-clones used for sequencing the regions of <u>Xenopus</u> <u>laevis</u> mitochondrial genomes which encompass the origins of H- and L-strand replication and their flanking genes.

	120	CAGTTTGGA				360 Ftacaacat Aatgtttgta	480 Attaatctat Taattagata	600 1440111441 111644114	720 CTCGCATATC GAGCGTATAG	840 TTATAATATA AATATTATAT	960 AAGATTTTT TTCTAAAAAA	1080 TACCATACGT ATGGTATGCA	1200 Gaacaaaata Cttgtttat	1320 TAATGTGTA Atttacacat	1440 Aggattatta Tcctaataat
	110	CTTTCTTCC 1				350 NTAAATTTA Fattttaat	470 ATAATTTAC TATTAAATG	590 84TTTCTATG TTAAGGATAC	710 CCTCATTTCT GGAGTAAAGA	830 TCCATCAAAA AGGTAGTTTT	950 Tatgaaaaat Atactttta	1070 CTATGTTTA GATACAAAAT	1190 Catataataa Gtatattatt	1310 Atacgtactt Tatgcatgaa	1430 AATTCAATTA TTAAGTTAAT
	100	LACTAAATCG A LTGATTTAGC 1 A	20 20 20 21 21 20 20 20 20 20 20 20 20 20 20 20 20 20			340 TCAACAATTT 1 Agttgttaaa	460 Acaagaaaat Tgttctttta	580 TTTTACAAAC AAAATGTTTG	700 Atataattaa Tatattaatt	820 Agcaaacaaa Tcgtttgttt	940 TTCCAAAAAT AAGGTTTTTA	1060 TGTACATCTC Acatgtagag	1180 1417446474 4744776747	1300 TAATGTATGA Attacatact	1420 AATAACTGTA TTATTGACAT
	6	ACAATCGTTG A	4 3444676 210 210 ACATTCTATG TGTAAGATAC			330 Catcaagaat Gtagttctta	450 Atgaaacat Tacttttgta	570 570 55655555555555555555555555555555	690 Atcaaccaaa Tagttggttt	810 Atataactg Tatattgac	930 Gaaatatatt Ctttatataa	1050 Atattccagt Tataggtca	1170 1414447764 414777467	1290 Gatcaataaa Ctagttattt	1410 AACTTTAGCT TTGAAATCGA
	80	ATTGGGGGGGTCA TAACCCGTCA	Repeat 200 200 76CTCTATAT ACGAGATATA			320 Tacaattaaa Atgttaattt	440 CCATATTATT GCTATAATAA	560 TTTCTCAATA AAAGAGTTAT	680 Caacaataaa Gttgttattt	800 Acatacttta Tgtatgaaat	920 Ctaaaactt Gattttgaa	1040 Ataccatttc Tatggtaagg	1160 CGTATTAACT GCATAATTGA	12 8 0 AAACATATAA TTTGTATAT	1400 Tgataaacaa Actattgtt
•	70	CCTCAAGACT GGAGTTCTGA	190 6104160144 6461406411		•	310 GAATAATATT CTTATTATAA	430 431 43144444444444444444444444444444	550 Acaccaataa Tgtggttatt	670 CCCAACACTA GGGTTGTGAT	790 Atgcatgaac Tacgtacttg	910 TAAAAATTAT ATTATATATA	1030 Acaattataa Tgttaatatt	1150 Catattatac Gtataatatg	1270 Tataacaat Atatttgtta	1 390 Acastctccs Tstcasassc
	09	CTAAAACCCT Gattttggga	180 180 1816cactaa Atacotgatt		**********	300 ATTTCCCCAC TAAAGGGGTG	420 44171744 7774747777	540 540 77774675	660 Agcatacatc Tcgtatgtag	780 Aagaaaatga TTCTTTTACT	900 Cataattact Gtattaatga	1020 AAAATACTAA TTTATGATT	1140 TCCAAAACAA Aggttttgtt	1260 Tacataaaaa Atgtatttt	1380 1767738788 8868877877
	50	AAGATTGAGG TTCTAACTCC	170 170 170 170 170 170 170	2		290 Cataaattta Gtatttaaat Gtatttaaat	410 ATAATAAGT TATTATTGA	530 804777480 7674447767	630 TAATATATCA Attatatagt	770 Tatttcttca Ataagaagt	890 CCACTAATCC GGTGATTAGG	1010 ATGAATATGC TACTTATACG	1130 Atcgtatact Tagcatatga	1250 Attataattt Taatattaaa	1370 TAATATATA Atttatata
	40	TTGTAAGCCG Aacattcggc	160 1150 160 160 160 160	CRH		280 GTATAATGAG Catattagtg Catattagtg	400 1111466161 111116161	520 520 50100000000000000000000000000000	640 66taaaaaa Ccatttttt	760 Cataataact Gtattatga	880 Caattctcaa Gttaagagtt	1000 644774444C 6774477776	1120 1176 1120 120 120 120 120	1240 Ataacattc Tatttgtaag	1360 1347145 1350
	30	AGCATCGGTC TCGTAGCCAG	150 Caacattcta 61161a641		ted Sequenc	270 TACATTCTAT Atgtaagata	390 AATCAATCTT TTACTTAGAA	510 Cacatatata Gtgtatatat	630 Taaccataac Attggtattg	750 Acatcagtta Tgtagtcaat	870 Attccactac Taaggtgatg	990 CTCACTAACT Gagtgattga	1110 1110 1114010000 1111010100000	1230 1230 121100000000000000000000000000	1350 Ataatactaa Tattatgatt
•	20	CTTAATTTAA Gaattaaatt	140 CCCCCCAAAGC CCCCCAAAGC		le yet	260 TATGGTTATA ATACCAATAT	380 AAACTGAGCA	500 500 514TATCATAA	620 AAGTATATGT TTCATATACA	740 CCATTATAAC GGTAATATG	860 111111111111111111111111111111111111	980 844447477 777777777777777777777777777	1 100 TCCATATATC AGGTATATAG	1220 AAAATACAAA TTTATGTTT	1340 Caaatccata Gtttaggtat
Ĩ	ERNA 10	GTCCTGATAG CaggaCtatC	1 30 CCACTATTGA GGTGATAACT			250 TAACATATTT ATTGTATAAA	370 TAATTAACTA	11000000000000000000000000000000000000	610 TCTCTATATA Agagatatat	7 30 TGTATAATAA Acatattatt	850 AAAACAATTT TTTTGTTAAA	970 Ataacaacaa Tattgttgtt	1090 Actgtttcac Tcaccaagtg	1210 TTAATTAATGA AATTATTGC	1 3 3 0 ATCGTACATA TAGCATGTAT

1560 Staccataac Catgetatte	1680 CTCGAAATAC CAGCTTTATC		AATAGGGTGG TTATCCCACC	1920 CGTTTCTCTA GCAAAGAGAT		2040 CAAAACCGAA GTTTTGGCTT BBBBBBBB	2160 TAGAATAGGG Atcttatccc	2280 2260 2280 2280 2280		2398 TTGAATTC Aacttaac
1550 rgaatctgcc acttagacgg	1670 CTCTCCTATG Gagaggatac		CAGATTACGT GTCTAATGCA	1910 Sttcttcagt Caaggagtca	C98-1	2030 CGTCAAACCC GCAGTTTGGG AAAAAAAA	2150 Atgaactgac Tacttgactg	2270 Tacacattat Atgtgtaata		2390 Getectage Ccaggatege
1540 CAAGATGCCC GTTCTACGGG	1660 1660 1660 1660		CTTAACAGTG CAATTGTCAC	1900 Atcagaacgt Tagtcttgca		2020 TTGTATTTTC Aacataaag	2140 TAGCCGTGAT ATCGGCACTA	2260 Ctatttccat Gataaaggta	-44 12	2380 ATAAAGGTTT TATTTCCAAA
1530 15704700 15404700	1650 167674477C Acacattaag	lication	TCAAAGTTGG AGTTTCAACC	1890 TTTACCCTTA AAATGGGAAT		2010 CTGTAAAACC Gacattttgg	2130 CCTGGACTGT GGACCTGACA	2250 Tagcacaaca Atcgtgttgt		2370 CTCCGTAAGC GaggCattCg GaggCattCg
1 520 CCTGGAGTGA GGACCTCACT	1640 Attagtccct Taatcagga	-etrand Rep	1/60 AATAGCATCT TTATCGTAGA	CSB-1 000000 1580 1580 1580 1580 1580 1580 1		2000 TTAGCTTTTC AATCGAAAAG	2120 Agacatttgt Tctgtaaaga	2240 6tctttccac cacaaa66t6		2360 CCTACGAAG GGATGCTTTC GGATGCTTTC
1510 CTTTACCCAA GAATT000TT	1630 TTAACCTAAG AATTGGATTC	Origin of N	1 / 50 TGGGATCTTC ACCCTAGAAG	1870 AAATGCAATG TTTACGTTAC	-2 1-2	1990 Accccccaaa T000000777	2110 60 T6T6T8G8 668C8T8TCT	2230 AACTATATTA TTGATATAAT		2350 TGAGATGAGC Actctactgg
1 500 CCCCTCCTCA CCCCTCCTCA	1620 T 67767 6466 Acaacact66	tion of the	1740 TTTTTTTCT AAAAAAAGA	1860 Cttaaaatg Caattttag	CSI	1980 TAAACCCCCT ATTTGGGGGA	2100 TCCTAACTTA AGGATTGAAT	2220 Acgtgataaa Tgcactattt		2340 CTGAAAATGC Gacttttagg
1490 CT00TTTCAA GACCAAAGTT	1610 6ta6t6aatc Catcacttag	21	1730 CTTCATGTTT GAAGTACAAA	1850 1850 1850 1850 1850		1970 TTATTGTGTG AATAAGAGAG	2090 TCCTAAGAGT AGGATTCTCA	2210 ATCGTATATA TAGCATATAT		2330 AA6CACA6CA TTC6T6T6T
1480 1480 1480 1480 1480	1600 80atct6t66 00ta6aca60		1720 ATTTAA60C TAAATTTCCG	1840 CGAAGTATTA GCTTCATAAT		1960 Taaagacgt Attttctgca	2080 Taaccaaaac Attggtttg	2200 TTGTATATAT AACATATATA		2320 29077144074 7694477647
1470 ACCTAATTCT TGGATTAAGA	1590 1590 1590 150417777		1710 CTTGTTTCTA GAACAAAGAT	1830 AAGACGTACA TTCTGCATGT		1950 AATTTTCTAA TTAAAGATT	2070 CAATTTATTC CTTAATTAG	2190 TTATACATTT AATATGTAAA		2310 CTGCTTACGT GACGAATGCA
1460 TCCACAATTA AGGTGTTTAAT	1580 17ACTTATGA 4ATGAATAGT		1700 TAACGCCAAGC ATTGCGTTCG	1820 CTTATCGGCT GAATAGCCGA		1940 Accessetes Tescccacc	2060 Agtaaaaac Tcattiig	2180 Caaacccatg Gtttgggtac		2300 TCTTTTGTA AGA AAAACAT AGA AAAACAT
1450 1450 1450	1570 CTAGGCCGGA GATCCGGCCT		1690 ATCTTGTCCT TAGAACAGGA	1810 1810 6464t4t66t 676t4t4666		1930 TTATTATCTC AATAATAGAG	2050 2050 2050 2050	2170 TGCCATTACC ACGGTAATGG		2290 TTAATTAATA AATTAATTAT AATTAATTAT

Figure 2. DNA sequence of the region of the <u>Xenopus</u> <u>laevis</u> mitochondrial genome encompassing the D-loop and its flanking genes.

mitochondrial genome were isolated by phenol extraction of the polyethylene glycol concentrated phage (27).

All DNA sequences were obtained by the dideoxynucleotide chain termination method (27-28). The M13 sub-clones were selected first at random and later by hybridization to overlapping fragments cloned in M13-mp8 or M13-mp9 in the opposite orientation. All inserts were sequenced using a flanking universal primer (29). The complete sequence of contiguous regions was assembled from individual, overlapping sub-clones containing fragments in either orientation, using the programs described by Staden (30-32) but modified to run on an IBM-3081 computer. Copies of these modified programs are available to others upon request. The optimal nucleotide sequence homology among the higher eucaryote mitochondrial DNA D-loop regions was determined by the NUCALN program of Wilbur and Lipman (33). All recombinant DNA experiments were performed in accordance with NIH guidelines.

RESULTS AND DISCUSSION

<u>Sequencing approach</u>. The restriction maps and sequencing strategies for the segments of the <u>Xenopus</u> <u>laevis</u>, whose sequences now are reported, are shown in Figure 1. The restriction endonuclease digestion sites predicted by the nucleotide sequence data shown in Figures 2 and 4 were confirmed by restriction enzyme mapping experiments (data not shown) and are in agreement with those presented in Figure 1 and reported earlier (5).

<u>H-strand replication origin and its flanking genes</u>. The complete nucleotide sequence of the <u>Xenopus</u> <u>laevis</u> mitochondrial DNA segments corresponding to the region surrounding the origin of H-strand replication is shown in Figure 2. Here the genes for $tRNA^{Pro}$ and $tRNA^{Phe}$ are separated by 2134 nucleotides, an amount almost twice as much as that separating these genes in other vertebrate mitochondrial genomes (2-4). The genes for $tRNA^{Thr}$ and $tRNA^{Pro}$ at the 5' end of this region are separated by 27 nucleotides in the <u>Xenopus</u> <u>laevis</u> mitochondrial genome, while they either overlap or are separated by only two nucleotides separating these two tRNA genes and the larger size of the non-coding region between the $tRNA^{Pro}$ and the $tRNA^{Phe}$ genes account for over half of the increased size of the <u>Xenopus</u> <u>laevis</u> mitochondrial genome to the mammalian mitochondrial DNAs (2-4).

An analysis of the D-loop region also reveals the presence of two large repeated sequences. The first occurrence of this repeated sequence begins 28

	Thr								•					
1		0	20	30	40	50		60	70	80	06	100	110	
н ј														
						111								
à	• GTCTTTGTA	BTACA-T-CI	FATATACT	GOTCTTGTA	AACCAGAG	AAGGAGAAA	CAAC-TAAC	CTCCCTAA0	ACt			TCAAGGAI	AGAAAC	-160
á	· GTCTTGATA	STATAAAC	ATTACTCT	COTCTOTA	AACCTGAA	ATGAAGAT	TTCT	CTTCTCAAG	ACA			TCAAGAAG	SAAGGAG	C-TA
					14 14									
	120	130	140	150	160	-	70	180	190	200	210	220	230	
Ň	. ACCTCCACTA	ATTOACCCCC	AAAGCCAA	CATTCTAAT	TTAATTC	TCTCCCGA	CGTATGCAC	TAAGTCATG	CTAATGCT	CTATATACA1	TCTATGTAT	ATGAGCATA	ATTTATI	LCCC
Ă	" AACTCCACCA	VTTAGCACCC	AAGCTAA	GATTCTAAT	TTAACTA	TTCTCT0-								
		111111 11	1 1111	1 1 1 1 1 1	1 11111	1 1 1								
ŏ	D AVICICACCA		AAACCTCA 11111 AAACCTCC	TATTCTAA -1 1111111 TATTCTAA -	TTAACTA									
i	240	250	260	270	280	7	06	300	310	320	330	340	350	
ň	. CACGTAACAT	ATTTATOG	TTATATAC.	ATTCTATGT	ATAATGAG	CATAAAT	FATTCCC	CACGAATAA	TATTACA	ATTAACATC	AAGAATTCA	ACAATTTATA	AATTTAT	FACAA
à	1													
Å														
	360	370	380	390	404	4	10	1						
Ř	ACATTAAATA	ACTAAAGT	GAGCAAAT	CAATCTTTA 11	TAACATGT	ATAATAAA i i	CTAAATATA : :	I						
ž				¥1	TTCTCTC1	TCTTTCAT	BGGGAAGCA							
å							1	TCCATAAAT	ACAAGAG	CCTTATCAGI	CATTAAATTT	ATCAAAATC	CCAATAAC	TCAAC
ŝ														
,	420	430 777474466	440	45	TACAGAA	460 447477	470 7774CATTA	480 ATCTATTTC	490 4400	500 ATATCATAA	510 510	520 TGAAGACTTA	53 ACATTTAA	0 Catat
						11				11 1		1		
		ULACCACC	LANUALT	SACTURACCU	ATCANCAN						111111	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
å	ACAGAATTTO	CACCCTAAC	CAATATT	ACAAACACC	ACTAGCTA	ACATAACA			CATAC	ACAGACCACI	1GAATGAATT	ACCTACGCAA 111		
â												AGTA	-CATAA	
Ř	540 AAATCTCACA(550 CCAATAATT	560 TCTCAATA/	570 LACATAATA	TTTTACA	80 AACAATTT	590 CTATGTAAC	600 TTTAATTCT	610 CTATATAA	620 Agtatatgt1	630 Faaccataac	640 667aaaaaa	65 .taatatat	0 CAAGC
h u														
å														
ì														

	<pre>x AIACATCCCCGAAGACTACAACAACCAACCAAAATATAATTAACCTCATTCTCTCCCCGCATATC 111111111111111111111111111111111</pre>	FUTATAATAGCGATTATAACGGGATCAUTAGGTAA 1111 111 111 Asccccctccccatactacgaacaatacagcaat	FCAACCTCAACTATCAAC
		ITTATATOCCCCATGCATATAACCAAGTACTCACC	CTCTATAG
2	1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1	ICTATTTTCCCCAAGCATATAAGC4AGTACATTAAL	I III ATCAATGGTTCAGGTCAT
	780 790 810 820 810 820 8	10 840 850 850 860	870 880 TCCACTACCAATCTCAA
	<pre># AAAATGAATGGAAGGCAGCGCGCGCGCGCGCGAGAAGCTAGCGAGCAGCGCGCGC</pre>	OTACATAGTATAAGCCATTTACCGTACATAG	CACATTACAGTCAAATCC
• •	• •	NOTACATAATACATAATTATTGACTOTGCATAG 111 1 1 111 FATAAGGACATATCTGGGTTATCTGACATAG	TACATTAT-GTCAAATTC 111 111 Accatacagtcataaact
		950 960 970 970 970 970	980 990 Atatetcactaac
	<pre>% CTTCTCGTCGTCGTCGCCCCCCTCGGTAGGGGGTCCCTTGACCGCCACCATCCTCCGT % 111 1 111 1 1111 1 1111 1 1111 1 1111 1</pre>	g AATCA	ATATCCCGCACAAGAGTG
•	<pre>> ATTCTTGATAGTATATCTTATATTCCTTACCATTAGATCACGAGCTTAATTACCATGCCGCGT :::::::::::::::::::::::::::::::</pre>	8	CCAACCCCCTAGGCAGGG
	CTTCTCTTCCATATGACTATCCCCTTCCCCATTTGGTCTATTAATCTACCATCCTCCGT	3AAACCA	ACAACCCGCCCCACCAATG
	1000 10 	LO 1020 1030 1040 Focaaaatactaaacaattaccatttcat	1050 1060 Attccagttgtacatctc
	<pre>i ctl-ctctcctcgctcgggcccataacacttggggggtaggctaaage-tgaa-ctgtatccgacatcatc</pre>	FI FIFF FIFFFFFFFFFFFFFFFFFFFFFFFFFFFF	
	IIII IIIIIIIIIIIIIII I IIIIII I IIII		£
	: :::: :::::::::::::::::::::::::::::::	IIIII IIIIIIIIII III II Fggttgttagtgggggggtgatgaatgggttatgg	
-	111 1111 1 11111 111 1 111 1 111 1 1 1 1	IIIIII IIIIIIIII I I I Egettettagegggecateattggtteate	
	1070 1070 1080 1090 1100 1110 1120 CTATOTITATACCATAGTACTOTTCACTCATAATACCAGTTCATTATTATCGT	1130 1140 1150 1160 Atacttccaaaacaacatattataccgtattaact	1170 Tataattcatattaaca
	111 1111 1 111 1 1 1 1 1 1 1 1 1 1 1 1	:: : : : : : : : : : : : : : : : : : :	1 11 11 11 11 11 11 11 11 11 11 11 11 1
		I III I IIII I IIII IIIIIIIIIIIIIIIIII	
		CATGACCAACATAACTGTGGGGGGGTGTCATGCATTIGG 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	EXECTITIATITEGCCE
	1190 1220 1210 1220	1230 1240 1250 126	0 1270
	TACATATAATAATAACAAAA	atgaataataagattgattataattttagataaa ::: ::::	LAATATAAACAATAAACAT
	TOCACGCGATAGCATTGCGAGACGCGGGAGCAGCACCCTATOTCGCAGTATCTGTCTT-TGATT	CCTGCCTCATCCTATTATTATCGCACCTACGTTC	CATATTACAGGCGAACAT
-	GATOCTTOGACTCAGCTATGGCCGTCAAAGGCCCTGACCCGGGAGCATCTATTGTAG	CTGGACTTAACTGCATCTTGAGCACCAGCAT	TAATGATAAGCATGGACAT
_	I II I I I I I I I I I I I I I I I I I	I I I I I I I I I I I I I I I I I I I	::: [AATT
	111-0476	111 1111 1 11 1 1 1 1 1 1 1 1 1 1 1 1	I I I I I I I I I I I I I I I I I I I

• ATA	1 2 90 1464 TCAATAATA/	1 300 1tgtatgaa1	1310 PACGTACTT	1320 Laatgtgtaa	1330 FCGTACATACA	1340 AATCCATAATA	1350 Atactaatat	1360 Taataaataa	1370 ATATATATTC	1380 : TTAATAAACA(1 3 90 STCTCCGTGATA
ACA.		1420 1420	1430 1770AATTAA	1440 1664TTATTAA	1450 1450	1460 LAGAATTAACC	1470 1470 TAATTCTAGC			1 500 CTCCTCACTT	
616			1550	1560	1570	1580 1580				1620	
555	16767AATTCAGG	1660 AAAGCCTCT	1670 CTCCTATGC	1680 1680	1690 CTTGTCCTTAA	1700 LCGCAAGCCTT	1710 577757AATT	1 7 2 0 T A A G G C C T T	1730 CATGTTTTT	1740 17777766	1750 BATCTTCAATAG
ATC1	1700 1700 1700 1700	1780 AACAGTGCA	1 7 90 6 ATTACGTA	1800 1800	1810 Catatggtgt	1820 Atcgctaag	1830 Acgtacacga	1840 AGTATTATGT	1850 TCATGGTGT1	1860 1860 1.1860	Lectron Composition Com
										TACAGTCA.	ATGCTTGTAGGA ATGGTCACAGGA ATGGTCACAGGA
					CCTTTA	GTTAAGGGTC.	ATTTATCCTC	ATAGACAAAG	CTCGAAAGAG	ATTC. 11111111111111111111111111111111111	ATGUTTGTTAGA

B0 GAGATATAATATAACATCAACCCTATGTCCTGATCAATTCTAGTACCAAAATATGACTTATATTTTAGTACTTGTAAAAATTTTACAAAATCAAAATCAATGTCCGTGAACCAAAACT -----TATTAAAGA CGTTTATTGTGTGTAAACCCCCCAAATTAGGTTTTCCTGTAAACCTTGTTTTGC-GTCAAACCCCAAAACCGGAAAAAATTTTACAGTAAAA be ATAATTATATATATCGCGCTTCATAAAATTTGCCGCTTAAATATCTACCACGCGCTTTAACAGACTTTTCGCTAGATACTTTTTAAATTTTCACGCGTTTGAATACTTGAATACTTTACAATACT .. :: 2170 2050 :: ... actecttacetace trage ---t-aage acae acte adatecte agate teae cetace aage tece trage ataage tteete ctage ette a ---GTTGATGTAGGTTAAC-CG--AAAGGAAGGCAGTGAAAATGGCTAGATGAGTCTGG--G-AAGTCGATAAGGATATGGTTGGTGCCAGGGTTGGTGT ---GTTAATGTAGCTTAATAAC--AAAGCAAAGCACTGAAAAGTCTTAGATGATAATTG---TATCCCATAAACACAAAGGTTTGGTCCTGGCCTTATAATT :: 2398 2160 ************** 2390 ••• CSB-3 2150 2380 2140 •• 2020 2370 1940 2130 --..... 2010 an CTAATCATACTCTATTACGCAATAAACATT------2360 1930 2120 •• •• 2000 •• • 2350 1920 2110 .. : • :: *********** ********** 2340 1910 CSB-2 :: •• 2100 •• 2330 1900 2090 1970 :: 2320 •• 1890 BO ATAAATGCTACTCA------2080 ••••• 1960 Phe :: : 2070 1950 •• :: ••• :: Ž 2 Ē . H ĥ 2 ě â Ľ, Z 2 ŝ ě å š ĥ å ê 2

nucleotides from the 5' terminus of the tRNA^{Pro} and extends to position 240. The identical sequence also occurs at position 257 through 301 (see Figure 2). It has been proposed that a series of repeated sequences in the mammalian mitochondrial genome, which are also within 100 nucleotides of the 5' end of the tRNA^{Pro} gene, may serve as terminators for synthesis of the mammalian 7S D-loop DNA (22). The <u>Xenopus laevis</u> mitochondrial 14S D-loop DNA's synthesis has been reported to begin at position 1730 to 1770 (24). Since the two major 14S DNAs are approximately 1350 and 1510 nucleotides long (21), the two repeated sequences beginning at positions 196 and 257 may be the terminators for 14S DNA synthesis.

Figure 3 shows an alignment of the Xenopus laevis, human, bovine and mouse mitochondrial genome sequences (2-4) in the regions from the tRNA^{Thr} gene through the initial portion of the 12S ribosomal RNA gene. In addition, the sequenced region of the rat mitochondrial genome (9) is also aligned with its mammalian and amphibian counterparts. The optimal alignment of each sequence pair was obtained using the NUCALN program described by Wilbur and Lipman (33) and aligned with the human sequence to produce the results shown in Figure 3. This alignment of the human and bovine sequences differs slightly from that reported earlier (34), as it is based on five different mitochondrial DNA sequences rather than two. As shown in Figure 3, all five DNAs have regions of high homology in their coding sequences for the tRNA and Although these four mammalian mitochondrial DNAs have some rRNA genes. sequence homology near the center of the non-coding D-loop region, the greatest homology is observed in the region between the putative origin of H-strand replication and the tRNA gene. In contrast, the Xenopus laevis and mammalian mitochondrial DNAs have low homology in the central region between the tRNAPro gene (position 167) and the putative origin of H-strand replication (approximately position 1750). However, some sequence homologies can be observed in the region between the origin of H-strand replication and the tRNA^{Phe} gene. These regions of sequence homology include the conserved sequence blocks (CSB) described earlier (3,22) and indicated in Figures 2 and 3.

The human, mouse, rat and <u>Xenopus</u> <u>laevis</u> mitochondrial genomes contain all three CSBs, while the bovine genome lacks CSB-3 in the region 3' to the

Figure 3. Alignment of the <u>Xenopus laevis</u>, human (2), bovine (3), mouse (4), and rat (9) DNA sequences in their D-loop regions. A colon (:) indicates that an identical nucleotide is present in the human DNA sequence while a dash (-) represents a gap introduced to maximize the sequence alignment.

100 106 100 106 114644 107007 1041011 404664	216 226 216 226 16at titagttaac 16ta aaatcaattg	336 346 1566 Catgtcaaac 1566 Catgtcaaac	8 Ser Thr Asn 440 447 2 TCA ACA AAC 2 AGT TGT TTG	
0 9 9 8 GCCCTAACCA GG 7 GOGGATTCGT CC	206 206 20645 CCCGCA 20646574 GCGCGCA	326 326 Aggatt tgcaat Aggttaa aggtta	AFE TEP Leu Ph 440 667 T64 TTA TT 662 ACT AAT AAT	igin of
70 70 11440 4600110 11110 100046111	196 196 1970111010 101111010	316 1100111001 10 1 AAGCAAAGGA AG	CO I SENE (AL AIA II- Thr 430 (A30) 110 GCA ATT ACT 140 GGT TAA TGA	passes the or
.Р 60 гаартаа саасас 177саатт стесте	176 186 176 186 186 1886 1860 1886 1960 188	ERMA 296 306 4000 000004AA00 1000 0000077700	416 416 8064 CTTACCTG A 8064 CTTACCTG A 7974	which encom
2 End T1 r * * tRMA 5 50 A T= AGAGA T11 T A TCTCT AAI	166 166 Aaacact ttaatti Tttgtga aattaa:	Replication 	406 406 GGGGGTA TTACTC GCGCGAT AATGAG	ndrial genome nes.
URF Leu Thr Leu Th 40 TTA ACT TTA AC AAT TGA AAT TG	156 156 1764A16CAA CTC AACTTACGTT 640	a of L-strand 216 216 214 214 214 214 214 214 214 214 214 214	396 396 396 396 396 396 396 396 396 396	<u>evis</u> mítochol flanking gei
ro Ile Ser Pro 30 24 ATT TCA CCA 37 TAA AGT GGT	136 146 145 ccaacatcaa 145 ggttgtagtt	01121 01121 01121 011210 01120 01120 01120 01120 01120 01120 01120 01120 01120 01120 01120 01120 0100 0100000000	76 386 386 386 11 Coltotada 11 Colocatot	e <u>Xenopus la</u> ition and its
F Phe Ile Ile F 20 1 TTT ATT ATT C(1 AAA TAA TAA G(126 126 16CTTGC 466ATTT 16CAACG 76CTAAA	246 2 246 2 247567 2 277667 2646741	366 366 366 710710 100710 110710 1007010	region of th trand replica
Ile Leu Ser Sei 10 ATC CTA TCC TCA TAG GAT AGG AG1	 116 AATCTCTGAA TAAG TTAGAGACTT ATTC *	236 236 26677AA6576 AA76 7664777646 7746	356 356 356 366657 7641 766657666 A67A * 674* 680A	Figure 4. The L-s

reported origin of H-strand replication. The occurrence of homologous sequences in this region of the D-loop may represent initiation points for the RNA primer implicated in the synthesis of the single stranded D-loop DNA initiation segment (21). Finally, these conserved sequence blocks occur in regions which can be drawn into very large hairpin structures (22,34). Since these regions in the <u>Xenopus</u> <u>laevis</u> mitochondrial D-loop segment can also be drawn as large hairpin structures, they may be important in the D-loop's biological function.

Recent evidence demonstrates that both strands of the vertebrate mitochondrial genome are completely transcribed from points at or near the tRNA^{Phe} gene (35,36,37). After transcription, the large polycistronic RNA is processed to its shorter mature RNA species by a series of steps which is not well understood. The promoters for Xenopus laevis mitochondrial transcription most likely are contained in the regions adjacent to the tRNA^{Phe} gene (37). Although these regions contain several AT rich segments, an examination of the sequence similarities (see Figure 3) does not immediately reveal which regions may function as the mitochondrial transcription promoters.

Origin of L-strand replication and its flanking genes. The segment of the Xenopus laevis mitochondrial genome shown in Figure 4 most likely encompasses the putative origin of L-strand replication because its surrounding tRNA and protein coding segments have high sequence homology with the other vertebrate mitochondrial genomes. As shown in Figure 5A, this region may form the hairpin loop structure similar to other vertebrate mitochondrial origins of L-strand replication. Its stem may be either nine nucleotides long with continuous base pairings, or fifteen nucleotides long but with two mismatched base pairings. The former structure would contain a nineteen nucleotide long A and T rich loop while the latter structure would contain a smaller T rich loop, as indicated by the additional dashes in the loop region of Figure 5A. This putative origin of L-strand replication is flanked on the 5' side by the genes for tRNA^{Tyr} and tRNA^{Cys} and on the 3' side by the genes for tRNA^{Trp}, tRNA^{Ala} and tRNA^{Asn}. As shown in Figure 5B, this loop and stem region has a high sequence homology with its mammalian counterparts. The spacing between genes in this region is summarized in Figure 5C. Here the distance between the tRNA^{Asn} and tRNA^{Cys} genes flanking the putative L-strand replication origin is the same as in the mammalian mitochondrial genomes. In all but the human mitochondrial genome, the distances between the other structural genes are very similar.

Α.	-	GGC T T	в.	Xe	st CTTC	em TCCCG-GT	1 TTTTTTG	GCTTA/	TAAA-C	sten GGGAGAAG
	T- T-	A		Hu	CTTC	rcccgcc-	*****	TCCCG	GCGGC	GGGAGGAG
	T T-	т т		Bo	CTTC	TCCCGCC-	TTTT1	TTTCT	rgc-ggc	GGGAGAAG
	Cys trna AAGGTTTGCCGG	G - C G - C C - G C - G T - A *C - G *T - A *T - A *T - A *C - G *T - A	Asn tRNA ******** ATGAAAGCT	Мо Т	CTTC	TACCGCC-	&TTT	****	FTC-GGC	GGTAGAAGA
c.	Species	URF2	Trp trna	A1 trna	a 	Asn tRNA	0r1	Cy: -tRNA	s tRN	Tyr A Col
	Xe	-2	2		0		32		-1	2
	Hu	- 2	7		۱		31		-1	12
	Во	-2	1		1		32		0	l
	Mo	-2	1		2		32		2	ι

Figure 5. A. Postulated secondary structure in the region of the origin of L-strand replication. B. Comparison of the primary structure of the <u>Xenopus</u> <u>laevis</u>, human, bovine and mouse L-strand replication origin regions. C. Distances between genes in this region.

As presented in Figure 4, the region we have sequenced begins with the last 45 nucleotides of the putative URF2 reading frame. Here, as has been reported for several other mitochondrial reading frames, the 3' end of the URF2 overlaps the tRNA^{Trp} gene. Thus, once the full length tRNA has been cleaved from the primary transcript, the URF2 mRNA must be polyadenylated to produce the required stop codon. This mechanism for post-transcriptional introduction of stop codons originally described for the human mitochondrial system (2) and later observed for other mammalian mitochondrial systems (3-4) may also occur in the <u>Xenopus laevis</u> system. Thus, the use of post-transcriptional polyadenylation to produce stop codons in mitochondrial m-RNAs may be a universal phenomenon in vertebrate mitochondria.

<u>tRNA genes</u>. The nucleotide sequences of the regions corresponding to the eight sequenced <u>Xenopus laevis</u> mitochondrial tRNA genes are shown in their cloverleaf forms in Figure 6 and are compared to their mammalian mitochondrial DNA counterparts in Figure 7. As has been observed with other vertebrate mitochondrial tRNAs (2-4,9,11,34,41-46), the <u>Xenopus laevis</u> mitochondrial tRNAs lack many of the features usually associated with cytoplasmic tRNAs. Only four of the eight tRNA genes reported here contain a T at position 8, and only three contain the two T's at the 5' end of loop IV.



		A.A.	р.	D.	D.	A.C.	A.C.	A.C.	٧.	т.	т.	т.	A.A.
		stem	sten	100p	stem	stem	100p	stem	100p	stem	100p	sten	stem
							•						
Thr	Xe	GTCCTGAT	TAGCTTA	A-TT	TAAAGC	TCGGT	CTTGTA	AGCCGA	AGAT	TGAGG-	-CTAAA-	ACCCTCC	TCAAGACT
ACN	Hu	****TG	**TA**	*AC	***TA*	*C*A**	*****	*A***G	-****	GA*AA-	*C-T	TT*T*	CA*G***A
	Bo	***T*TG	***TAC*	*C	***TAT	CT***	*****	****	-***	G***A/	***C-T	******	CT******
	Mo	***T***	***TA**	AA	C*TTA*	CT***	******	*A**TG	-***	GA**A1	[*T		*******
Pro	τ.	COCCACAC											
CCN	Hu	*******	******	**	TAGAAI	JTTGGC	TTTGGG	GGTCAA	T-AGT	GGAGGI	TT-GAG	-TCCTTC	TTTCTCGA
		*****		* *	*****		*******	TTCTT	-G##	****T	*****	<u>A***T</u>	**CTCT**
	Mo	*****		*		L.CV.	******	TGATTG		"AGAC"	GCA	GT###	**C**T**
	по		1-6	******	******	ACCA**	******	TTCTGG	;=-G==	##GGAG	;*A-*C-	T*C*T	C****T**
Phe	Xe	GCTTACG	TAGCTTA	A-GT	-AAAGC	ACAGC	ACTGAAA	ATGCTG	-AGAT	GAGCCO	CTA-CGA	A-AGCTC	
UUY	Hu	******	******	CCTCC-*	C*****	*****	******	*****	-***C	*G**T	AC-A-T	CA*C*	*******
	Bo	*T*G*T*	******	*C-CC		*AG***	******	****C1	-****	*****	0	****	*******
	Mo	*T*A*T*	******	*TAAC	-*****	*****	******	*****1	_****	*GATA	*T-G	TATC*	******
Trp	Xe	AGAGATT	TAAGTTA	AC-	AAGACT		CTTCAA	ACCCCT	AAG-C			T-07007	
UGR	Hu	*******	**G****	*-ATA	C****C	*****	******	******	C**-T	*****	-CC#4		*******
	Bo	**GA***	**G****	*	C****C	*****	******	*****	***_*	*****	-C*ATT		*******
	Mo	***AG**	**G*A**	T-A*-	T**T*C	GC****	******	*****	***-A	*AACA-	C*C*-	AGTT*	**CT****
41.	•												
ALA	ve Te	AAGGCTT	FAGCTTA	AT	TAAAGT	GTTTTA	\GTTGCA	TTCAAT	T-GAT	GTTGG <i>i</i>	TAAAAT	CCTGC	AAGCCTTA
GCN	Hu	****GC*	******	**	*****	*GC*G*	*T****G	;***G*	****	*CA*A0	S*GGGGG*	TT***	*GT****
	Bo	G***A**	******	**	******	*G**G*	•T****	*****	****	*****	G*GT*G*	*T***	**T****
	Mo	G***TC*	******	**	*****C	A.&**G*	•T****	*****	·A-***	*****	**G**G*	*T*A*	*GT****
Asn	Xe	TAGAATG	AAGCTC-	GTTGG	ATTGA	GTTTAG	GCTGTTA		ATGTT	GCGGGA	TCGAG-	G-CCCGT	CTTTCTAG
AAY	Hu	*******	****CA-	***-*AT	T*GG*T	*C****	******	*****G	TGT**	*T***1	*TA**T	*****	TGG****
	Bo	*******	****CA-	*****-0	T*GG*T	*****	******	*****	G*T*C	*T***C	GT*GA-	*-***AC	*****
	Mo	***T**	****CA-	**AAT	-*GG*T	*****	*****	*****	T*T*C	*TA**1	TT*ATT	**T*C	*****
Cys	Xe	AAGCCTG	CGGTG	-TT-G	ACAT	GCCAGA	TTGCAA	ATCTCO			G-AA-G	G-TTTGC	CGGGGCTTC
UGY	Hu	*GCT*C*	A****A-	-**T		ATTGAS	******	**TCGA	-****	***601	TC**A-		****6C*T
•	Bo	*GC****	*****	-**T	***C	TTGA	******	*****		***601	TC**-T	*C***	*******
	Mo	GGT*T*A	A***A-	-*AT	T***	T*GA*	*****	**TCGA	-**GT	*T*G*0	-A**-T	C-*C*A*	TAA*AC*T
Tv-	Xe	GGTAAGT		^-									*****
TAV	Hu	*******	******	**	-64***		1101A0	A**TAA	-4C**	**C***	TURA-G	*****	*******
	Ro	*******		*0	-++++				- 10 -		AUU		******
	Ma	*****							-AGT				******
	по		Lasalas	*-	-TANC	ATTA " I		A-TAA	-AC##		TAWAL		=T===C==

Figure 7. Nucleotide sequence comparison of vertebrate mitochondrial tRNA.

As with other mitochondrial tRNAs, the <u>Xenopus laevis</u> mitochondrial tRNAs also have several instances of mis-matched base pairings in their stem regions. The most prevalent mis-match is the G+T pairings which do not cause significant distortions in these helical regions (47) and are allowed under the wobble hypothesis (48). Other mismatched pairings occur in the stem of loop I in tRNA^{Asn} and tRNA^{Tyr}. These loosely paired regions may be stabilized by base stacking and/or tertiary structural interactions, rather than by classical base pairings, as in the case of the truncated mammalian tRNA^{Ser} (AGY) (46).

As can be seen in Figure 7, the primary sequences of these eight <u>Xenopus</u> <u>laevis</u> mitochondrial tRNA genes are quite similar to those of the mammalian mitochondria. The greatest homology occurs in the anticodon loop and stem and in the amino acid acceptor stem regions. In contrast, the least conservation of sequence homology occurs in loops I and IV and in their corresponding stem regions. It might be that the lack of conserved elements in loops I and IV is due partially to the removal of any evolutionary pressure for maintaining the internal promoters required for eucaryote tRNA gene transcription (38), since the mitochondrial tRNAs are transcribed as large polycistronic species from a single promoter on each strand which is located in the D-loop region (39,40).

In summary, the primary structures of the <u>Xenopus laevis</u> mitochondrial DNA segments reported in this communication show regions which are homologous to the mammalian mitochondrial genomes. This observation indicates that elements of the amphibian and mammalian mitochondrial DNAs nucleotide sequences and their overall gene organization have been conserved during evolution.

ACKNOWLEDGEMENTS

We thank Dr. Igor Dawid for providing the pXlm-31 chimeric plasmid containing the entire <u>Xenopus laevis</u> mitochondrial genome and M. Stevens for technical assistance in the early stages of the experiments. This work was supported in part by grants GM-30399 and GM-30400 from the NIH and a Biomedical Research Support grant from the University of Oklahoma.

REFERENCES

.

- Altman, P.L. and Katz, D.D. (1976) Biological Handbooks. I. Cell Biology. pp. 217-219. Federation of American Societies for Experimental Biology. Bethesda, Md.
- Anderson, S., Bankier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457-465.
- Anderson, S., deBruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) J. Mol. Biol. 156, 683-717.
- Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Cell 26, 167-180.
- 5. Cordonnier, A.M., Vannier, P.A. and Brun, G.M. (1982) Eur. J. Biochem, 126, 119-127.
- Droon, A.M., Pepe, G., Bakker, H., Holtrop, M., Bollen, J.E., Van Bruggen, E.F.J., Cantatore, P., Terpstra, P. and Saccone, C. (1977) Biochim. Biophys. Acta 478, 128-145.
- 7. Kobayashi, M., Seki, T., Yaginuma, K. and Koike, K. (1981) Gene 16, 297-307.
- Koike, K., Kobayashi, M., Yaginuma, K., Taira, M., Yoshida, E. and Imai, M. (1982) Gene 20, 177-185.
- Sekiya, T., Kobayashi, M., Seki, T. and Koike, K. (1980) Gene 11, 53-62.
 Saccone, C., Cantatore, P., Gadeleta, G., Gallerani, R., Lanave, C., Pepe, G. and Kroon, A.M. (1981) Nucleic Acids Res. 9, 4139-4148.

11.	Wolstenholme, D.R., Fauron, C.M.R. and Goddard, J.M. (1982) Gene 20, 63-69.
12.	Brown, G.C. and Simpson, M.V. (1982) Proc. Natl. Acad. Sci. U.S.A. 39, 3246-3250.
13.	Klukas, C.K. and Dawid, I.B. (1976) Cell 9, 615-625.
14.	Goddard, J.M. and Wolstenholme, D.R. (1978) Proc. Natl. Acad. U.S.A. 75, 3886-3890.
15.	Goddard, J.M. and Wolstenholme, D.R. (1980) Nucleic Acids Res. 8, 741-757.
16.	Clary, D.O., Goddard, J.M., Martin, S.C., Fauron, C.M.R. and Wolstenholme, D.R. (1982) Nucleic Acids Res. 10, 6619-6637.
17.	Clayton, D.A. (1982) Cell 28, 693-705.
18.	Bogenhagen, D., Gillum, A.M., Martens, P.A. and Clayton, D.A. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 253-262.
19.	Bogenhagen, D., Lowell, C. and Clayton, D.A. (1981) J. Mol. Biol. 148, 77-93.
20.	Martins, P.A. and Clayton, D.A. (1979) J. Mol. Biol. 135, 327-351.
21.	Gillum, A.M. and Clayton, D.A. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 677-681.
22.	Walberg, M.W. and Clayton, D.A. (1981) Nucleic Acids Res. 9, 5411-5421.
23.	Upholt, W.B. and Dawid, I.B. (1977) Cell 11, 571-583.
24.	Ramirez, J.L. and Dawid, I.B. (1978) J. Mol. Bio. 119, 133-146.
25.	Thuring, R.W.J., Sanders, J.P.M. and Borst, P. (1975) Analytical Biochem. 66, 213-217.
26.	Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
27.	Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A.
	(1980) J. Mol. Biol. 143, 161-178.
28.	Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
29.	Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. 8, 1731-1743.
30.	Staden, R. (1979) Nucleic Acids Res. 6, 2601-2610.
31.	Staden, R. (1978) Nucleic Acids Res. 5, 1013-1015.
32.	Staden, R. (1977) Nucleic Acids Res. 4, 4037-4051.
33.	Wilbur, W.J. and Lipman, D.L. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 726-730.
34.	Anderson, S., Bankier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F.,
	Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1982) in
	Mitochondrial Genes, Slonimski, P., Borst, P. and Attardi, G., Eds. pp. 5-43. Cold Spring Harbor Press. Cold Spring Harbor.
35.	Montova, J., Ojala, D. and Attardi, G. (1981) Nature 290, 465-469.
36.	Montaya, J., Gaines, G.L. and Attardi, G. (1983) Cell. in press.
37.	Rastl, E. and Dawid, I.B. (1979) Cell 18, 501-510.
38.	Hall, B.D., Clarkson, S.G. and Tocchini-Valentini, G. (1982) Cell 29, 3-5.
39.	Aloni, Y. and Attardi, G. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 1757-1761.
40.	Murphy, W., Attardi, B., Tu, C. and Attardi, G. (1975) J. Mol. Biol. 99, 809-814.
41.	Roe, B.A., Wong, J.F.H., Chen, E.Y., Armstrong, P.W., Stankiewicz, A.,
	Ma, D.P. and McDonough, J. (1982) in Mitochondrial Genes, Slonimski, P., Borst, P. and Attardi, G., Eds. pp. 45-49, Cold Spring Harbor Press, Cold Spring Harbor
42.	Roe. B.A., Wong, J.F.H., Chen, R.Y. and Armetrong P.W. (1981) in
	Recombinant DNA, Proc. Third Cleveland Symp. on Macromolecules, Walton, A.G., Ed. pp. 167-176, Elsevier Pub. Co., Amsterdam.

- Randerath, E., Agrawal, H.P. and Randerath, K. (1981) Biochem. Biophys. Res. Comm. 103, 739-744.
- 44. Randerath, K., Agrawal, H.P. and Randerath, E. (1981) Biochem. Biophys. Res. Comm. 100, 732-737.
- 45. Agrawal, H.P., Gupta, R.C., Randerath, K. and Randerath, E. (1981) FEBS Letters 130, 287-290.
- 46. DeBruijn, M.H.L., Schreier, P.H., Eperon, I.C., Barrell, B.G., Chen, E.Y., Armstrong, P.W., Wong, J.F.H. and Roe, B.A. (1980) Nucleic Acids Res. 8, 5213-5222.
- 47. Rich, A. and RajBhandary, U.L. (1976) Ann. Rev. Biochem. 45, 805-860.
- 48. Crick, F.H.C. (1966) J. Mol. Biol. 19, 548-555.