Structure and sequence of the mitochondrial 20S rRNA and tRNA tyr gene of *Paramecium primaurelia*

Jeffrey J. Seilhamer and Donald J. Cummings

Department of Microbiology, University of Colorado Health Sciences Center, Denver, CO 80262, USA

Received 21 September 1981

ABSTRACT

The sequence and structure of the large (20s) mitochondrial (mt) rRNA gene and flanking regions from *Paramecium primaurelia* have been determined. The gene contains two regions of strong homology with other large mt rRNAs: one 44-base region near the 5' end and a 321-base region near the 3' end. Another region of strong homology to both ends of *E. coli* 23s RNA exists at loci consistent with these regions. The *Paramecium* gene appears to be 2204 bases in length and contains slightly more homology to *E. coli* rRNA than its mammalian or fungal counterparts. The gene, located about 1200 bp from the replicative terminal end of the linear mt DNA, is transcribed in the same polarity as replication. Previous R-looping studies detected no large introns within the gene. Here we describe sequences resembling degenerate rRNAs, one of which could represent a small intron. A tRNA tyr gene was found on the same DNA strand, 127 bp downstream from the large rRNA presumptive 3' end. The tRNA is flanked on both sides by short DNA regions of ~90% A + T content.

INTRODUCTION

Much attention has been focused recently upon mt rDNA gene sequences for three reasons: First, they have been used to quantitate evolutionary sequence divergence (1-6). Second, knowledge of RNA secondary structure, along with consideration of the alterations associated with resistance to ribosome-affecting antibiotics may help elucidate rRNA role(s) in ribosome structure and function (7-11). Third, sequences flanking rDNA have yielded important information about mt RNA transcription and processing (8,12-22). Of particular interest has been the finding that mouse and human rRNA genes are found immediately adjacent to tRNAs, the latter of which probably act as processing punctuation by directing cleavage from a much larger multigene transcript (19-22).

We were interested in examining *Paramecium* mt DNA for all of the above reasons, but particularly the gene organization and transcriptional aspects. *Paramecium* mt DNA is a linear 14 μ molecule, a characteristic limited apparently to ciliates (23-24), and has a unique mode of replication (25-26). We wanted to see if the organization of rDNA and its flanking regions resembled mammalian

gene organization (i.e. interspersed tRNA as transcriptional processing signals), fungal gene organization (i.e. genes separated by noncoding A + T rich regions), or whether its gene organization as well was unique. We report here the gene sequence, its homology to other known mt RNAs, and present several observations about the flanking DNA regions.

METHODS

Paramecium primaurelia strain 513 was obtained from the Edinburgh stocks. Cells were grown and DNA was prepared as described (27). RNA was extracted from mitochondrial pellets by the technique of Parish and Kirby (28). RNAs were examined by electrophoresis on 5% acrylamide/8M urea (not shown) or 0.75% methyl mercury/1.8% agarose gels as described (29).

Hind III fragment 5 was cloned into pBR325 in our lab by Jane Laping for large-scale propagation. Restriction mapping was done via single and double digests on acrylamide or agarose gels, and in some cases by fragment isolation from acrylamide preparative gels with subsequent subdigestion. Southern hybridizations were performed as described (30). Cloned DNA was end-labeled, midcut with restriction enzymes, and the halves were isolated by acrylamide gel electrophoresis and sequenced according to the method of Maxam and Gilbert (31). All enzymes were obtained from Bethesda Research Labs except calf alkaline phosphatase (Boeringer/Mannheim). Crude γ -³²P-ATP for end labeling was obtained from ICN.

RESULTS

Mitochondrial RNAs

Mt RNAs from Paramecium grown on Klebsiella as a food source are shown in Fig. 1. Axenic-grown species lack the 23s and 16s bands, and the 25s and 17s bands are cytoplasmic rRNA contaminants. The two mt rRNAs, 20s and 13s, are the other major species present. Note that the 20s and 13s bands are about equal in stain intensity, suggesting (due to their 2-fold difference in relative size) that less intact 20s rRNA is present. Some 20s rRNA molecules may contain discrete nicks such that one of the cleavage fragments comigrates with the 13s peak (Seilhamer, J.J., unpublished data). Similar behavior was observed when sucrose gradients were used to separate rRNAs (27). The two bands marked X might be discrete cleavage products of the rRNAs. tRNAs were not visible on this gel but could be visualized on 5-8% acrylamide/8M urea gels.

Restriction Map and Polarity of Hind III Fragment 5

The fragment of mt DNA containing the 20s rRNA gene has been identified

Fig. 1: Cytoplasmic and Mitochondrial RNAs. 1.8% agarose/0.75% methyl mercury gel of cytoplasmic and mitochondrial RNAs. (See Methods)

previously by Southern hybridization of ^{125}I -labeled rRNA (27) as Hind III fragment 5 (H5). The position and orientation of H5 was determined in Fig. 2, taking advantage of an asymmetric Bgl II cut within H5 (Fig. 3). The genomic terminal Eco RI fragment 3 (E3, 6400 bp) containing H5 (Lane 2) and the genomic terminal Bam HI fragment 1 (B1, 21 kbp) containing E3 (Lane 1) were sub-digested with Bgl II and electrophoresed on a 1.8% agarose gel. In Lane 3, E3 DNA was end-labeled prior to digestion with Bgl II and the autoradiograph of the gel is shown. Lane 4 shows a Southern blot of Lane 2, challenged with a nick-translated probe of the 292 bp Hae III fragment of H5 (represented by tail on arrow in map).

The map in Fig. 2 shows the only orientation of the four Bgl II fragments

Fig. 2: Determination of H5 Polarity. Map shows positions of Bgl II fragments determined by agarose gel electrophoresis (lanes 1 and 2), end-labelling of Eco RI fragment 3 (E3; lane 3), and by hybridization of a portion of H5 (lane 4).

of E3 possible from the data: The 800 bp fragment is at the left (inner) end since it end-labels very heavily (Lane 3) and is not present in Lane 1. The 3600 bp fragment is at the right (terminal) end of the genome (and E3) since it also end-labeled (Lane 3), although less than the inner end, and appears diffuse in Lane 2. The replicative terminal end always labels poorly and appears heterogeneous in length; therefore, it can easily be identified (fragments labeled T in map). The 1600 bp fragment contains the left 553 bases of H5 (Lane 4), so the only piece(s) long enough to contain the remaining 2350 bases of H5 is the 3600 bp terminal fragment of E3. Similar arguments were used for mapping Msp I and Tha I digests (data not shown). H5, then, is located 1.2 kb from the right (replicative-terminal) end of the genome, oriented in

Fig. 3: Fine Structure Restriction Map of H5 Within Complete Sp. 1 Map.

the direction shown in Figs. 2 and 3, and about 12kb to the right of the region containing the small (13s) rRNA gene (27,32).

H5 Sequence

Cloned H5 DNA (2903 bp) was subdigested with fine-cutting restriction enzymes and a restriction map was constructed (Fig. 3). Two regions of inter-species sequence conservation, HR1 and HR2 (see below), are shown by arrows. Below the map is the strategy used to sequence H5. The non-transcribed strand with respect to rRNA is shown in Fig. 4. The entire sequence has been covered (at least one strand) and most regions have been verified from overlapping and/or opposing strand sequences. Restriction cuts at bases 2113 and 2445 have not been crossed, leaving open the possibility that small gaps could exist at these points. The Taq I site at base 403 cuts only in the presence of 100-fold excess enzyme and 10% glycerol, due perhaps to methylation or secondary structure, and is not shown in Fig. 3. Base 1585 could not be determined (indicated by N). Based on homology to other rRNA sequences (see below), we estimate the gene 5' end at base 53 and 3' end at base 2256 (arrows, Figs. 3 and 4).

Features of the rDNA Sequence: Homologous Region 1 (HR1)

Comparisons of rDNA sequence from mouse mt (10,20), *E. coli* (10,33), *Aspergillus* mt (15), and corn chloroplast (2) showed a 44 base homologous region beginning 100-200 bases from the 5' terminus (Fig. 5). The number of divergent bases relative to *Paramecium* was roughly equal, with chloroplast being the most divergent. Note that this region consists of smaller, very highly

conserved blocks, and most divergence occurs in the gaps between those blocks. Also, most alterations are conservative (i.e., transitions rather than transversions).

Homologous Region 2 (HR2)

Previous studies (7-11) have identified a region of strong sequence conservation near the 3' end of large subunit rDNAs from other organisms. We have aligned the *Paramecium* sequence with homologous regions from *E. coli* (10, 33), yeast mt (8), and mouse mt (10,20) in Fig. 6, extending as far as homology is clear. This region includes 321 bases for *E. coli* and yeast mt, and 218

```

1  AACCTTTTGA GTTAAATTAC TTTT TTTTGT GCAGCAAGGG TGGTCTCAAC
51  CTGGTTTCAT TCCCPAGCTA ATAAAGCTTCG ATTCTTTGTA CCGGCTCAGA
101 GTATATATTA TTAACAAAAT AATATCGAAA TGT TTTTCAA AAAAAAAGC
151 TAATAAAACC AAATTTTGAT AAAAAATAAG TACCGTGAGG GAAAGGGTGA
201 AAAGAAATTT GTAGAATGCT TAAAGATTC TGAATCTAG TGCAGTGAAA
251 CAGTTAAAGC GTGTGTTTT AACGTACCTT TTGTATAATG GCGCAACTAG
301 TTTATAAAAT TAGCGAGCTT TATCAATCG CGTAATGAAA ATGTATAGT
351 TAATTTTATA AGACCCCGAA GTCAAGTGAT CTAATCATGG CTAGGTAGAA
401 GATCGAACCC ATAAATGTTG CAAAATTTCC GGAGAAGCTG TGATTAGGGG
451 TGAAAGGCTA ATCAAACTTG ACGATAGCTG GTTTTTCCGG AAATCTATCT
501 ACGTAGAGTA TTTTTTTTCT TTTGTCCGGG GTAGTGAAT CTITCTTAA
551 GAAGATCTGA TCTTTTTTCC TGAAGAATCT TCCAATACGC ATTAAATAA
601 TAATAAAAC GGACTGAGAG TGCAAGATT CTGTGTCGAG AGGGAACAG
651 CCCAGACCGA ACGATAAAGT GCATAACAA TCGAAGTAA AAGAATTTT
701 TTTAAAAAAA TATTGGGAGG TAGGCTTAGA ATCAGCCAGC CTTTAAAGAA
751 AGCGAAACAG CTCACCAATA GCTACGACTA TTTGAATAAA ACTCGTTATT
801 AAAAACAAAT TTTAAATTAT AGCGCTAAGC ATTGTACTGA AGCGGCGGGT
851 AACTCGGTAG CGAAACCTTT TGTAGTCTGT TGAAGCTTTA TTGAAAAATA
901 GGCTGGAGAT ATCAAAATTC ATAATGTTGG CATGAGTAAT GGACAAAATG
951 TTCAAATCAT TTTTCTTTGA TAAGTTAGCG TTGCTTTCTT TTGATCATCT
1001 TACAAAGTGT GATTCCGCCCT CTAATTTTTT AAATATAAAT TTTTAAAGAT
1051 GAGGTTTGTA TTTTTTCTG GCTTCCTTTG CGTTCTTCTT TATTTTTTTA
1101 GGAGAGTGTG GGGAGGCGGA CTGAAAAAT TTTACAAGGG GCACCGTACT
1151 AACACTAACA CAAGTACTTT AGTCGAGCAG AGACGACAGA AGAGCTAAAG
1201 ATATTGAAGG AACTCGGCAA AATTACTTTG TAAGTTAGCG ATAAAAAGTG
1251 CGTCAGCAAA CAAAAAATG GGGGTAGCG ACTGTTTACT AAAAACATAA
1301 GATTTTGCAA AATTTAATTA TGATGTATAA AATCTGACTC CTGCCCCGTG
1351 TTGTCATCCA AAGTTTAAAC GGTAGCGCT GTTGATAACA AGCAGCAATA
1401 AACGGCGGCC ATAAGTCTGA TGTCTCTAAG GTAGCAAAAT CCCTTGACGG
1451 GTAAGTTCCG TCCTGCACGA ATGAGTAAC GACTGCCCTA CTGTCTCCAA

```

```

1501  TATCAGCTCT ATGAAATTGA ATTTGCTGTG AAGATGCACC TTTTACAAC
1551  TAGACGGAAA GACCTATGC ACCTTTACTG ATGCAAGGAA CTAAGAGAT
1601  ATACTGGAGA TAAATTAAGG TAGGAGTTAA ATCGCAATTG AAAAATACT
1651  TCTCTTTTAC GTCTCTTCAT AAAAATGAGT AGATAATTTT TACTTTTCTG
1701  GCTGTTAGTT TAACTGGGGC GGTGGCCACC TAAAAAGTAA CGGAGGTGAG
1751  CATAAGTTA CGCTTTGTGA GAAATTTTTT TTCGCAAAAT GAGTTAATAA
1801  AACTGCGTAA TTTGATTAAA TTACAGACTA GTAATTTAGG GCCTTATTGC
1851  CTGCTATAAT GATCCGGTGT TTTTTTTTGA ATAAGACATC GCTCAACGAA
1901  TAAAAGGTAC GCTAGGGTAA ACAGGCTTAT AAATTCTGAG AGTTCTTATT
1951  AAAGAATTTG TTTGGCACCT CGATGTCGGC TCATCACATC CTGGTGGTGC
2001  AGAATCTGCC AAGGGTTTGG HR2 CTGTTCCGCA ATTAAAGTGG TACGTAGCT
2051  GGGTTTAAAA CGTTGTGAGA CAGTTTGGTC CCTATCTGTT GTAGAATTAA
2101  GAAACGAGGC TGAGATCGGA TGCCAGTACG AGAGGACCGA TTAGAAGTAG
2151  CCTCTGGTTC GGCATCTACT TTAATAAAG GTATAGAAGC TACGTAAGC
2201  TGGATAGCCT AATTTTTTAA TTAAGTTT TTACTTGAA TAATTAGAA
2251  CCTTTAGCA TCTCGTTTCT AATAATATTT TAGAAGAACT ACTAAATTAA
2301  TTTGAAGTTT GCTGTTATAG GGCTCTTTAC GAGCGTACCA GCTTCTTTA
2351  AAAAAATATT AAGACTCTTA TTATTATTAT ATCAACTAAT GCCTGAGTGG
2401  TTAAGCGGGC AGACTGTAAA TCTGTTGGTA GTACCTCGCT TCGTTTCAAT
2451  CCAACTTACT TCATACCAAT AAATCAAAAT AAATTTTATC AATAATAAG
2501  TTACAAAAAT TTTTTTTATA AAAAAAATTT TAGAACATCT TTGTTTTTTT
2551  TTTTGGGAA CTAAGCTAGC CCCTGATTTT TAAAATTTGA GAGCAGCTCT
2601  TTTTGTATC TTTAATAAGC GCTTCTCTAT CTTTAACITT TTTTATATA
2651  GAGGTTTGT TCTTATTAA TTTTTTTTTT TAAAAACATT GACAAGTCTA
2701  AACTTATTTT TTTAGCAGAT AAAAGGATTC ACTATAGGCT TCAATACGTA
2751  AAGTACTTTA ACTATCTTTT CGATAGGGAG GGTTTTATA AAAAATATAC
2801  AATTATAGG TTGATATTC GAGAGCGGT GAGCTCTTT AATAAATACC
2851  CAATCTTTA TTTTTTTTTT TTTACAAAA ATAAACCTCT ATTAATACAG
2901  CTT

```

Fig. 4: Sequence of H5 (Non-transcribed Strand).

bases for mouse mt. The sequences show high degrees of sequence conservation, most changes again being conservative and distributed nonrandomly as in HR1. Degrees of homology were calculated separately for each of the three domains delimited by the mouse sequence. For bases 31-248, where all sequences show homology, the pairings of *Paramecium* mt with *E. coli*, yeast mt, and mouse mt sequences show 80%, 74%, and 64% conservation, respectively. Although the regions to the left and right of the mouse-homologous region generally show considerably less homology, the ranking of their degrees of homology are consistent with the results observed within the middle 218 bases. The similarity

		Divergent bases
Mouse mt	5'-(110 bs)...G A A - - - C T - A G T A C C G C A A G G G A A A G G - - A T T G A A A - G A C T A A T T	12
Paramécium mt	5'-(1197 bs)...G A A - - - C A - C T A A G T A C C C C G A A G G A A A A G G - T T G A A A - G A - A A T T	(0)
E. coli	5'-(193 bs)...G A A A C A T C T T A G T A C C C C G A A G G A A A A G G - T T C A A C C G A - G A T T	13
Zea mays cp	5'-(182 bs)...G A A A C A T C T T A G T A G C C A A G A G G A A A A G A A A G C A A G - C - G A T T	16
Aspergillus mt	5'-(?? bs)...G A A A T A T C T T A G T A A G C T T A G G A A A A G A A A T T C A A A G A - G A T T	13

Fig. 5: 5'-End Homology Region (HR1).

of the Paramecium and E. coli sequence is most striking within these outer regions.

The central region of HR2 can be folded as shown in Fig. 7 presenting a possible secondary structure for the region. Shown is a proposed structure for E. coli and the analogous structure built from the Paramecium sequence showing only the altered bases. We have modified the structure proposed by Baer and Dubin (10) to include possible base pairing of four more nucleotides within the central (largest) loop, evidence for which was obtained by comparison of mouse and Paramecium sequence structure. Note that most changes in stem sequence are compensated (i.e. alter both pairs), strongly supporting

E. coli 5'-(2385 bs)...
 Parametium mt 5'-(18037 bs)...
 Yeast mt 5'-(2116 bs)...
 Mouse mt 5'-(1251 bs)...

E- A C G G A T A A A A G G T A C T C G G G G A T A A C A G G C T G A T A A G G G C C A T C G C T C A
 P- A C G A A T A A A A G G T A C G C T A G G G G A T A A C A G G C T T A T A A A T T C C A G A G A G T T C
 Y- A C G A A T A A A A G G T A C G C T A G G G G A T A A C A G G G T A T A T A A A C G A A A G A G T A G
 M- A C G G A - C C A A A G T T A C T C T A G G G A T A A C A G G C G C A T C C A T T T A A G A G T T C

E- A T A T C G A C G C G G T G T T T G G C A C C T C G A T G T C G G C C A T C A C A T C C T - G G
 P- C T A T T A A A A A G T T G T T T G G C A C C T C G A T G T C G G C C A T C A C A T C C T G T
 Y- A T A T G T A A A A G T T G T T T G G C A C C T C G A T G T C G A C T C A C A C T T T C C T C T
 M- A T A T C G A C A A T A A G G G T T T A C G A C C T C G A T G T T G G A T C A G G A C A T C C A C A T

E- G G C T G A A G T A G G T C C C A A G G G T A T G G C T G T T C G C C A T T T A A A G T G G T A C G
 P- G G T T G C A A A A A A C C T T G C C A A G G G T T T G C C T G T T C G C C A A T T A A A G T G G T A C G
 Y- G G T T G T A A A A A A C C T T A A G A A G G G T T T G A C T G T T C G T C A A T T A A A A T G T T A C G
 M- G G T - G T A A A A G C T A T T A T A A G G T T C G T T T G T T C A A C G A T A T A A A G T C C T A C G

E- C G A G C T G G G G T T T A A A C A G T C G T G A G A C A G T T C G G T C C C A T C T G C C G T G G
 P- T G A C T G G G G T T T A A A A C G T T G T G A G A C A G T T C G G T C C C T A T C T G T T G T A G
 Y- T G A G T T G G G G T T A A A A A C G T T G T A A C A G T A T G G T T C C C T A T C T G C T G A A G
 M- T G A T C T G G T T C A G A C C G A G C A A T C C A G G T C G T T T C T A T C T A T T T A C G

E- C G C G T G A A G A C T G A G G G G G C T G C T C T A G T A - C G A G A G G A C C G G A G T A G
 P- A A T T A A A A A C T T G A G - - A T C G G A T G C C A G T A - C G A G A G G A C C G A T T A G
 Y- A T T T A T T T A T C A A A T T A A A T C C A T T A T T A G T A T C G T C A A G A A C C A T A A T
 M- A T T T C T C

E- G A C G C A T C A T C G G T G T C G G G T T G T C A T G C C A . . . (199 bs) - 3'
 P- A A C T A G C A C T C G G T T C G G C A T T A C T T T A A A T . . . (807 bs) - 3'
 Y- G A A T C A A C C G A T G G T G T A T C T A T T G A T T A A T A A . . . (3107 bs) - 3'
 M- . . . (114 bs) - 3'

Fig. 6: 3'-End Homology Region (HR2). Several bases are deleted from the figure (*).

DNA Origin and Terminus

6399

Fig. 8: Possible Large Subunit rRNA Terminal Structures. Included are bases 1 to 25 and 2887 to 2904 of E. coli 23s rRNA and bases 53 to 75 and 2227 to 2256 from Fig. 4.

would suggest that rDNA begins at base 53 and terminates at base 2256, giving a total rDNA length of 2204 bases, a size consistent with observed 20s rRNA molecular weight estimates (Fig. 1). This would place the Paramecium gene in a unique size class intermediate between E. coli and mammalian mt, closer to the size of yeast 21s mt RNA.

Other Features of the 20s rDNA Sequence

Beginning around bases 24 and 888 in Fig. 4 are sequences from which can be drawn structures that strongly resemble portions of tRNA genes (Fig. 9) for formyl-methionine (A) and isoleucine (B). Base-pairing alterations and deviations with respect to general E. coli tRNA structure are indicated. Paramecium tRNAs seem much more similar to E. coli tRNAs than they do to other mt counterparts (see below), and hence we felt that E. coli tRNAs made more

here, then part of structure B would represent an exon which is also an intron within rDNA. Although previous studies using R-looping (27) found no large introns within 20s rDNA, small introns (<100 bp) would have gone undetected. Interestingly, structure "B" resides at a different location within rRNA than the introns observed in fungi (7,8,14,18). Mapping with the single strand specific nuclease S_1 should help resolve this question.

tRNA tyr

Beginning 127 bp downstream from the proposed 3' end of 20s rDNA is a tRNA tyr gene. Its sequence and structure as deduced from the DNA sequence is illustrated in Fig. 10. Note in Fig. 4 that it is coded by the same strand as 20s rRNA, and that it is flanked by 40 bp on the left and 88 bp on the right by sequences consisting of 90-91% A + T. Also note that although mt tRNA genes usually do not contain the 3'-terminal CCA within the sequence (15,20,21) a CCA sequence is present here, located two bases downstream from its usual position. Whether this DNA-coded CCA is included within the mature tRNA via excision of the two intervening nucleotides (UA), or whether their presence is merely coincidental merits further study.

We have assumed its tyrosine designation based on its anticodon (GUA), however, its specificity has yet to be verified by in vitro charging reactions. Human, mouse and rat mt rDNAs are flanked at the 3' end with leucine tRNAs (6,19-21) while yeast, Neurospora, and Aspergillus contain threonine tRNAs at

Fig. 10: tRNA tyr as Deduced from DNA Sequence.

the analogous loci (15,36,37). Therefore, an active tRNA gene here represents a significant divergence. Transcription and function of this gene has not yet been demonstrated; however, its location and completeness of structure suggest that it probably is utilized by the mt. Its identity and function is supported further by its homology to tRNA tyr sequences from other organisms. In Fig. 11, *Paramecium* tRNA tyr is compared with its counterparts from *E. coli* (38), *Neurospora* mt (39), and human mt (21). Sequence conservation, indicated by boxed regions, is exhibited particularly within the anticodon and T ψ C regions, and the distal portions of the aminoacyl stem. Some conservation is seen also within the more variable D region and in the length of the variable stem. Note that as before, the *Paramecium* sequence shows markedly greater homology to the *E. coli* sequence (60%) than to the *Neurospora* mt tRNA tyr (41%), a sequence particularly described as "prokaryote-like" (39). Interestingly, the *Paramecium* sequence shows slightly greater homology to human than *Neurospora* mt, although the figure for human could be inaccurate due the uncertainty of nucleotide placement within the highly deleted variable and D loop regions (deletions have been counted as a mismatch).

Downstream Region

The region 3' to the tRNA tyr gene has not been characterized. Although a transcript has yet to be mapped within this region, three AUG's can be found in Fig. 4 on the same strand as the rRNA and tRNA genes, at positions 2623, 2740, and 2829. The first two reading frames terminate within 16 and 4 triplet codons, respectively. The 2829 "start", however, constitutes a potentially open reading frame which runs 25 codons before reaching the end of H5. This possible ORF is preceded by three "TATA" sequences at -15, -24, and -33, and a complete "Goldberg/Hogness" box at -42, some or all of which could con-

	Aminoacyl Stem				D Stem		D Loop		D Stem		Acodon S		Anticodon L	
<i>E. coli</i>	5'-	<u>G</u>	<u>A</u>	U G G G G	<u>U</u>	<u>U</u>	C C C G	<u>A</u>	<u>G</u>	<u>G</u>	A	G C <u>A</u> <u>G</u> <u>A</u>	<u>C</u> <u>U</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>A</u>	
<i>Paramecium</i> mt	5'-	<u>G</u>	<u>A</u>	A G U A A	<u>U</u>	<u>U</u>	G C U G	<u>A</u>	<u>G</u>	<u>U</u>	G	G C <u>A</u> <u>G</u> <u>A</u>	<u>C</u> <u>U</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>A</u>	
<i>Neurospora</i> mt	5'-	<u>A</u>	<u>G</u>	G A G G G	<u>U</u>	<u>U</u>	C C G U	<u>U</u>	<u>U</u>	<u>G</u>	G	U U <u>A</u> <u>A</u> <u>G</u>	<u>C</u> <u>U</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>A</u>	
Human mt	5'-	<u>G</u>	<u>U</u>	A A A A A	<u>U</u>	<u>G</u>	C C U G	<u>A</u>	<u>G</u>	<u>-</u>	<u>-</u>	<u>A</u> <u>U</u> <u>G</u> <u>G</u> <u>A</u>	<u>C</u> <u>U</u> <u>G</u> <u>U</u> <u>A</u> <u>A</u> <u>A</u>	

	Acodon S	Variable Stem	T ψ C Stem	T ψ C Loop	T ψ C Stem	Aminoacyl Stem	Bases conserved	
e	<u>U</u> <u>C</u> <u>U</u> <u>G</u> <u>C</u>	<u>C</u> <u>G</u> <u>U</u> <u>C</u> <u>A</u> <u>U</u> <u>C</u> <u>G</u> <u>A</u> <u>C</u> <u>U</u> <u>U</u> <u>C</u>	<u>G</u> <u>A</u> <u>A</u> <u>G</u> <u>G</u>	<u>T</u> <u>U</u> <u>C</u> <u>G</u> <u>A</u> <u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>U</u> <u>U</u> <u>C</u>	<u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>A</u> <u>C</u> <u>C</u>	<u>A</u>	49 60
p	<u>U</u> <u>C</u> <u>U</u> <u>G</u> <u>U</u>	<u>U</u> <u>G</u> <u>U</u> <u>G</u> <u>A</u> <u>U</u> <u>A</u> <u>C</u> <u>C</u> <u>G</u> <u>U</u> <u>C</u>	<u>G</u> <u>U</u> <u>U</u> <u>G</u> <u>G</u>	<u>T</u> <u>U</u> <u>C</u> <u>G</u> <u>A</u> <u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>A</u> <u>A</u> <u>C</u>	<u>U</u> <u>U</u> <u>A</u> <u>C</u> <u>U</u> <u>U</u> <u>C</u>	<u>A</u>	(82) --
n	<u>C</u> <u>U</u> <u>U</u> <u>A</u> <u>A</u>	<u>U</u> <u>G</u> <u>A</u> <u>C</u> <u>U</u> <u>A</u> <u>A</u> <u>U</u> <u>A</u> <u>G</u> <u>U</u> <u>C</u>	<u>G</u> <u>A</u> <u>A</u> <u>G</u> <u>G</u>	<u>T</u> <u>U</u> <u>C</u> <u>A</u> <u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>U</u> <u>U</u> <u>U</u>	<u>U</u> <u>C</u> <u>U</u> <u>C</u> <u>C</u> <u>U</u>	<u>A</u>	34 41
h	<u>U</u> <u>C</u> <u>U</u> <u>A</u> <u>A</u>	<u>A</u> <u>G</u> <u>A</u> <u>C</u> <u>-</u> <u>-</u> <u>-</u> <u>-</u> <u>-</u> <u>-</u> <u>-</u>	<u>A</u> <u>G</u> <u>G</u> <u>G</u> <u>G</u>	<u>T</u> <u>U</u> <u>A</u> <u>G</u> <u>G</u> <u>-</u> <u>C</u>	<u>C</u> <u>U</u> <u>C</u> <u>U</u> <u>U</u>	<u>-</u> <u>U</u> <u>U</u> <u>U</u> <u>A</u> <u>C</u> <u>C</u>	<u>A</u>	40 49

*Indicates the position of bases deleted from the sequence shown here.

Fig. 11: tRNA tyr Interspecies Homology. Four tRNA tyr genes are compared; homologous regions are boxed and quantitated at right. *Paramecium* T within T ψ C loop was assumed. D = dihydrouridine.

stitute transcriptional control signals. We hope to locate any gene within this region by S_1 mapping and Northern blotting techniques.

DISCUSSION

The present work suggests that Paramecium mt rRNA gene organization may be distinct from both fungal and mammalian mt systems, but also may retain some features of both. The large 25 μ circular yeast mt DNA molecule affords the luxury of suspending its genes and gene pieces (i.e. exons) between large spacers consisting mostly of very A+T-rich DNA; so much, in fact, that the overall genomic A+T ratio is 82% (23,36,40). In contrast, mammalian mt genomes have compressed a similar amount of genetic information into a 5 μ circle by elimination of almost all intergenic and intragenic (i.e. introns) spacer DNA sequences (19-22). Consequently, a Paramecium mt DNA of 14 μ length and 63% A+T (25) might well be expected to represent an intermediate case (i.e. some gene spacer sequences present, but perhaps fewer, shorter, and less rich in A+T than in yeast).

The H5 fragment is probably a valid representative of the Paramecium mt genome: it represents 7% of the DNA sequence and has a similar (66%) A+T content. It contains two complete genes (and perhaps part of a third). No long stretches of extremely high (>90%) A+T content such as those found in yeast are present, although shorter stretches of 40 and 88 bp length immediately surround the tRNA gene. Taken as a whole, the 127 bp 77% A+T spacer connecting the rRNA and tRNA tyr genes is indeed shorter and less A+T-rich than analogous spacers in yeast (assuming correct placement of the rDNA 3' end). This spacer resembles somewhat more closely the 114 bp 86% A+T rDNA 3' spacer of Aspergillus mt, whose total genome is even smaller (11 μ , 15). Unfortunately, not enough of the rDNA 5' end spacer has been sequenced that transcriptional signals and/or tRNAs to the left of the gene could be found. Nevertheless, it appears from the sequences examined here that Paramecium has undergone, to an extent, the same mt genomic reduction process of which mammalian mt genomes are such an extreme example. This is also reflected by the reduction in rRNA gene size. This process, however, clearly has stopped short of removing all intergenic spacers, making the single transcript and tRNA-directed cleavage model proposed for mammalian cell mt transcription (18-22) seem unlikely here.

The evolution of large mt rDNA sequence (Figs. 5,6) appears roughly as expected; however the small degree of change between Paramecium mt and E. coli seen within HR2 and the 5'/3'-end structure of rDNA and within the tRNA tyr

gene is somewhat surprising. It will be interesting to see if the genetic code utilized by *Paramecium* mt contains the alterations associated with other mt systems (21,36,41,42), or whether the more conventional prokaryotic and eukaryotic cytoplasmic code is used. In other respects, particularly large rRNA size, and genome conformation and replication (23,27), *Paramecium* mt are quite unlike bacteria. We therefore caution that although these results contain implications toward endosymbiotic origin of organelles, these implications may reflect individual gene evolution rather than organelle evolution, making any deductions of evolutionary origin of organelles drawn directly from these data uncertain at best.

ACKNOWLEDGEMENTS

We would like to thank Gary Olsen and Sheldon Nicholl for help with computer analysis, and Bernard Dujon for helpful suggestions. This work was supported by NIH GM 21948 (DJC) and 1-F32-AG-05221-01 (JJS).

REFERENCES

1. Jacq, B. (1981) *Nucl. Acids Res.* 9, 2913-2932.
2. Edwards, K., and Kossel, H. (1981) *Nucl. Acids Res.* 9, 2853-2869.
3. Delius, H., and Koller, B. (1980) *J. Mol. Biol.* 142, 247-261.
4. Bohnert, H.J., Gordon, K.H.J., and Crouse, E.J. (1980) *Mol. Gen. Genet.* 179, 539-545.
5. Eperon, I.C., Anderson, S., and Nierlich, D.P. (1980) *Nature* 286, 460-467.
6. Saccone, C., Cantore, P., Gadaleta, G., Gallerani, R., Lanave, C., Pepe, G., and Kroon, A.M. (1981) *Nucl. Acids Res.*, 9, 4139-4150.
7. Kearsey, S.E. and Craig, I.W. (1981) *Nature* 290, 607-608.
8. Dujon, B. (1980) *Cell* 20, 185-197.
9. Blanc, H., Wright, C.T., Bibb, M.J., Wallace, D.C., and Clayton, D.A. (1981) *Proc. Natl. Acad. Sci.* 78, 3789-3793.
10. Baer, R.J. and Dubin, D.T. (1981) *Nucl. Acids Res.* 9, 323-337.
11. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K., and Kossel, H. (1981) *Nucl. Acids Res.* 9, 3287-3306.
12. Veldman, G.M., Klootwijk, J., deJonge, P., Leer, R.J., and Planta, R.J. (1980) *Nucl. Acids Res.* 8, 5179-5192.
13. Dubin, D.T., Timko, K.D., and Baer, R.J. (1981) *Cell* 23, 271-278.
14. Bos, J.L., Osinga, K.A., VanderHorst, G., Hecht, N.B., Tabak, B.F., Van Ommen, G.J., Borst, P. (1980) *Cell* 20, 207-214.
15. Kocheil, H.G., Lazarus, C.M., Basak, N., and Kuntzel, H. (1981) *Cell* 23, 625-633.
16. Green, M.R., Grimm, M.G., Goewert, R.R., Collins, R.A., Dole, M.D., Lambowitz, A.M., Heckman, J.E., Yin, S., RajBhandary, U.L. (1981) *J. Biol. Chem.* 256, 2027-2034.
17. Levens, D., Ticho, B., Ackerman, E., and Rabinowitz, M. (1981) *J. Biol. Chem.* 256, 5226-5232.
18. Lazarus, C.M., Lundsford, H., Hahn, U., Stepien, P.O., and Kuntzel, H. (1980) *Mol. Gen. Genet.* 177, 389-397.
19. Ojala, D., Merkel, C., Gelfand, R., and Attardi, G. (1980) *Cell* 22, 393-403.

20. Van Etten, R.A., Walberg, M.W., and Clayton, D.A. (1980) *Cell* 22, 157-170.
21. 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., Sander, F., Schrier, P.H., Smith, A.J.H., Standen, R., and Young, I.G. (1981) *Nature* 290, 457-465.
22. Battey, J., and Clayton, D.A. (1980) *J. Biol. Chem.* 255, 11599-11606.
23. Borst, P. (1972) *Ann. Rev. Biochem.* 41, 333-376.
24. Goddard, J.M., and Cummings, D.J. (1975) *J. Mol. Biol.* 97, 593-609.
25. Goddard, J.M., and Cummings, D.J. (1977) *J. Mol. Biol.* 109, 327-344.
26. Pritchard, A., and Cummings, D.J. (1981) *Proc. Natl. Acad. Sci.* (in press).
27. Cummings, D.J., Maki, R.A., Conlon, P.J., and Laping, J. (1980) *Mol. Gen. Genet.* 178, 499-510.
28. Parish, J.H., and Kirby, K.S. (1966) *Biochem. Biophys. Acta* 129, 554-560.
29. Bailey, J.M., and Davidson, N. (1976) *Anal. Biochem.* 70, 75-85.
30. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
31. Maxam, A.M., and Gilbert, W. (1980) in *Methods in Enzymology*, Vol. 65, L. Grossman and K. Moldave, eds, New York: Academic Press, 499-595.
32. Cummings, D. J., and Laping, J. (1981) *Mol. Cell. Biol.* (in press).
33. Brosius, J., Dull, T.J., and Noller, H.F. (1980) *Proc. Natl. Acad. Sci.* 77, 201-204.
34. Orozoco, E.M., Rushlow, K.E., Dodd, J.R., and Hallick, R.B. (1980) *J. Biol. Chem.* 255, 10997-11003.
35. Koch, W., Edwards, K. and Kossel, H. (1981) *Cell* 25, 203-213.
36. Bonitz, S.G., Berlan, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F., Nobrega, M.P., Thallenfeld, B.E., and Tzagoloff, A. (1980) *Proc Natl. Acad. Sci.* 77, 3167-3170.
37. Heckman, J.E., and RajBhandary, U.L. (1980) *J. Biol. Chem.* 17, 583-595.
38. Goodman, H.M., Abelson, J., Landy, A., Brenner, S., and Smith, J.D. (1968) *Nature* 217, 1019-1024.
39. Heckman, J.E., Alzner-DeWeerd, B., RajBhandary, U.L. (1979) *Proc. Natl. Acad. Sci.* 76, 717-721.
40. Borst, P., and Grivel, L.A. (1978) *Cell* 15, 705-723.
41. Heckman, J.E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S., and RajBhandary, U.L. (1980) *Proc. Natl. Acad. Sci.* 77, 3159-3163.
42. Barrell, B.G., Anderson, S., Bankier, A.T., de Bruijn, M.H.L., Chen, E., 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. (1980) *Proc. Natl. Acad. Sci.* 77, 3164-3166.