Phylogenetic relationships and altered genome structures among *Tetrahymena* mitochondrial DNAs

Gregg B. Morin and Thomas R. Cech

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

Received August 13, 1987; Revised November 25, 1987; Accepted December 2, 1987

ABSTRACT

Tetrahymena thermophila mitochondrial DNA is a linear molecule with two tRNAs, large subunit β (LSU β) rRNA (21S rRNA) and LSU α rRNA (5.8S-like RNA) encoded near each terminus. The DNA sequence of ~ 550 bp of this region was determined in six species of *Tetrahymena*. In three species the LSU β rRNA and tRNA^{Leu} genes were not present on one end of the DNA, demonstrating a mitochondrial genome organization different from that of *T. thermophila*. The DNA sequence of the LSU α rRNA was used to construct a mitochondrial phylogenetic tree, which was found to be topologically equivalent to a phylogenetic tree based on nuclear small subunit rRNA sequences (Sogin et al. (1986) *EMBO J.* 5, 3625-3630). The mitochondrial rRNA gene was found to accumulate base-pair substitutions considerably faster than the nuclear rRNA gene, the rate difference being similar to that observed for mammals.

INTRODUCTION

All eukaryotic organisms have at least two genetic systems. The vast majority of the coding capacity is contained in the nuclear genetic system, but a small set of important functions is encoded in the mitochondrial genetic system. Besides genes for rRNAs and tRNAs, mitochondrial DNA (mtDNA) generally encodes proteins involved in the electron transport chain. In contrast to the conserved genotype of mtDNA, the structure of mtDNA varies dramatically. Most mtDNAs are circular; higher eukaryotes have circular mitochondrial genomes of about 16 kbp, fungi have circular genomes of 30-75 kbp and plants have circular genomes ranging from 250 to 2500 kbp (1). However, the ciliated protozoa *Tetrahymena* and *Paramecium* have linear mtDNAs of about 50 kbp (2,3) and *Hydra* mtDNA consists of two linear plasmids of about 7.5 kbp each (4). The widely varying structure of mtDNA has led to the proposal that the different genomes arose from different endosymbiotic events (5).

The structure of *Tetrahymena pyriformis* mtDNA has been studied in some detail. The mtDNA has a near-terminal inverted duplication (the ribosomal duplication) of about 3 kbp containing large rRNA genes, and there is one centrally located copy of the small rRNA gene (6). The ribosomal duplication encodes the

LSU α rRNA (5.8S-like RNA) and the LSU β rRNA (21S rRNA) separated by a tRNA^{leu} gene (7). External to the ribosomal duplication is a region of DNA heterogeneous in length. This telomeric region is separated from the ribosomal duplication by tRNA^{tyr} and tRNA^{met} genes on the left and right sides, respectively (7). The heterogeneous telomeric region consists of 32 ± 9 tandem repeats of a 31 bp sequence (8). These characteristics also apply to *T. thermophila* mtDNA, except its telomeric repeat is composed of a different sequence, 53 bp in length (9). Electron microscopy has indicated that other species of *Tetrahymena* may have a different mtDNA structure, especially in the organization of the ribosomal duplication (10). The same report indicated a low degree of sequence similarity among the mtDNAs studied.

mtDNA has become a popular tool in phylogenetic experiments because of its small size and sequence conservation. The evolutionary rate-of-change in mammalian mtDNA has been shown to be about 5-10 times faster than that of nuclear DNA (11-13). This high divergence rate has allowed researchers to finely map evolutionary relationships in closely related species (14,15), which is difficult to do with nuclear markers. Using mtDNA as a molecular clock researchers have estimated divergence times for organisms and tracked ancient migration routes (16). The validity of using mtDNA as a molecular clock has recently been questioned on the basis that the nuclear and mitochondrial evolutionary divergence rates might not be linked (17). In *Drosophila*, mitochondrial and nuclear phylogenetic trees have been reported to have different branching patterns, suggesting mitochondria can evolve separately from nuclei (18,19); however, nucleotide sequence comparisons of functionally equivalent genes were not reported. Nuclear phylogenetic trees have been constructed for *Tetrahymena* based on electrophoretic mobilities of certain enzymes and the sequences of rRNAs (20-22). Studying the evolutionary relationships of *Tetrahymena* mtDNA helps address the generality of mtDNA as a useful molecular clock, and provides information on nuclear/mitochondrial evolution for an organism that diverged long before the organisms for which similar information is known.

We studied the structure and sequence of the terminal regions of the mtDNA from six species of *Tetrahymena*. We found that the LSU α rRNA gene is present at both ends of all six mtDNAs, but the LSU β rRNA gene is present on only one side in three species. Thus, the mitochondrial genomes of *Tetrahymena* are relatively plastic in the organization of the ribosomal duplication, and gross organizational changes are tolerated. The DNA sequence of the LSU α rRNA gene was used to construct a phylogenetic tree. The mtDNA phylogeny was found to be a sensitive tool to evaluate phylogenetic relationships among different species of

Tetrahymena. The rate of nucleotide substitution in the mtDNA is much higher than that of the nuclear DNA.

MATERIALS AND METHODS

Nucleotides and Enzymes

Labeled and unlabeled nucleoside triphosphates were obtained from New England Nuclear and P-L Biochemicals, respectively. 7-deaza-dGTP was obtained from Boehringer Mannheim. Restriction enzymes were purchased from New England Biolabs, T4 polynucleotide kinase from U.S. Biochemicals, and AMV reverse transcriptase from Life Sciences. T7 RNA polymerase was provided by Art Zaug.

Tetrahymena Strains, mtDNA Isolation and Cloning of mtDNA Termini

All *Tetrahymena* stocks were maintained in 1% proteose-peptone as described by Zaug and Cech (23). mtDNA was isolated as described in Morin and Cech (8). The cloning of terminal fragments is described in Morin and Cech (8,9). The following table details the *Tetrahymena* strains and plasmids used.

DNA Sequencing

DNA sequencing was performed using a modified dideoxynucleotide method (24). ³²P label was incorporated into oligonucleotide primers by 5'-end labeling with γ -³²P-ATP and T4 polynucleotide kinase (25). All oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizer. In a 0.6 ml eppendorf tube, 2 pmol of a 5' end-labeled primer was mixed with 0.2 pmol of linearized or supercoiled plasmid DNA in 12 μ l of annealing buffer (50 mM Tris-HCl, pH 8.3, 60 mM NaCl and 10 mM dithiothreitol). The mixture was heated to boiling (95°C) for 5 min and immediately placed on ice. 2 μ l of this solution was put in each of five tubes containing 1 μ l of one of the dideoxynucleoside triphosphates (400 μ M),

Species	Strain	Mating Type	Telomeric DNA Plasmid
<i>thermophila</i>	B	VII	pTmt-Rs1 (R) ^a & Ls1 (L)
<i>malaccensis</i>	MP75	II	pTmt-MP75endA (L) & B (L)
<i>pyriformis</i>	ST	NA ^b	pTmt-STendA (R) & B (L)
<i>pigmentosa</i>	UM1060	II	pTmt-UM1060endA (R) & B (R) & D (L)
<i>pigmentosa</i>	UM1091	I	pTmt-UM1091endA (R) & B (R)
<i>hyperangularis</i>	EN10	I	pTmt-EN10endA (R) & B (R)
<i>hegewischi</i>	KP7	II	pTmt-KP7endA (R) & B (R)

^a(R), right end; (L), left end.

^bNA, not applicable (*T. pyriformis* is amiconucleate)

the last tube containing 1 μ l of annealing buffer. 2 μ l of enzyme/nucleotide buffer (125 mM Tris-HCl, pH 8.3, 150 mM NaCl, 6.25 mM dithiothreitol, 75 mM Mg acetate, 1 mM of each deoxynucleoside triphosphate, and 1 unit of AMV reverse transcriptase) was added to each tube. The tubes were incubated at 37°C for 30 min, and stopped with the addition of 5 μ l of deionized formamide with 0.06% each of bromophenol blue and xylene cyanol. The mixtures were heated to 95°C for 2 min and 3-5 μ l of each mixture was fractionated on a 6% polyacrylamide, 8 M urea gel. The sequencing was accomplished using six primers, three for each strand, complementary to conserved regions of the *T. pyriformis* and *thermophila* α -fragment sequences, whose sequences were known (7; unpublished data), and a primer complementary to the 5' region of the tRNA^{Leu} gene. In this way, both strands of the α -fragment were sequenced, except for the approximately 25 3'-terminal nucleotides and the regions upstream and downstream of the α -fragment, which were sequenced on one strand only. Two independent plasmid DNA clones of each species were sequenced, the second sequencing being performed with 7-deaza-dGTP substituted for dGTP to minimize sequence anomalies due to compression of G+C-rich regions during polyacrylamide gel electrophoresis (26). In some cases plasmid clones were not available and the sequence was obtained by directly sequencing gel-purified mtDNA restriction fragments (see Figure 3).

Gel Blotting and Hybridization

Blotting, hybridization and T7 RNA polymerase transcriptions were performed as described in Morin and Cech (9). ³²P-labeled T7 RNA polymerase transcripts of *Bam*H I digested pTmt-Rs1 (9), a 462 bp α -fragment clone, and *Bam*H I digested pTmt-Eef, a 170 bp β -fragment clone, were used as hybridization probes for the experiment in Figure 1.

RESULTS

mtDNA Restriction Site Analysis

The *Eco*R I and *Hind* III restriction fragment sizes for mtDNA from 13 strains representing 6 species of *Tetrahymena* are shown in Table 1. The restriction patterns for all the species are easily distinguishable; there is little conservation of restriction sites. *T. thermophila* and *T. malaccensis* have a common *Eco*R I fragment as do *T. hyperangularis* and *T. hegewischi*. No other restriction fragments appear to be common among the six species. This suggests low sequence similarity among the mtDNAs. This same result has been previously observed for other species of *Tetrahymena* (10), and is in contrast to the relatively high similarity among four sibling species of *Paramecium* reported by Cummings (27). Within the *Tetrahymena* mitochondrial rRNA genes, on the other hand, restriction sites are

Table 1. Restriction enzyme fragment sizes of *Tetrahymena* mitochondrial DNAs

species:	<i>the</i> ^a	<i>mal</i>	<i>pyr</i>	<i>pig</i>	<i>pig</i>	<i>hyp</i>	<i>heg</i>
strain ^b :	BVII C3-368 PB9R	MP75	ST GL nGL UM918	8ALP	UM1060	EN10	KP7
enzyme: <i>EcoR</i> I							
	13500 ^c	34000	14600E	26000E	—	15800	13200
	13000	—	9600	17500	17500	7400E	8600
	7000	7000	8100E	7300	7300	7200	7300E
	6500	3100*,E	6500	3400	6500	5700E	5400E
	3500	3100	4000	3150	—	3150	4250
	3000	2800	2750	—	6500E	2400	3100
	1500E,D	1330	1950	—	2200E	1300	2450
		1280*,E	1900	2050	2050	1130	—
			300	1950E	—	630	630
				1650	1650		
				1320	1320		
totals:	49500	52610	49700	64320	45020	44710	44900
enzyme: <i>Hind</i> III							
	12800	15000	13500	~38000E	—	15700	19500
	8900	11800	8600	24000	24000	11800	7900
	4480D	8200	5350	5900	5900	6500E	7600
	4050	5000	4000	—	2850E	5900	3300
	3000E,D	4250*,E	3650	620	620	2550E	2750
	2480	2630*,E	2500E,D	380	380	2400	2600E
	1900	1080	2180				2200
	1550		1400				600
	900		940				
	700		810D				
	550D		780				
	300		690				
			620				
totals:	49640	47960	48330	68900	33750	44850	46450

^aAbbreviations: *the*, *T. thermophila*; *mal*, *T. malaccensis*; *pyr*, *T. pyriformis*; *pig*, *T. pigmentosa*; *hyp*, *T. hyperangularis*; *heg*, *T. hegewischi*; E, end fragment; D, doublet; ~, size is approximate ($\pm$ 5 kbp); *, detected by blotting and hybridization with a telomeric repeat probe.

^bStrains are described in Materials and Methods and in ref. 8.

^cFragment sizes are in base-pairs.

often conserved (not shown), possibly indicating higher sequence conservation within the rRNA cistrons than the rest of the mtDNA.

The total size of the genome varies from 45 to 64 kbp based on the *EcoR* I data. The total size of the *Hind* III fragments for the two *T. pigmentosa* strains are inaccurate since only one end fragment was found for each strain. The ends of *Tetrahymena* mtDNA are heterogeneous in length (10), and sometimes cannot be detected by ethidium bromide staining since they are too diffuse. Digestion with

other restriction enzymes detects both ends in *T. pigmentosa* (not shown), giving total sizes that agree well with those obtained using *EcoR* I. Other than *T. pigmentosa* the size totals determined by the two enzymes are in close agreement. *T. thermophila*, *malaccensis* and *pyriformis* mtDNAs are approximately 50 kbp, *T. pigmentosa* UM1060, *hyperangularis* and *hegewischi* mtDNAs are approximately 45 kbp, and *T. pigmentosa* 8ALP mtDNA is approximately 64 kbp.

No restriction site polymorphism was observed among the several strains of *T. thermophila* and *T. pyriformis* examined. There is an additional *EcoR* I site in *T. pigmentosa* 8ALP that splits the *T. pigmentosa* UM1060 6500 bp fragment into fragments of 3400 and 3150 bp; this is the only restriction site polymorphism observed within a species. The mtDNA of *T. pigmentosa* 8ALP is approximately 19 kbp larger than that of *T. pigmentosa* UM1060, because the right end fragment is larger in the 8ALP strain (data not shown). The origin of this 19 kbp is unknown. The non-terminal restriction fragments are the same size. *T. pigmentosa* UM1091 mtDNA has identical restriction fragments and genome size to those of UM1060 (not shown). Thus, 8ALP appears to be an unusual strain.

Organization of the Ribosomal Duplication

Figure 1 shows *T. hegewischi*, *hyperangularis*, *pigmentosa*, and *thermophila* mtDNAs digested with various restriction endonucleases, separated by agarose gel electrophoresis, blotted and hybridized with either a LSU α rRNA (α -fragment) probe (left panel) or a LSU β rRNA (β -fragment) probe (right panel). The α -fragment is present on both terminal fragments of all the digests (see also below). The α -fragment is also present on both ends of *T. pyriformis* and *malaccensis* mtDNAs (7; see also below). The sub-stoichiometric band in *T. pigmentosa* 8ALP mtDNA (~ 10.5 kbp) represents a sub-population of mtDNA with a DNA insertion or rearrangement on the left end (unpublished data). Only one side of *T. hegewischi*, *hyperangularis* and *pigmentosa* mtDNAs are recognized by the β -fragment probe. Both sides of *T. thermophila* mtDNA are recognized by the β -fragment probe, as is the case in *T. pyriformis* and *malaccensis* (not shown). The simplest explanation is that the ribosomal duplication is much smaller in *T. hegewischi*, *hyperangularis* and *pigmentosa* (the *pigmentosa* family) than in the remaining three species. Sequencing experiments confirmed this result (below) and showed both the LSU β rRNA and the tRNA^{leu} genes are missing from the left sides of the mtDNAs in the *pigmentosa* family, and only the LSU α rRNA gene is duplicated (Figure 2).

Sequences of the LSU α rRNA Genes

To study the evolutionary relationships between the mtDNAs of the six species, the terminal fragments were cloned. The LSU α rRNA represents a gene that is present in all genetic systems and has a conserved function, and thus it is

Figure 1. Nucleic Acid Hybridization Shows Absence of a Complete Ribosomal Duplication in mtDNA from the *T. pigmentosa* family. Purified mtDNA from the indicated species and strains was digested with the marked restriction enzymes, fractionated by electrophoresis on 0.8% agarose gels, blotted and hybridized with an α -fragment (α) or β -fragment (β) probe (see Materials and Methods). The species abbreviations are the same as in Table 1. The restriction enzyme abbreviations are: E, *EcoR* I; H, *Hind* III; N, *Nco* I; P, *Pst* I. The Markers (m) are 32 P-end-labeled λ *Xho* I restriction fragments and 1 kbp ladder (Bethesda Research Laboratory); the fragment sizes are indicated.

Figure 2. *Tetrahymena* mtDNA Genome Organizations. (A) Genome organization for *T. thermophila*, *malaccensis* and *pyriformis*. (B) Genome organization for *T. pigmentosa*, *hyperangularis* and *hegewischi* (the *pigmentosa* family). Only the end regions are depicted. α and β stand for the LSUα rRNA (α-fragment) and LSUβ rRNA (β-fragment) genes, respectively. Tyr, leu and met represent the respective tRNA genes. The direction of transcription is indicated. The vertical and angled lines represent telomeric repeats. The angled lines indicate a different telomeric repeat than is present on the other side (8). The right side (R) is defined by the presence of the SSU rRNA gene closer to this end (not shown).

a good marker for studying phylogenetic relationships. The sequences of the α-fragments and surrounding regions are presented in Figure 3. For *T. thermophila*, *malaccensis* and *pyriformis* mtDNAs (the genomes with A-type organization) the left and right side sequences are presented as one, but downstream of the α-fragment the sequences on the two sides separate. The *T. pyriformis* sequence was previously known (7); our sequence is in complete agreement with that obtained previously and extends it both upstream and downstream. In *T. pigmentosa*, *hyperangularis* and *hegewischi* mtDNAs (the genomes with B-type organization) the left and right side α-fragment sequences are presented as one; upstream and downstream the sequences on the two sides separate. Downstream from the terminal tRNA genes are the telomeric repeats, the sequences of which are presented elsewhere (8). The sequence joining the terminal tRNA gene and the telomeric repeats varies from 0–48 bp among the different species (Figure 3). Upstream of the α-fragment in *T. pigmentosa* and *hegewischi* mtDNAs, the sequence showed that the tRNA^{leu} gene is not present. Furthermore, even though there were ambiguities at some positions in the sequence, the data clearly showed the LSUβ rRNA gene to be missing. This information supports the B-type genome organization shown in Figure 2. The coding regions in the ribosomal duplication are very tightly spaced, separated by 2 to 17 bp. These spacer regions are much less conserved than the coding regions.

The 5' and 3' endpoints of the α-fragment in *T. pyriformis* were established by direct RNA sequencing (7). The endpoints in the other five species were placed by aligning the sequences with the *T. pyriformis* sequence and in some cases by

noting where the ribosomal duplication ends, which places the 3' end. The α -fragment contains 277 nucleotides in *T. thermophila*, 280 nucleotides in *pyriformis*, and 281 nucleotides in *T. malaccensis*, *pigmentosa*, *hyperangularis* and *hegewischi*. The 3' end of the β -fragment was determined by comparing to the known 3' end of the *T. pyriformis* β -fragment (7). The tRNA genes are discussed below.

Figure 3 also shows the positions of proposed secondary structures in the α -fragment RNA above the DNA sequence as a set of palindromes. The structures are those proposed by Heinonen et al. (7) for the *T. pyriformis* α -fragment, and conform with structures proposed for the *E. coli* LSU rRNA (28). We numbered those structures that form palindromes within the α -fragment and lettered those sequences that base-pair with sequences in the β -fragment. Phylogenetic comparisons are useful in analyzing presumed secondary structure. Sequence a in the α -fragment can base-pair with sequence a' at the 3' terminus of the β -fragment (Figure 3) in all cases even though there is considerable sequence variation. The sequences b, c, d and e presumably base-pair with portions of the β -fragment that have been sequenced only in *T. pyriformis*. All the proposed palindromes were conserved in all six species and compensatory base changes were observed in palindromes 1 and 5. However, palindrome 1 contains one or two mismatches in all six species and palindrome 3 contains a T-T mismatch in the *pigmentosa* family. There are also compensatory base changes in palindrome 6, but occurrence of A-A as well as G-C, A-T and G-T pairs complicates interpretation. Therefore, the phylogenetic comparison presented here supports some features of the proposed secondary structure.

mtDNA Phylogenetic Tree

Table 2 shows the sequence similarity values and number of nucleotide changes between all the pair-wise combinations of mitochondrial α -fragments. Below the α -fragment data the equivalent data for the nuclear SSU rRNA gene (22) is shown for comparison. The SSU rRNAs contain 1752-1753 nucleotides while the α -fragments contain 277-281 nucleotides, yet in most cases there are more nucleotide changes in the α -fragment comparisons than the SSU rRNA comparisons. Thus, in all comparisons the similarity between mitochondrial α -fragments is significantly lower than the similarity between the corresponding nuclear SSU rRNAs. The third line below the similarity values shows how many times faster the mitochondrial nucleotide rate-of-change is compared to that of the nuclear SSU rRNA. Some of the nucleotide rate-of-change values are high or low because there are few nucleotide changes between those species, making them statistically unreliable. Excluding those values the average rate of nucleotide substitution in the mitochon-

[illegible]

```

      tRNAtyr <---          --> telomeric repeats/3' terminal side
L the GGTTCGATTCCGGTCCACCTCA/AAATTTAATTAGCAA/-->/3'
L mal ...../AATT/-->/3'
L pyr .....G...A...T.../AATTATTGTTAATATATAAATTATTTTTTAAATTTTATAATTATC/-->/3'
L pig .....G.....T.../ATTTTATTATTAACATAGTACATATAATAT/-->/3'
L hyp .....G.....T.../AAATATTATTAAATATTATTTTATAGTATAAAAAATAATATAA/-->/3'
L heg .....G.....T.../-->/3'
-----
      tRNAmet <---          --> telomeric repeats/3' terminal side
R the AGTGGCT/TTCGTT/-->/3'
R mal ...../AGTTATAAAATTTTATAATATTTAATTAT/-->/3'
R pyr ..C...C/TTATTATTAATATTATAATAACACATTATTACAAATG/-->/3'
R pig ..C...C/-->/3'
R hyp ..C...C/-->/3'
R heg ..C...C/TTTATTC/-->/3'

```

Figure 3. DNA Sequences of the Terminal Regions of *Tetrahymena* mtDNAs. The sequence determined for the *T. thermophila* BVII end region is shown. The sequences of the other species are shown only where they differ from that of *T. thermophila*; the dots indicate positions of nucleotide identity. The species abbreviations are the same as in Table 1. The coding regions are indicated and are separated by slashes. The duplicated regions of both sides are presented as one sequence. Downstream of the α -fragments in all species and upstream of the α -fragments in the *pigmentosa* family, the sequences of the two sides diverge. L and R indicate the left and right sides, respectively; when they are capitalized it indicates that side was sequenced. The lower case indicates that side was not sequenced, but inferred from the other side where the sequence is duplicated. The lower case nucleotides in the *T. pyriformis* and *pigmentosa* α -fragments are positions where the left and right sides have different sequences. The regions upstream of the left-side α -fragment in *T. hegewischi*, and downstream of the α -fragment on the left sides of *T. hyperangularis* and *hegewischi* and the right side of *T. malaccensis* were determined by directly sequencing gel-purified mtDNA restriction fragments, since these regions were not cloned. The lower case "n"s indicate places where the sequence was not determined because these regions were too close to the primer. The lettered and numbered regions above the α -fragment sequence refer to proposed secondary structures (see text). The 181 nucleotides of the α -fragment that correspond to the *E. coli* LSU rRNA gene are from the 5' end to the end of region d, from the start of stem 1 to the end of region e, and from the start of stem 4 to the 3' end of the α -fragment.

drial gene exceeds that of the nuclear gene by a factor of 11.

The similarity data were used to construct a phylogenetic tree according to the method used to construct the *Tetrahymena* nuclear SSU rRNA phylogenetic tree (21,22). The mitochondrial tree is shown in Figure 4, along with the relevant portion of the nuclear tree for comparison. The evolutionary distances were calculated from the data in Table 2 and are shown on the mitochondrial tree. The nuclear tree is drawn to a scale approximately 10 times larger than that of the mitochondrial tree. The nuclear tree was rooted by comparison to *Paramecium* SSU rRNA. The mitochondrial tree has not been rooted, but has been arranged to facilitate direct comparison with the nuclear tree. According to the mitochondrial tree, *T. thermophila* and *malaccensis* are grouped together, *T. pyriformis* is by itself and *T. pigmentosa*, *hyperangularis* and *hegewischi* form another group. The

Table 2. Sequence similarities among *Tetrahymena* mitochondrial LSU α rRNAs and macronuclear SSU rRNAs

species ^a		<i>mal</i>	<i>the</i>	<i>pyr</i>	<i>pig</i>	<i>hyp</i>	<i>heg</i>
<i>mal</i>	mt	---	17 ^b	46	57	56	53
	nuc		4	21	30	30	33
<i>the</i>	mt	.948 ^c	---	53	54	55	54
	nuc	.998 26x		19	30	30	33
<i>pyr</i>	mt	.845	.824	---	42	41	38
	nuc	.988 13x	.989 16x		22	22	25
<i>pig</i>	mt	.804	.819	.856	---	1	4
	nuc	.983 12x	.983 11x	.987 11x		0	9
<i>hyp</i>	mt	.808	.815	.859	.996	---	3
	nuc	.983 11x	.983 11x	.987 11x	1.000 -		9
<i>heg</i>	mt	.819	.819	.870	.986	.989	---
	nuc	.981 10x	.981 10x	.986 9x	.995 3x	.995 2x	

^aSpecies abbreviations are the same as Table 1.

^bThe upper right half of the table gives the number of base changes between the DNA sequences.

^cThe lower left half of the table gives the similarity values as determined by the method of Elwood and Sogin (21). The third line is how many times greater the mitochondrial divergence is relative to the nuclear divergence:

(1 - mt similarity)/(1 - nuc similarity).

mt: Mitochondrial LSU α rRNA values.

nuc: Macronuclear SSU rRNA values as determined by Sogin et al. (22).

mitochondrial tree is topologically equivalent to the nuclear tree.

The mitochondrial tree was able to differentiate *T. pigmentosa* and *hyperangularis*, while the nuclear tree could not. The nuclear SSU rRNAs of *T. pigmentosa* and *hyperangularis* have identical sequences, while the α -fragments have a single nucleotide change. However, another strain of *T. pigmentosa* (UM1091) has an α -fragment sequence different from that of the strain presented (UM1060) and identical to that of *T. hyperangularis*.

tRNA Gene Locations and Secondary Structures

The α -fragment and the β -fragment are separated by a putative tRNA^{leu} gene, and downstream of the ribosomal duplication there are putative tRNA^{tyr} and tRNA^{met} genes on the left and right sides, respectively. The positions of the tRNA genes are conserved in all six species (Figure 3). The location of the

Figure 4. **Mitochondrial and Nuclear Phylogenetic Trees.** (A) The phylogenetic tree constructed from the mitochondrial α -fragment sequence similarity values of Table 2 by the method of Elwood and Sogin (21). The evolutionary distance values are shown for each branch. The evolutionary distance is the horizontal component of the distance between nodes of the tree. The nuclear tree was rooted by comparison to the nuclear phylogenetic tree (see text). (B) The phylogenetic tree constructed from macronuclear SSU rRNA sequences by Sogin et al. (22). Only the species used in the present study are included. The nuclear tree is drawn to about a ten times larger scale than the mitochondrial tree (see text).

tRNA^{leu} gene was first found in *T. thermophila* (9; see also ref. 29) and the locations of the tRNA^{tyr} and tRNA^{met} genes were first found in *T. pyriiformis* (7; see also ref. 29). The tRNA^{tyr} has been isolated and sequenced (30), and found to

Figure 5. Phylogenetic Conservation of Proposed tRNA Secondary Structures. The sequences of the putative *T. thermophila* BVII tRNA genes from Figure 3 are drawn to indicate proposed RNA secondary structure. The nucleotide changes for each of the other species are indicated. The species abbreviations are the same as in Table 1. The dashed lines indicate possible base-pairing interactions.

correspond to the DNA sequence shown in Figure 5. RNA has not been sequenced for the tRNA^{leu} and tRNA^{met} genes, but stable transcripts from these genes were detected (7). The tRNAs are identified using the universal genetic code to translate the anticodons. The tRNA^{tyr} is 80 nucleotides, except in *T. malaccensis* where it is predicted to be 81 nucleotides. (Values do not include the non-encoded CCA at the 3' end.) The tRNA^{leu} and tRNA^{met} RNAs are predicted to be 81 and 67 nucleotides, respectively. The proposed secondary structures of the tRNAs are shown in Figure 5. The nucleotide changes are shown in context of their effect on the structures.

DISCUSSION

Mitochondrial Genome Organizations

T. pyriformis and *thermophila* have an approximately 3 kbp near-terminal inverted duplication called the ribosomal duplication (6,9). The ribosomal duplication encodes the LSU α rRNA, LSU β rRNA and tRNA^{leu} genes. External to the ribosomal duplication are tRNA^{tyr} and tRNA^{met} genes on the left and right sides, respectively (7,9). All the coding regions are separated by short regions that are highly variable in sequence. This feature is reminiscent of the organization of animal mtDNA, whose protein reading frames are separated by tRNA genes, and whose genome is transcribed as one polycistronic unit. The tRNAs have been postulated to act as signals in the processing of the major RNA transcript (31). It is possible that the *Tetrahymena* ribosomal duplication is transcribed as a single unit and that the tRNAs play a similar role.

Hybridization experiments showed all six species have the α -fragment present near both termini (Figure 1; see also ref. 8). However, hybridization and subsequent sequencing experiments indicated that the left sides of the mtDNAs of *T. pigmentosa*, *hyperangularis* and *hegewischi* (the *pigmentosa* family) are missing the expected copies of the β -fragment and tRNA^{leu} genes and that only the α -fragment gene is duplicated. These results lead us to propose the B-type genome organization (Figure 2B) for the *pigmentosa* family. The A-type genome organization shown in Figure 2A has been previously proposed for *T. thermophila* and *pyriformis* (7,9), and the results in this paper include *T. malaccensis* in this category. The B-type organization is missing the β -fragment and the tRNA^{leu} genes on the left side of the genome. The *pigmentosa* family also has different telomeric repeats on either side of the mtDNA in contrast to the A-type species (8). The B-type genomes are about 45 kbp and the A-type genomes are about 50 kbp; the size difference is partially accounted for by the loss of one copy of the β -fragment and tRNA^{leu} genes. More of the size difference can be explained by the observation that the *pigmentosa* family mtDNAs have a shorter average length of telomeric DNA repeats than the A-type members (unpublished data); 1-2 kbp remains to be explained.

Electron microscopy studies of renatured single-stranded mtDNAs (10) gave structures that can be interpreted as supporting both the A- and B-type genome organizations. The mtDNA structure of *T. pyriformis* W, phenoset C (now *T. furgasoni*, ref. 32) can be interpreted as having only the α -fragment terminal duplication and the same telomeric repeat on both sides. The *T. pyriformis* PP, phenoset E (now *T. Iwoffi*, ref. 32) mtDNA structure can be interpreted as having the full ribosomal duplication but with different telomeric repeats on the two sides. These structures are permutations of the A- and B-type organizations. Thus there appear to be at least four mtDNA organizations, with different arrangements of the ribosomal duplication and telomeric repeats. These arrangements might have arisen by recombinational deletion or duplication events.

T. pigmentosa 8ALP mtDNA is 19 kbp larger than that of *T. pigmentosa* UM1060 or UM1091. The right-side terminal restriction fragments are the only fragments that differ in size between these strains, and account for the entire size difference. The size difference might be due to very large numbers of telomeric repeats on the right side of 8ALP mtDNA, as seen for *T. malaccensis* and *hyperangularis* (unpublished data), or to a duplication of one end of the mtDNA as postulated for *T. pyriformis* GL (33).

Mitochondrial Phylogeny Based on rRNA and tRNA Sequences

Ribosomal RNA sequences have proven to be an accurate basis for establishing

and quantifying phylogenetic relationships (21,34). The LSU α rRNA gene, present near the mtDNA termini in all six species of *Tetrahymena*, contains sequences analogous to the 5' end of prokaryotic LSU rRNA (7). The mtDNA termini, including the α -fragment, had been previously cloned to study the telomeric DNA (8,9), making the α -fragment convenient for use in constructing a mtDNA phylogeny. Phylogenies based on 5S rRNAs and tRNAs can be criticized because these genes are small and have limited information content. The α -fragment is somewhat larger, and considering the number of nucleotide sequence changes accumulated, the accuracy of the phylogenetic tree is expected to be higher than that based on nuclear SSU rRNA genes.

The α -fragment and surrounding regions were sequenced (Figure 3) and the pair-wise α -fragment sequence similarities were used to construct the phylogenetic tree shown in Figure 4A. The relevant portion of a phylogenetic tree constructed using nuclear SSU rRNAs (22) is seen in Figure 4B. There is no difference in the branching pattern of the trees. A phylogenetic tree constructed using biochemical characteristics and sequences of the nuclear 5S, 5.8S and LSU rRNAs is essentially the same as that in Figure 4A (20). This differs from the situation in *Drosophila*, where a discordance of nuclear and mitochondrial trees has been reported (18,19). The *Tetrahymena* mtDNA tree is consistent with many other features of the mtDNAs. The A- and B-type genome organizations and the two basic genome sizes occupy separate branches of the tree. The only amiconucleate species, *T. pyriformis*, occupies its own branch. The base changes in the tRNA genes are consistent with the branching pattern of the tree (see below). The sequences and characteristics of the mitochondrial telomeric repeats from all six species are consistent with the phylogeny (8). The telomeric repeats evolve 6- to 60-fold faster than the α -fragment, and provide a better basis for differentiating between closely related species such as *T. pigmentosa* and *hyperangularis* (8).

A secondary structure has been proposed for the α -fragment RNA of *T. pyriformis* (7, see Figure 3). The sequences of the α -fragment from the five other species support some features of this structure. All the proposed duplex regions are conserved; when base changes occur, they usually maintain base-pairing or are accompanied by a second-site change that restores base-pairing. The α -fragments vary in size from 277 to 281 nucleotides, and all the insertions/deletions are in regions that vary highly in sequence and do not affect the proposed secondary structure.

The locations of the tRNA genes are conserved in all six species, and the proposed RNA structures are supported by the observed base changes. The predicted tRNA^{Leu} and tRNA^{Tyr} structures conform to the conserved tRNA

secondary structure, and all the invariant and semi-invariant nucleotides are present (7). The putative tRNA^{met} is small and has an abnormal D loop and stem. A compensatory base-pair change in the D stem supports the 3 bp stem formation, which would leave only a 4 base loop. The Levitt base-pair is not present, and the base following the anticodon is not a purine. There is no thymidine residue in position 8. However, the gene is not unusual in the context of other mitochondrial tRNA genes (35-37). Transfer RNA genes do not usually provide a good basis for phylogeny, because they are short and their sequences are highly constrained. Nevertheless, the three *Tetrahymena* mitochondrial tRNA sequences presented here give qualitatively the same phylogenetic tree as the more extensive comparison of LSU α rRNA sequences.

Using the sequences of the mitochondrial α -fragments, *Tetrahymena* phylogenetic trees can be constructed that, based on the number of nucleotide substitutions found, appear to be somewhat more accurate than trees constructed using nuclear SSU rRNA as a marker. The sequences of the mitochondrial telomeric repeats can be used to delineate very closely related species that are otherwise indistinguishable. Mitochondrial DNA sequences should prove useful in future studies of *Tetrahymena* evolution.

Faster Rate of Nucleotide Substitution in Mitochondrial than in Nuclear rRNA Genes

All the pair-wise sequence similarity values of the six α -fragments are presented in Table 2. When compared to similar data for the nuclear SSU rRNAs of *Tetrahymena* (19), the mitochondrial α -fragment appears to evolve approximately 11 times faster than the nuclear SSU rRNA. The mitochondrial α -fragment can be aligned with three regions of *E. coli* LSU rRNA: nucleotides 1-45, 184-268 and 295-347 (7,28). These three regions are highly conserved in evolution relative to LSU rRNA as a whole, and they accumulate nucleotide substitutions only 2.1 times faster than the 950 positions of SSU rRNA that can be aligned among all organisms (G. Olsen, pers. commun.). When the *Tetrahymena* nuclear SSU rRNA comparison of Sogin et al. (22) is restricted to these same 950 positions, there is either 0 or 1 nucleotide change in each pair-wise comparison (G. Olsen, pers. commun.). In contrast, the mitochondrial α -fragment pair-wise comparisons show as many as 35 nucleotide changes in the 181 nt that can be directly aligned with the *E. coli* LSU rRNA (Fig. 3). Comparing the 1 change in 950 positions for the nuclear SSU rRNA with the 35 changes in 181 positions for the mitochondrial α -fragment, and correcting for the 2.1-fold rate difference expected for these two rRNAs, the α -fragment accumulates nucleotide substitutions 90 times faster. The α -fragment corresponds to sequences in the nuclear 5.8S rRNA (7,38,39). Comparing the 1

change in 154 nucleotides for *Tetrahymena* nuclear 5.8S rRNA (*the* vs *pyr*; ref. 38; *the* vs *pig*; ref. 20) with the 26 or 27 changes in the corresponding 129 nucleotides of the α -fragment, the α -fragment accumulates substitutions 30 times faster. Thus, the 11-fold higher rate-of-change calculated for the mitochondrial α -fragment compared to the entire nuclear SSU rRNA gene underestimates the rate of substitution in mitochondrial rRNA compared to nuclear rRNA genes. The rate difference we observe in *Tetrahymena* is of the same order as that of mammals, where mitochondrial DNA (and mitochondrial rRNA genes) evolve 5-10 times faster than single-copy nuclear DNA (11-13), which in turn is estimated to evolve about 40-fold faster than nuclear rRNA genes (40). Therefore, in mammals the mitochondrial rRNA genes are estimated to evolve about 200-400 times faster than nuclear rRNA genes.

In the pair-wise comparisons of the α -fragment sequences, transversions exceeded transitions by a ratio of about 70:30. This was true even for the most closely related species, such as *T. thermophila* and *malaccensis*, or *T. hyperangularis* and *hegewischi*, and contrasts to the preponderance of transitions seen in comparisons of mammalian mtDNA (12) and closely related *Drosophila* species (41). Most of the transversions were A $\longleftrightarrow$ T, suggesting that there might be selective pressure to maintain the A + T-content of the gene. Excluding the A $\longleftrightarrow$ T transversions, the number of transversions was substantially lower than the number of transitions. The high frequency of A $\longleftrightarrow$ T transversions is similar to that found in *Drosophila* mtDNA (42).

Recently it has been reported that in *Drosophila* and sea urchins, in contrast to mammals, there is an approximately equal divergence rate for mitochondrial and single-copy nuclear DNAs (17,19,43). Since a generalized molecular clock hypothesis requires a constant evolutionary rate between mitochondrial and nuclear DNAs, Vawter and Brown (17) propose that the vertebrate - sea urchin comparison invalidates the hypothesis. They suggest the mitochondrial and nuclear DNAs evolve independently, as a result of fluctuations in their relative divergence rates. In contrast, our data indicate that in *Tetrahymena* the evolutionary rate difference between mitochondria and nuclei is relatively constant, and that the nuclear and mitochondrial DNAs evolved in concert. Although fluctuations were observed in the relative rate, they were interpreted as statistically unreliable, because the nuclear markers are >99.5% identical. In *Tetrahymena* we see no reason to invalidate the molecular clock hypothesis. Comparison of functionally equivalent DNA sequences, e.g. nuclear and mitochondrial rRNA genes, is needed to clarify the constancy of the molecular clock in sea urchins and *Drosophila*.

ACKNOWLEDGEMENTS

Our sincere appreciation is extended to Dr. Mitch Sogin who ran our DNA sequences on his computer programs to generate the phylogenetic tree used in this paper. We thank Gary Olsen, Allan Wilson and Jeff Sampson for helpful discussions, H. Sasnor, Z. Stabinsky and M. Caruthers for the synthesis of oligonucleotides, and A. Sirimarco for preparation of the manuscript. This work was supported by grant GM25273 from the National Institutes of Health. T.R.C. is an American Cancer Society Research Professor.

REFERENCES

1. Gray, M.W. (1982) *Can. J. Biochem.* **60**, 157-171.
2. Suyama, Y. and Miura, K. (1968) *Proc. Natl. Acad. Sci. USA* **60**, 235-242.
3. Goddard, J.M. and Cummings, D.J. (1975) *J. Mol. Biol.* **97**, 593-609.
4. Warrior, R. and Gall, J. (1986) *Arch. Sci. (Geneva)* **38**, 439-445.
5. Gray, M.W., Sankoff, D. and Cedergren, R.J. (1984) *Nucleic Acids Res.* **14**, 5837-5852.
6. Goldbach, R.W., Borst, P., Bollen-De Boer, J.E. and Van Brauggen, E.F.J. (1978) *Biochim. Biophys. Acta* **521**, 169-186.
7. Heinonen, T.Y.K., Schnare, M.N., Young, P.G. and Gray, M.W. (1987) *J. Biol. Chem.* **262**, 2879-2887.
8. Morin, G.B. and Cech, T.R. (1988) *Cell*, in press.
9. Morin, G.B. and Cech, T.R. (1986) *Cell* **46**, 873-883.
10. Goldbach, R.W., Arnberg, A.C., Van Bruggen, E.F.J., Defize, J. and Borst, P. (1977) *Biochim. Biophys. Acta* **477**, 37-50.
11. Brown, W.M., George, Jr., M. and Wilson, A.C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1967-1971.
12. Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) *J. Mol. Evol.* **18**, 225-239.
13. Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M. and Koike, K. (1982) *J. Mol. Evol.* **19**, 28-35.
14. Ferris, S.D., Wilson, A.C. and Brown, W.M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2432-2436.
15. Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. and Cavalli-Sforza, L.L. (1983) *J. Mol. Evol.* **19**, 255-271.
16. Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) *Nature* **325**, 31-36.
17. Vawter, L. and Brown, W.M. (1986) *Science* **234**, 194-196.
18. DeSalle, R. and Giddings, L.V. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6902-6906.
19. Solignac, M., Monnerot, M. and Mounolou, J.-C. (1986) *J. Mol. Evol.* **23**, 31-40.
20. Preparata, R.M., Meyer, E.B., Nanney, D.L., Preparata, F.P., Vossbrinck, C.R. and Simon, E.M., manuscript in preparation.
21. Sogin, M.L. and Elwood, H.J. (1986) *J. Mol. Evol.* **23**, 53-60.
22. Sogin, M.L., Ingold, A., Karlok, M., Nielsen, H. and Engberg, J. (1986) *EMBO J.* **5**, 3625-3630.
23. Zaug, A.J. and Cech, T.R. (1980) *Cell* **19**, 331-338.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
25. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527-2538.
26. Barr, P.J., Thayer, R.M., Laybourn, P., Najarian, R.C., Seela, F. and Tolan, D.R. (1986) *BioTechniques* **4**, 428-432.
27. Cummings, D.J. (1980) *Molec. Gen. Genet.* **180**, 77-84.

28. Noller, H.F. (1984) *Ann. Rev. Biochem.* **53**, 119-162.
29. Suyama, Y., Jenney, F. and Okawa, N. (1987) *Curr. Genet.* **11**, 327-330.
30. Schnare, M.N., Heinonen, T.Y.K., Young, P.G. and Gray, M.W. (1985) *Curr. Genet.* **9**, 389-393.
31. Ojala, D., Montoya, J. and Attardi, G. (1981) *Nature* **290**, 470-474.
32. Nanney, D.L. and McCoy, J.W. (1976) *Trans. Amer. Microsc. Soc.* **95**, 664-682.
33. Arnberg, A.C., Goldbach, R.W., Van Bruggen, E.F.J. and Borst, P. (1977) *Biochim. Biophys. Acta* **477**, 51-69.
34. Pace, N.R., Olsen, G.L. and Woese, C.R. (1986) *Cell* **45**, 325-326.
35. de Bruijn, 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.
36. Arcari, P. and Brownlee, G.G. (1980) *Nucleic Acids Res.* **8**, 5207-5212.
37. Wolstenholme, D.R., Marfarlane, J.L., Okimoto, R., Clary, D.O. and Wahleithner, J.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1324-1328.
38. Van Bell, C.T. (1985) *J. Mol. Evol.* **22**, 231-236.
39. Walker, T.A. and Pace, N.R. (1983) *Cell* **33**, 320-322.
40. Cann, R.L., Brown, W.M. and Wilson, A.C. (1984) *Genetics* **106**, 479-499.
41. DeSalle, R., Freedman, T., Prager, E.M. and Wilson, A.C. (1987) *J. Mol. Evol.*, in press.
42. Wolstenholme, D.R. and Clary, D.O. (1985) *Genetics* **109**, 725-744.
43. Powell, J.R., Caccone, A., Amato, G.D. and Yoon, C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9090-9093.