J Mol Evol (1994) 39:598–613

Journal of Molecular Evolution

The Mitochondrial Ribosomal RNA Genes of the Nematodes *Caenorhabditis elegans* and *Ascaris suum*: Consensus Secondary-Structure Models and Conserved Nucleotide Sets for Phylogenetic Analysis

Ronald Okimoto, Jane L. Macfarlane, David R. Wolstenholme

Department of Biology, University of Utah, Salt Lake City, UT 84112 USA

Received: 20 August 1993 / Accepted 20 June 1994

Abstract. The small- and large-subunit mitochondrial ribosomal RNA genes (mt-s-rRNA and mt-l-rRNA) of the nematode worms *Caenorhabditis elegans* and *Ascaris suum* encode the smallest rRNAs so far reported for metazoa. These size reductions correlate with the previously described, smaller, structurally anomalous mt-tRNAs of *C. elegans* and *A. suum*. Using primer extension analysis, the 5′ end nucleotides of the mt-s-rRNA and mt-l-rRNA genes were determined to be adjacent to the 3′ end nucleotides of the $tRNA^{Glu}$ and $tRNA^{His}$ genes, respectively. Detailed, consensus secondary-structure models were constructed for the mt-s-rRNA genes and the 3′ 64% of mt-l-rRNA genes of the two nematodes. The mt-s-rRNA secondary-structure model bears a remarkable resemblance to the previously defined universal core structure of *E. coli* 16S rRNA: most of the nucleotides that have been classified as variable or semiconserved in the *E. coli* model appear to have been eliminated from the *C. elegans* and *A. suum* sequences. Also, the secondary structure model constructed for the 3′ 64% of the mt-l-rRNA is similar to the corresponding portion of the previously defined *E. coli* 23S rRNA core secondary structure. The proposed *C. elegans/A. suum* mt-s-rRNA and mt-l-rRNA models include all of the secondary-structure element-forming sequences that in *E. coli* rRNAs contain nucleotides important for A-site and P-site (but not E-site) interactions with tRNAs. Sets of apparently homologous sequences within the mt-s-rRNA and mt-l-rRNA core structures, derived by alignment of the *C. elegans* and *A. suum* mt-rRNAs to the corresponding mt-rRNAs of other eukaryotes, and *E. coli* rRNAs were used in maximum-likelihood analyses. The patterns of divergence of metazoan phyla obtained show considerable agreement with the most prevalent metazoan divergence patterns derived from more classical, morphological, and developmental data.

Key words: *Caenorhabditis elegans* — *Ascaris suum* — Mitochondrial ribosomal RNA genes — Secondary structure models — Phylogenetic analysis

Correspondence to: D.R. Wolstenholme

Introduction

The mitochondrial (mt) genomes of most multicellular animals (metazoa) are single circular, double-stranded DNA molecules that have species-specific sizes in the range of 14–42 kb (Wolstenholme 1992). Complete nucleotide sequences and gene contents have been determined for a number of vertebrate mtDNA molecules including eight mammals: human, cow, mouse, rat, blue and fin whales, and grey and harbor seals (Anderson et al. 1981, 1982; Bibb et al. 1981; Gadaleta et al. 1989; Arnason et al. 1991, 1993; Arnason and Johnsson 1992; Arnason and Gullberg 1993); a bird, *Gallus domesticus* (Desjardins and Morais 1990); an amphibian, *Xenopus laevis* (Roe et al. 1985); a fish, *Crossostoma lacustre* (Tzeng et al. 1992), and for a number of phylogenetically diverse invertebrates, including two sea urchins, *Strongylocentrotus purpuratus* and *Paracentrotus lividus*

(Jacobs et al. 1988; Cantatore et al. 1989), and a sea star, *Asterina pectinifera* (Asakawa et al. 1991); two insects, *Drosophila yakuba* and *Apis melifera* (Clary and Wolstenholme 1985a; Crozier and Crozier 1993); and two nematodes, *Caenorhabditis elegans* and *Ascaris suum* (Okimoto et al. 1992). Each completely sequenced metazoan mtDNA molecule contains the genes for the same 12 or 13 proteins, all concerned with oxidative phosphorylation, and the genes for the RNA components of the mitochondrion's own protein-synthesizing machinery: 22 tRNAs and two rRNAs.

The mtDNA-encoded tRNAs and the mtDNA-encoded rRNAs of metazoa have diverged to various degrees from their prokaryotic counterparts. Because of an extended wobble mechanism, the 22 mt-tRNAs encoded by each of the sequenced metazoan mtDNAs are apparently sufficient to decode all of the mtDNA-contained protein genes. Most metazoan mt-tRNA genes can be folded into secondary structures similar to those of prokaryotic and eukaryotic nuclear-encoded tRNAs (standard tRNAs) except that usually the sizes and sequences of the dihydrouridine (DHU) and TΨC loops are highly variable (Wolstenholme 1992). In all $tRNA^{Ser(AGN)}$ genes so far identified in metazoa, the dihydrouridine (DHU) arm is replaced with a short loop of nucleotides (Arcari and Brownlee 1980; de Bruijn et al. 1980; Garey and Wolstenholme 1989). However, the complete set of mt-tRNAs of the two nematodes *C. elegans* and *A. suum* are structurally anomalous. In each species, the $tRNA^{Ser(UCN)}$ as well as the $tRNA^{Ser(AGN)}$ contains a DHU arm-replacement loop, and in all of the 20 remaining mt-tRNAs of each species, the TΨC arm and variable loop have been replaced with a simple loop of between 6 and 12 nucleotides (Wolstenholme et al. 1987; Okimoto and Wolstenholme 1990; Okimoto et al. 1992).

Metazoan mt-s-rRNAs and mt-l-rRNAs have been considerably reduced in size relative to the homologous 16S and 23S rRNAs of *Escherichia coli,* and these reductions are mainly due to loss of sequences that can be folded into distinct secondary structural elements, rather than due to deletions of small numbers of nucleotides throughout the two genes (Glotz et al. 1981; Zweib et al. 1981; Gutell et al. 1985; Roe et al. 1985; Clary and Wolstenholme, 1985, 1987). The secondary structures of the 16S and 23S rRNAs of *E. coli* have been thoroughly established from data obtained using chemical protection and crosslinking and from considerations of phylogenetically conserved helices and loops (reviewed in Noller et al. 1986). Also, extensive work has been done with *E. coli* rRNAs to establish which nucleotides and secondary-structure elements are involved in ribosome functions (reviewed in Dahlberg, 1989; Moazed and Noller, 1990; Noller, 1993). Based on the *E. coli* 16S and 23S rRNA primary sequences, and secondary-structure models, phylogenetic sequence analyses with various degrees of completeness have been proposed for the s-rRNA and l-rRNA of a divergent range of organisms and organelles (Glotz et al. 1981; Zweib et al. 1981; Clary and Wolstenholme, 1985, 1987; Gutell et al. 1985; Gutell and Fox, 1988).

From comparative sequence analyses of eight s-rRNAs (two prokaryotic, two eukaryotic nuclear, three mitochondrial, and one chloroplast) Steigler et al. (1981) identified a core of highly conserved secondary structures. Almost all sites identified as being essential to ribosome function have been located within this core (Dahlberg et al. 1989; Noller, 1993). Sequences within the s-rRNA core structure and within a similarly derived core secondary structure of the l-rRNAs have been used as the basis of phylogenetic analysis of rRNA sequences (Gray et al. 1984; Cedergren et al. 1988).

In this paper we present detailed consensus secondary-structure models for the mt-s-rRNAs and for the 3′ two-thirds of the mt-l-rRNAs of *C. elegans* and *A. suum,* the smallest metazoan mt-rRNAs so far described (Okimoto et al. 1990, 1992). Using our nematode secondary-structure models as a basis, we have defined new sets of homologous nucleotide sequences within metazoan mt-s-rRNAs and mt-l-rRNAs that, under specific conditions of phylogenetic sequence analysis, yield patterns of divergence of metazoa that have considerable agreement with phylogenies based on more classical, nonsequence methods.

Materials and Methods

Details regarding culturing of *C. elegans* and preparation, cloning, and sequencing of *C. elegans* and *A. suum* mtDNA are given or referred to in Okimoto et al. (1992).

RNA Isolation. C. elegans whole-cell RNA was isolated as follows. Approximately 9.0 ml of packed volume worms were resuspended in an equal volume of TES (20 mM EDTA, 50 mM Tris-HCl [pH 8.0], 150 mM NaCl), 20 ml of phenol (prepared as in Maniatis et al. 1982) was added, and the mixture was periodically shaken and kept on ice until lysis of the worms was complete. The supernatant was reextracted with phenol; then it was extracted once with chloroform and subjected to cesium chloride centrifugation using a Ti50 rotor, as in Maniatis et al. (1982), but using 2 mg rather than 10 mg ethidium bromide per tube. The RNA pellet was resuspended in TES and sequentially extracted with phenol (twice) and chloroform (once) to remove the ethidium bromide. The RNA was aliquoted, ethanol precipitated, rinsed once with 70% ethanol, and stored at −20°C under 70% ethanol.

Primer Extension Analysis. Oligonucleotide primers corresponding to the complements of the *C. elegans* s-rRNA gene position 72–92 (Fig. 2A) 5′ TTATTGCCTACTCTGTAAATC (21 mer) and the l-rRNA gene position 69–88 (Fig. 2B) 5′ TTTACTCTACCCTTTGTTTC (20 mer) were synthesized. These oligonucleotide primers were 5′ end-labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase as in Okimoto et al. (1990). To 10 μg *C. elegans* whole-cell RNA, 0.8 ng of ^{32}P-labeled primer was annealed for 30 min at 42°C in a solution containing 4 μl 5× reverse transcriptase buffer (BRL: manufacturer's recommended reaction buffer): total volume, 14 μl. The mixture was cooled to room temperature, and 1 μl RNase inhibitor (BRL), 1 μl (20 units) RNase H-reverse transcriptase (BRL: M-MLVH-RT), 2 μl 0.1 M di-

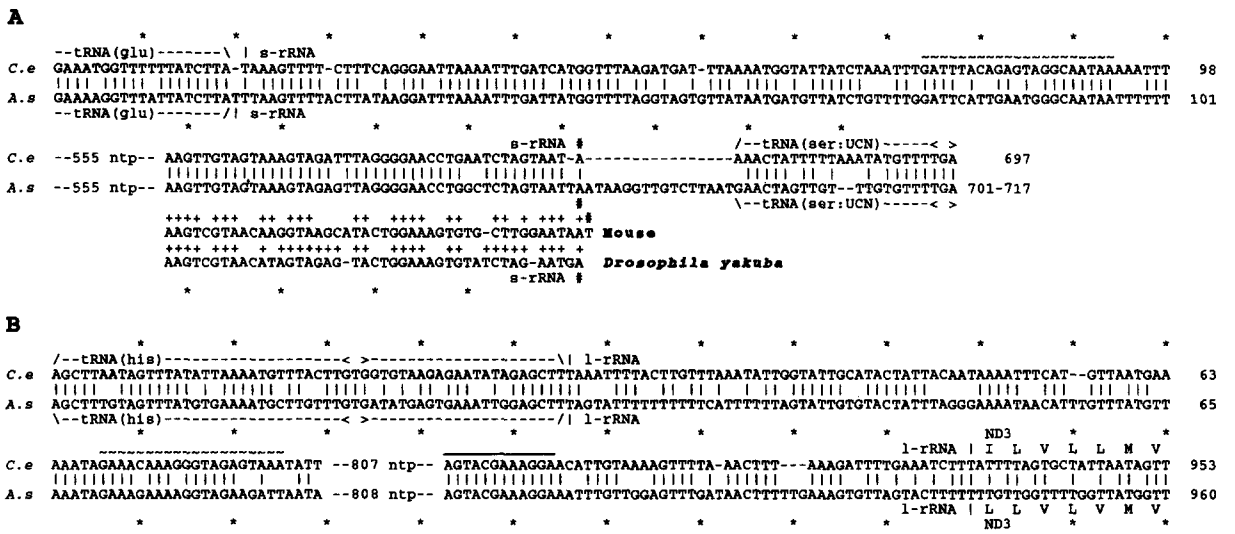

Fig. 1. Comparisons of corresponding nucleotide sequences of *Caenorhabditis elegans* (*C.e*) and *Ascaris suum* (*A.s*) mtDNAs that contain the 5′ end-proximal and 3′ end-proximal regions of small-subunit rRNA genes (s-rRNA) and the large-subunit rRNA genes (l-rRNA). The *numbers to the right* refer only to the rRNA gene sequences. *Vertical lines* indicate similarities between the *C. elegans* and *A. suum* nucleotide sequences. A *dash* indicates the absence in one nematode sequence of a nucleotide that occurs in the corresponding nematode sequence. Predicted amino acid sequences of the ND3 gene are shown in the *one-letter code.* Sequences of tRNA genes are identified by *dashed lines* within which *brackets* identify the anticodons. Transcription of all genes is from left to right. In **A**, the sequences of the 3′ end-proximal regions of the mouse and *Drosophila yakuba* mt-s-rRNA genes are aligned with the corresponding regions of the nematode sequences. The terminal nucleotide of each s-rRNA gene is indicated by a # sign. A *plus sign* (+) indicates a match with the *A. suum* sequence. Sequences complementary to the oligonucleotides used in the primer extension experiments are indicated by *wavy overlines.* The *solid line* above the *C. elegans* sequence in **B** indicates a 13-ntp sequence that is 100% conserved between the two nematodes, mouse, and *D. yakuba* mt-l-rRNA genes. The complete *C. elegans* and *A. suum* mt-s-rRNA and mt-l-rRNA sequences are given in Okimoto et al. (1992). EMBL Data Library accession numbers X54252 and X54253.

thiothreitol, and 2μl of dXTPs (dATP, dCTP, dGTP, and dTTP (each 2.5 mM)) were added. This mixture was incubated at 37°C for 40 min. The reaction was stopped by the addition of 30 μl of formamide sequence loading dye and boiled for 3 min. Reference DNA sequences were obtained using the s-rRNA-specific and l-rRNA-specific primers described above (0.3 ng total/sequencing reaction) and M13 bacteriophage DNA clones containing the sense strand of the presumed 5′ end-proximal region of the respective *C. elegans* mt-rRNA gene.

Results and Discussion

Gene content and arrangement are identical in the *C. elegans* and *A. suum* mtDNA molecules except for the relative locations of the AT (putative control) region (Okimoto et al. 1992). All genes are transcribed in the same direction. Both nematode mtDNA molecules contain a single s-rRNA gene and a single l-rRNA gene that were identified from nucleotide sequence similarities to the corresponding genes of other mtDNAs and the *E. coli* 16S and 23S rRNAs (Figs. 7 and 8). In the *C. elegans* and *A. suum* mtDNA molecules, the two rRNA genes are separated from each other by protein genes, as has been found in sea-urchin mtDNAs. This contrasts with the situation in *Drosophila* mtDNA and vertebrate mtDNAs where the two rRNA genes are separated only the $tRNA^{Val}$ gene.

In *C. elegans* and *A. suum,* the 5′ and 3′ ends of the s-rRNA genes are located next to the $tRNA^{Glu}$ and $tRNA^{Ser(UCN)}$ genes, respectively, and the 5′ and 3′ ends of the l-rRNA genes are located next to the $tRNA^{His}$ and ND3 genes, respectively (Fig. 1).

5′ and 3′ End Locations of the rRNA Genes

Primer extension analyses were used to locate the 5′ termini of the *C. elegans* s-rRNA gene and l-rRNA gene (Fig. 2). Two bands of different intensity are visible in the lane containing the products of the s-rRNA reaction; they represent extension products differing by one nucleotide (Fig. 2A). The size of the product represented by the major band indicates the 5′ end of the s-rRNA gene to be immediately adjacent to the 3′ terminal nucleotide of the $tRNA^{Glu}$ gene. The minor band represents a product that would overlap the $tRNA^{Glu}$ gene by a single nucleotide. Alternatively, this minor, one-nucleotide-longer product could indicate the presence within the collection of RNA molecules used in the primer extension experiment of mt-s-rRNAs that have an internal, single nucleotide insertion. This explanation seems tenable because the RNA sample used was isolated from a large population of worms.

In the lanes containing the product of the l-rRNA reaction (Fig. 2B), there is a major band representing an extension product of a size indicating the 5′ end of the l-rRNA gene to be immediately adjacent to the 3′ terminal nucleotide of the $tRNA^{His}$ gene. However, there are also three minor bands indicating the presence of one

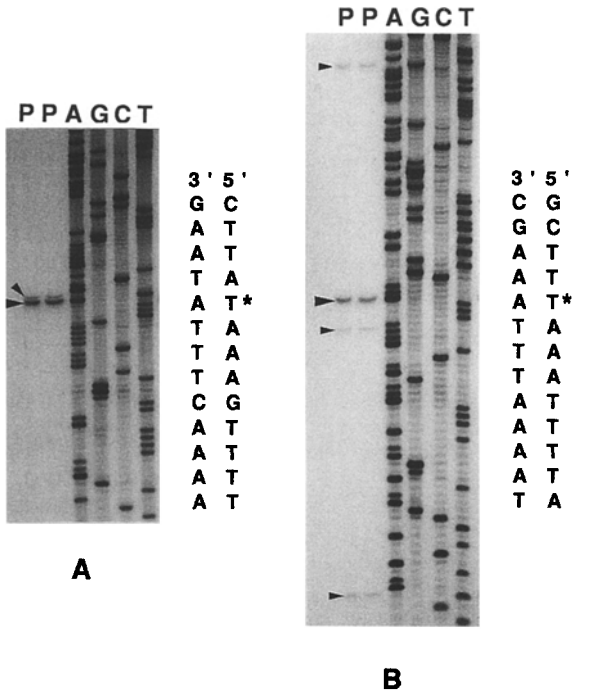

Fig. 2. Primer extension analysis to determine the nucleotide locations of the 5′-ends of the *C. elegans* mt-s-rRNA (**A**) and mt-l-rRNA (**B**) gene transcripts. The two autoradiographs shown each result from electrophoresis through a 6.5% polyacrylamide gel. In each autoradiograph, P identifies two (duplicate) lanes that contain extension products which resulted from reverse transcriptase synthesis, using whole-cell RNA, from the primers whose locations are indicated in Fig. 1A and B. Lanes *A, C, G,* and *T* contain the extension-dideoxy-chain termination products using these same primers and rRNA gene-containing-M13 DNA clones. In each *P* lane in **A,** there are two bands that represent a major extension product (*large arrowhead*) and a one-nucleotide-larger, minor-extension product (*small arrowhead*). In each *P* lane in **B,** there are four bands that represent one major extension product (*large arrowhead*) and three minor extension products (*small arrowheads*): one larger (above) and two smaller (below) than the major product. The 5′-end nucleotide locations of the major mt-s-rRNA and mt-l-rRNA extension products are indicated on the sense strand sequence (*) to the right of the respective autoradiograph.

extension product that terminates within the $tRNA^{His}$ gene, 46 nt from the 3′ end, and two other products that terminate 5 nt and 40 nt downstream from the 3′ end of the $tRNA^{His}$ gene. These minor bands could have resulted from secondary site annealing of the primers, or, in the case of those representing shorter extension products, from premature termination at two specific sites.

From considerations of primary sequence similarities, and of secondary-structure potential between the 3′ end-proximal regions of the s-rRNA genes of *C. elegans* and *A. suum,* and *D. yakuba* and mouse (Figs. 1, 3, and 4), it seems likely that in *C. elegans,* the 3′ terminal nucleotide of the s-rRNA gene is adjacent to the 5′ terminus of the $tRNA^{Ser(UCN)}$ gene. However, the *A. suum* s-rRNA gene has either a 16- or 17-nt 3′ extension relative to the *C. elegans* s-rRNA gene, or there are these numbers of nucleotides separating the *A. suum* s-rRNA and $tRNA^{Ser(UCN)}$ genes (Fig. 1A; Okimoto and Wolstenholme 1990).

Beginning 41 and 44 nt upstream from the 5′ end of the predicted translation initiation codon of the *C. elegans* and *A. suum* ND3 genes (ATT and TTG; Okimoto et al. 1990) is a segment of 13 nt that is 100% conserved between the two nematode, mouse, and *D. yakuba* mt-l-rRNA genes. Similarity of the nematode nucleotide sequences that lie between this conserved region and the beginning of the ND3 gene and the nucleotide sequence of the 3′ end-proximal region of the mouse and *D. yakuba* l-rRNA genes is insufficient to indicate the possible 3′ end nucleotide of the nematode l-rRNA genes. Therefore, our designation of the nucleotide 5′ to the putative ND3 translation initiation codon as the 3′ terminus of each nematode mt-l-rRNA gene is tentative. From the above given definitions of the nematode mt-rRNA gene termini, the *C. elegans* and *A. suum* s-rRNA genes would encode products of 697 nt and 700 (717) nt, respectively, and the *C. elegans* and *A. suum* l-rRNA genes would encode products of 953 nt and 960 nt, respectively. These are the shortest s-rRNAs and l-rRNAs encoded by any metazoan mtDNA so far recorded, as we have noted previously (Okimoto et al. 1992). By comparison, the mt-s-rRNA genes of mouse and *D. yakuba* are 955 and 789 ntp, and the mt-l-rRNA genes of these species are 1,581 and 1,325 ntp, respectively (Bibb et al. 1981; Clary and Wolstenholme 1985b).

Secondary Structure Model of the Nematode mt-s-rRNA Genes

A consensus secondary-structure model for the *C. elegans* and *A. suum* s-rRNA genes was constructed (Fig. 3), based on the secondary-structure model of the *E. coli* 16S rRNA (Noller et al. 1986), on the phylogenetically inferred mt-s-rRNA secondary structure models of Gutell et al. (1985), and on the positions within these models of short segments of highly conserved nucleotides. Within the nematode model, when secondary-structure elements similar to those of the phylogenetically inferred models could not be formed, the most consistent secondary structure that included compensatory nucleotide changes shared by the *C. elegans* and *A. suum* sequences was chosen. In general, secondary-structure stability (change in free energy) was not considered in constructing the nematode model. However, in some cases a choice was made in favor of the more stable of an alternative pair of possible structures when this structure showed a minimal conflict with the phylogenetically defined structure. As noted by Gutell et al. (1985) for other metazoan mt-rRNAs, in the nematode s-rRNAs there are some instances where more nucleotide pairs can be made in a potential stem than is required to match the standard, inferred structure.

In Fig. 4, the nematode consensus secondary-structure model is compared with that of the *E. coli* 16S rRNA. In the *E. coli* diagram (Fig. 4B) are shown the 48 second-

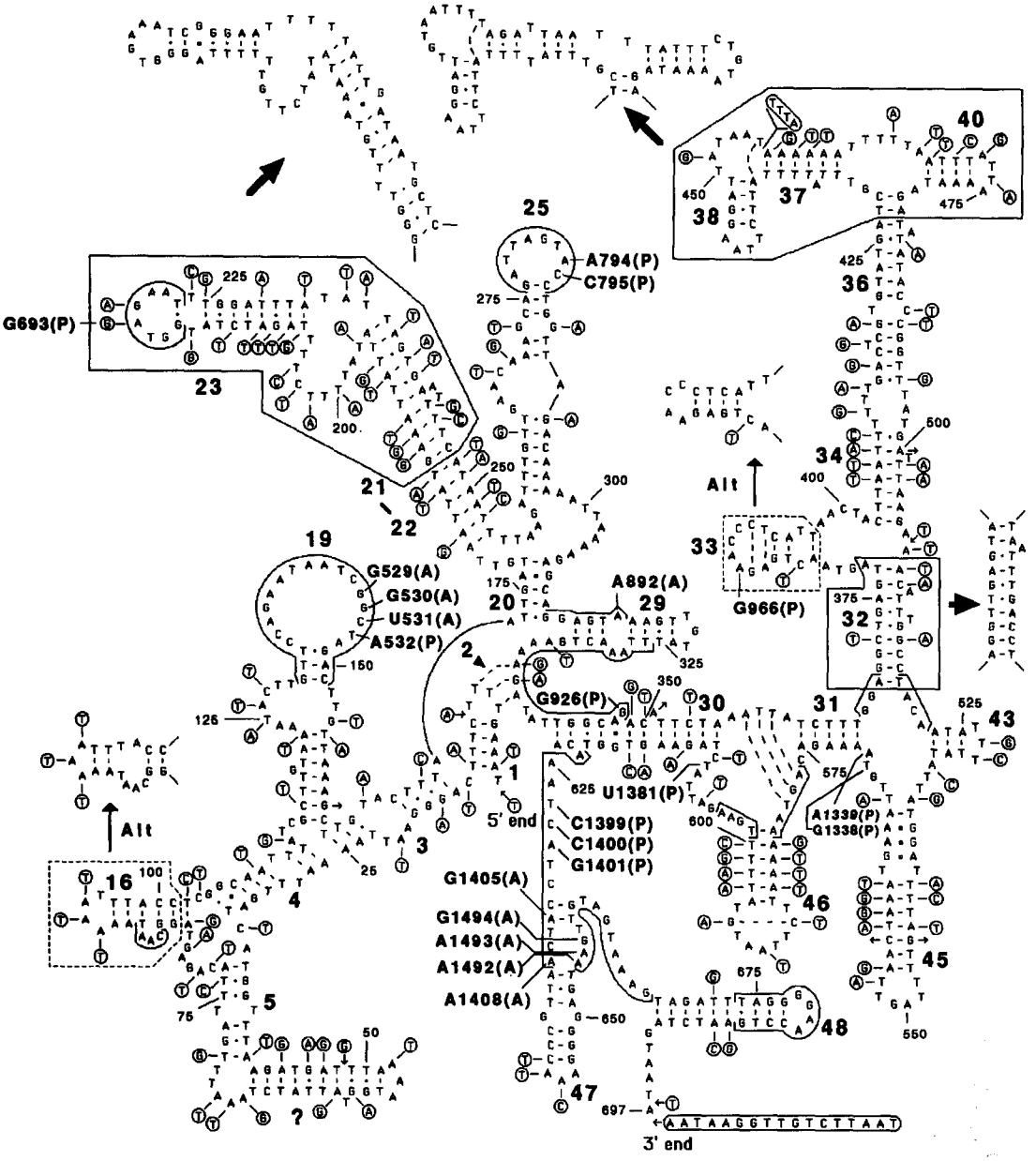

Fig. 3. Consensus secondary-structure model for the nematode mt-s-rRNA genes. The continuous sequence shown is that of the *C. elegans* gene, and this sequence is numbered every 25 nt from the 5′ end. Nucleotide substitutions in the *A. suum* mt-s-rRNA gene sequence are indicated by *circled* nucleotides, and insertions/deletions are indicated by *arrows*. The encircled sequence of 17 nt at the 3′ end of the model may be an extension of the *A. suum* s-rRNA sequence relative to the *C. elegans* s-rRNA sequence. (See text.) Watson–Crick base pairings are indicated by a dash between nucleotides. Presumed pairing of G and T (U in the RNA sequence) is indicated by G•T, and possible pairing of G and A by G○A. Extra nucleotide pairing possibilities are shown by *dashed lines*. Short runs of nucleotides that are highly conserved at similar locations in the general secondary structure models of Gutell et al. (1985) are outlined by *solid lines*. More plausible secondary structures for some regions of the *A. suum* sequence relative to the boxed secondary structures of the corresponding region of the *C. elegans* sequence are indicated by *bold arrows*. Two alternative secondary structures for corresponding regions of both nematode s-rRNA sequences shown *boxed with broken lines* are indicated by *thin arrows* (*Alt*). Bold numbers (*1–48*, as defined by Dams et al. 1988), identify those secondary-structure elements that appear to have been conserved relative to the *E. coli* 16S rRNA model (Fig. 4A). Nucleotides in the nematode sequence to which a boldface letter–number combination is connected indicate nucleotide locations in the *E. coli* s-rRNA sequence (+ nucleotide positions in Figs. 4A and 4B) that are involved in binding to either the aminoacyl site (A) or the peptidyl-transferase site (P) of the ribosome (Noller et al. 1986; Dahlberg 1989; Noller 1993).

ary-structure elements (stems or stem-loops) defined by Dams et al. (1988), the universally conserved blocks of nucleotides (U1–U8) defined by Steigler et al. (1981) and Gray et al. (1984) and, within the entire secondary-structure model, nucleotides that have been classified as universal, semiconserved, and variable in regard to their presence in other s-rRNAs (Gray et al. 1984). In the nematode diagram (Fig. 4A) are shown the secondary structure elements, the universally conserved blocks of nucleotides, and the variously classified sorts of nucleotides (see Dahlberg 1989) that have been conserved. The major difference between the *E. coli* and nematode

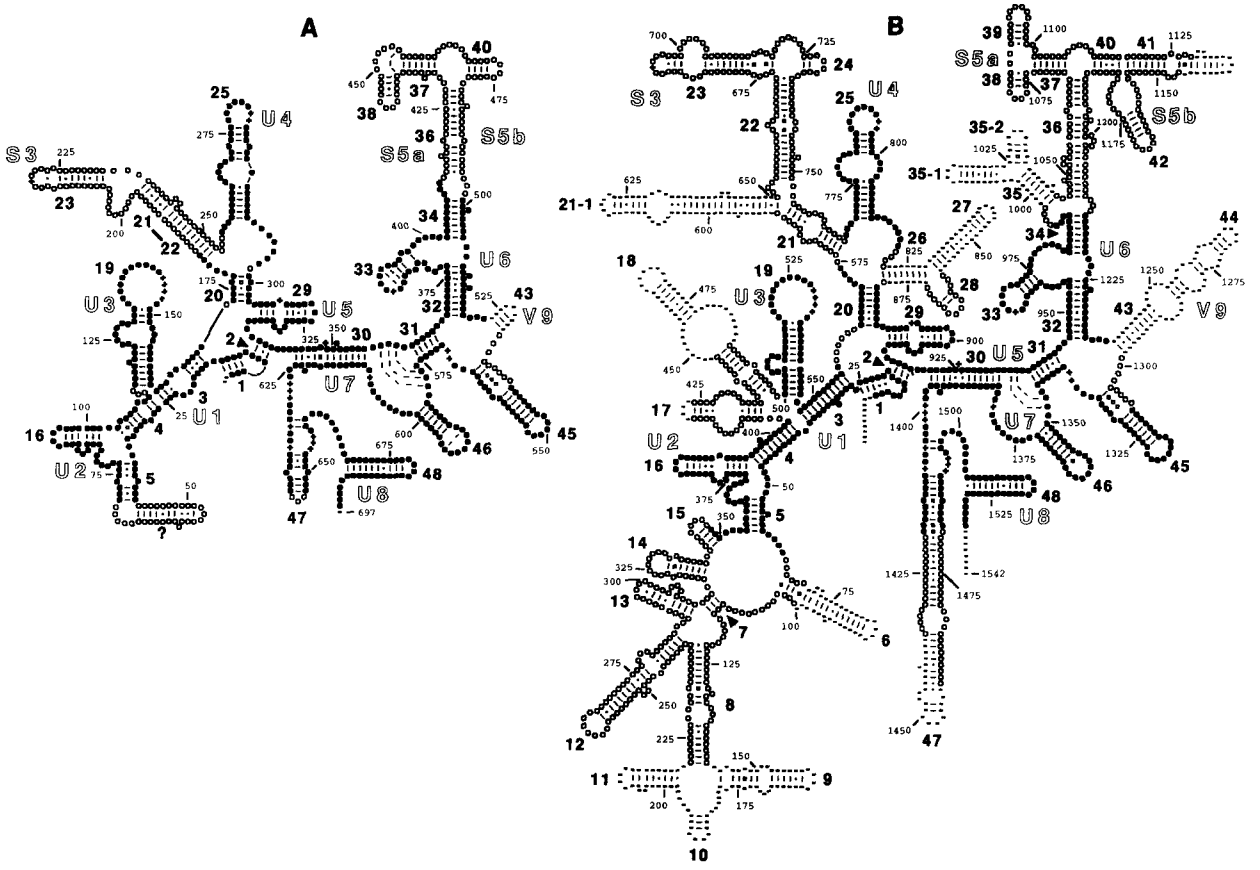

Fig. 4. Comparison of the proposed *C. elegans/A. suum* mt-s-rRNA consensus secondary-structure model **(A)** and the secondary-structure model **(B)** of the *E. coli* 16S-rRNA given by Dahlberg (1989), modified from Noller et al. (1986). The different classes of nucleotide pairs are as in Fig. 3. Numbering of each sequence is from the 5′ end. In both models, classification of the universal (●), semiconserved (○), and variable (–) nucleotides is that given by Gray et al. (1984). Bold numbers (*1–48*) in the *E. coli* 16S-rRNA model identify those secondary-structure elements defined by Dams et al. (1988). Bold numbers in the nematode mt-s-rRNA model identify those secondary-structure elements that appear to have been conserved relative to the *E. coli* 16S-rRNA model. In both models, *open blocked U1–U8* identify the universally conserved blocks of contiguous nucleotides in the s-rRNAs (Gray et al. 1984) and *open blocked S3, S5a,* and *S5b* identify the only extensive regions of semiconserved nucleotides that have been retained in the nematode secondary structure. In the nematode model are shown nucleotides (+) corresponding in position to nucleotides in the *E. coli* model that have been shown to be involved in the decoding site.

models is the elimination from the latter of single secondary-structure elements, or groups of secondary-structure elements. Similar kinds of reductions have been noted for secondary-structure models of other metazoan mtDNAs (Glotz et al. 1981; Zweib et al. 1981; Gutell et al. 1985; Roe et al. 1985; Clary and Wolstenholme 1985b, 1987), but these are never as extensive as in the nematode model.

In the *C. elegans/A. suum* model, the majority of nucleotides classified as variable or semiconserved in the *E. coli* model are not found: of the 423 variable and the 589 semiconserved nucleotides recognized in the *E. coli* 16S rRNA model, only approximately 12 and 208 respectively, remain in the nematode mt-s-rRNA model. The universal core (U1–U8) of the nematode s-rRNA model (Fig. 4) contains 479 nt (*C. elegans*) and 476 nt (*A. suum*), compared to 530 nt in the *E. coli* model. Taking into consideration possible insertions in the sequence of the nematode universal core relative to that of *E. coli* (seven and four for *C. elegans* and *A. suum* respectively, see Fig. 3), this region of the nematode model may lack as many as 69 nt relative to the *E. coli* model, which is the greatest diminution so far recorded for any of the published metazoan mt-s-rRNAs (Fig. 7).

Recently, a secondary-structure model of the *C. elegans* s-rRNA sequence alone has been presented by Gutell (1993). Although there is considerable agreement in detail between much of this model and our consensus nematode model, differences in folding of three segments of the s-rRNA sequence warrant comment. The first difference involves nucleotides 1–103. In the Gutell model, nt 1–23 are shown as an unfolded 5′ tail. The following 17 nucleotides are folded into helix 1, followed by 18 nucleotides that include the 5′ strands of helices 4 and 5. The remaining 45 nucleotides are shown as an unstructured loop. In contrast, in our consensus model (Fig. 3), only the first nucleotide is unpaired, and the following 12 nt form helical element 1. Nucleotides 14–32 include the 5′ strands of helices 3 and 4, and the majority of nucleotides 33–103 form a secondary structure comprising three helical elements.

In the *E. coli* 16S rRNA secondary-structure model,

there are only eight 5′ terminal, unpaired nucleotides, and this number is reduced to between one and three in other metazoan mt-s-rRNA secondary-structure models. Also, it seems to us unlikely that in the greatly shortened nematode s-rRNAs, the 5′ terminus-proximal 23 nucleotides would be excluded from secondary-structure involvement. The validity of each of the helical regions formed by nt 33–103 in our model is supported by compensatory nucleotide changes in the *C. elegans* and *A. suum* sequences that maintain secondary structure (Fig. 3). Also, the first and third of these helices resemble in overall structure and relative location *E. coli* helices 5 and 16, respectively. Also, the 5′ strand of helix 16 of the *C. elegans* and *A. suum* sequence contains a conserved CAA loop-out. In our model, the second helical element is formed by 29 nt that appear to be the remnant of a 183-nt segment in the *E. coli* 16S rRNA that is folded into ten helical elements (6–15). However, it is not clear to which of these *E. coli* elements the single *C. elegans*/*A. suum* helix might correspond (? in Figs. 3 and 4).

A second difference between our consensus model and the Gutell model concerns nt 178–206 and 233–255. These nucleotides, which connect the nematode equivalent of *E. coli* helical elements 20 and 25 with helical element 23 (Fig. 4B), are shown mostly unpaired in the Gutell model. In contrast, in our model (Fig. 3) the majority of these nucleotides are used to construct two partially alternative (partially slipped) helices, the validity of which are very strongly supported by compensatory changes in the *C. elegans* and *A. suum* sequences. These helices are interpreted as the equivalent of stacked *E. coli* helical elements 21 and 22 (formed by elimination of helix 21–1; Fig. 3).

There is also a difference between the two s-rRNA models regarding the secondary structure of nucleotides 422–428. We have folded this sequence to form the distal portion of helical element 36 and a complex stem and loop structure containing the *E. coli* equivalents of helices 37, 38, and, tentatively, 40 (Figs. 3, 4). A similar alternative overall folding arrangement is proposed for the *A. suum* model that takes into consideration a 4-nt insertion and noncompensatory substitutions (Figs. 3 and 4A and B). It has been proposed that the corresponding regions of the *D. yakuba* and human mt-s-rRNAs have been folded into a similar three-helix structure (helices 37, 38, and 40; Gutell et al. 1985). In the Gutell *C. elegans* model, helix 36 is extended (following a bulged A) by 4 ntp, which makes this helix closer in length to the *D. yakuba* and mouse models, but this excludes helix 40. Also in the Gutell model, 2 out of 5 ntp in helix 37 could not be formed by the *A. suum* sequence.

Further Discussion Regarding the Nematode s-rRNA Secondary-Structure Model

In spite of the severe size reductions of the nematode s-rRNA relative to *E. coli* 16S rRNA and other metazoan mt-s-rRNAs, all nucleotide positions that in *E. coli* 16S rRNA are interpreted as being associated with decoding (reviewed by Dahlberg 1989; Noller 1993) appear to be present in the nematode s-rRNA secondary-structure model (Figs. 3 and 4).

In *E. coli* 16S rRNA, the phylogenetically highly conserved, so-called 530 loop (*E. coli* nt 518–533; *C. elegans* nt 133–149, Figs. 3 and 4) has been shown to be associated with the activities of two types of antibiotics, streptomycin and neomycin-related aminoglycosides that induce coding errors in *E. coli* (Moazed and Noller 1987; Melancon et al. 1988). However, as the 530 loop maps to the surface of the 30S subunit it cannot be directly involved with the decoding region. Rather its position appears to be crucial in conformational changes in the ribosome resulting from the interaction of tRNAs with the decoding site (Moazed and Noller, 1990). In the 5′ domain of the nematode secondary-structure model, all of the helices that can be identified are shorter, relative to the corresponding helices of *E. coli* 16S rRNA, by between 1 and 3 ntp, except for helices 1 and 5.

These helix shortenings in nematodes may constitute a concerted change that results in a more compact structure that maintains the 530 loop and the decoding site in the same relative arrangement that is found in the *E. coli* 30S subunit. An alternative rationale for helix shortening in the 5′ domain and in other phylogenetically conserved secondary structures of the nematode mt-s-rRNA may be associated with the aberrant structure of nematode mt-tRNAs (Wolstenholme et al. 1987; Okimoto and Wolstenholme 1990; Okimoto et al. 1992). All 22 of the *C. elegans* and *A. suum* mt-tRNAs lack one of the side arms. For each species, in two of the mt-tRNA genes (for $tRNA^{Ser(AGN)}$ and $tRNA^{Ser(UCN)}$) the DHU (dihydrouridine) arm is replaced with a loop of 5–8 nt, and in each of the remaining 20 mt-tRNAs the variable loop and TΨC arm (TV) are together replaced with a loop of 6–12 nt. The tertiary folded functional form of the DHU arm replacement loop-containing $tRNA^{Ser(AGN)}$ of cow has been interpreted from chemical probing data to have an overall similar L-shaped structure, as do standard tRNAs (de Bruijn and Klug, 1983). Also it has been recently argued from considerations of patterns of nucleotide correlations in the TV-replacement loop-containing *C. elegans* and *A. suum* mt-tRNAs that these tRNAs also have a final L-shaped folded form (Wolstenholme, et al. 1994). Conservation of overall tertiary configuration of the different secondary-structural sets of mt-tRNAs would seem to minimize the necessity for dramatic changes in the mt-s-rRNA core structure.

Secondary-Structure Model of the Nematode mt-l-rRNA Genes

A consensus secondary-structure model for the 3′ 64% of the *C. elegans* and *A. suum* mt-l-rRNA genes, again

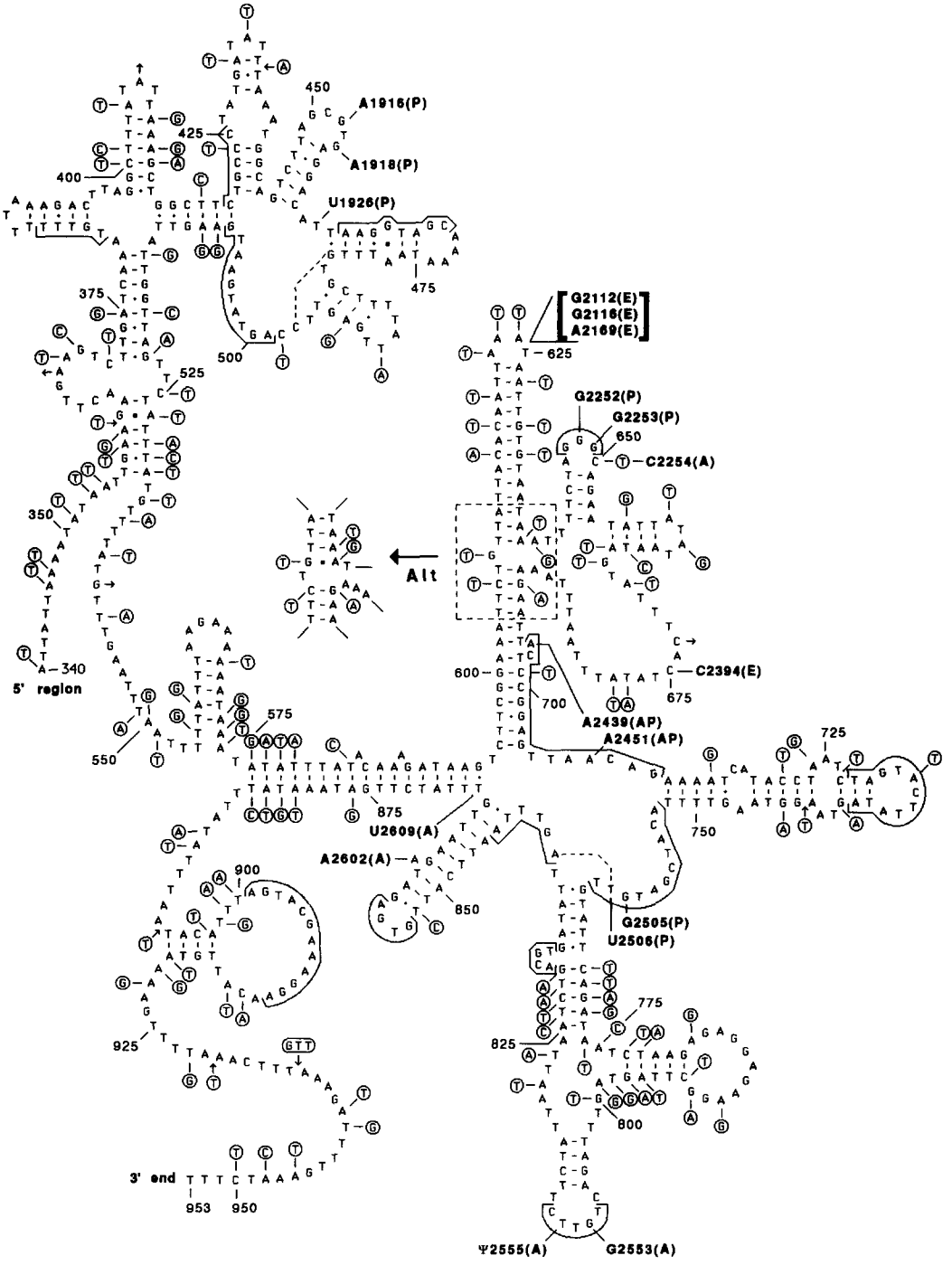

Fig. 5. Consensus secondary-structure model for the 3′, 614 nt of the nematode mt-l-rRNA genes. The continuous sequence shown is that of the *C. elegans* gene and this sequence is *numbered* every 25 nt (5′–3′). Nucleotide substitutions in the *A. suum* mt-l-rRNA sequence are indicated by *circled* nucleotides and insertions/deletions are indicated by *arrows*. Watson–Crick base pairings are indicated by a *dash* between nucleotides. Presumed pairing of G and T (U in the RNA sequence) is indicated by *G•T*, and possible pairing of G and A by *GoA*. Extra base-pairing possibilities are shown by *dashed lines*. Short runs of nucleotides that are highly conserved at similar locations in the secondary-structure model of Gutell and Fox (1988) are outlined with *solid lines*. An alternative secondary structure for one region of the model (*boxed*) is indicated by the *thin arrow* (*Alt*). Nucleotide or internucleotide spaces in the nematode sequence to which a *boldface letter–number combination* is connected indicate nucleotide locations in the *E. coli* 23S rRNA sequence (e.g., a G at position 2252 indicated by *G 2252*) that are involved in binding to the aminoacyl site (*A*), the peptidyl-transferase site (*P*), or both (*AP*), or to the exit site (*E*) of the ribosome, as given by Noller et al. (1986).

based on primary and secondary structural considerations (Gutell and Fox 1988), is shown in Fig. 5. This model contains 614 nt for *C. elegans* (nt 340–953) and 618 nt for *A. suum* (nt 343–960). In Fig. 6, the nematode consensus l-rRNA secondary structure model is compared with that of the corresponding portion (1,258 nt) of the *E. coli* 23S rRNA. In both models are shown conserved core nucleotides given for the *E. coli* 23S rRNA

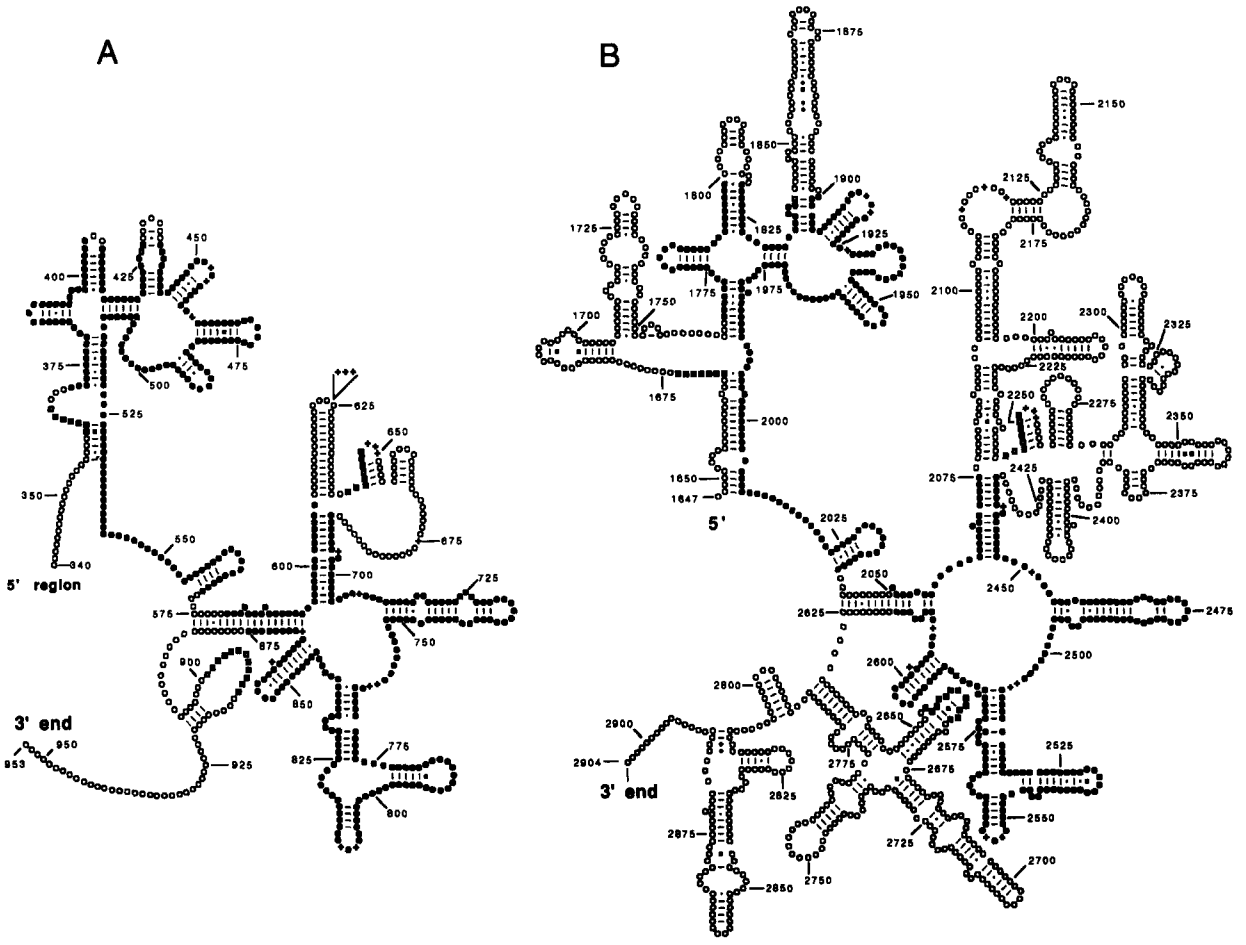

Fig. 6. Comparison of the proposed consensus secondary-structure model of the 3′, 614 nt of the *C. elegans/A. suum* mt-l-rRNA genes (Fig. 5A) and the corresponding segment of the secondary structure model of the *E. coli* 23S rRNA (Fig. 5B; Gutell et al. 1993). The different classes of nucleotide pairs are as in Fig. 3. Numbering of each sequence is from the 5′ nucleotide indicated. In both models, solid dots are conserved nucleotides as given for the *E. coli* 23S rRNA by Cedergren et al. (1988) (see Fig. 8); *open squares* are other nucleotides used for phylogentic comparisons (also shown in Fig. 8); and *open circles* are nucleotides that are considered variable. In the nematode model are shown nucleotides (+) corresponding in position to nucleotides in the *E. coli* model (also shown by +) that appear to be involved in the decoding site (Fig. 5).

by Cedergren et al. (1988). Although there is a large overall size difference between the nematode and *E. coli* sequences, compared to the 404 nt in the conserved core sequences defined by Cedergren et al. (1988) (Fig. 8) there are 398 nt and 399 nt in the corresponding core sequences of *C. elegans* and *A. suum* mt-l-rRNAs, respectively. Taking into account apparent insertions in the *C. elegans* and *A. suum* sequences relative to the *E. coli* sequence, the nematode core sequences each have 9 nt less than the *E. coli* core sequences. By comparison, in the corresponding regions of the mouse and *D. yakuba* l-rRNAs (788 nt and 723 nt) the core sequences each have 2 nt less than the *E. coli* core sequences (Fig. 8).

Most of the sequence differences between the 3′ portions of the nematode mt-l-rRNAs and the *E. coli* 23S rRNA have resulted from elimination of blocks of sequences (Fig. 6), many of which correspond to sequences eliminated from the mouse and *D. yakuba* l-rRNAs relative to the *E. coli* 23S rRNA (Fig. 8). The high degree of conservation of core sequences is reflected in overall general similarity of the nematode mt-l-rRNA secondary-structure model to those of the corresponding portions of the mt-l-rRNA secondary-structure models of *D. yakuba* and mosquito (Gutell and Fox 1988).

In the nematode mt-l-rRNA model are found all of the secondary-structure elements and nucleotides that in *E. coli* 23S rRNA are associated with tRNA binding to the ribosomal A and P sites (Moazed and Noller 1989, 1990) (Figs. 5 and 6). Although the C675 nt in the *C. elegans* (and *A. suum*) mt-l-rRNA sequence could correspond to the E (exit) site-associated C2394 nt in the *E. coli* 23S rRNA sequence (Fig. 6) there are no apparent equivalents in the *C. elegans/A. suum* model (Fig. 5) to the three nucleotides—G2112, G2116 and A2169—that in *E. coli* 23S rRNA are associated with E-site function. This suggests that the proposed E site (Moazed and Noller 1989) may not be present in nematode mt-ribosomes.

A secondary-structure model of the *C. elegans* l-rRNA sequence has been presented recently by Gutell et al. (1993). Again there is considerable agreement in

detail between the 3′ section of this model and our consensus nematode model. The more noticeable differences between the two proposed structures are as follows.

In our model we have constructed a 6-ntp stem between nt 356–361 and 526–531 that corresponds to the upper half of a similarly located stem (nt 1,648–1,667: 1,993–2,009) in the *E. coli* 23S rRNA model (Fig. 6). The validity of this stem is supported by compensatory substitutions in the *C. elegans* and *A. suum* sequences, and also by similarity of six of seven nucleotides extending from the 5′ strand of the stem into a loop-out in the nematode and *E. coli* models (nt 360–366 and 1,667–1,673, respectively). In the Gutell et al. *C. elegans* model, this stem is omitted.

In the nematode consensus l-rRNA model, nt 640–690 appear to be the remnant of a 190-nt sequence (nt 2,244–2,433) in the *E. coli* 23S rRNA model that is folded into a complex secondary structure. In both our model and that of Gutell et al., nt 642–655 are folded into an identical stem-and-loop structure that includes three highly conserved nucleotides that in *E. coli* interact with tRNAs (Figs. 5 and 6). However, a second 12-nt stem and loop constructed within this sequence in the nematode consensus model (Fig. 5), supported by compensatory substitutions between the *C. elegans* and *A. suum* sequences, is not shown in the Gutell et al. model.

In the central region of the nematode consensus l-rRNA model, we have added (relative to the *E. coli* 23S rRNA model) 4 ntp and 2 ntp to two adjacent stems (nt 588–593: 867–872; and 844–845: 865–866, Fig. 6). Although neither of these additions could be made completely in the *E. coli* model, in the nematode model they make uninterrupted extensions of apparently stable helices in both the *C. elegans* and *A. suum* sequences. In the Gutell et al. *C. elegans* model, the folding of this region adheres to the *E. coli* pattern.

Within the terminal 70 nt (nt 884–953, Fig. 6) of the nematode consensus model, we have constructed only a single helical element. This corresponds to a stem and loop in *E. coli* 23S rRNA with respect to relative location and inclusion of a highly conserved 12-nt sequence (11/12 to *E. coli*, Figs. 5 and 6). In *E. coli* part of the distal portion of the loop has been folded into a bulged stem (Fig. 6). The corresponding helical elements in the Gutell et al. *C. elegans* model appear to have been constructed to best fit the *E. coli* bulged stem. Also, in the Gutell et al. model, nt 921–953 have been folded into simple stem and loops. Similar structures could be formed from the *A. suum* sequence, but it is not clear to which helical elements they correspond in the *E. coli* 23S rRNA model.

The 5′ 37% of the *C. elegans* and *A. suum* mt-l-rRNA sequences (nt 1,339 and 1,342, respectively) are shorter by approximately 1,305, 445, and 270 nt than the corresponding segments of *E. coli* 23S rRNA and mouse and *D. yakuba* mt-l-rRNAs, respectively. We have not constructed a consensus secondary-structure model for this part of the nematode l-rRNAs. However, such a structure showing eight conserved helical elements was included in the Gutell et al. *C. elegans* l-rRNA secondary-structure model.

Conserved Primary Sequences in Nematode s-rRNA and l-rRNA Genes

Gray et al. (1984) derived a set of primary sequences of small-subunit rRNAs (s-rRNAs), judged by them to be homologous, that they used in phylogenetic analyses. Selection of these sequences was based on their locations in the universally conserved core of secondary-structure models. A set of sequences based on similar criteria were derived for the large-subunit rRNAs (l-rRNAs) by Cedergren et al. (1988). Alignments of the respective *C. elegans* and *A. suum* mt-s-rRNA and mt-l-rRNA sequences with the corresponding sequences of *E. coli* 16S and 23S rRNAs and other mt-s-rRNAs and mt-l-rRNAs from a protozoan (*Paramecium tetraurelia*), a fungus (*Saccharomyces cerevisiae*), and three metazoa, the fly *Drosophila yakuba,* the sea urchin *Paracentrotus lividus,* and mouse (*Mus musculus*), are given in Figs. 7 and 8.

Because there have been an unusually high number of nucleotide substitutions and deletions (and some apparent insertions) in the universal core sequences of the *C. elegans* and *A. suum* rRNA genes relative to the corresponding rRNA gene sequences of other metazoa (see above, and Fig. 9), we have been unable to make a reliable assessment of homology for a considerable number of nucleotides in the nematode core regions. This problem is greater for the core sequences of the s-rRNA genes than for the core sequences of the 3′ 64% of the l-rRNA genes because of the higher overall conservation of the nematode mt-l-rRNA secondary structure. Therefore, for the purpose of examining the interrelationships of the *C. elegans* and *A. suum* mtDNAs with those of other metazoa, we have derived new sets of s-rRNA and l-rRNA core sequences. These sequences (given in Figs. 7 and 8) are a subset of the s-rRNA core sequences of Gray et al. (1984) and the l-rRNA core sequences of Cedergren et al. (1988), respectively, and in each case comprise sequence blocks that appear to us to represent homologous positions in all of the rRNAs considered. Our s-rRNA sequence set includes 203 of the 545 nucleotide positions in the sequence set of Gray et al. (1984) (Fig. 7). Our l-rRNA sequence set comprises a total of 284 nucleotide positions that include 253 of the 404 nucleotide positions in the sequence sets for the 3′ 64% of the l-rRNAs used by Cedergren et al. (1988) as well as 31 nucleotide positions from other conserved regions (Fig. 8).

Phylogenetic Sequence Comparisons

Similarity comparisons for the s-rRNA and l-rRNA genes represented in Figs. 7 and 8, based on our newly

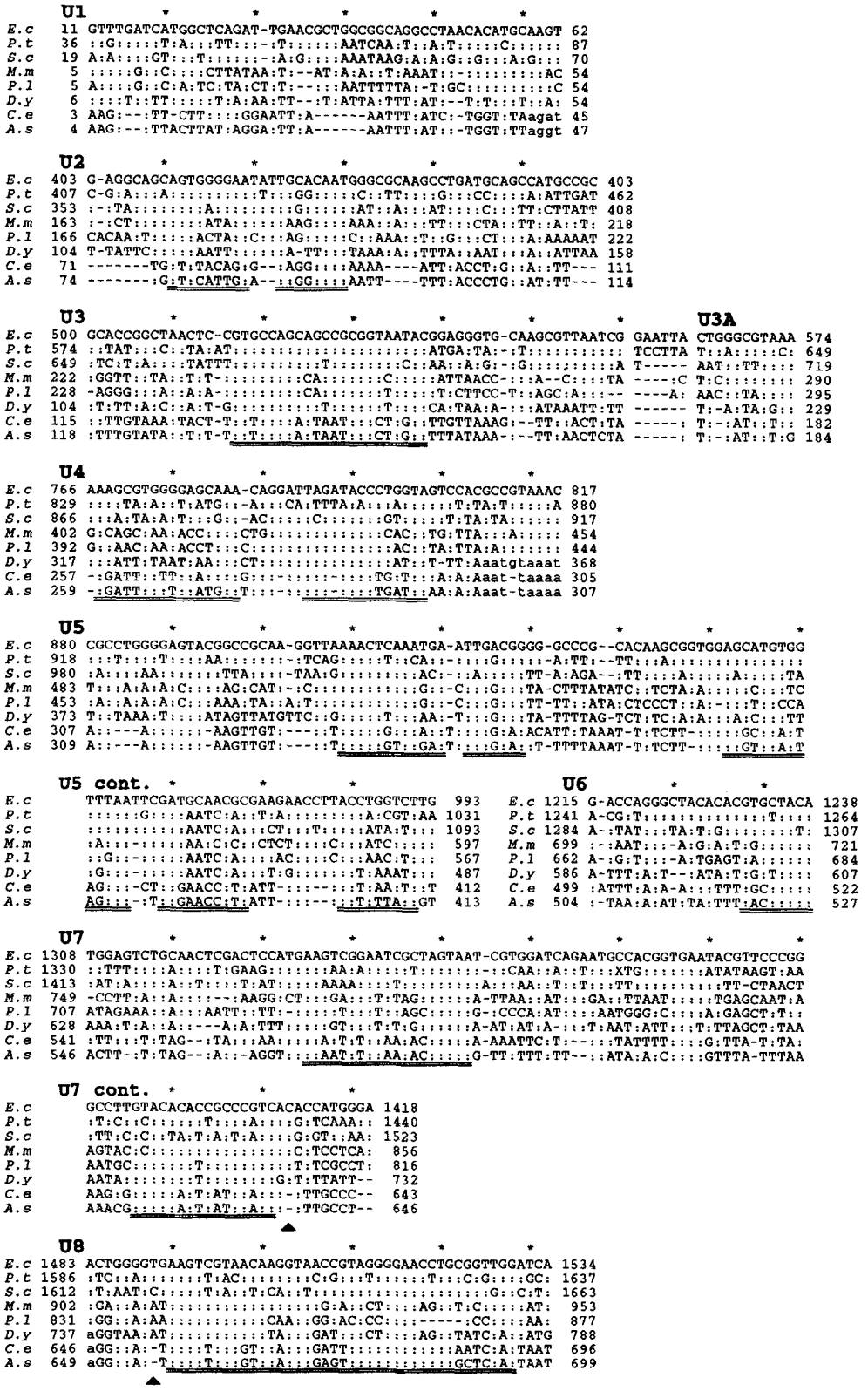

Fig. 7. Alignments of the universally conserved blocks of nucleotide sequences (U1–U8; Gray et al. 1984) of the s-rRNA genes of various mtDNAs and 16S rRNA of *E. coli*. All sequences are written 5′–3′, left to right. The *numbers* at the beginning and end of each sequence indicate the locations of the sequences in the respective s-rRNA and 16S RNA sequences referenced below. Identity and origin of the sequences are: *E.c, Escherichia coli* (Brosius et al. 1978); *P.t, Paramecium tetraurelia* (Seilhamer et al. 1984a); *S.c, Saccharomyces cerevisiae* (Li et al. 1982); *M.m, Mus musculus* (Bibb et al. 1981); *P.l, Paracentrotus lividus* (Cantatore et al. 1989); *D.y, Drosophila yakuba* (Clary and Wolstenholme 1985b); *C.e, Caenorhabditis elegans; A.s, Ascaris suum* (Fig. 3; Okimoto et al. 1992). A *dash* indicates the absence in a sequence of a nucleotide that is present in one or more other sequences. A *colon* indicates identity between a nucleotide in an mtDNA sequence and the *E. coli* sequence. *Lowercase letters* indicate nucleotides that lie within expected universal regions but which lack expected secondary-structure involvement. *Double underlines* identify sequences that compose the s-rRNA data set used for phylogenetic analysis (Fig. 10). *Solid triangles* beneath the U7 and U8 sequences indicate the presumed sites of nucleotide insertions/deletions in the interstem loops between the coaxial stack helix and helix 47 (Fig. 3). It should be noted that the alignments of the U5 and U8 sequences of *P. lividus* shown differ from those of Neefs et al. (1990).

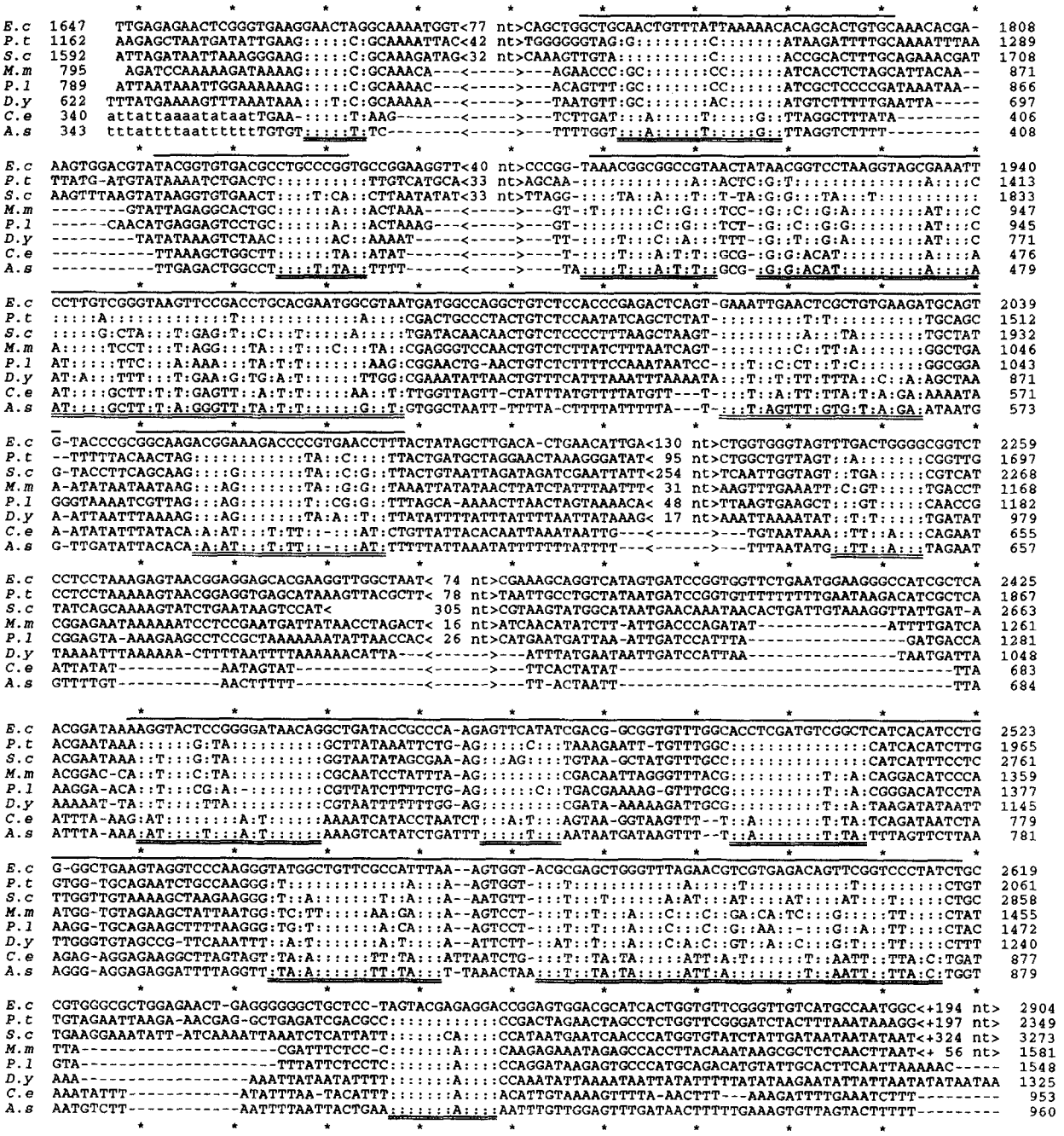

Fig. 8. Alignments of the 1258 nt 3′ end-proximal region of the *E. coli* 23S rRNA (the 3′ region (nt 1647–2904), as defined by Gutell and Fox, 1988) with the corresponding region of the mt-s-rRNA genes of *C. elegans* (*C.e*), *A. suum* (*A.s*) (Fig. 3; Okimoto et al. 1992), *D. yakuba* (*D.y,* Clary and Wolstenholme 1985b); *Paracentrotus lividus* (*P.l,* Cantatore et al. 1989); *Mus musculus* (*M.m,* Bibb et al. 1981); *Saccharomyces cerevisiae* (*S.c,* Sor and Fukuhara 1983); and *Paramecium tetraurelia* (*P.t,* Seilhamer et al. 1984b). *Lowercase letters* at the 5′ ends of the nematode sequences indicate nucleotides that are contiguous with, but not considered part of, the designated 3′ end-proximal region. A *dash* indicates the absence in a sequence of a nucleotide that is present in one or more other sequences. A *colon* indicates identity between a nucleotide in an mtDNA sequence and the *E. coli* sequence. The lengths of segments of nucleotides present in some sequences but not others are shown in *brackets. Single overlines* indicate the nucleotide sequences classified as universal, and used in the phylogenetic analyses of Cedergren et al. (1988). *Double underlines* identify sequences that compose the l-rRNA data set used for phylogenetic analysis (Fig. 10).

derived sequence sets, are summarized in Fig. 9. The *S. cerevisiae* and *P. tetraurelia* mt-s-rRNAs and mt-l-rRNAs have diverged less from the *E. coli* 16S and 23S rRNAs than have any of the corresponding metazoan mt-rRNAs. The *C. elegans* and *A. suum* mt-rRNA sequences have diverged more from the corresponding *E. coli* rRNA sequences than have any of the other mt-rRNA sequences considered. Divergence of the two nematode mt-l-rRNAs from the mt-l-rRNAs of *P. tetraurelia, S. cerevisiae, D. yakuba,* and *P. lividus* and mouse are greater than the divergence between any pair of mt-l-rRNAs of these five organisms.

	E.c	P.t	S.c	M.m	P.l	D.y	C.e	A.s
E.c		77.8	76.4	75.4	70.9	69.5	62.6	60.6
P.t	90.5		72.9	69.5	67.0	69.0	58.6	60.6
S.c	78.2	81.7		68.5	65.5	68.5	62.1	61.1
M.m	74.3	75.7	73.6		81.3	75.8	60.6	58.6
P.l	74.3	73.9	71.1	87.7		74.4	61.1	59.6
D.y	71.5	73.9	72.2	82.7	83.8		69.5	66.5
C.e	65.8	66.9	69.7	63.7	65.5	68.7		89.7
A.s	65.5	65.1	67.3	63.4	65.8	67.6	95.1	

s-rRNA

l-rRNA

Fig. 9. Similarities (as percentages) among conserved regions of nucleotide sequences of s-rRNAs (*above the diagonal*) and l-rRNAs (*below the diagonal*) of various mtDNAs and *E. coli*. The nucleotide sets used in the comparison were those indicated by *doubled underlines* in Figs. 7 and 8 and used in phylogenetic analyses (Fig. 10). *E.c, Escherichia coli; P.t, Paramecium tetraurelia; S.c, Saccharomyces cerevisiae; M.m, Mus musculus; P.l, Paracentrotus lividus; D.y, Drosophila yakuba; C.e, Caenorhabditis elegans; A.s, Ascaris suum.* References to sequence origins are given in the legends to Figs. 7 and 8.

With the exception of the divergences from *D. yakuba* mt-s-rRNA, the divergences of the *C. elegans* and *A. suum* mt-s-rRNAs from mt-s-rRNAs of other metazoa are also greater than any of the divergences found between the five mt-s-rRNAs considered (Fig. 9). The noticeably higher-similarity values for the *C. elegans/D. yakuba* (69.5%) and the *A. suum/D. yakuba* (66.5%) comparisons may reflect the unusually high A+T content of the sequence sets of all three of these mt-s-rRNAs (63.8%, 62.3%, 67.5% for *C. elegans, A. suum,* and *D. yakuba,* respectively).

We have used the maximum likelihood analysis of Felsenstein (PHYLIP, version 3.3) to calculate the most plausible phylogeny (Fig. 10) of the species represented in the 203-nt mt-s-rRNA and 284-nt mt-l-rRNA data sets (Fig. 6 and 8). The PHYLIP maximum likelihood program (DNAML) allows for specification of the transition:transversion ratios (s:v) for the various comparisons. Using computer simulations, Fukami-Kobayashi and Tateno (1991) showed that the DNAML topology estimation was robust when the data included varying s:v ratios and varying nucleotide compositions. In regard to the present data, these factors are important for the following reasons. The observed s:v ratios for comparisons of the *C. elegans* and *A. suum* entire mt-s-rRNAs and the conserved regions are 0.8:1 and 1.9:1, respectively (Table 1). The corresponding values for components of the *C. elegans* and *A. suum* mt-l-rRNAs are 0.6:1 and 1.6:1, respectively. For a number of vertebrate and invertebrate mtDNAs, it has been argued that transitions are accumulated in great excess to transversions but that with increase in divergence time the observed ratio of transitions to transversions decreases due to a masking effect (Brown et al. 1982; DeSalle et al. 1987). The s:v ratios for s-rRNAs of closely related primates was found to average 12:1 (Hixson and Brown 1986). Although information regarding s:v ratios for rRNA genes of closely related nematodes is lacking, it has been shown that the s:v ratio for substitutions between COII genes in strain comparisons of *C. elegans* and *C. briggsae* was 12:1. (The s:v ratio for the *C. elegans* and *A. suum* COII genes is about 0.9:1 [76.7% sequence similarity, see Okimoto et al. 1992]).

Nucleotide compositions for the rRNAs included in our data sets are shown in Table 2. There is considerable variation between the species, as has been noted previously for entire mtDNA sequences (for example, Wolstenholme 1992). However, for each of the four nucleotides in both genes, the variations between species are reduced in the regions we have defined as conserved (Figs. 7 and 8). In further support of our choice of the DNAML method, Hasegawa et al. (1991) found that this method was more efficient in obtaining the correct phylogeny than was the distance method used by Saitou (1988) in cases where the evolutionary rate differs among lineages. Such rate differences seem likely from our simple sequence similarity comparisons (Fig. 9).

Using the DNAML method for the mt-s-rRNA and mt-l-rRNA data sets, we obtained a consistent topology regarding the placement of the nematode branch point, irrespective of order of species introduction, and at s:v values varying from 2:1 (the default value) to 25:1. The branches directly relating to the position of the nematode branch never included zero lengths within their confidence limits. The relative branch points of the nematodes for the two data sets are consistent with the most prevalent classical phylogeny (see, for example, Wilson et al. 1978), but they are not in total agreement with the nematode branch points determined by Cedergren et al. (1988) for the larger nuclear s-rRNA and l-rRNA data sets using maximum parsimony and bootstrapping analysis. In the trees derived by Cedergren et al. (1988) from nuclear s-rRNA and l-rRNA data sets, the branch points of *C. elegans* deviate from the position expected from classical phylogenies. In the tree derived from nuclear s-rRNA data, *C. elegans* branches closer to vertebrates than does the arthropod *Artemia salina.* In the tree derived from nuclear l-rRNA data, the *C. elegans* branch is closer to the protist *Crithidia fasciculata* than to vertebrates, fungi, and plants. Interestingly, Cedergren et al. (1988) attributed these phylogenetic inconsistencies to an abnormally high rate of mutation (nucleotide substitution) in the nematode nuclear rRNA genes.

In the phylogenetic trees constructed from our mt-s-rRNA and mt-l-rRNA data sets the relative branch points concerning *Paramecium* and yeast differ (Fig. 10). However, the same relative branch points are found for *E. coli, Paramecium,* yeast, and metazoa as were calculated by Cedergren et al. (1988) from their larger mt-rRNA data sets. Trees derived using both mt-l-rRNA data sets have *Paramecium* branching closer to *E. coli* than does yeast, while trees derived using both mt-s-rRNA data

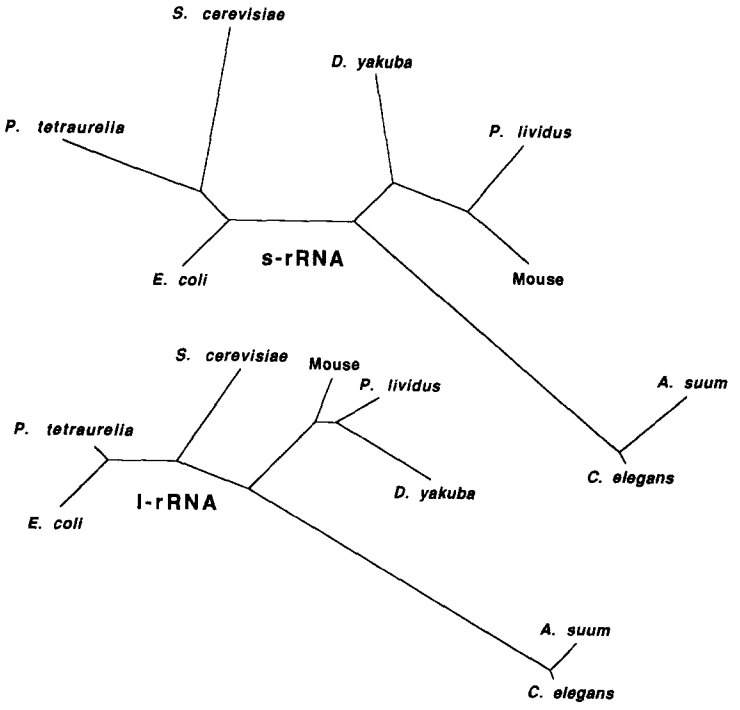

Fig. 10. Phylogenies resulting from applying the maximum-likelihood program DNAML, of the PHYLIP package version 3.3 by Joseph Felsenstein to the s-rRNA and l-rRNA data sets shown in Figs. 7 and 8. The following options were used: T option: Repeat runs were first made in which the transition:transversion (s:v) ratio was varied from the default 2.0 to 5.0, 10.0 and 25.0. T values of 3.0 and 4.0 were also used in the case of the l-rRNA data set. Based on the data obtained (see text) the s:v ratio was set at 10 (these figures). F option: The program used the empirical base frequencies calculated from the nucleotide content of the sequences provided for the analysis. G option: the global search and replace function was used to "improve the results." J option: the jumble option was used to get a random order of introduction of species using at least ten different random seed numbers for each T option value. In the case of T = 25 where the order of introduction of the species was found to significantly alter the branch lengths the same random seed numbers (orders of species introduction) were used at the lower T values to determine their affect.

sets have *Paramecium* and yeast branching from a more recent, common branch. These observations indicate that the DNAML method we used is subject to the same irregularities, with regard to yeast and *Paramecium*, between our mt-s-rRNA and mt-l-rRNA data sets as occurred using maximum parsimony and bootstrapping in the corresponding, larger rRNA data sets (Cedergren et al. 1988).

Varying s:v ratios for our mt-l-rRNA data had only one effect on branching: that of *D. yakuba*, sea urchin, and mouse. At s:v ratios of 4:1 or greater, the branching of these species was as shown in Fig. 10, but at s:v ratios of 2:1 or 3:1 it was similar to that shown in the tree derived from our mt-s-rRNA data (Fig. 10). In both topologies, the differing branches included a zero-branch-length possibility within their confidence limits. It may be possible to resolve this discrepancy by increasing the size of the metazoan data set.

Up to s:v ratios of 10:1, branch length showed differences that were essentially zero with change in order of species introduction. However, at s:v ratios of 25:1, although the topology remained constant with change in order of species introduction, branch lengths varied as much as tenfold.

It should also be noted that because branch lengths tend to increase with increase of s:v ratios, using our data sets the relative branch lengths for *E. coli*, *Paramecium*, and yeast shown in Fig. 10 may be overestimates.

From the common metazoan branch point, branch-length estimates for the nematodes average about twice the branch-length estimates for the other metazoan species (2.0 and 2.1 for the s-rRNA and l-rRNA data sets, Fig. 10), indicating a two-times-greater rate of nucleotide substitution in the nematode lineages. Interestingly, Cedergren et al. (1988) also noted a relatively long branch leading to *C. elegans* in their phylogenetic tree based on nuclear s-rRNA sequences.

Acknowledgments. We thank Kirk Thomas for oligonucleotides. D.R.W. is an Associate of the Program in Evolutionary Biology of the Canadian Institute for Advanced Research. This work was supported by National Institutes of Health Grant GM-18375 and has been submitted by R. Okimoto in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Biology), College of Science, University of Utah.

Table 1. Transitions and transversions between the entire sense strands, and between the conserved regions of the sense strands of the corresponding rRNA genes of *C. elegans* and *A. suum*

	Numbers of nucleotide substitutions			
	n	Transitions (s)	Transversions (v)	s:v
s-rRNA				
Total	697/717	67	86	0.8:1
Conserved regions	202	13	7	1.9:1
l-rRNA				
Total	953/960	85	145	0.6:1
Conserved regions	284	8	5	1.6:1

Table 2. Nucleotide composition data for various mitochondrial rRNA genes and 16S and 23S rRNAs of *E. coli*

	Nucleotide composition as percentages (sense strand)									
	Total gene					Conserved regions				
	A	C	G	T	nt	A	C	G	T	nt
s-rRNA										
C. elegans	35.9	9.8	16.5	37.9	697 nt	38.1	14.9	21.3	25.7	202 nt
A. suum	31.9	9.1	19.0	40.0	717 nt	36.1	14.9	22.8	26.2	202 nt
D. yakuba	40.4	7.6	13.1	38.9	789 nt	36.6	14.4	19.3	29.7	202 nt
P. lividus	36.8	22.1	19.1	22.0	878 nt	35.9	24.8	19.7	19.7	198 nt
M. musculus	37.8	19.7	16.1	26.4	955 nt	34.0	24.4	22.2	19.2	203 nt
S. cerevisiae	41.7	9.2	13.4	35.8	1669 nt	37.0	18.7	22.2	22.2	203 nt
P. tetraurelia	33.5	16.2	20.7	29.6	1662 nt	32.5	18.2	26.6	22.7	203 nt
E. coli	25.2	22.8	31.6	20.4	1542 nt	29.6	21.7	29.6	19.2	203 nt
Maximum difference (%)	16.5	15.2	18.5	17.5		8.5	10.4	10.3	10.5	
l-rRNA										
C. elegans	36.5	7.6	13.4	42.5	953 nt	31.9	11.2	19.0	37.9	284 nt
A. suum	26.6	6.3	17.0	49.2	960 nt	30.2	10.2	21.1	38.6	284 nt
D. yakuba	41.8	6.0	10.7	41.5	1325 nt	30.2	15.4	22.5	31.9	284 nt
P. lividus	37.4	18.8	17.7	26.1	1548 nt	28.3	20.9	25.4	25.4	282 nt
M. musculus	38.3	19.3	16.1	26.4	1581 nt	28.1	20.7	25.3	26.0	284 nt
S. cerevisiae	42.9	8.6	12.8	35.8	3273 nt	30.5	16.8	24.6	28.1	284 nt
P. tetraurelia	32.3	15.1	21.3	31.3	2349 nt	27.0	20.4	27.0	25.6	284 nt
E. coli	26.2	22.0	31.4	20.4	2904 nt	26.3	21.8	28.1	23.9	284 nt
Maximum difference (%)	15.6	16.0	20.7	28.8		5.6	11.6	9.1	14.7	

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

Arnason U, Gullberg A, Widegren B (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J Mol Evol 33:556–568

Arnason U, Gullberg A (1993) Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. J Mol Evol 37:312–322

Arnason U, Gullberg A, Johnsson E, Ledje C (1993) The nucleotide sequence of the mitochondrial DNA molecule of the grey seal *Halichverus grypus*, and a comparison with mitochondrial sequences of other true seals. J Mol Evol 37:323–330

Arnason U, Johansson E (1992) The complete mitochondrial DNA sequence of the harbor seal, *Phoca vilulina*. J Mol Evol 35:493–505

Arcari P, Brownlee GG (1980) The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucleic Acids Res 8:5207–5212

Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J Mol Evol 32:511–520

Bibb MJ, van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Brimacombe R, Atmadja J, Stiege W, Schuler D (1988) A detailed model of the three-dimensional structure of *Escherichia coli* 16S ribosomal RNA *in situ* in the 30S subunit. J Mol Biol 199:115–136

Brosius J, Dull TJ, Noller HF (1980) Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci USA 77:201–204

Brosius J, Palmer ML, Kennedy PS, Noller HF (1978) Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci USA 75:4801–4805

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Cantatore P, Roberti M, Rainaldi G, Gadaleta MN, Saccone C (1989) The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J Biol Chem 264:10965–10975

Cedergren R, Gray MW, Abel Y, Sankoff D (1988) The evolutionary relationships among known life forms. J Mol Evol 28:98–112

Clary DO, Wahleithner JA, Wolstenholme DR (1983) Transfer RNA genes in *Drosophila* mitochondrial DNA: related 5′ flanking sequences and comparisons to mammalian mitochondrial tRNA genes. Nucleic Acids Res 11:2411–2425

Clary DO, Wolstenholme DR (1983) Genes for cytochrome *c* oxidase subunit I, URF2 and three tRNAs in *Drosophila* mitochondrial DNA. Nucleic Acids Res 11:6859–6872

Clary DO, Wolstenholme DR (1985a) The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Clary DO, Wolstenholme DR (1985b) The ribosomal genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res 113:4029–4045

Clary DO, Wolstenholme DR (1987) *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J Mol Evol 25:116–125

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honey bee *Apis melifera:* complete sequence and genomic organization. Genetics 133:97–117

Dahlberg AE (1989) The functional role of ribosomal RNA in protein synthesis. Cell 57:525–529

Dams E, Hendriks L, Van de Peer Y, Neefs J-M, Smits G, Vandenbempt I, De Wachter R (1988) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 16s:r87–r173

de Bruijn MHL, Schreier PH, Eperon IC, Barrell BG, Chen EY, Arm-

strong PW, Wong JFH, Roe BA (1980) A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" loop and stem. Nucleic Acids Res 8:5213–5222

de Bruijn MHL (1983) *Drosophila melanogaster* mitochondrial DNA: a novel organization and genetic code. Nature 304:234–241

de Bruijn MHL, Klug A (1983) A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire 'dihydrouridine' loop and stem. EMBO J 2:1309–1321

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila.* J Mol Evol 26:157–164

Desjardins P, Morais R (1990) Sequence and organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. J Mol Biol 212:599–634

Fukami-Kobayashi K, Tateno Y (1991) Robustness of maximum likelihood tree estimation against different patterns of base substitutions. J Mol Evol 32:79–91

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisa E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516

Garey JR, Wolstenholme DR (1989) Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a $tRNA^{Ser}AGN$ that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. J Mol Evol 28:374–387

Glotz C, Zweib C, Brimacombe R (1981) Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucleic Acids Res 9:3287–3306

Gray MW, Sankoff D, Cedergren RJ (1984) On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit rRNA. Nucleic Acids Res 12:5837–5852

Gutell RR (1993) Collections of small subunit (16S and 16S-like) ribosomal RNA structures. Nucleic Acids Res 21:3051–3054

Gutell RR, Weiser B, Woese CR, Noller HF (1985) Comparative anatomy of 16S-like ribosomal RNA. Prog Nucleic Acid Res Mol Biol 32:155–216

Gutell RR, Noller HF, Woese CR (1986) Higher order structure in ribosomal RNA. EMBO J 5:1111–1113

Gutell RR, Gray MW, Schare MN (1993) A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. Nucleic Acids Res 21:3055–3074

Gutell RR, Fox GE (1988) A compilation of large subunit rRNA sequences presented in a structural format. Nucleic Acids Res 16s:r175–r269

Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32:443–445

Hixson JE, Brown WM (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol Biol Evol 3:1–18

Jacobs HT, Elliot DJ, Math VB, Farquharson A (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185–217

Li M, Tzagoloff A, Underbrink-Lyon N, Martin N (1982) Identification of the paromomycin-resistance mutation in the 15S rRNA gene of the yeast mitochondria. J Biol Chem 257:5921–5928

Melancon P, Lemieux C, Brakier-Gingras L (1988) A mutation in the 530 loop of *Escherichia coli* 16S ribosomal RNA causes resistance to streptomycin. Nucleic Acids Res 16:9631–9639

Moazed D, Noller HF (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. Nature 327:389–394

Moazed D, Noller HF (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. Cell 57:585–597

Moazed D, Noller HF (1990) Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16S rRNA. J Mol Biol 211:135–145

Moazed D, Noller HF (1991) Sites of interaction of the CCA end of peptidyl-tRNA with 23S rRNA. Proc Natl Acad Sci USA 88:3725–3728

Neefs JM, Van de Peer Y, Hendriks L, De Wacher R (1990) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 18s:2237–2317

Noller HF (1993) tRNA-rRNA interactions and peptidyl transferase. FASEB J 7:87–89

Noller HF, Asire M, Barta A, Douthwaite S, Goldstein T, Gutell RR, Moazed D, Normanly J, Prince JB, Stern S, Triman K, Turner S, Van Stolk B, Wheaton V, Weiser B, Woese CR (1986) Studies on the structure and function of ribosomal RNA. In: Hardesty B, Kramer G (eds) Structure, function and genetics of ribosomes. Springer-Verlag, New York, pp 143–163

Okimoto R, Wolstenholme DR (1990) A set of tRNAs that lack either the TΨC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J 9:3405–3411

Okimoto R, Macfarlane JL, Wolstenholme DR (1990) Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans.* Nucleic Acids Res 18:6113–6118

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum.* Genetics 130:471–498

Roe BA, Ma DP, Wilson RK, Wong JFH (1985) The complete nucleotide sequence of the *Xenopus levis* mitochondrial genome. J Biol Chem 260:9759–9774

Saitou N (1988) Property and efficiency of the maximum likelihood method for molecular phylogeny. J Mol Evol 27:261–273

Seilhamer JJ, Olsen GJ, Cummings DJ (1984a) Paramecium mitochondrial genes. 1 Small subunit rRNA gene sequence and microevolution. J Biol Chem 259:5167–5172

Seilhamer JJ, Gutell RR, Cummings DJ (1984b) Paramecium mitochondrial genes. 2 Large subunit rRNA gene sequence and microevolution. J Biol Chem 259:5173–5181

Sor F, Fukuhara H (1983) Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria. Nucleic Acids Res 11:339–348

Stern S, Weiser B, Noller HF (1988) Model for the three-dimensional folding of 16S ribosomal RNA. J Mol Biol 204:447–481

Stiegler P, Carbon P, Ebel JP, Ehresmann C (1981) A general secondary-structure model for procaryotic and eukaryotic RNAs of the small ribosomal subunits. Eur J Biochem 120:487–495

Thomas WD, Wilson AC (1991) Mode and tempo of molecular evolution in the nematode *Caenorhabditis:* cytochrome oxidase II and calmodulin sequences. Genetics 128:269–279

Tzeng C-S, Hui C-F, Shen S-C, Huang PC (1992) The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res 20:4853–4858

Wilson EO, Eisner T, Briggs WR, Dickerson RE, Metzenberg RL, O'Brien RD, Susman M, Boggs WE (1989) Life on earth, 2nd ed. Sinauer, Sunderland, MA

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. In: DR Wolstenholme, KW Jeon (eds) Mitochondrial genomes. Int Rev Cyto, vol 141. Academic Press, New York, pp 173–216

Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc Natl Acad Sci USA 84:1324–1328

Wolstenholme DR, Okimoto R, Macfarlane JL (1994) Nucleotide correlations that suggest tertiary interactions in the TV-replacement loop-containing mitochondrial tRNAs of the nematodes, *Caenorhabditis elegans* and *Ascaris suum.* Nucl Acids Res 22:4300–4306

Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res 9:3621–3640