The Mitochondrial Genomes of Two Nematodes, *Caenorhabditis elegans* and *Ascaris suum*

Ronald Okimoto, Jane L. Macfarlane, Douglas O. Clary[1] and David R. Wolstenholme

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

Manuscript received July 10, 1991
Accepted for publication November 22, 1991

ABSTRACT

The nucleotide sequences of the mitochondrial DNA (mtDNA) molecules of two nematodes, *Caenorhabditis elegans* [13,794 nucleotide pairs (ntp)], and *Ascaris suum* (14,284 ntp) are presented and compared. Each molecule contains the genes for two ribosomal RNAs (s-rRNA and l-rRNA), 22 transfer RNAs (tRNAs) and 12 proteins, all of which are transcribed in the same direction. The protein genes are the same as 12 of the 13 protein genes found in other metazoan mtDNAs: Cyt *b*, cytochrome *b*; COI–III, cytochrome *c* oxidase subunits I–III; ATPase6, F_o ATPase subunit 6; ND1–6 and 4L, NADH dehydrogenase subunits 1–6 and 4L: a gene for ATPase subunit 8, common to other metazoan mtDNAs, has not been identified in nematode mtDNAs. The *C. elegans* and *A. suum* mtDNA molecules both include an apparently noncoding sequence that contains runs of AT dinucleotides, and direct and inverted repeats (the AT region: 466 and 886 ntp, respectively). A second, apparently noncoding sequence in the *C. elegans* and *A. suum* mtDNA molecules (109 and 117 ntp, respectively) includes a single, hairpin-forming structure. There are only 38 and 89 other intergenic nucleotides in the *C. elegans* and *A. suum* mtDNAs, and no introns. Gene arrangements are identical in the *C. elegans* and *A. suum* mtDNA molecules except that the AT regions have different relative locations. However, the arrangement of genes in the two nematode mtDNAs differs extensively from gene arrangements in all other sequenced metazoan mtDNAs. Unusual features regarding nematode mitochondrial tRNA genes and mitochondrial protein gene initiation codons, previously described by us, are reviewed. In the *C. elegans* and *A. suum* mt-genetic codes, AGA and AGG specify serine, TGA specifies tryptophan and ATA specifies methionine. From considerations of amino acid and nucleotide sequence similarities it appears likely that the *C. elegans* and *A. suum* ancestral lines diverged close to the time of divergence of the cow and human ancestral lines, about 80 million years ago.

THE mitochondrial (mt-) genomes of multicellular animals (metazoa) are contained in a single, circular molecule with a species-specific size that varies from 14 to 39 kb (MORITZ, DOWLING and BROWN 1987; SNYDER *et al.* 1987). The only known exceptions are found in the cnidarian genus *Hydra* where the mt-genomes occur as two unique 8-kb linear molecules (WARRIOR and GALL 1985). Complete nucleotide sequences and gene content have been determined for four mammals; human, cow, mouse, and rat (ANDERSON *et al.* 1981, 1982a,b; BIBB *et al.* 1981; GADALETA *et al.* 1989); a bird, *Gallus domesticus* (DESJARDINS and MORAIS 1990); an amphibian, *Xenopus laevis* (ROE *et al.* 1985); two sea urchins, *Strongylocentrotus purpuratus*, and *Paracentrotus lividus* (JACOBS *et al.* 1988; CANTATORE *et al.* 1989); and an insect, *Drosophila yakuba* (CLARY and WOLSTENHOLME 1985a). Also, partial mtDNA sequences have been obtained from a number of other vertebrates and invertebrates (references in GAREY and WOLSTENHOLME 1989).

Each completely sequenced metazoan mtDNA contains the genes for the structural RNAs of the mitochondrion's own protein synthesizing machinery (2 rRNAs and 22 tRNAs), and 12 or 13 proteins. These proteins are all components of the oxidative phosphorylation system: cytochrome *b* (Cyt *b*), subunits I–III of cytochrome *c* oxidase (COI–III), subunits 6 and 8, of the F_o ATPase complex (ATPase6 and ATPase8), and subunits 1–6 and 4L of the respiratory chain NADH dehydrogenase (ND1–6 and 4L) (CHOMYN and ATTARDI 1987). In each metazoan mtDNA molecule there is a region of variable length [125 nucleotide pairs (ntp) to about 20 kb; JACOBS *et al.* 1988; BOYCE, ZWICK and AQUADRO 1989] that lacks genes, and in some cases has been shown to contain signals for the initiation of replication and transcription (the control region; MONTOYA *et al.* 1982; CLAYTON 1982, 1984). In some mtDNA molecules segments of various sizes, often within the control region, are duplicated (references in OKIMOTO *et al.* 1991).

The genetic codes used by metazoan mt-protein genes contain various modifications (BARRELL, BANKIER and DROUIN 1979; BARRELL *et al.* 1980). In all

[1] Current address: Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, California 94143.

metazoan mtDNAs, TGA specifies tryptophan rather than being a stop codon. ATA has been interpreted as specifying methionine rather than isoleucine in all metazoan mt-genetic codes except those of echinoderms (JACOBS *et al.* 1988; CANTATORE *et al.* 1989). In vertebrate mt-protein genes, AGA and AGG are absent (BIBB *et al.* 1981) or are used as rare stop codons (ANDERSON *et al.* 1981, 1982b; ROE *et al.* 1985). However, in *D. yakuba* mtDNA, AGA (but not AGG) specifies serine and, in nematode, platyhelminth, and echinoderm mtDNAs both AGA and AGG specify serine (WOLSTENHOLME and CLARY 1985; WOLSTENHOLME *et al.* 1987; HIMENO *et al.* 1987; GAREY and WOLSTENHOLME 1989).

Both translation initiation and translation termination of metazoan mt-protein genes have unusual features [see OKIMOTO, MACFARLANE and WOLSTENHOLME (1990) for references]. Among many of these protein genes, triplets other than ATG (AUG) are used as translation initiation codons. These include ATA, ATT, ATC, GTG, GTT and TTG. Also ATAA has been suggested as the translation initiation codon of the Drosophila COI gene. Some mt-protein genes in organisms from different metazoan phyla end in T or TA rather than a complete termination codon. UAA codons in mature transcripts of these genes are generated by precise cleavage from multicistronic primary transcripts, followed by polyadenylation (OJALA, MONTOYA and ATTARDI 1981).

Because of unusual wobble rules the 22 tRNAs that are encoded in all metazoan mtDNAs are apparently sufficient to decode all of the mtDNA-encoded protein genes (BARRELL *et al.* 1980). Modifications in structure are found among metazoan mt-tRNA genes. Variation in both size and sequence of the dihydrouridine (DHU) and TΨC loops are found in many mt-tRNA genes in organisms ranging from platyhelminths to mammals (discussion and reference in WOLSTENHOLME *et al.* 1987; GAREY and WOLSTENHOLME 1989). In all metazoan mtDNAs so far examined the $tRNA^{ser}$(AGY/A/G) gene lacks a DHU arm (references in GAREY and WOLSTENHOLME 1989). The mt-$tRNA^{ser}$(UCN) gene of nematodes, but not other metazoa, has a similar structure, and the remaining 20 mt-tRNA genes of nematodes all lack a TΨC arm (WOLSTENHOLME *et al.* 1987, 1989; OKIMOTO and WOLSTENHOLME 1990; OKIMOTO *et al.* 1991).

In this paper we present and compare the nucleotide sequences of the free-living nematode *Caenorhabditis elegans* and the pig intestinal parasitic nematode *Ascaris suum*. Various aspects of these sequences regarding protein, tRNA, and rRNA gene structure, modifications of the nematode mt-genetic code, codon usage, and some evolutionary considerations are discussed.

MATERIALS AND METHODS

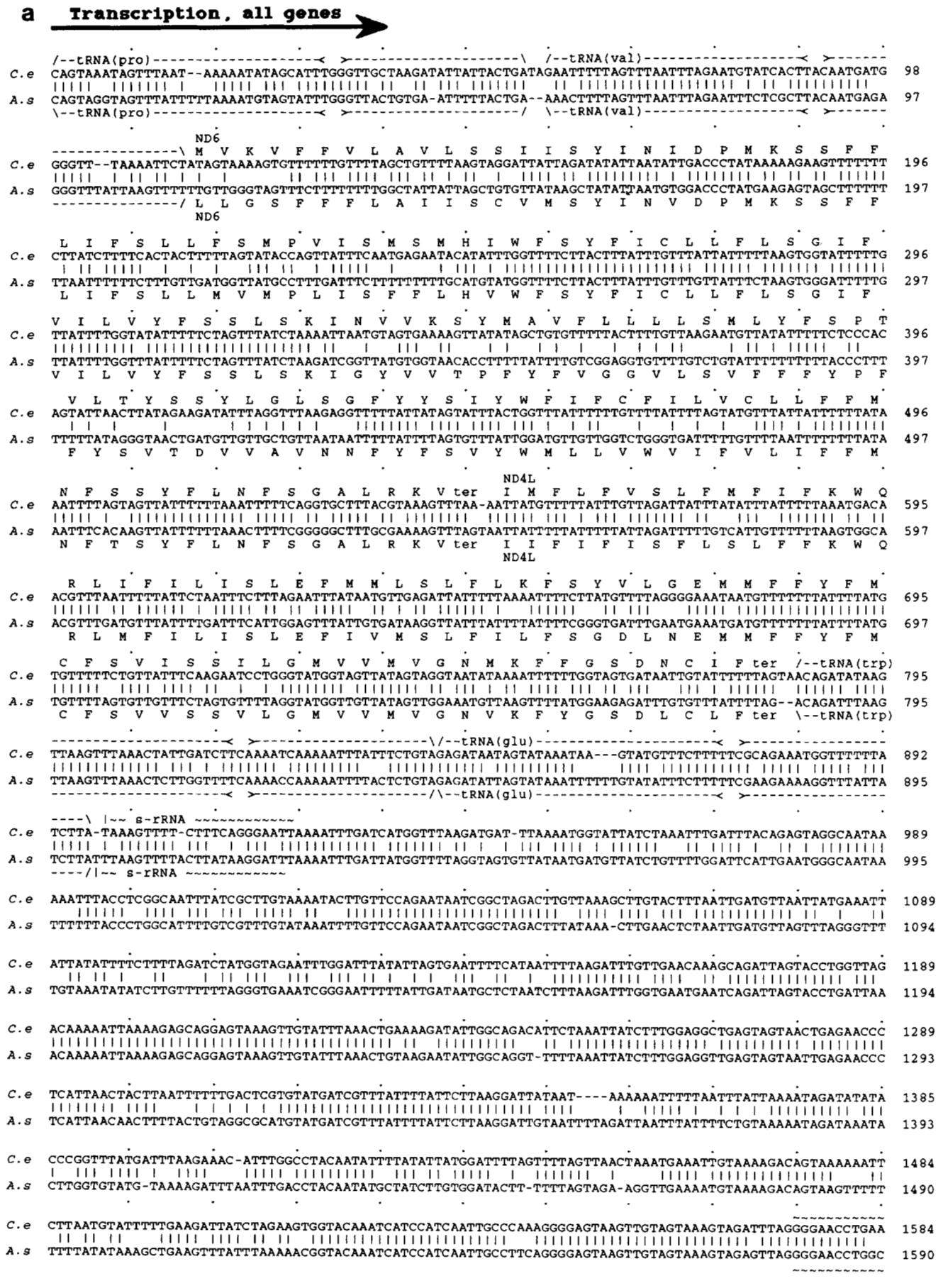
Animals, and mtDNA isolation: Adult *A. suum* were obtained from pig intestines at a local slaughterhouse. Mitochondria were isolated from body wall muscle or from mature eggs by methods previously used to isolate Drosophila mitochondria (WOLSTENHOLME and FAURON 1976). *C. elegans* (Bristol, N2 strain) were maintained, amplified and harvested as given in BRENNER (1974) and SULSTON and BRENNER (1974), except that *Klebsiella aerogenes* was used as the food source. Worms were ruptured with a Dounce homogenizer (pestle A) and mitochondria were isolated as for *A. suum*, except that mannitol was used instead of sucrose, and all solutions contained 0.1–0.2% bovine serum albumin. Mitochondria from both species were lysed with 10% Sarkosyl and mtDNAs were isolated by cesium chloride-ethidium bromide centrifugation (WOLSTENHOLME and FAURON 1976).

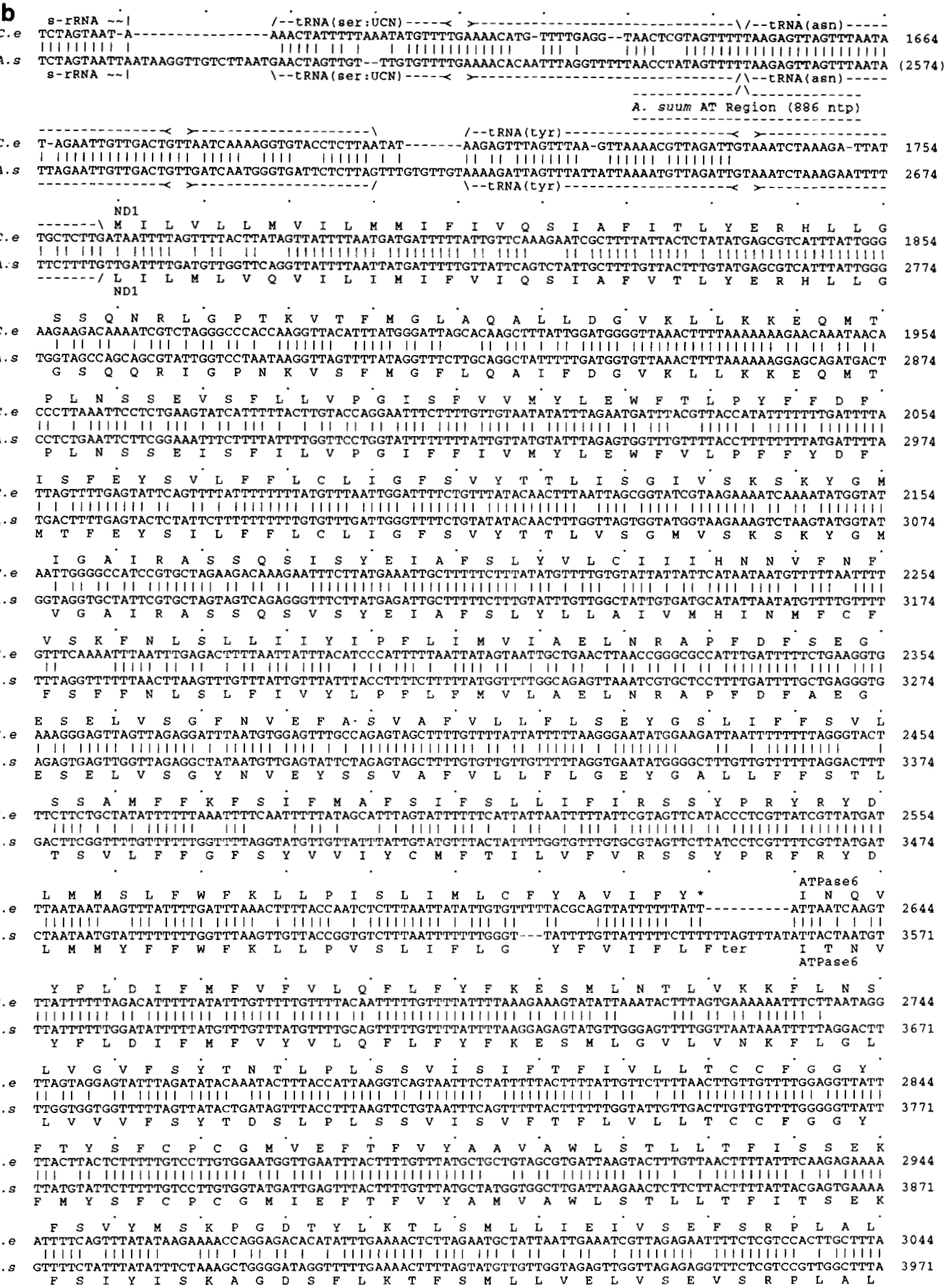
Restriction enzyme digestions and cloning: *A. suum* and *C. elegans* mtDNA restriction fragments were cloned into the plasmids pBR325 or pUC9 and amplified using as hosts *E. coli* K12 HB101 and JM101, respectively. Preparation and identification of primary clones, restriction enzyme digestions, electrophoresis, cloning of fragments into bacteriophage M13 vectors (M13 mp8, M13 mp9, M13 mp18 and M13 mp19), purification of single-stranded and double-stranded M13 DNAs, and preparation of viral DNAs containing partial deletions of cloned restriction fragments of mtDNAs were as given or referred to in CLARY *et al.* (1982) and WAHLEITHNER and WOLSTENHOLME (1987).

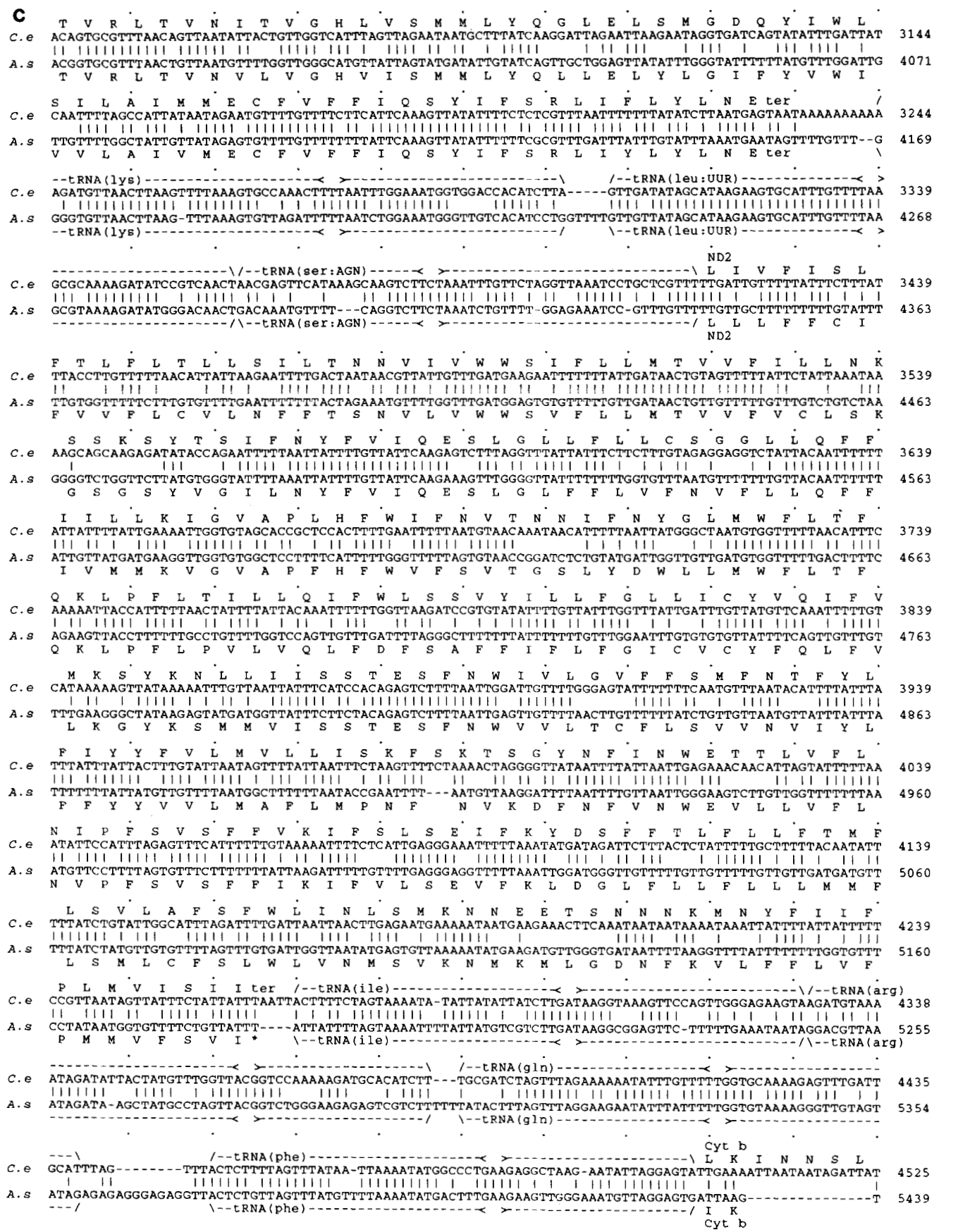
Sequencing and sequence analyses: DNA sequences were obtained by the extension-dideoxy-termination procedure (SANGER, NICKLEN and COULSON 1977) from sets of deletion clones (HONG 1982; DALE, MCCLURE and HOUCHINS 1985) containing overlapping segments of the entire sequence of each complementary strand of the *A. suum* and *C. elegans* mtDNA molecules. Consensus sequences were assembled from individual sequences using the compiling program of STADEN (1982). Nucleotide sequences were analyzed by the SEQ program (BRUTLAG *et al.* 1982). Nematode mt-protein genes were identified by comparing predicted amino acid sequences with amino acid sequences of mouse and *D. yakuba* mt-protein genes (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a) using the TYPIN and SEARCH programs (JUE, WOODBURY and DOOLITTLE 1980; DOOLITTLE 1981) and, in some cases, by hydropathic profile comparisons (KYTE and DOOLITTLE 1982). mt-rRNA genes were identified from sequence similarities to mouse and *D. yakuba* mt-rRNA genes, and mt-tRNA genes were identified by eye, from their ability to fold into specific, consensus secondary structures. The nucleotide sequences of the *C. elegans* and *A. suum* mtDNA molecules have been submitted to the EMBL Data Library under the accession numbers X54252 and X52453.

RESULTS AND DISCUSSION

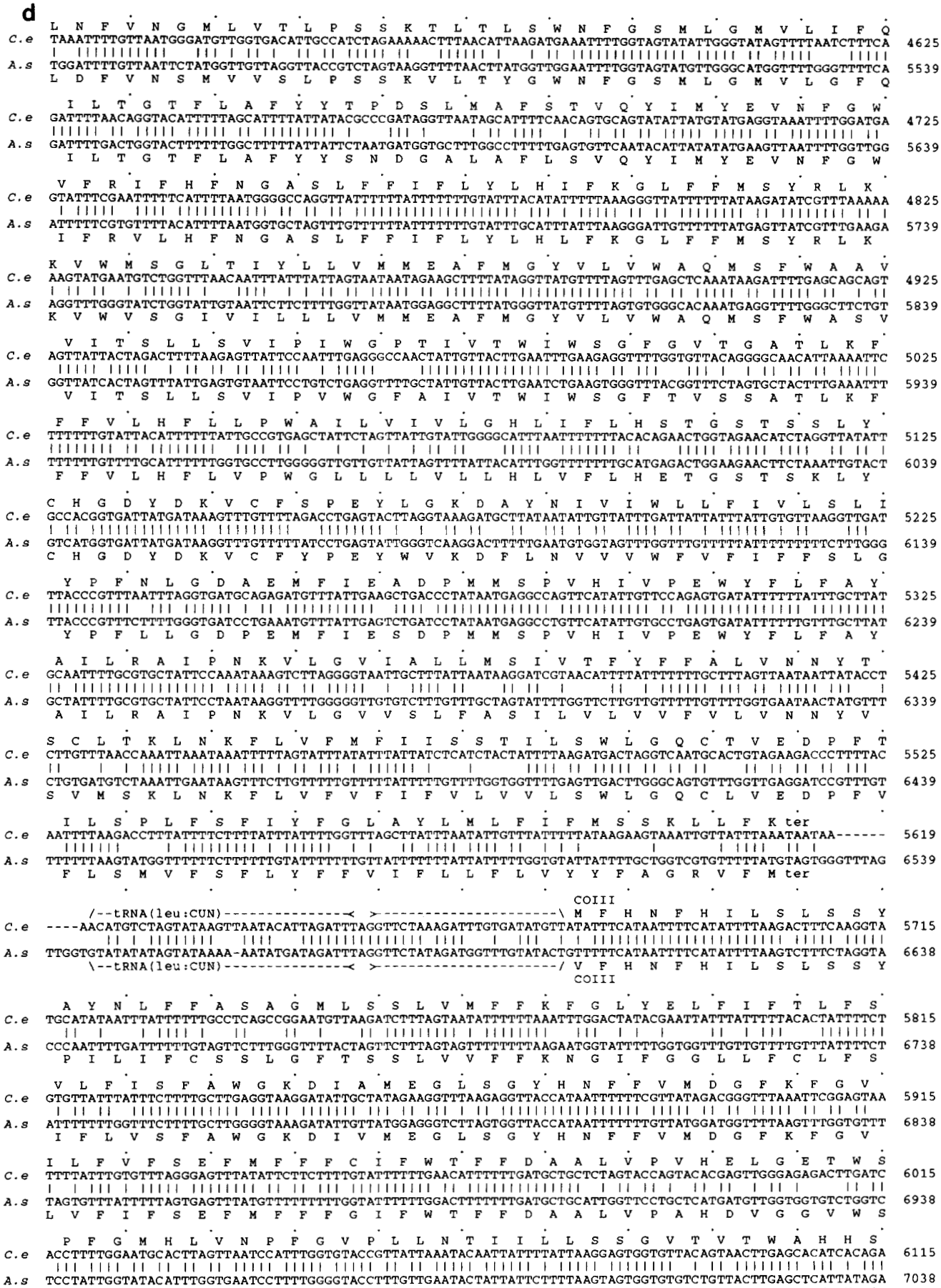
Genome structure and organization: The entire nucleotide sequences of the mtDNA molecules of *C. elegans* (13,794 ntp) and *A. suum* (14,284 ntp) are given and compared in Figure 1; gene content and organization in the two molecules are summarized in the maps shown in Figure 2. These are the smallest metazoan mtDNA molecules so far recorded. Each of the two nematode mtDNAs contains the genes for 12 proteins, 2 rRNAs and 22 tRNAs. The protein genes

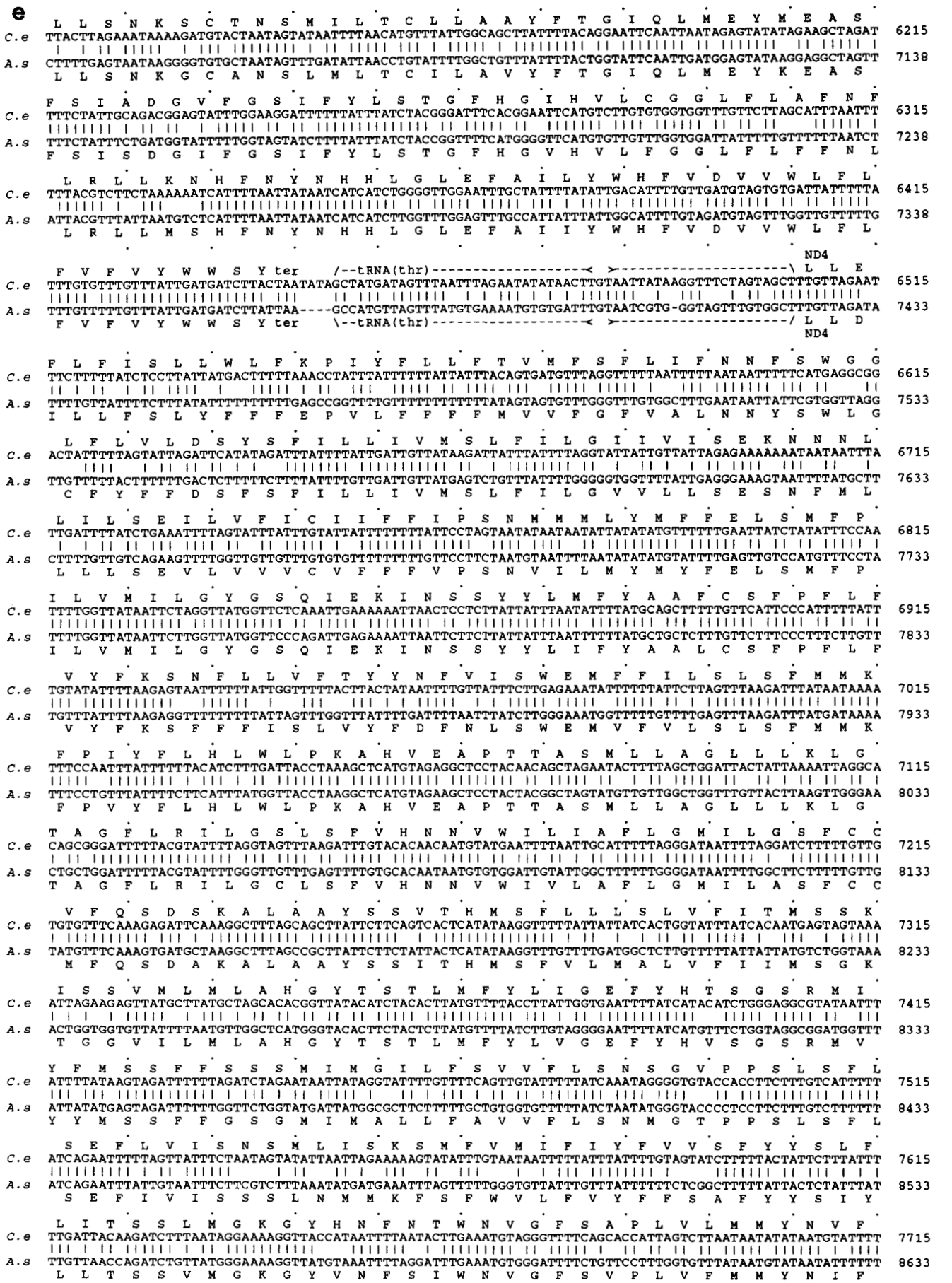

a Transcription, all genes
/--tRNA(pro)
\--tRNA(val)
C.e
A.s
ND6
ND4L
s-rRNA

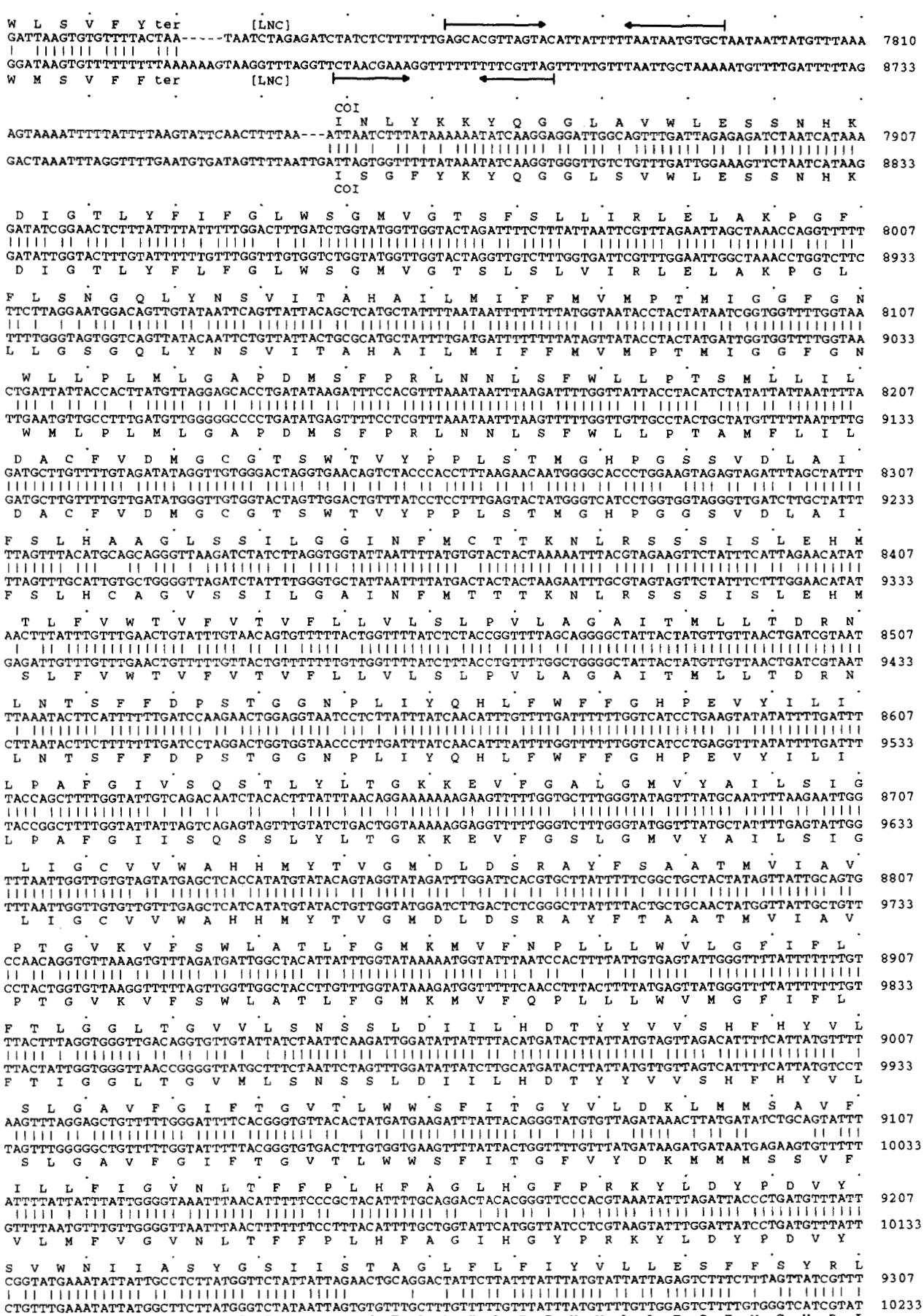

b
s-rRNA
/--tRNA(ser:UCN)
\--tRNA(ser:UCN)
\/--tRNA(asn)
/\--tRNA(asn)
A. suum AT Region (886 ntp)
/--tRNA(tyr)
\--tRNA(tyr)
ND1
ATPase6
C.e
A.s

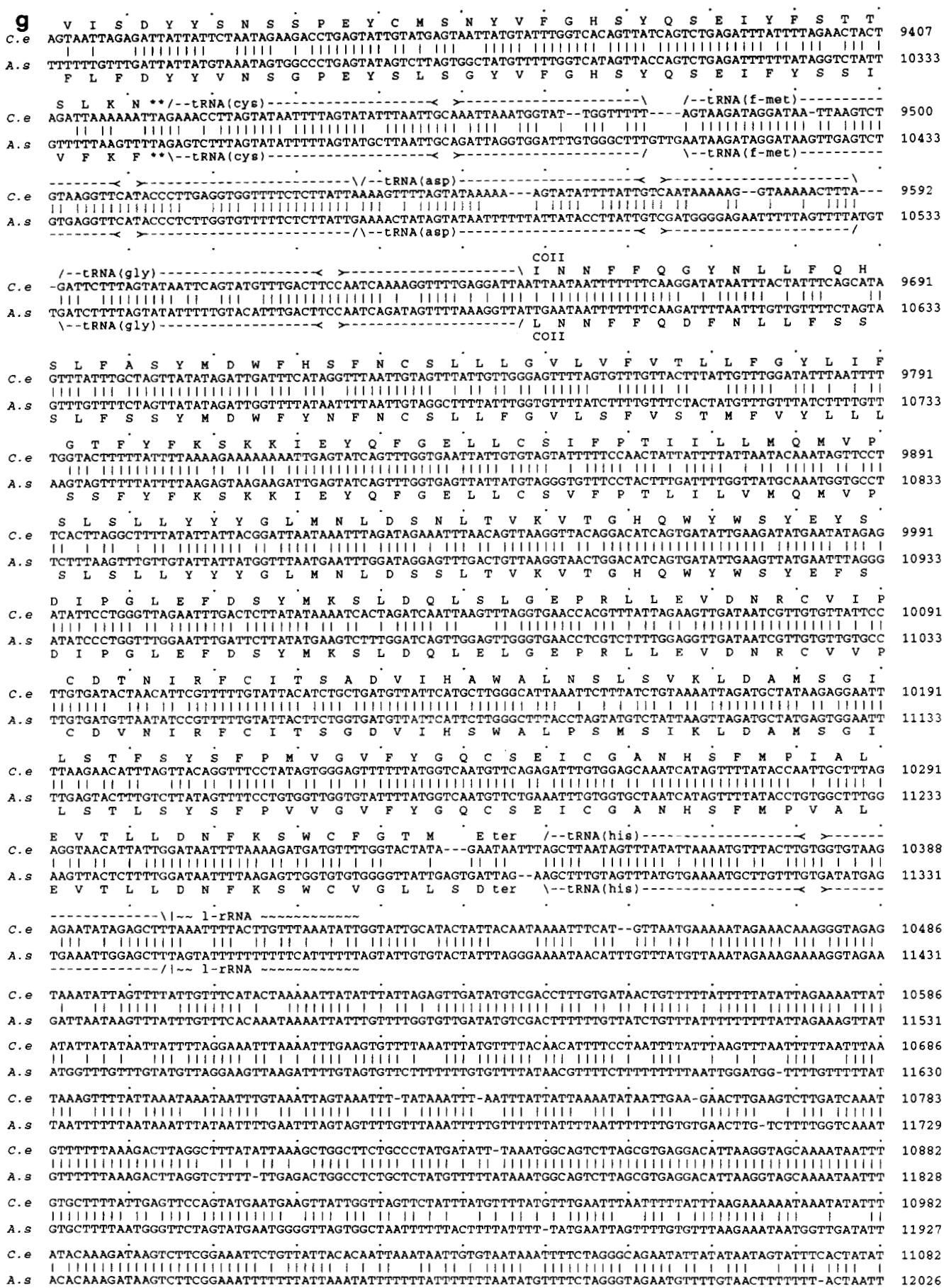

C
C.e
A.s
tRNA(lys)
tRNA(ser:AGN)
tRNA(leu:UUR)
tRNA(ile)
tRNA(gln)
tRNA(phe)
tRNA(arg)
ND2
Cyt b
ter

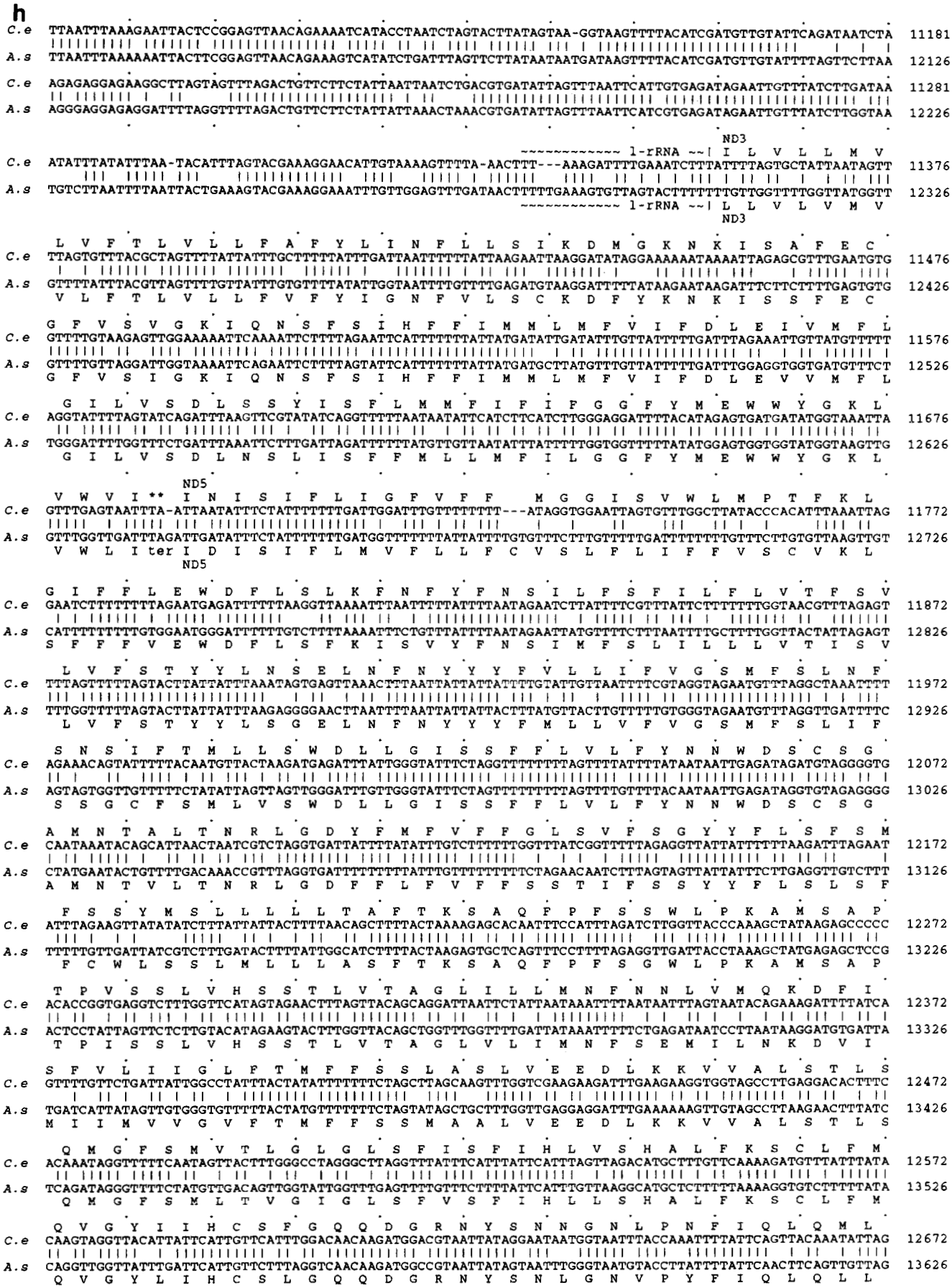
d




e
C.e
A.s
/--tRNA(thr)
\--tRNA(thr)
ND4
ND4


f
C.e
A.s
[LNC]
COI


g
C.e
A.s
/--tRNA(cys)
/--tRNA(f-met)
/--tRNA(asp)
/--tRNA(gly)
COII
\--tRNA(gly)
\--tRNA(asp)
\--tRNA(cys)
\--tRNA(f-met)
/--tRNA(his)
\--tRNA(his)
E ter
l-rRNA


h
C.e
A.s
ND3
1-rRNA
ND5
ter

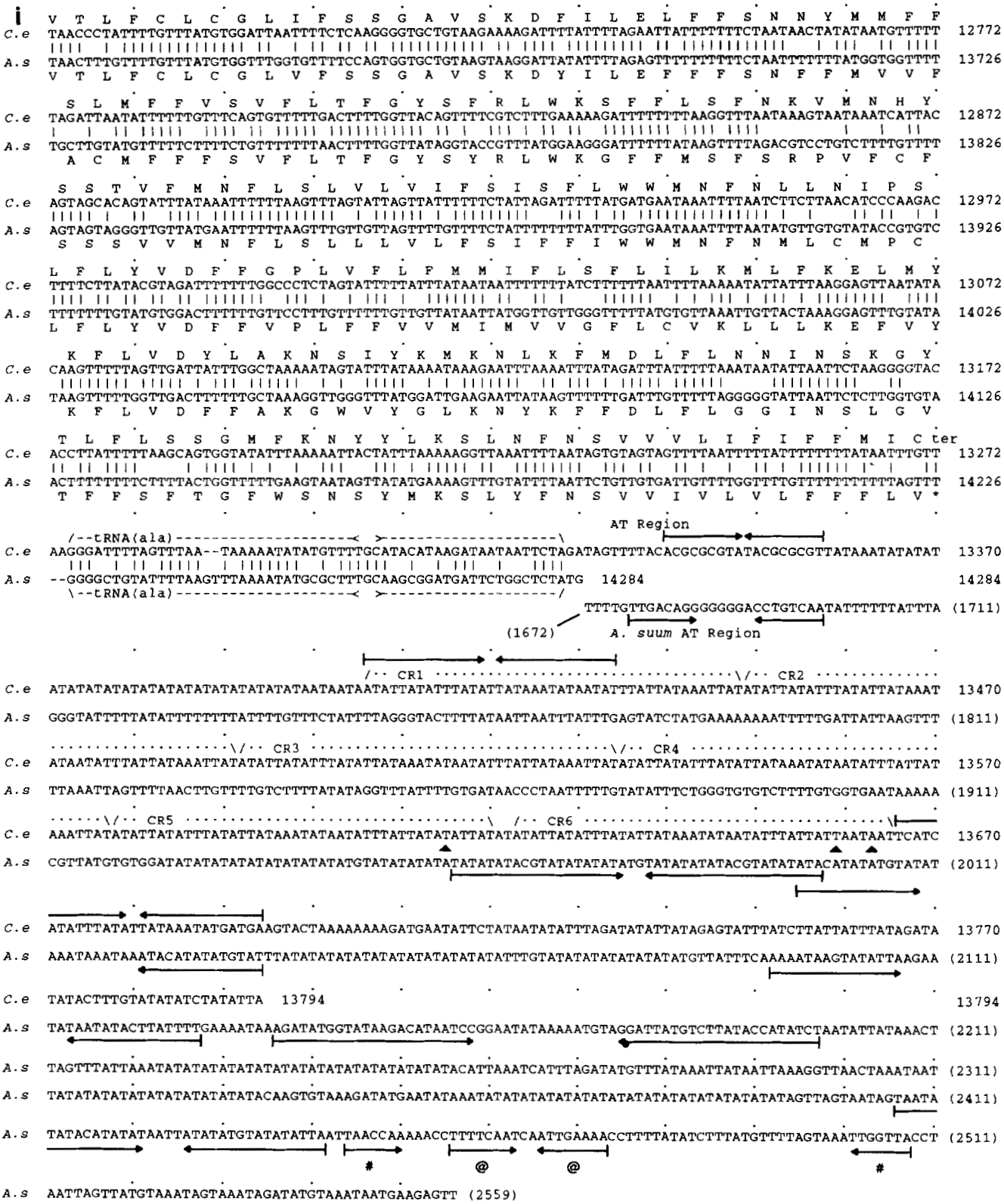


FIGURE 1.—Comparison of the nucleotide sequences of the mtDNA molecules of the two nematodes, *Caenorhabditis elegans* (*C.e*; 13,794 ntp) and *Ascaris suum* (*A.s*; 14,284 ntp). All of the 12 protein genes (Cyt *b*, cytochrome *b*; ATPase6, subunit 6 of the F_o ATPase; COI–III, cytochrome *c* oxidase subunits I–III; ND1–6 and 4L, subunits 1–6 and 4L of the respiratory chain NADH dehydrogenase), two rRNA genes and 22 tRNA genes in each sequence are transcribed in the same direction, from left to right (5′-3′). The strand of each gene shown is therefore the sense strand. In each sequence the numbering (right) begins with the first nucleotide of the $tRNA^{pro}$ gene, proceeds through that gene continuously around the entire molecule and terminates with the last nucleotide of the AT region in the *C. elegans* molecule, and with the second intergenic nucleotide following the $tRNA^{ala}$ gene in the *A. suum* molecule. A dash indicates the absence in one nematode

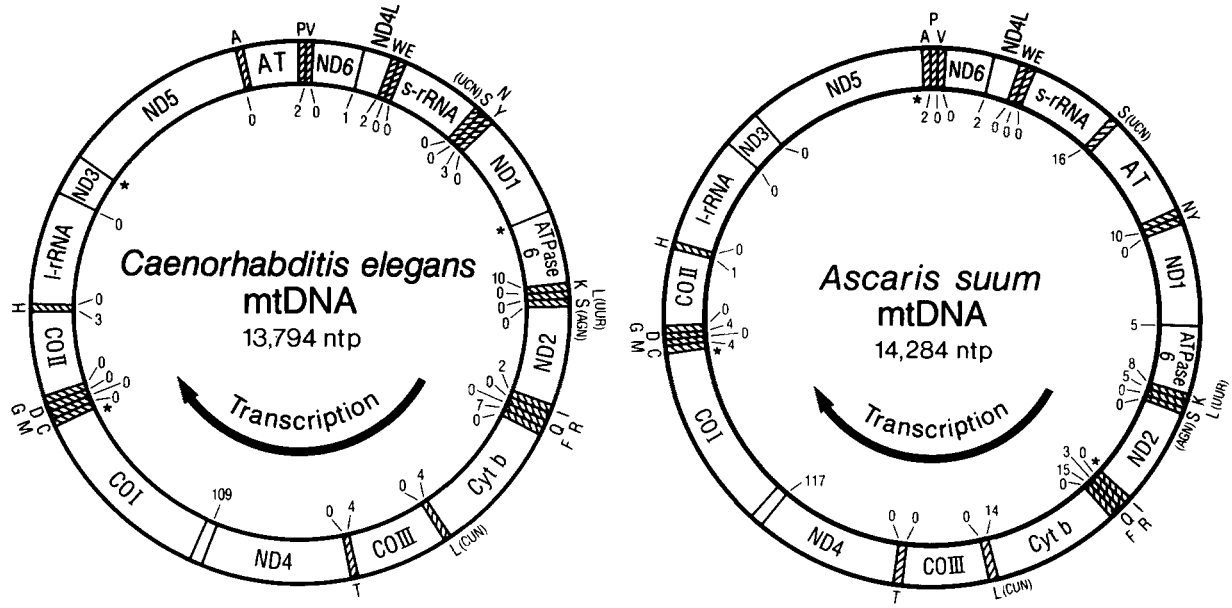


FIGURE 2.—Gene maps of the *C. elegans* and *A. suum* mtDNA molecules. Both molecules contain the genes for twelve proteins, the small and large subunit rRNAs (s-rRNA and l-rRNA) and 22 tRNAs (hatched areas). Each tRNA gene is identified by the one letter amino acid code. Serine and leucine tRNA genes are also identified by the codon family that the tRNAs recognize. All genes are transcribed in the direction shown by the arrows. Numbers at gene boundaries on the inner side of the map indicate apparently noncoding nucleotides between genes. An asterisk indicates an incomplete termination codon: T or TA. The AT region is a sequence of 466 ntp (93% A+T) in *C. elegans* mtDNA, and 886 ntp (85% A+T) in *A. suum* mtDNA, in which genes have not been identified. The region between the ND4 and COI genes (109 ntp in *C. elegans* and 117 ntp in *A. suum*) includes a sequence with the potential to form a stable hairpin structure (Figure 3), but again this region lacks genes. Gene order is identical in the two nematode mtDNA molecules, except for the relative locations of the AT regions.

are the same as 12 of the 13 protein genes found in vertebrate, *Drosophila yakuba*, and sea urchin mtDNAs. A gene for ATPase8 that precedes the ATPase6 gene in other metazoan mtDNAs has not been located in either of the two nematode mtDNAs. Both the *C. elegans* and *A. suum* mtDNA molecules include a sequence that contains runs of AT dinucleotides, and direct and inverted repeats, but in which genes have not been identified. This sequence (466 ntp in *C. elegans*, and 886 ntp in *A. suum*) has been designated the AT region, and may be the control region of the molecule (see below). Between the ND4 and COI genes is a sequence (109 ntp in *C. elegans* and 117 ntp in *A. suum*) that includes a region with the potential to form a hairpin structure (Figures 1 and 3). Gene arrangement is identical in the *C. elegans* and *A. suum* mtDNA molecules, except that in *C. elegans* the AT region is located between the $tRNA^{ala}$ and $tRNA^{pro}$ genes and in *A. suum* it is located between the $tRNA^{ser}$(UCN) and $tRNA^{asn}$ genes.

In *A. suum* and *C. elegans* mtDNAs all genes are transcribed in the same direction. This again contrasts to the situations found in other totally sequenced metazoan mtDNAs where different proportions of protein, and tRNA genes are transcribed in opposite directions (ANDERSON *et al.* 1981; ROE *et al.* 1985; CLARY and WOLSTENHOLME 1985a; JACOBS *et al.* 1988; CANTATORE *et al.* 1989; DESJARDINS and MORAIS 1990). As in all other previously sequenced metazoan mtDNAs, none of the *C. elegans* and *A. suum* mt-genes contain introns, and other than the AT region and the 109 ntp/117 ntp hairpin-containing regions, there are very few nucleotides between genes: 38 in *C. elegans* and 89 in *A. suum* (Figures 1 and 2).

nucleotide sequence of a nucleotide that occurs in the corresponding, other nucleotide sequence. Vertical lines indicate similarities between the *C. elegans* and *A. suum* nucleotide sequences. The predicted amino acid sequence of each of the 12 protein genes in each molecule are shown above the *C. elegans* sequence and below the *A. suum* sequence. Complete termination codons are indicated by ter, and presumptive incomplete termination codons, T or TA, are indicated by * and **, respectively. Modifications of the standard genetic code used in translation of the nematode mt-protein gene sequences are: AGA and AGG = serine; TGA = tryptophan; ATA = methionine. Sequences of tRNA genes are identified by dashed overlines and underlines within which brackets identify the anticodons. The terminal regions of the small and large rRNA genes are indicated with wavy overlines and underlines. The location of the AT region in the *A. suum* molecule (between $tRNA^{ser}$(UCN) and $tRNA^{asn}$: nucleotides 1672–2559) is indicated, and the 886 ntp sequence of this segment is shown at the end of the *A. suum* molecule in the location corresponding to the *C. elegans* (466 ntp) AT region. Six copies of a 43 ntp repeated sequence (CR1-CR6) within the *C. elegans* AT region are identified by dotted overlines. Nucleotides within repeats CR5 and CR6 that are variant relative to the identical CR1 to CR4 repeats are indicated by inverted arrowheads below the sequence. Arrow pairs above and beneath the *C. elegans* and *A. suum* sequences, respectively, identify inverted repeat sequences in both the AT regions and the 109 ntp (*C. elegans*) and 117 ntp (*A. suum*) long, noncoding (LNC) regions located between the ND4 and COI genes. The *C. elegans* and *A. suum* AT regions and LNC regions have not been aligned with regard to nucleotide sequence similarity.

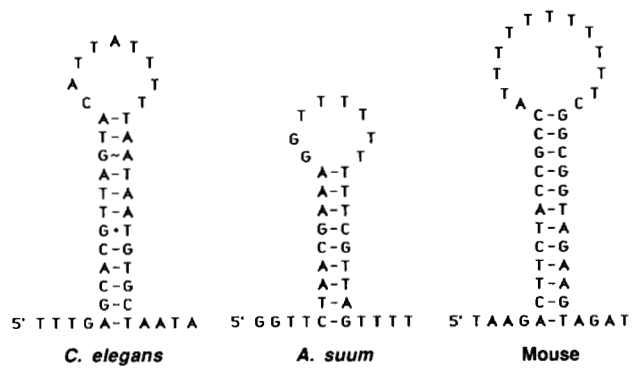


FIGURE 3.—Potential hairpin structures found within the long, noncoding (LNC) regions located between the ND4 and COI genes of *C. elegans* and *A. suum* (Figure 1), compared to the hairpin structure in which L strand synthesis originates in mouse mtDNA (CHANG *et al.* 1985).

Also, it is not necessary to predict overlaps between any of the genes in either of the nematode mtDNA molecules.

Gene arrangements: The various protein, rRNA and tRNA genes have identical arrangements in amphibian, bird and mammalian mtDNAs, except that in bird mtDNA, segments of the molecule comprising the Cyt *b*, $tRNA^{pro}$ and $tRNA^{thr}$ genes, and the ND6 and $tRNA^{glu}$ genes have been transposed relative to each other (DESJARDINS and MORAIS 1990). Limited protein-rRNA gene rearrangements and more extensive tRNA gene rearrangements have occurred in *D. yakuba* and sea urchin mtDNA molecules relative to vertebrate mtDNAs (CLARY and WOLSTENHOLME 1985a; JACOBS *et al.* 1988; CANTATORE *et al.* 1989). In contrast, in *C. elegans* and *A. suum* mtDNAs (Figure 2) there is a profound paucity of conservation of gene arrangements relative to those found in vertebrate, sea urchin and *D. yakuba* mtDNAs. The most notable exception is that the COI and COII genes are adjacent to each other and transcribed in the same direction in *C. elegans* and *A. suum* mtDNAs as they are in *D. yakuba* and vertebrate mtDNAs.

The AT (putative control) regions: The AT regions of the *C. elegans* and *A. suum* mtDNA molecules are 93.1% A+T and 84.7% A+T respectively, and in the AT region of each species there is a run of 18 AT dinucleotides. In addition, in the *A. suum* AT region, there are eight other runs of between four and seventeen AT dinucleotides (Figure 1).

Approximately half (55.4%) of the AT region of the *C. elegans* mtDNA molecule comprises six copies of a directly repeated sequence of 43 ntp (that does not include the 18 AT dinucleotide run). This repeated sequence is 100% A+T, and the first 29 ntp of each copy is a pair of 14 ntp inverted repeat sequences separated by a single ntp. The six, 43 ntp copies are strictly tandemly arranged except that two ntp (AT) separate copies five and six (CR5 and CR6, Figure 1). The first four copies (CR1–CR4) have 100% similarity to each other, but the fifth and sixth copies have one ntp and two ntp substitutions, respectively. Two inverted repeat sequence pairs that include C and G nucleotides occur in the *C. elegans* AT region outside of the direct repeat-containing region: a 9-ntp pair upstream from the repeat region, and a 14-ntp pair separated by 1 ntp immediately downstream from the direct repeat region, that has an 11/14 match to the inverted repeat within the 43-ntp direct repeats. Other inverted repeats of various sizes that lack G and C nucleotides also occur in the *C. elegans* AT region.

The *A. suum* AT region contains eight pairs of inverted repeat sequences that include G and C nucleotides. The two sequences of each repeat pair vary in length from 8 to 23 ntp, and are either immediately adjacent to each other or are separated by up to 50 ntp. There is an overlap between two of the repeat pairs, and one inverted repeat pair is located within the sequence that separates a second pair. A set of directly repeated sequences similar to that found in the *C. elegans* AT region does not occur in the *A. suum* AT region.

Both inverted and directly repeated sequences have been found in a large, apparently noncoding segment of a variety of metazoan mtDNA molecules, including those of the root knot nematode *Meloidogyne javanica* (see OKIMOTO *et al.* 1991 for references). In some cases this segment has been shown to be the control region (CLAYTON 1982, 1984). At this time we have no information beyond that discussed above relating to whether or not the AT regions of the *C. elegans* and *A. suum* mtDNA molecules contain sequences necessary for the initiation of transcription and/or of replication.

The sequence between the ND4 and COI genes: The second long, apparently non-coding region (LNC, Figure 1) in each of the *C. elegans* and *A. suum* mtDNA molecules, that lies between the ND4 and COI genes (109 and 117 ntp, respectively), contains a sequence that can fold into a hairpin structure (Figure 3). The loop in both of these potential hairpins has a run of T nucleotides: six in *A. suum* and four in *C. elegans*. A similar hairpin structure with a run of Ts in the loop, has been located in an apparently noncoding segment of mtDNAs from three *Drosophila* species, *Xenopus laevis* and a number of mammals (CHANG *et al.* 1985; ROE *et al.* 1985; CLARY and WOLSTENHOLME 1987; MONNEROT, SOLIGNAC and WOLSTENHOLME 1990). In mouse and human mtDNA, synthesis of the second (L) strand of the molecule has been shown to initiate within the run of Ts (acting as the template strand) of the loop (CHANG *et al.* 1985). These observations regarding *C. elegans* and *A. suum* mtDNA leave open the possibility that

the association of initiation of second strand synthesis of mtDNA within a T-rich loop of a hairpin structure may have been highly conserved throughout much or all of the evolution of the metazoa.

Protein genes: Nine of the mt-protein genes of *C. elegans* and *A. suum* were identified solely from similarities of their predicted amino acid sequences to those of mouse and *D. yakuba* mt-protein genes (Figure 4). The remaining three nematode mt-protein genes, ATPase6, ND6 and ND4L were identified from considerations of amino acid sequence similarities and hydropathic profile comparisons with mouse and *D. yakuba* ATPase6, ND6 and ND4L genes (Figures 4 and 5), and size (Table 1). Between the corresponding *C. elegans* and *A. suum* mt-protein genes, nucleotide sequence and predicted amino acid sequence similarities averaged 72.3% [range 64.1% (ND6) to 79.3% (COI)], and 73.7% [range 57.5% (ND2) to 88.2% (COI)], respectively. Eight of the homologous pairs of *C. elegans* and *A. suum* mt-protein genes have similar lengths, three pairs each differ by 3 ntp, and one pair (Cyt *b*) differ by 15 ntp. However, ten of the nematode mt-protein genes are considerably smaller (range 2.4% to 18.6%, Table 1) than the homologous genes of mouse and *D. yakuba*. The exceptions are the COI and COII genes that are both approximately 2% larger in *C. elegans* than in mouse (Table 1).

Translation initiation and termination codons: We have reported evidence obtained from alignment of the *C. elegans* and *A. suum* nucleotide sequences indicating that in six of the mt-protein genes of *A. suum* (COII, ND1, ND2, ND3, ND4 and ND6) and three of the mt-protein genes of *C. elegans* (Cyt *b*, ND2 and ND4), TTG is used as the translation initiation codon (OKIMOTO, MACFARLANE and WOLSTENHOLME 1990). Also, GTT seems to be the translation initiation codon of the *A. suum* COIII gene. The translation initiation codons of all of the five remaining *A. suum* mt-protein genes appear to be ATT, and the remaining nine *C. elegans* mt-protein genes seem to begin with either ATT or ATA. In many of the protein genes of mtDNAs from other metazoan phyla, triplets other than ATG are often used as translation initiation codons. Some or all ATN codons are used in this way among mammals, Drosophila and echinoderms (ANDERSON *et al.* 1981, 1982a,b; BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a; JACOBS *et al.* 1988; CANTATORE *et al.* 1989). GTG and GTT have been reported as rare mt-protein gene translation initiation codon in various metazoa (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a; JACOBS *et al.* 1988; GADALETA *et al.* 1989), and ATAA has been suggested as a translation initiation codon of the Drosophila COI gene (DE BRUIJN 1983; CLARY and WOLSTENHOLME 1983b). Only in *X. laevis* do all of the mt-protein genes begin with ATG (ROE *et al.* 1985). *C. elegans* and *A. suum* mt-protein genes are the first to be reported to use TTG as a putative translation initiation codon, and to be the only sets of protein genes among completely sequenced metazoan mtDNAs that totally lack ATG translation initiation codons.

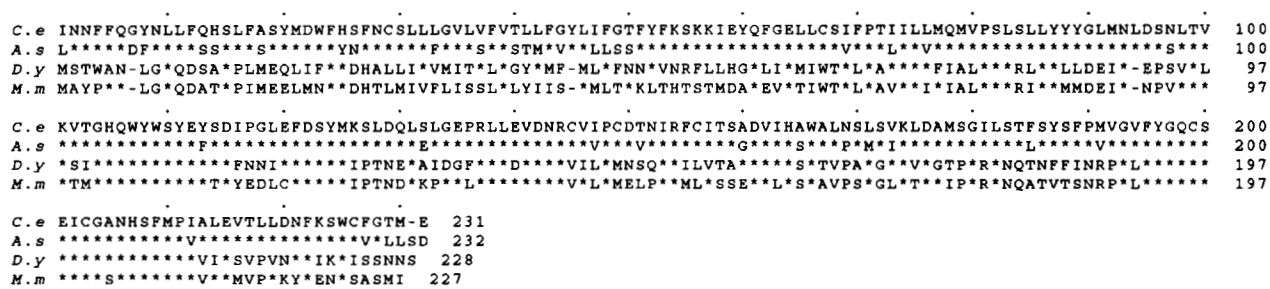
Nine of the *C. elegans* mt-protein genes and nine of the *A. suum* mt-protein genes end with a complete translation termination codon; TAA or TAG (Figure 1). Each of the remaining *A. suum* mt-protein genes ends with either a T (ND2 and ND5) or with TA (COI) which in each case is followed by a tRNA gene (Figure 1). The *C. elegans* COI gene also ends with TA, followed by a tRNA gene. TA occurs at the end of the *C. elegans* ND3 gene but this TA is immediately adjacent to the putative ATT initiation codon of the downstream ND5 gene. The *C. elegans* ND1 gene appears to terminate with a single T that is followed by the putative ATT translation initiation codon of the ATPase6 gene. Similar, apparent partial termination codons (T or TA) are found at the ends of some mt-protein genes from other metazoa (ANDERSON *et al.* 1981; CLARY and WOLSTENHOLME 1985a; CANTATORE *et al.* 1989). In mammals it is known that the primary transcription products of mtDNA are multicistronic RNAs, that individual gene transcripts are produced by precise cleavage, and that those protein gene transcripts ending in U acquire a complete termination codon by polyadenylation (OJALA, MONTOYA and ATTARDI 1981). The finding that some of the *C. elegans* and *A. suum* mt-protein genes end with T or TA is therefore consistent with the view that the nematode mitochondrial transcription mechanism also involves production of mature protein gene transcripts from primary, multicistronic transcripts, followed by polyadenylation.

It has been argued that in those metazoan mtDNAs (mammals, amphibia, Drosophila) in which tRNA genes are located between most pairs of protein genes, the secondary structure of the tRNA is important for the precise cleavage of the mature protein gene transcript from primary multicistronic transcripts (OJALA *et al.* 1980; OJALA, MONTOYA and ATTARDI 1981). Consistent with this view is the finding that in these mtDNAs the sequence adjacent to the 3′-terminus of a protein gene, followed immediately by another protein gene (rather than a tRNA gene), may have the potential to form a hairpin structure (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a). In *C. elegans* and *A. suum* mtDNAs there are three cases in which a protein gene is followed directly by another protein gene (Figures 1 and 2). For two of these, the ND3 gene followed by the ND5 gene and the ND6 gene followed by the ND4L gene, the sequences upstream from the translation termination codon-transcription initiation codon junctions have the potential to form

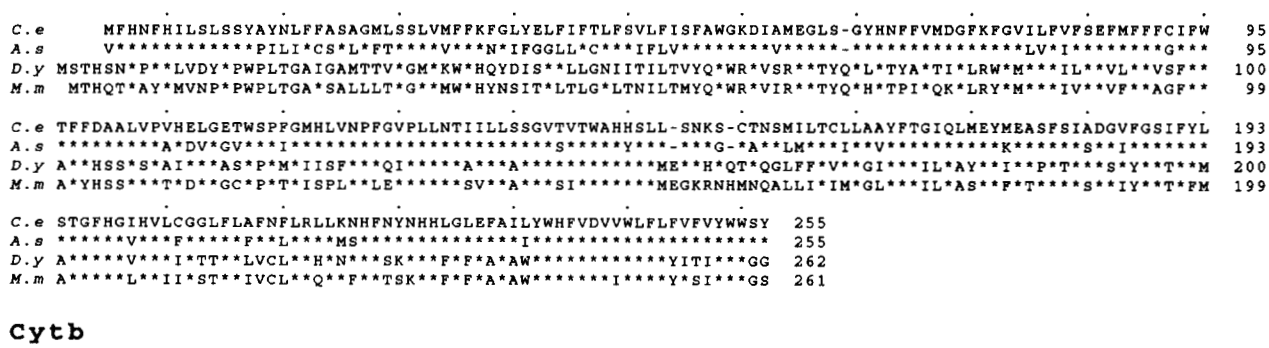
a

COI

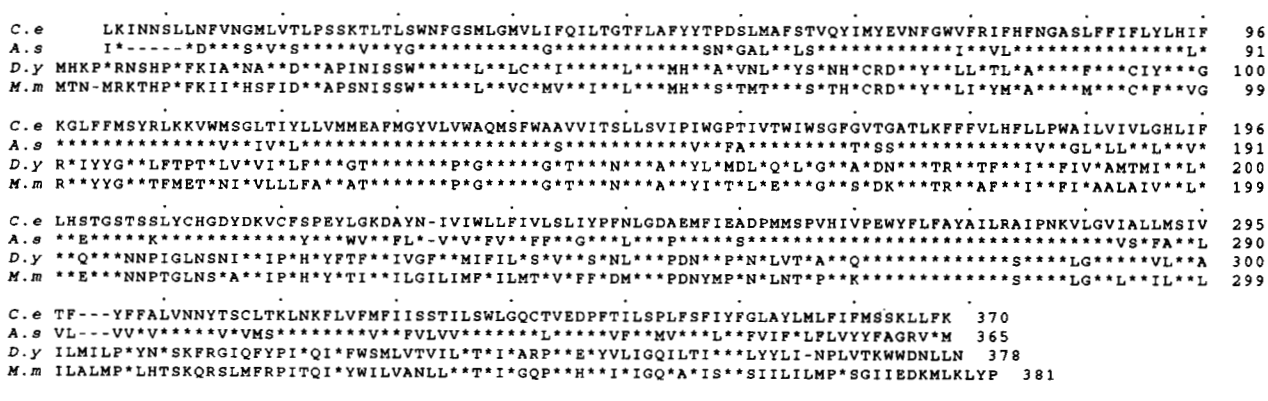
COII



COIII



Cytb



b

ATPase6

```
C.e             INQVYFLDIFMFVFVLQFLFYFKESMLNTLVKKFLNSLVGVFSYTNTLPL-------SSVISIFTFIVLLTCCFGGYFTYSFCPCGMVE   82
A.s             *TN**********Y*************GV**N***GL**V*****DS***-------*****V***L***********M********I*   82
D.y MMTNLFSVFDPSAIFNLSLNWLST*LGLLMIPSIYWLMPSRYNI*W**ILLTLHKEFKTL*GPSGHNG*TF*F*SL*SLI*FNN*M*L*P*I*TSTSHLT  100
M.m MNENLFASFITPTPMMG*PIVVAIIMFPSI**PSSKRLI*NRLHS*QHW**KLIIKQMM*IHTPKGRTWTLM*-VSLIMFIGSTNLL*LLPHT*T*TTQLS   99

C.e FTFVYAAVAWLSTLLT-FISSEKFSVYMSKPGDTYLKTLSML-LIEIVSEFSRPLALTVRLTVNITVGHLVSMMLYQGLELSMGDQY------IWLSILA  174
A.s ******M*********-**T*****I*I**A**SF***F***-*V*L***V**************VL***VI******L***YL*IF*------V*IVV**  174
D.y L*LSL*LPL**CFM*YGW*NHTQHMFAHLV*QG*PAILMPFMVC**TI*NII**GT*A****A*MIA***LLTL*GNTGP---SMS*LLVTFLLVAQ*AL  197
M.m MNLSM*IPL*AGAVI*G*RHKL*S*LAHFL*QG*PISLIP**II**TI*L*IQ*M**A****A***A***LMHLIGGATLVL*NISPPTATITFIILL*L  199

C.e IMMECFVFFIQSYIFSRLIFLYLNE  199
A.s *V*****************Y*****  199
D.y LVL*SA*TM****V*AV*ST**SS*VN  224
M.m TIL*FA*AL**A*V*TL*VS***HDNT  226
```

ND1

```
C.e   MILVLLMVILMMIFIVQSIAFITLYERHLLGSSQNRLGPTKVTFMGLAQALLDGVKLLKKEQMTPLNSSEVSFLLVPGISFVVMYLEWFTLPYFFDFI   98
A.s   L**M*VQ***I***VI*****V*********G**Q*I**N**S***FL**IF*******************I**I*****F*I*******V**F*Y**M   98
D.y MEF**S*IGSL*LI*CVLV*V**L**L**KV**YI*I*K**N**GL**IP*PFC*AI**FT***TY**L*NYL*YYIS*IF*LFLSLFV*MCM*F*VKLY  100
M.m  IN**T**VP**I------AM**L**V**KI**YM*L*K**NI*GPY*IL*PFA*AM**FM**P*R**TT*MSL*IIA*TL*LTLALSL*VP**MPHPL*   93

C.e SFEYSVLFFLCLIGFSVYTTLISGIVSKSKYGMIGAIRASSQSISYEIAFSLYVLCIIIHNNVFNFVSKFNLS----LLIIYIPFLIM----VIAELNRA  190
A.s T****I***************V**M********V*********V*********L*A*VM*I*M*C*F*F****----*F*V*L***F*----*L******  190
D.y **NLGG*****CTSLG***VMVA*WS*N*N*ALL*GL**VA*T****VSLA*IM*SF*FLIGSY*MIYF*YYQIYMWF***LF*MSLVWLTISL**T**T  200
M.m NLNLGI**I*ATSSL***SI*W**WA*N***SLF**L**VA*T****VTMAIIL*SVLLM*GSYSLQTLITTQEHMW**LPAW*MAM*WFISTL**T***  193

C.e PFDFSEGESELVSGFNVEPASVAFVLLFLSEYGSLIFFSVLSSAMFFKFSIFMAFS-------------IFSLLIFIRSSYPRYRYDLMMSLFWFKLLPI  277
A.s ****A*********Y***YS*********G***A*L***T*T*VL**G**YVVIYC-------------M*TI*V*V******F******YF*******V  277
D.y ****A**************YS*GG*A*I*MA**A*IL*M*M*FCVI*LGCDV*NLLFYVKLT-------F*SFVF*WA*GTL**F***KL*Y*A*KCF*SF  293
M.m ***LT**************Y*AGP*A*F*MA**TNI*LMNA*TTII*LGPLYYINLPELYSTNFMMEALLLS*TFLW**A****F***QL*H*L*KNF**L  293

C.e SLIMLCFYAVIFYY  291
A.s ***F*G-*F***LF  290
D.y **NY*L*FIGFKILLFSFLLWIFFSKKLMEN  324
M.m T*-A**MWHISLPIFTAGVPPYM  315
```

ND2

```
C.e    LIVFISLFTLFLTLLSIL---TNNVIVWWSI-FLLMTVVFILLNKSS-KSYTSIFNYFVIQESLGLLFLLCSGGLLQF-------------------   73
A.s    *LL*FCI*VV**CV*NFF---*S**L****V-********VC*S*G*-G**VG*L**********F**VFNVF****-------------------   73
D.y IFYNSSK*LFT*IMIIGTL*TVTSNSWLGA*MGLEIN*LSFIPL*SDNNNLM*TEASLK**LT*ALASTVL*FS*IL--LMLANNLNNEINESFTSMIIM   98
M.m MNPITLA*IY**I*-LGPV*TMSS**LMLM*VGLE*S*LAIIPM*I**KNPR*TEAATK***T*ATASMII**AIVLNYKQLGTWMFQQQTNGLILNMTL   99

C.e FIILLKIGVAPLHFWIFNVTNNIFNYGLMWFLTFQKLPFLTILLQIFWLSSVYILL--------FGLLICYVQIFVMKSYKNLLIISSTESFNWIVLGVF  165
A.s **VMM*V****F***V*S**GSLYDWL*************PV*V*L*DF*AFF*F*---------**ICV**F*L**L*G**SMMV*********V**TC*  165
D.y SAL***S*A**F***FP*MMEGLTWMNALMLM*W**IAP*---ML*SY*NIKNL**ISVILSVII*AIGGLN*TSLR*----*MAF**INHLG*MLSSLM  191
M.m MALSM*L*L**F***LPE**QG*PLHMGLIL**W**IAP*S**I**YP*LNST*I*MLAITSIFM*AWGGLN*TQMR*----IMAY**IAHMG*MLAILP  195

C.e FSMFNT-FYLFIYYFVLMVLLISK--------------FSKTSGYNFINWETTLVFLNIPFSVSFFVKIFSLSEIFKYDSFFTLFLLFTMFLSVLAFSFW  249
A.s L*VV*V-I***F**V***AF*MPN--------------*-NVKDF**V***VL*****V*******I***V***V**L*GL*L****LM****M*C**L*  248
D.y I*ESIWLI*FIF*S*LSF**TFMFNIFKLFHLNQLFSW*VNSKILK*SLFMNF*SLGGL*PFLG*LP*WLVIQQLTMCNQY*L*T*MMMST*IT*F*YLR  291
M.m YNPSLTLLNLM**IILTAPMFMALMLNNSMTINSISLLWN**PAMLTMISLML*SLGGL*PLTG*LP*WIIIT*LM*NNCLIMAT*MAM*A*LN*F*YIR  295

C.e LINLSMKNNEETSNNNKM------------------NYFIIFPLMVISII  282
A.s *V*M*V**MKMLGD*F*V------------------LF*LV**M**F*V*  281
D.y ICYSAFML*YFEN*WIMEMNMNSNN-TNLYLIMTFFSI*GL*LISLFFFML  341
M.m **YSTSLTMFP*N**S**MTHQTKTKPNLMFSTLAIMSTMTL**APQL*T  345
```

ND3

```
C.e  ILVLLMVLVFTLVLLFAFYLINFLLSIKDM-GKNKISAFECGFVSVGKIQNSFSIHFFIMMLMFVIFDLEIVMFLGILVSDLSSYISFLM----MFIFI   94
A.s  L***V**VL*******V**IG**V**C**F-Y*****S*******I*************************V***********N*L***F*----LLM**   94
D.y IFSIIIIAS*IL*ITTVVMF*-ASI**K*ALIDRE*S*P*****DPKSSSRLP**LR**LITII*L***V**ALI*PMIIILKY*NIMIWTITSII**L*   99
M.m IN*YTVIFINIL*S*T--LI*VA*W*PQMNL-YSE-ANPY****DPTSSARLP**MK**LVAIT*LL*****ALL*PLPWAIQTIKT*TM*IMAFILVT*   96

C.e LG-GFYMEWWYGKLVWVI  111
A.s F*-*************L*  111
D.y *LI*L*H**NQ*M*N*SN  117
M.m *SL*LAY**TQKG*E*TE  114
```

ND4

```
C.e     LLEFLFI-SLLWLFKPIYFLLFTVMFSFLIF------------NNFSWGGLFLVLDSYSFILLIV----MSLFILGIIVISEKNNN----LLILSE   75
A.s     **DI*LF-**YFF*E*VL*FF*M*V*G*VAL------------**Y**L*C*YFF**F*******----*******VVLL**S*FM----**L***   75
D.y MLKII*FL**LTPVCFINNMYWMVQIMLF*ISF**L-----LMNNFM*YW*EISY**GC*ML*YG*VLLSLWIC**ML*ASES-IN*Y**YKNLF*LNIV   95
M.m MLKII*PS*MLLP*T**SS*KKTWTNVTSY****SLTSLTLLWQTDE*YKNFSNM*-SS*PL*TP*I*LTAWLLP*MLMASQNHLK*D**VLQK*Y*SML   99

C.e I-LVFICIIFFIPSNMMMLYMFFELSMFPILVMILGYGSQIEKINSSYYLMFYAAFCSFPFLFVYF--------KSNFLLVFTYYNFVISWE---MFFIL  163
A.s V-**VV*VF**V***VILM**Y*****************************I****L*********--------**F*FISLV*FD*NL***---*V*V*  163
D.y *-*LLLLVLT*SSMSLF*F*L***S*LI*T*FL***W*Y*P*RLQAGV**L**TLLV*L*M*IGI*YVMNKTGSMNFY*MNNFMF*YD------LLY*C*  188
M.m *S*QILL*MT*SATELI*F*IL**ATLI*T*II*TRW*N*T*RL*AGI*FL**TLIG*I*L*IALILIQNHVGTLNLMI*S**THTLDA**SNNLLWLAC  199

C.e SLSFMMKFPIYFLHLWLPKAHVEAPTTASMLLAGLLLKLGTAGFLRILGSLSFVHNN----VWILIAFLGMILGSFCCVFQSDSKALAAYSSVTHMSFLL  259
A.s *********V*****************************************C*******----***VL*******A****M****A*********I*****V*  259
D.y LCA*LV*M*MFLV************VSG**I***IM****GY*L**VINF*QLMNLKY-SF***S*SLV*GV*M*LV*LR*T*L***I*****A**GIV*  287
M.m MMA*LI*M*L*GV************IAG**I**AI*****SY*MI**SII*DPLT-KYMAYPF**LSLW***MT*SI*LR*T*L*S*I*****S**ALVI  298
```

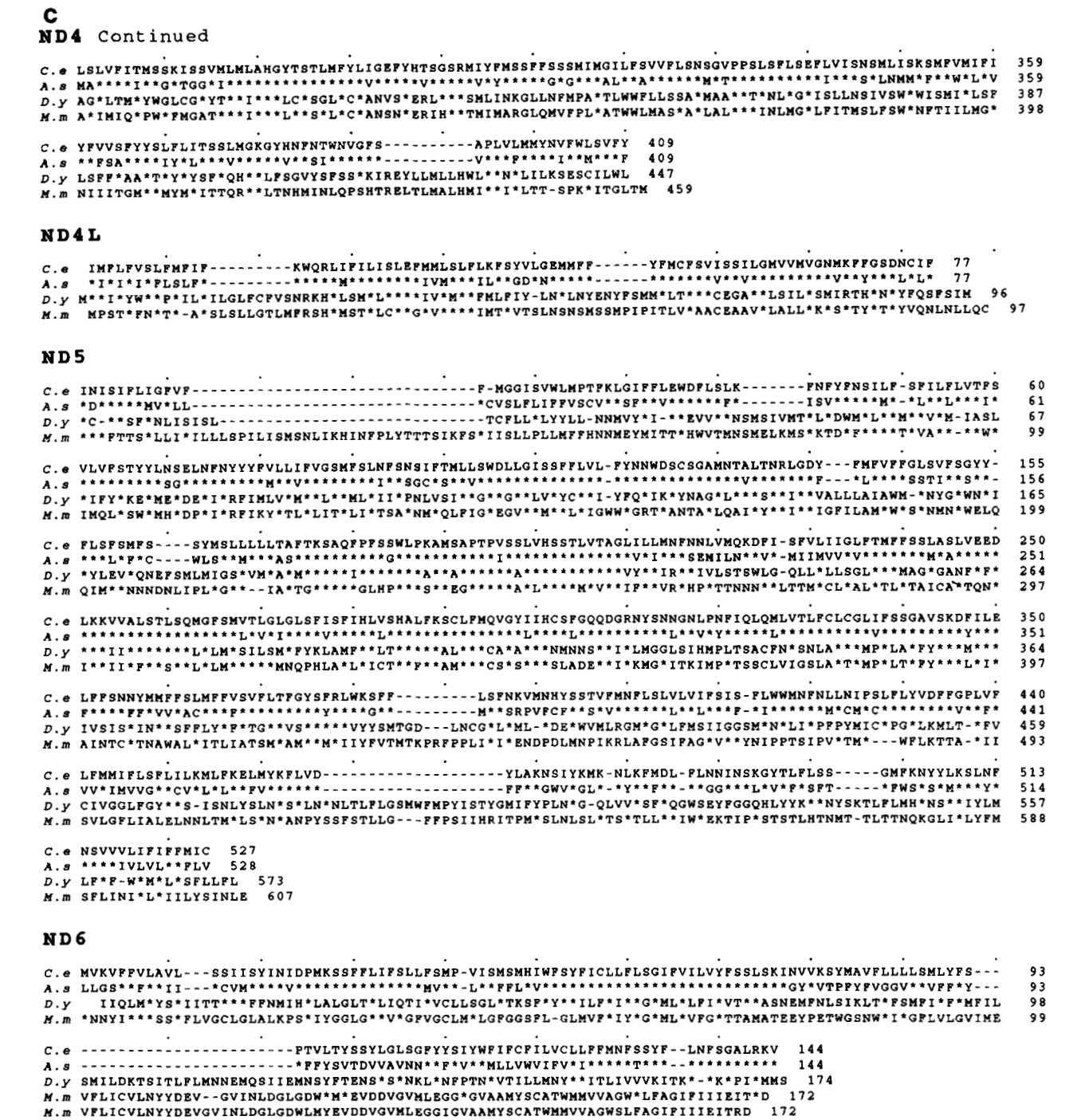


FIGURE 4.—Comparisons of the amino acid sequences of the 12 mtDNA-encoded proteins of *C. elegans* (*C.e*) with the amino acid sequences of the corresponding proteins of *A. suum* (*A.s*), *D. yakuba* (*D.y*; CLARY and WOLSTENHOLME 1985a) and mouse (*M.m*; BIBB *et al.* 1981). All amino acid sequences are the predictions of nucleotide sequences. An asterisk indicates an amino acid that is conserved relative to *C. elegans*. A dash indicates the absence in one sequence of an amino acid that occurs in one or more of the corresponding, other amino acid sequences. Genetic code modifications used for translation of mtDNA nucleotide sequences are as follows. ATA and TGA specify methionine and tryptophan, respectively, in all species; AGA and AGG specify serine in *C. elegans*, *A. suum*; AGA specifies serine in *D. yakuba* [AGA and AGG do not specify an amino acid in mouse and AGG does not specify an amino acid in *D. yakuba* (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a)].

a hairpin structure (Figure 6). However, in the remaining case, the ND1 gene followed by the ATPase6 gene, although a stem of 7 ntp and a loop of 4 nt can be formed by a sequence separated by 11 ntp from the 3′-terminus of the *C. elegans* ND1 gene, no equivalent or other stable secondary structure potential is found near the 3′-end of the *A. suum* ND1 gene. Also, although the large sequences that separate the *C. elegans* and *A. suum* ND4 and COI genes each include a hairpin-forming sequence, beginning 27 ntp and 19

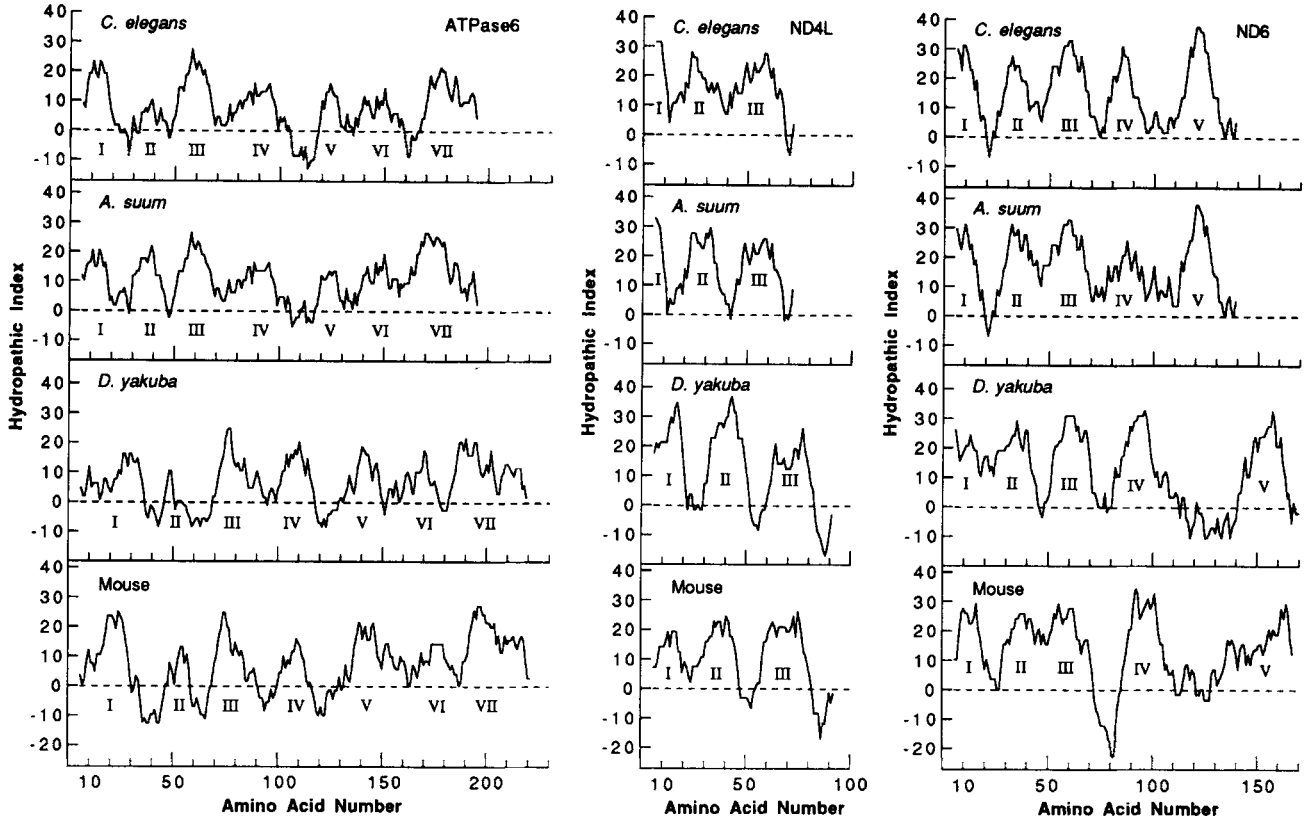


FIGURE 5.—Comparisons of the hydropathic profiles of the proteins predicted from the ATPase6, ND4L and ND6 genes of *C. elegans*, *A. suum*, *D. yakuba* and mouse. Each profile was calculated by the method of KYTE and DOOLITTLE (1982) using an 11 amino acid window. Hydrophobic regions have hydropathy of >0, and hydrophilic regions have hydropathy of <0. Corresponding hydrophobic domains in homologous proteins from the different species are indicated by Roman numerals.

TABLE 1

Amino acid and nucleotide sequence comparisons concerning the *C. elegans* and *A. suum* mt-protein genes

Gene	No. of amino acids				Percentage nucleotide sequence similarity[b] *C. elegans*/*A. suum*	Percentage amino acid sequence similarity[a,c]			
	C. elegans	*A. suum*	*D. yakuba*[a]	Mouse[a]		*C. elegans*/ *A. suum*	*C. elegans*/ *D. yakuba*	*C. elegans*/ mouse	*D. yakuba*/ mouse
COI	525	525	512	514	79.3	88.2	61.6	58.6	75.0
COII	231	232	228	227	76.7	83.2	39.0	39.4	56.6
COIII	255	255	262	261	74.5	79.6	46.6	42.9	64.5
Cyt *b*	370	365	378	381	68.8	72.4	43.1	41.7	67.5
ATPase6	199	199	224	226	74.3	79.4	22.5	22.5	34.4
ATPase8	NF[d]	NF	53	67	NF	NF	NF	NF	26.0
ND1	291	290	324	315	71.9	73.2	36.4	35.1	45.3
ND2	282	281	341	345	66.0	57.5	16.9	20.6	34.0
ND3	111	111	117	114	74.1	80.2	30.5	25.6	43.2
ND4	409	409	446	459	73.7	72.9	29.5	28.1	42.2
ND4L	77	77	96	97	74.4	70.1	26.8	21.4	35.7
ND5	528	528	573	607	70.1	69.1	31.3	25.4	39.7
ND6	145	144	174	172	64.1	58.3	18.6	12.6	17.0

[a] Data for *D. yakuba* and mouse are from CLARY and WOLSTENHOLME (1985a) and BIBB *et al.* (1981), respectively.
[b] These values are calculated from data in Figure 1, and include all deduced insertions/deletions.
[c] These values are calculated from Figure 4, and include all deduced insertions/deletions.
[d] NF = not found.

ntp, respectively, from the 5′-end (Figure 1), the 3′-ends of both of these intergenic sequences and the 3′-ends of the ND4 genes again lack stable secondary structure potential.

Ribosomal RNA genes: The *C. elegans* and *A. suum* mt-s-rRNA and mt-l-rRNA genes were identified from similarities to the corresponding genes in other metazoan mtDNAs and in prokaryotic organisms (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985b; NOLLER *et al.* 1986). In both nematode mtDNAs the s-rRNA and l-rRNA genes are separated from each other by protein genes, as is found in sea urchin

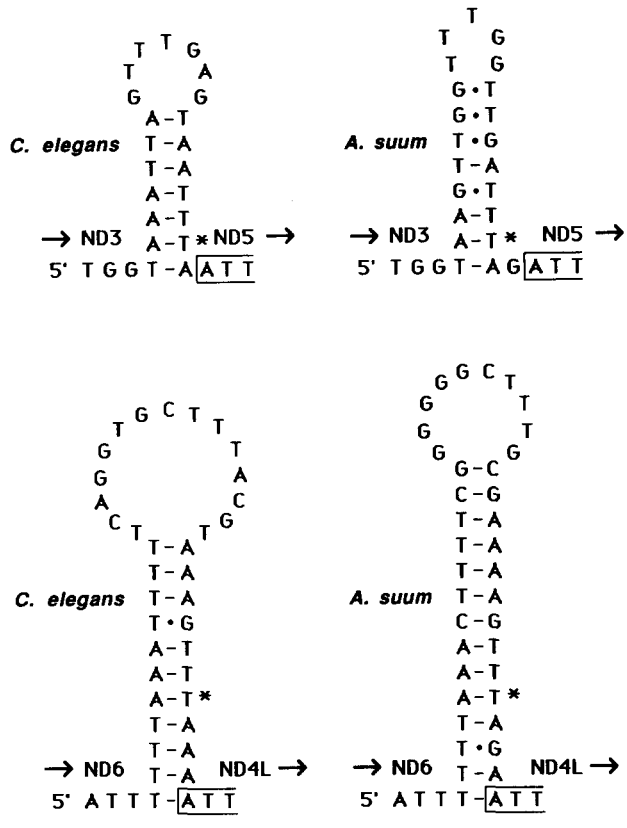


FIGURE 6.—Potential hairpin configurations of the nucleotides of the junction sequences of the ND3 and ND5 genes, and of the ND6 and ND4L genes of *C. elegans* and *A. suum*. Arrows indicate the direction of transcription. Translation initiation codons are boxed, and the first nucleotide of each translation termination codon or incomplete termination codon (TA, at the end of the ND3 gene) is identified with an asterisk. In transcripts, the predicted free energy changes [calculated using the free energy increments given by FREIER *et al.* (1986)], for the hairpin structures between the ND3 and ND5 gene are −1.7 kcal/mol (*C. elegans*) and −0.4 kcal/mol (*A. suum*) and for the hairpin structures between the ND6 and ND4L genes are −2.7 kcal/mol (*C. elegans*) and −11.3 kcal/mol (*A. suum*).

mtDNAs. In contrast, in vertebrate and Drosophila mtDNAs, the two rRNA genes are found adjacent to each other (BIBB *et al.* 1981; CLARY and WOLSTENHOLME 1985a). In both nematode mtDNAs, the $tRNA^{glu}$ and $tRNA^{ser}$(UCN) genes are located next to the 5′- and 3′-ends, respectively, of the s-rRNA, and the $tRNA^{his}$ and ND3 genes are located next to the 5′- and 3′-ends, respectively, of the l-rRNA gene. That the nucleotides immediately adjacent to the 3′-end of the $tRNA^{his}$ gene and the $tRNA^{glu}$ gene in *C. elegans* mtDNA are the 5′-terminal nucleotides of the s-rRNA and l-rRNA genes, respectively, was confirmed by primer extension analysis (R. OKIMOTO, J. L. MACFARLANE and D. R. WOLSTENHOLME, unpublished results).

The nucleotide immediately proceeding the $tRNA^{ser}$(UCN) gene in *C. elegans* mtDNA, and the corresponding nucleotide in *A. suum* mtDNA (separated from the $tRNA^{ser}$(UCN) gene by 16 ntp, Figure 1) are tentatively defined as the 3′-ends of the nematode s-rRNA genes. In both *C. elegans* and *A. suum* mtDNAs the nucleotide immediately preceding the translation initiation codon of the ND3 gene (ATA in *C. elegans* and TTG in *A. suum*) (Figure. 1)) is interpreted as the 3′-end of the l-rRNA gene. The *C. elegans* and *A. suum* mt-s-rRNA genes (697 and 701 ntp, respectively) and mt-l-rRNA genes (953 ntp and 960 ntp, respectively) as defined above are the smallest metazoan mt-rRNA genes so far recorded. Similarity between the two nematode mt-s-rRNA genes is 76.2% and between the two nematode mt-l-rRNA genes is 74.4%.

We have constructed a consensus secondary structure model for the entire *C. elegans* and *A. suum* mt-s-rRNA genes (R. OKIMOTO, J. L. MACFARLANE and D. R. WOLSTENHOLME, unpublished results). This model, which is based on a comparison of the *C. elegans* and *A. suum* rRNA gene sequences, secondary structures previously proposed for other metazoan mt-s-rRNA genes (GUTELL *et al.* 1985; ZWIEB, GLOTZ and BRIMACOMBE 1981), and the locations within the model of short conserved, primary sequences, has a remarkable resemblance to the core structure model for the s-rRNAs used by GRAY, SANKOFF and CEDERGREN (1984).

From similar primary and secondary structural considerations we have also built a consensus secondary structure model for the 3′, 63% of the predicted *C. elegans* and *A. suum* mt-l-rRNAs (R. OKIMOTO, J. L. MACFARLANE and D. R. WOLSTENHOLME, unpublished results). This model again fits well the previously proposed models for the corresponding regions of other metazoan mt-l-rRNAs (GLOTZ, ZWIEB and BRIMACOMBE 1981; CLARY and WOLSTENHOLME 1985b; GUTELL and FOX 1988). However, although the 5′, 37% of the *C. elegans* and *A. suum* mt-l-rRNA sequences share 70% similarity, they are approximately 270 and 445 ntp shorter than the corresponding regions of the predicted *D. yakuba* and mouse mt-l-rRNAs, respectively, and we have been unable to construct a satisfactory secondary structure model for this portion of the nematode mt-l-rRNAs. In view of the unusual structure of the *C. elegans* and *A. suum* mt-tRNAs, it is interesting to note that, with one exception, the proposed secondary structures for the mt-s-rRNAs and the 3′, 63% of the mt-l-rRNAs of these organisms include all of the secondary structure element-forming sequences that in *Escherichia coli* rRNAs contain nucleotides important for interactions with tRNAs (NOLLER *et al.* 1986; DAHLBERG 1989; MOAZED and NOLLER 1989). The exception is the A2169-containing element that is also absent from secondary structure models of other metazoan mt-l-rRNAs.

Transfer RNA genes: In 20 of the mt-tRNA genes of *C. elegans* and *A. suum* the variable loop and TΨC arm are together replaced with a loop of between 6

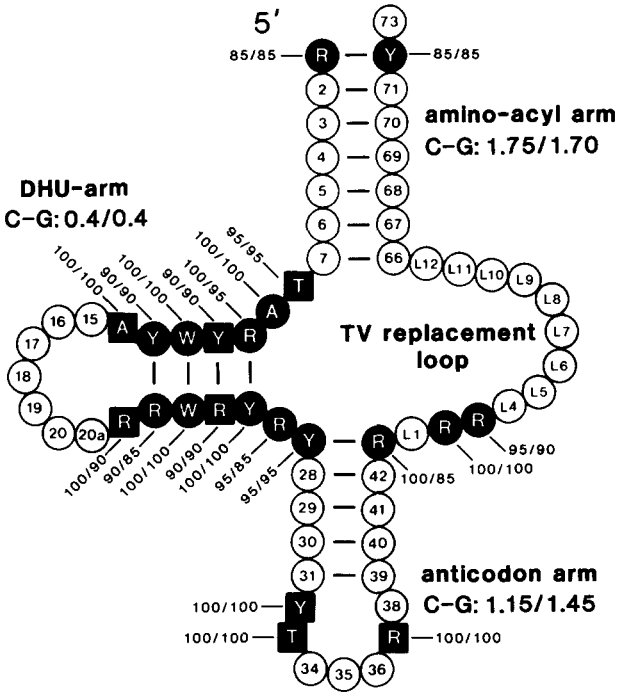


FIGURE 7.—The consensus secondary structure of the TV (TΨC arm-variable loop)-replacement loop-containing *C. elegans* and *A. suum* mt-tRNA genes. The number of nucleotides shown in the DHU (dihydrouridine)-loop and the TV-replacement loop are the maximum numbers observed. Letters in solid squares identify nucleotides or nucleotide combinations that occur among *C. elegans*/*A. suum* mt-tRNA genes with frequencies (as percentages) shown by the accompanying numbers and are also constant nucleotides in prokaryotic and eukaryotic nuclear-encoded tRNAs. Letters in solid circles identify other nucleotide or nucleotide combinations that occur in *C. elegans*/*A. suum* mt-tRNA genes with frequencies (as percentages) shown by the accompanying numbers. Ten of these nucleotides (1, 9, 10, 13, 22, 25, 26, 27, 43 and 72) are considered semi-invariant in prokaryotic and eukaryotic nuclear-encoded tRNAs (DIRHEIMER *et al.* 1979; SINGHAL and FALLIS (1979). The numbering system used (1–43 and 66–73, given in the open circles) reflects the conventional numbering system used for yeast $tRNA^{phe}$ (SPRINZL *et al.* 1989). L1–L12 denote the maximum of 12 nt that occur in the TV-replacement loop of the *C. elegans* and *A. suum* mt-tRNA genes (Figure 2). A, adenine; T, thymine; R, adenine or guanine; Y, cytosine or thymine; W, adenine or thymine. The mean number of C-G bonds/stem for each of the three stems of the *C. elegans* and *A. suum* mt-tRNA genes are shown.

and 12 nt (designated the TV-replacement loop; WOLSTENHOLME *et al.* 1987; OKIMOTO and WOLSTENHOLME 1990). The secondary structure potentials of these tRNA genes, which are between 54 and 63 ntp in *C. elegans*, and between 54 and 62 ntp in *A. suum*, are summarized in Figure 7. Each of the tRNAs predicted from these genes has an aminoacyl stem of 7 ntp, a DHU stem of 4 ntp (3 ntp in *A. suum*, $tRNA^{asn}$), a DHU loop of between 5 and 8 nt in *C. elegans* and 5 and 9 nt in *A. suum*, an anticodon stem of 5 ntp and an anticodon loop of 7 nt. In each nematode mtDNA there are two further sequences that have been interpreted as tRNA genes for ser(AGN) and ser(UCN). The predicted secondary structures of each of these sequences contains a variable loop (always 4 nt) and a TΨC arm (stem of 3 or 5 ntp; loop of 3–6 nt), but the DHU arm is replaced with a loop of between 5 nt and 8 nt. Nucleotide sequence similarity between the corresponding putative tRNA genes of *C. elegans* and *A. suum* are in the range 58% ($tRNA^{asp}$) to 91% [$tRNA^{leu}$(UUR)], with a mean of 75%.

The following considerations (discussed previously in detail for *A. suum*, see WOLSTENHOLME *et al.* 1987) supported our interpretation that the structures mentioned above are the complete set of tRNA genes of each of the *C. elegans* and *A. suum* mt-genomes. The number of putative tRNA genes found in each molecule, 22, is the same as that found in vertebrate and higher invertebrate (*D. yakuba* and sea urchin) mtDNAs. The trinucleotides in the anticodon positions of the *C. elegans* and *A. suum* tRNA genes are unique, are the same for each species, are compatible with codon usage in the *C. elegans* and *A. suum* mtDNAs (Table 2), and have a high correspondence to anticodons of vertebrate and higher invertebrate mt-tRNA genes. Eight nucleotide or nucleotide combinations, considered invariant in prokaryotic and nuclear-encoded eukaryotic tRNAs (standard tRNAs) (SPRINTZL *et al.* 1989; DIRHEIMER *et al.* 1979; SINGHAL and FALLIS 1979), are conserved with a frequency of between 90 and 100% in the *C. elegans* and *A. suum* mt-tRNA genes (Figure 7). Ten other nucleotides or nucleotide combinations that are considered semi-invariant in standard tRNAs occur in the *C. elegans* and *A. suum* mt-tRNA genes with a frequency of between 85 and 100%. In addition, of the 20 *C. elegans* and 20 *A. suum* mt-tRNA genes that contain a TV-replacement loop, all have an A-T pair or a T-A pair in the DHU arm at position 12–23, and a purine occurs at positions L2 and L3 in the TV-replacement loop in 100%, and 95% (*C. elegans*) or 90% and 90% (*A. suum*), respectively, of these genes. Eighteen of the *C. elegans* and *A. suum* mt-tRNA genes contain the $pyrimidine_{11}$-$purine_{24}$ pair which is highly conserved in standard tRNAs. However, in the $mt\text{-}tRNA^{f\text{-}met}$ and $mt\text{-}tRNA^{trp}$ genes of each nematode, a G-C pair is found in this position. A G-C pair is also found in place of a $pyrimidine_{11}$-$purine_{24}$ pair in the $mt\text{-}tRNA^{f\text{-}met}$ and $mt\text{-}tRNA^{trp}$ genes of other metazoa.

Each of the two DHU-replacement loop-containing mt-tRNA genes from *C. elegans* and *A. suum* contain the conserved R_1, Y_{86}, Y_{27}, R_{43}, Y_{32}, T_{33} and R_{37} nucleotides. In three of these structures (*C. elegans* $tRNA^{ser}$(AGN) and $tRNA^{ser}$(UCN) and *A. suum* $tRNA^{ser}$(UCN) there are possible secondary interactions between nucleotides in the DHU-replacement loop and the variable loop.

The frequencies of C nucleotides in the TV-replacement loop-containing mt-tRNA genes of *C. elegans*

TABLE 2

Codon usage in the 12 mt-protein genes of *C. elegans* and *A. suum*

		C.e.	*A.s*			*C.e*	*A.s*			*C.e*	*A.s*			*C.e*	*A.s*
Phe	TTT	429	484	Ser	TCT	84	144	Tyr	TAT	144	154	Cys	TGT	45	60
(GAA)	TTC	23	9	(UGA)	TCC	5	3	(GUA)	TAC	29	16	(GCA)	TGC	2	0
Leu	TTA	356	142		TCA	51	6	TER	TAA	8	2	Trp	TGA	62	27
(UAA)	TTG	81	304		TCG	5	9		TAG	1	7	(UCA)	TGG	9	48
Leu	CTT	51	57	Pro	CCT	27	72	His	CAT	45	53	Arg	CGT	29	30
(UAG)	CTC	1	0	(UGG)	CCC	7	0	(GUG)	CAC	13	1	(ACG)	CGC	0	0
	CTA	36	4		CCA	39	1	Gln	CAA	34	17		CGA	1	1
	CTG	5	6		CCG	8	10	(UUG)	CAG	10	26		CGG	1	2
Ile	ATT	257	200	Thr	ACT	62	86	Asn	AAT	139	107	Ser	AGT	61	127
(GAU)	ATC	23	10	(UGU)	ACC	7	7	(GUU)	AAC	12	5	(UCU)	AGC	6	4
Met	ATA	134	44		ACA	69	8	Lys	AAA	95	29		AGA	126	29
(CAU)	ATG	44	131		ACG	6	6	(UUU)	AAG	14	65		AGG	39	34
Val	GTT	118	263	Ala	GCT	62	92	Asp	GAT	55	70	Gly	GGT	93	144
(UAC)	GTC	9	9	(UGC)	GCC	9	5	(GUC)	GAC	8	6	(UCC)	GGC	5	6
	GTA	98	37		GCA	39	5	Glu	GAA	52	25		GGA	60	18
	GTG	28	67		GCG	4	2	(UCC)	GAG	26	50		GGG	29	39

The numbers of occurrences of each codon in the 12 mt-protein genes of *C. elegans* (*C.e*; total 3,430 codons) and *A. suum* (*A.s*; total 3,425 codons) are given. Assumed modifications relative to the standard genetic code are, AGA and AGG specify serine; ATA species methionine and TGA specifies tryptophan (see text). The anticodon corresponding to each two- and four-codon family is shown in parentheses.

TABLE 3

Nucleotide composition data for the mt-protein genes of *C. elegans*, *A. suum*, *D. yakuba*, P. *lividus*, X. *laevis*, mouse, cow and human

	Percentage nucleotide composition of the sense strand						Percentage of codons ending in:						Percentage of codons that specify leucine	Percentage of leucine codons beginning with	
	T	C	A	G	A + T	G + T	T	C	A	G	A + T	G + T		T	C
C. elegans	46.0	9.3	29.5	15.3	75.5	61.3	49.7	4.7	36.6	9.0	86.3	58.7	15.5	82.5	17.5
A. suum	51.9	8.0	18.5	21.5	70.4	73.4	62.7	2.4	11.5	23.4	74.2	86.1	15.0	86.9	13.1
D. yakuba	44.4	11.1	32.3	12.2	76.7	56.6	48.4	3.3	45.4	2.9	93.8	51.3	16.8	90.6	9.4
P. lividus	31.0	23.6	28.6	16.8	59.6	47.8	25.5	22.5	39.9	12.1	65.4	37.6	15.4	28.7	71.3
X. laevis	31.7	24.5	31.1	12.7	62.8	44.4	29.3	23.2	43.5	4.0	72.8	33.4	16.1	37.8	62.2
Mouse	29.4	26.0	33.2	11.4	62.6	40.8	22.9	27.4	46.6	3.2	69.5	26.1	15.5	21.9	78.1
Cow	28.0	27.7	31.9	12.4	59.9	40.4	19.8	31.9	43.7	4.7	63.5	24.5	16.0	19.0	81.0
Human	25.6	33.1	29.3	12.0	54.9	37.6	15.3	43.1	36.5	5.1	51.8	20.4	17.3	12.0	88.0

Data for *P. lividus* (CANTATORE *et al.* 1989), *X. laevis* (ROE *et al.* 1985), mouse (BIBB *et al.* 1981), cow (ANDERSON *et al.* 1982b) and human (ANDERSON *et al.* 1981) mtDNAs are derived from all protein genes except the ND6 genes. The ND6 gene in each of these mtDNAs is the only gene transcribed from the light (L) strand, and both the nucleotide composition of the sense strand and the frequencies of nucleotides in the third positions of codons are complementary to those of the other 12 mt-protein genes of each species. The *D. yakuba* (CLARY and WOLSTENHOLME 1985a), *C. elegans* and *A. suum* data are derived from all protein genes. In *D. yakuba* mtDNA nine protein genes are transcribed from one stand and four from the other, but there is no clear distinction between the nucleotide compositions of the sense strands of genes transcribed from the two complementary strands. In *C. elegans* and *A. suum* mtDNAs all protein genes are transcribed from the same strand (Figure 1). Nucleotides concerned with termination are excluded.

and *A. suum* are 8.6 and 7.9%, respectively, close to the frequencies of this nucleotide in the *C. elegans* and *A. suum* mt-protein genes: 9.3 and 8.0%, Table 3. However, in spite of this low use of C in tRNA genes, there are between one and three C-G nucleotide pairs in 100% (*C. elegans*) and 95% (*A. suum*) of the amino-acyl stems, and in 85% (*C. elegans*) and 90% (*A. suum*) of the anticodon stems. Also 35% of the DHU stems of these tRNA genes in both *C. elegans* and *A. suum* contain one or two C-G pairs. The differences in average frequencies of C-G pairs in the amino-acyl, anticodon, and DHU stems of *C. elegans* and *A. suum* mt-tRNA genes, 1.7 and 1.8/stem, 1.2 and 1.5/stem and 0.4 and 0.4/stem, respectively (Figure 7), suggest that C-G pairs contribute to differential stem stability that may be important to some aspect of tRNA function.

Ten of the 20 TV-replacement loop-containing tRNA genes of *A. suum* contain a single mismatched pair of nucleotides in the amino-acyl stem and one

(phe) contains two mismatched nucleotide pairs in this stem. Nine of the 12 mismatches are T T pairs at the base of the stem (position 7). Of the remaining three mismatches A G is found at position 7 of the $tRNA^{leu}$(UUR) gene and G A and T C occur at position 6 of the $tRNA^{phe}$ and $tRNA^{trp}$ genes, respectively. The rather constant location of these mismatches suggests that they may have some functional significance. However, although eleven of the 20 TV-replacement loop-containing tRNA genes of *C. elegans* also have a mismatched pair in position 6 or 7 of the amino-acyl stem, only eight of these *C. elegans* tRNA genes are the homologues of the mismatch pair-containing tRNA genes in *A. suum*, and only in three are both nucleotide and position specificity conserved.

Only two *A. suum* and one *C. elegans* tRNA genes (ala and f-met, and asn, respectively), contain a mismatch in the anticodon stem, and two *A. suum* tRNA genes (ala and asp) contain a mismatch in the DHU stem.

Data from Northern blot hybridization experiments has provided direct evidence that both the TV-replacement loop-containing tRNA genes and the DHU-replacement loop-containing tRNA genes of *C. elegans* and *A. suum* are transcribed (OKIMOTO and WOLSTENHOLME 1990). Further, these transcripts were shown to be the same size as the respective tRNA gene, to which three nucleotides (presumably CCA), are added following transcription. The only exception among those tested was *C. elegans* mt-$tRNA^{asn}$, most molecules of which had one nucleotide (plus CCA) more than predicted from the gene. These results fully support the view that the unusual mt-tRNA genes of nematode worms encode tRNAs that are active in mt-protein synthesis.

Codon usage and genetic code modifications: Codon usages among the 12 mt-protein genes of *C. elegans* and *A. suum* are given in Table 2. Comparisons of nucleotide compositions of the sense strand, and of nucleotides in the third positions of codons of mt-protein genes of *C. elegans*, *A. suum* and other metazoan mtDNAs are given in Table 3. In the *C. elegans* mt-protein genes all codons are used except CGC (Arg). This codon, together with three others [all also ending in C: CTC (Leu); CCC (Pro); and TGC (Cys)] are not found in *A. suum* mt-protein genes.

Both *C. elegans* and *A. suum* mtDNAs, like *D. yakuba* mtDNA are A+T rich; *C. elegans* 75.5%; *A. suum* 70.4%. Other metazoan mtDNAs have A+T contents in the range 54.9% to 62.8% (Table 3). Between *C. elegans* and *A. suum* mtDNAs both the nucleotide composition of the sense strand and the nucleotides in the third position of codons differ in that A is more frequent in *C. elegans* than in *A. suum*, and T and G are more frequent in *A. suum* than in *C. elegans*. The frequency of G in the sense strand (21.5%) and in the third position of codons (23.4%) of *A. suum* mtDNA is much higher than in other completely sequenced metazoan mtDNAs. Only for three sequenced protein genes of the platyhelminth, *Fasciola hepatica* have higher corresponding values (26.7 and 27.8%) been reported (GAREY and WOLSTENHOLME 1989).

TABLE 4

Correspondence in position between ATA and ATG codons, and between these codons and ATT, ATC and other codons in the 12 mt-protein genes of *C. elegans* and *A. suum*

Codon correspondence	No. of occurrences	Percentage of total *C. elegan*		Percentage of total *A. suum*	
C. elegans A. suum		ATAs	ATGs	ATAs	ATGs
ATA ↔ ATA	21	15.7		48.0	
ATA ↔ ATG	62	46.3			47.3
ATG ↔ ATG	24		54.5		18.3
ATG ↔ ATA	7		15.9	15.9	
ATT/C ↔ ATA	6			13.6	
Other ↔ ATA	10			22.7	
ATT/C ↔ ATG	9				6.9
Other ↔ ATG	36				27.5
ATA ↔ ATT/C	6	4.5			
ATA ↔ Other	45	33.6			
ATG ↔ ATT/C	3		6.8		
ATG ↔ Other	10		22.7		

ATA specifies methionine: The differential use of A and G in the third position of codons of *C. elegans* and *A. suum* mt-protein genes have provided the basis for arguments favoring the interpretation that in the nematode mt-genetic code ATA specifies methionine (as in most other metazoan mt-genetic codes) rather than isoleucine. The number of ATA and ATG codons in the 12 mt-protein genes of *C. elegans* and *A. suum* are 134 and 44, and 44 and 131, respectively. Therefore, if both ATA and ATG specify methionine then the percentage of methionine would be approximately equal in the *C. elegans* and *A. suum* mt-proteins (5.2 and 5.1%), and these percentages would be similar to those found in *D. yakuba* (5.7%) and vertebrate mt-proteins (5.2–6.9%). However, if ATA specifies isoleucine rather than methionine, then there would be 3.0 times less methionine in the *C. elegans* than in the *A. suum* mt-proteins, and the methionine content of these proteins would be only 1.3 and 3.8%, respectively.

Further data supporting the view that ATA specifies methionine rather than isoleucine in the nematode mt-genetic code was obtained by examining the correspondence in position between ATA and ATG codons, and between each of these codons and ATT, ATC and other codons in the *C. elegans* and *A. suum* mt-protein genes (Table 4). Of the 134 ATA codons in *C. elegans* mt-protein genes only 21 (16%) ATAs are found at corresponding position in the *A. suum* genes, but 62 (46%) are replaced with ATG. Although 51 (38%) of the ATA codons in *C. elegans* mt-protein

TABLE 5

Data concerning the specificity of codons in mouse and *D. yakuba* mt-protein genes that correspond to TGA and TGG codons in *C. elegans* and *A. suum* mt-protein genes

	Total TGA codons	Corresponding codons in mouse mtDNA specifying:		Corresponding codons in *D. yakuba* mtDNA specifying:	
		Tryptophan	Other amino acid[a]	Tryptophan	Other amino acid[a]
C. elegans	62	43 (69.4%)	19 [F(5), P(3), 1(3), T(2), R(2), L, N, V, K]	43 (69.4%)	19[F(5), V(2), I(2), N(2), R(2), L, M, Y, E.S, d]
A. suum	27	19 (70.4%)	8[P(2),F,I,L,R,T,N]	20 (74.1%)	7[F(2),L(2),R,I,M]
	Total TGG codons				
C. elegans	9	5 (55.6%)	4[Y,M,L,N]	4 (44.4%)	5[I,P,Y,L,V]
A. suum	48	28 (58.3%)	20[F(4), I(4), T(2), S(2), L(2), V(2), H, K, P, R]	29 (60.4%)	19[M(3), N(3), V(2), F(2), I(2), L(2), T, E, S, R, d]

[a] Specification of amino acids given in the one letter code. d indicates the absence in *D. yakuba* mtDNA of a codon corresponding in position to a TGA or TGG codon in a nematode mtDNA.

genes are replaced in the *A. suum* genes with codons other than ATG, isoleucine-specifying ATT and ATC codons account for only 6 (5%) of these. In comparison, of the 13 (30%) ATG codons in *C. elegans* mt-protein genes that are replaced in *A. suum* with codons other than ATA, 3 (7%) are replaced with ATT or ATC. Of the 131 ATG codons in the *A. suum* mt-protein genes, only 24 (18%) ATGs are found at corresponding positions in the *C. elegans* genes, but, as noted above, 62 (47%) are replaced with ATA. Of the 16 (36%) ATA codons in the *A. suum* mt-protein genes that are replaced in *C. elegans* with codons other than ATG, only 6 (14%) of these codons are ATT or ATC. This compares to the finding that of the 45 (34%) ATG codons in *A. suum* that are replaced in *C. elegans* with codons other than ATA, 9 (7%) are replaced with ATT or ATC codons.

Taken together these data strongly support the argument that in the nematode mt-genetic code, ATA specifies the same amino acid as ATG, presumably methionine.

TGA specifies tryptophan: The codon TGA, that specifies translation termination in the standard genetic code, is found internally in all of the nematode mt-protein genes except the *C. elegans* ND6 gene and the *A. suum* ND6 and ND3 genes. TGG is found in all *A. suum* mt-protein genes but only in the *C. elegans* COI, COII, ND1, ND5 and ND6 genes. The following considerations support the view that, as in other metazoan, protozoan and fungal mtDNAs examined to date, TGA, together with TGG, specify tryptophan in the nematode mt-genetic code.

Of the 62 and 27 TGA codons among the *C. elegans* and *A. suum* mt-protein genes, 43 (69%) and 19 (70%), correspond in position to tryptophan-specifying codons in mouse mt-protein genes, and 43 (69%) and 20 (74%) correspond in position to tryptophan-specifying codons in *D. yakuba* mt-protein genes (Table 5). [These data compare favorably with data from such comparisons for TGG codons in *C. elegans* and *A. suum* mt-protein genes (Table 5), in regard to supporting the interpretation of TGA as a tryptophan-specifying codon.] The frequencies of correspondence in position of TGA codons in nematode mt-protein genes, and tryptophan-specifying codons in the homologous mouse and *D. yakuba* mt-protein genes are even higher for the three overall most conserved genes, COI, COII and COIII. Twenty-two (92%) of the 24 TGA codons in the *C. elegans* COI, COII and COIII genes and all of the 17 TGA codons in these *A. suum* genes correspond in position to tryptophan-specifying codons in the homologous genes of both mouse and *D. yakuba*.

Of the 71 and 75 TGR codons that occur in the mt-protein genes of *C. elegans* and *A. suum*, (Table 2), 68 occur at corresponding positions in the two sets of genes, and TGA codons in the *C. elegans* mt-protein genes replace 79% of all TGG codons in the *A. suum* mt-protein genes (Figure 1). Also, if TGG alone specified tryptophan in the nematode mt-genetic code then the *A. suum* mt-protein genes would contain 5.3 times more tryptophan (48:9) than the *C. elegans* mt-protein genes, rather than 1.1 times more (75:71) if both TGA and TGG specify tryptophan (Table 2).

AGA and AGG specify serine: The codons AGA and AGG, which specify arginine in the standard genetic code, do not occur internally in any of the vertebrate mt-protein genes sequenced so far. It has been argued that in some vertebrate mtDNAs, AGA and AGG are used as rare translation termination codons (ANDERSON *et al.* 1981, 1982b; ROE *et al.* 1985). In contrast, substantial evidence has been presented favoring the conclusion that AGA specifies serine in the Drosophila mt-genetic code (CLARY and WOLSTENHOLME 1983a,b; DE BRUIJN 1983; WOLSTENHOLME and CLARY 1985). AGG codons do not occur in mt-protein genes of *D. yakuba* and *D. melanogaster* (CLARY and WOLSTENHOLME 1983a, 1985a; DE

BRUIJN 1983; GARESSE 1988). Also, it appears that in the mt-genetic codes of echinoderms and platyhelminthes both AGA and AGG specify serine (HIMENO *et al.* 1987; JACOBS *et al.* 1988; CANTATORE *et al.* 1989; GAREY and WOLSTENHOLME 1989). The interpretation of AGA and AGG as serine-specifying codons in the nematode mt-genetic code (mentioned in WOLSTENHOLME *et al.* 1987) is based on the following arguments.

AGA and AGG codons occur internally in all *C. elegans* and *A. suum* mt-protein genes with the exceptions that AGA codons are absent from the *A. suum* COII and ND6 genes, and AGG codons are absent from the *C. elegans* ND4L gene. Of the 42 AGA and 12 AGG codons in the *C. elegans* COI, COIII and Cyt *b* genes (the three most highly conserved mt-protein genes between *C. elegans* and mouse, and *C. elegans* and *D. yakuba*), none correspond in position to arginine-specifying codons in the homologous mouse and *D. yakuba* genes. Rather, 17 (41%) and 21 (50%) of the *C. elegans* 42 AGA codons correspond to serine-specifying codons in the mouse and *D. yakuba* genes, respectively. The remaining *C. elegans* AGA codons correspond to eleven other codons in the mouse genes and to twelve other codons and a deletion in the *D. yakuba* genes. The most prevalent of these other codons are those that specify alanine (five) in the mouse genes, and alanine and valine (four each) in the *D. yakuba* genes. Four of the twelve AGG codons in the *C. elegans* COI, COIII and Cyt *b* genes correspond to serine-specifying codons in both the homologous mouse and the *D. yakuba* genes. The remaining eight *C. elegans* AGG codons correspond to codons specifying five other amino acids in the mouse and the *D. yakuba* genes, the most prevalent of which are glycine and valine (two each) and glycine, valine and proline (two each), respectively.

Further evidence that AGA and AGG specify serine in the *C. elegans* and *A. suum* mt-genetic code was obtained by examining the frequencies of correspondence of AGA and AGG codons in the 12 *C. elegans* mt-protein genes with AGN, TCN, and other codons in the homologous *A. suum* genes (Table 6). Of 126 AGA codons in the *C. elegans* genes, only 18 (14%) and 16 (13%) are AGA and AGG, respectively, in the *A. suum* genes. However, 56 (44%) of these *C. elegans* AGAs are AGT (54) and AGC (2) in the *A. suum* genes. Also, of the 39 AGG codons in the *C. elegans* genes, only four (10%) and six (15%) are AGA and AGG, respectively, but 17 (43%) are AGY codons (all AGT) in the *A. suum* genes. The substitution of AGT in *A. suum* for AGA in *C. elegans* is in line with the general prevalence in the third position of codons of A nucleotides in the *C. elegans* genes and of T nucleotides in the *A. suum* genes (Table 3). Also, a high frequency of third position A ↔ T substitution is found for the TCN family of serine-specifying codons. Of the 56 TCR codons (51 TCA, 5 TCG) in the *C. elegans* genes, 40 are TCY codons (39 TCT, 1 TCC) in the *A. suum* genes.

TABLE 6

Correspondence in position of AGA and AGG codons in the 12 *C. elegans* mt-protein genes with the different AGN, TCN and other codons in the 12 homologous *A. suum* mt-protein genes

Codon correspondence	No. of occurrences	Percentage of total *C. elegans*:	
C. elegans A. suum		AGAs	AGGs
AGA ↔ AGA	18	14.3	
AGG	16	12.7	
AGT	54	42.9	
AGC	2	1.6	
AGG ↔ AGA	4		10.3
AGG	6		15.4
AGT	17		43.6
AGC	0		0
AGA ↔ TCN[a]	7	5.6	
Other[b]	29	23.0	
AGG ↔ TCN[a]	2		5.1
Other[c]	10		25.6

[a] All are TCT.

[b] Codons for glycine (10), phenylaline (4), asparagine (3), cysteine (3), alanine (2), tyrosine (2), glutamamic acid (1), methionine (1), threonine (1), valine (1) and one deletion.

[c] Codons for glycine (4), lysine (2), leucine (1), methionine (1) cysteine (1), asparagine (1).

Taken together the above data strongly support the conclusion that in the *C. elegans* and *A. suum* mt-protein genes AGA and AGG specify the same amino acid as AGT (and AGC) codons, which in all known genetic codes is serine. If it were the case that in nematode mitochondria AGR codons specify a different amino acid than AGY codons, then in the *C. elegans* mtDNA-encoded proteins the amino acid specified by AGR codons would substitute 44% of the time for the amino acid specified by AGY (serine) codons in the homologous *A. suum* proteins. Also, if AGR codons specified an amino acid other than serine then the relative amount of serine in the *A. suum* and *C. elegans* mt-protein genes would be 1.4 rather than 0.9, the value expected if both AGR and AGY codons specify serine in the nematode mt-genetic code (Table 2).

A further observation consistent with the view that AGR and AGY codons specify serine is that neither the *C. elegans* nor the *A. suum* mtDNA molecules contain the two tRNAs that would be necessary to decode AGR and AGY codons if they specified different amino acids. Rather, there is a single tRNA, the anticodon of which, TCT, could be expected to recognize all AGN codons. However, although the T nucleotide in the first position of this anticodon is the

same as that found in the first position of other anticodons of mt-tRNAs that recognize all members of a four codon family, it is noted that the anticodon GTC, that occurs in vertebrate mt-tRNAs that decode AGY (serine) codons, can apparently decode AGY and AGA codons in Drosophila mtDNAs (CLARY and WOLSTENHOLME 1985a) and all AGN codons in echinoderm mtDNAs (JACOBS *et al.* 1988; CANTATORE *et al.* 1989).

Differential nucleotide usage in nematode and human mt-protein genes: The frequency of leucine is high and remarkably constant, in the range 15.0 to 16.9%, among mt-proteins of different metazoa. Among mammalian, insect (*D. yakuba*) and platyhelminth (*F. hepatica*) mtDNAs the ratio of TTR to CTN triplets used as leucine-specifying codons is positively correlated with the differential use of T and C nucleotides in the third position of codons (Table 3; CLARY and WOLSTENHOLME 1985a; GAREY and WOLSTENHOLME 1987). Data from *C. elegans* and *A. suum* mt-protein genes add to the generality of this correlation; the ratio of TTR:CTN codons is 4.7:1 and 6.7:1 for the mt-protein genes of *C. elegans* and *A. suum*, respectively, and the corresponding ratios of T:C nucleotides in the third positions of codons are 10.1:1 and 26.1:1 (Table 3). However, this correlation is not found for either *P. lividus* or *X. laevis* mt-protein genes (Table 3) suggesting that in both of these cases the constraints on synonymous T and C nucleotides in the first positions of codons are different from those of synonymous T and C nucleotides in the third positions of codons.

Between nematode and human mt-protein genes, there is a T:C bias regarding both the first position of codons (other than leucine-specifying codons) and in the second position of codons, similar to the bias found for codon third positions and leucine-specifying codon first positions (Table 7): that is, a greater average use of C in humans and a greater average use of T in nematodes. The higher average C in human first (replacement) positions results from greater use of all codons that begin with C (and, therefore, the amino acids they specify): that is, codons for proline (CCN), histidine (CAY), glutamine (CAR) and arginine (CGN). The higher average T in the first (replacement) position of nematode codons results from greater use of codons specifying phenylalanine (TTY), tyrosine (TAY) and cystine (TGY). The differential occurrence of phenylalanine in nematode mtDNA-encoded proteins (*C. elegans*, 13.2%; *A. suum*, 14.4%) and the homologous human proteins (5.7%) is greater than for any other amino acid.

There is a higher frequency of TCN (serine) codons in human (5.9%) than in *C. elegans* (4.2%) and *A. suum* (4.7%) mt-protein genes (Table 2). However, the much higher frequency of serine-specifying AGN codons in nematode mt-protein genes (*C. elegans*, 6.8% and *A. suum* 5.7%), than AGY, serine-specifying codons (1.3%) in human mt-protein genes results in an overall serine ratio of between 1.4:1 and 1.5:1 for the nematode:human mt-protein genes.

TABLE 7

Frequencies of nucleotides in the different positions in codons of the mt-protein genes of *C. elegans*, *A. suum* and *H. sapiens*

Codon position		Percentage nucleotides T	C	A	G
1	*C. elegans*	38.7 (25.9)[a]	9.0 (6.2)[a]	32.0	20.3
	A. suum	41.2 (28.0)[a]	8.2 (6.2)[a]	26.1	24.5
	H. sapiens	20.5 (18.4)[a]	27.9 (12.7)[a]	31.8	19.8
2	*C. elegans*	49.5	14.2 (9.9)[b]	19.8	16.6
	A. suum	51.7	13.4 (8.6)[b]	18.3	16.7
	H. sapiens	41.0	28.2 (22.3)[b]	19.5	11.4
3	*C. elegans*	49.7	4.7	36.6	9.0
	A. suum	62.7	2.4	11.5	23.4
	H. sapiens	15.3	43.1	36.5	5.1

Nucleotides concerned with termination are excluded. The sets of mt-protein genes in the three species are the same except that the human set (ANDERSON *et al.* 1981) includes a gene for ATPase8 not found in the nematode mtDNAs, and excludes the gene for ND6 (see footnote to Table 3).

[a] The numbers in parentheses are frequencies that exclude leucine-specifying codons (see text).

[b] The numbers in parentheses are frequencies that exclude TCN (serine-specifying) codons.

The frequency of each codon family with a C in the second position (CCN, proline; ACN, threonine; GCN, alanine), except TCN (serine, as noted above), is at least 2.0 times greater in human than in nematode mt-protein genes (Table 7). The higher average T in the second position of nematode codons (other than leucine-specifying codons) results from remarkably higher frequencies in nematode than in human mt-protein genes of TTY (phenylalanine) and GTN (valine) codons: 2.3–2.5:1 and 1.7–2.5:1, respectively. The frequencies of the remaining two codon families with T in the second position: ATY (isoleucine) and ATR (methionine) are both higher in humans than in nematodes: 1.1–1.5:1 and 1.1:1, respectively.

In contrast to the biases for C and T nucleotides in positions 1 and 2 of codons of nematode and human mt-protein genes, no such overall bias exists regarding A and G nucleotides in position 1, or of A nucleotides in position 2 of nematode and human mt-protein genes (Table 7). The frequencies of A nucleotides in positions 1 and 2 in the human codons (31.8 and 19.5%) is in each case within the range of frequencies of A nucleotides in the corresponding positions in nematode codons (26.1–36.0% and 18.3–19.8%) and the frequency of G nucleotides in position 1 in human codons (19.8%) is close to the frequency of G nucleotides in position 1 (20.3 and 24.5%) in nematode codons (Table 7). The noticeably lower frequency of

TABLE 8

Comparisons of mean nucleotide sequence and amino acid sequence similarities of the 12 mt-protein genes of *C. elegans* and *A. suum*, and of the homologous 12 mt-protein genes of human and cow (Anderson et al. 1982a,b)

	Mean nucleotide sequence similarity		Mean amino acid sequence similarity	
	C. elegans/A. suum	Human/cow	*C. elegans/A. suum*	Human/cow
Nonadjusted[a]	72.3	72.1	73.7	75.2
Adjusted[b]	72.6	72.5	74.5	78.0

[a] The means of the percentage similarities for the nucleotide sequences and amino acid sequences of the 12 mt-protein gene comparisons for *C. elegans* and *A. suum* and for human and cow.

[b] The mean similarities for the nucleotide sequences and amino acid sequences of the gene sets joined end to end.

G nucleotides in position 2 of human codons than in nematode codons mainly reflects the much higher use of AGN (serine) codons in nematodes than of AGY (serine) codons in humans, as noted above.

It can be argued that the bias in the use of T and C nucleotides at synonymous positions in codons of nematode and human mt-protein genes has resulted from differential selection of these nucleotides. However, both differential nucleotide selection and selection for the amino acids that have C and T in the first and/or second codon position could have contributed to the similar T and C bias found at replacement positions.

Evolutionary considerations: In Table 8, mean nucleotide sequence and predicted amino acid sequence similarities of the twelve *C. elegans* and *A. suum* mt-protein genes are compared with those of the homologous twelve human and cow mt-protein genes (ANDERSON *et al.* 1982a,b). Interestingly, both the non-adjusted and adjusted (for gene length) means of nucleotide sequence similarity for the *C. elegans/A. suum* and the human/cow comparisons differ by only 0.1%. However, the nonadjusted and adjusted means of predicted amino acid sequence similarities are 1.5% and 3.5%, respectively, higher for the human/cow comparison than for the *C. elegans/A. suum* comparison. This indicates that, in spite of the apparently greater nucleotide biases at codon third positions between the *C. elegans* and *A. suum* genes than between the human and cow genes (Table 3), third position differences are of higher frequency between the human and cow genes.

It has been proposed that the ancestral lines of human and cow diverged approximately 80 million years ago (WILSON, CARLSON and WHITE 1977). If, therefore, it is assumed that the mean rates of mt-protein gene amino acid sequence evolution have been the same in the *C. elegans* and *A. suum* lines, and in the human and cow lines, then the data in Table 8 indicate that divergence of the *C. elegans* and *A. suum* lines occurred close to, but before, the divergence of the human and cow lines. This conclusion is supported by data from comparisons of nucleotide sequence similarities of s-rRNA genes and of l-rRNA genes between these nematodes and mammals: the mt-s-rRNA and mt-l-rRNA genes of *C. elegans* and *A. suum* are 76.2% and 74.4% similar, respectively, compared to corresponding values of 78.0% and 77.0% for human and cow. However, data from comparisons of amino acid sequences of the most highly conserved mt-protein genes of organisms in different metazoan phyla indicate that the rate of evolution of the mt-protein genes have been greater in nematodes than in vertebrates (R. OKIMOTO and D. R. WOLSTENHOLME, unpublished results). The magnitude of this difference suggests the possibility that the divergence of the *C. elegans* and *A. suum* lines occurred more recently than that of the human/cow lines, rather than earlier.

The time of separation of the *C. elegans* and *A. suum* ancestral lines has been a point of contention among nematode systematists. One view is that these lines diverged in the early Cenozoic, less than 65 million years ago (MAGGENTI 1983). A contrasting view is that they diverged much earlier, in the mid-Cambrian, 550 million years ago (POINAR 1983). The above compared similarities of nucleotide and amino acid sequences between *C. elegans* and *A. suum* mt-genes clearly support the more recent divergence time. In further support of this conclusion is the observation that the mean nucleotide and amino acid sequence similarities of the mt-protein genes of mouse and *D. yakuba*, whose ancestral lines diverged about 600 million years ago (WILSON *et al.* 1978), are 53.2 and 44.7% respectively, compared to 72.3 and 73.7% for *C. elegans* and *A. suum*.

We thank JOHN F. ATKINS, RAYMOND F. GESTELAND, MARTIN C. RECHSTEINER and JILL A. WAHLEITHNER for discussion, JOHN S. PARKINSON for a computer program and KIRK THOMAS for oligonucleotides. This work was supported by National Institutes of Health grants GM 18375 and RR 07092.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B.

A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1982a Comparison of the human and bovine mitochondrial genomes, pp. 5–43 in *Mitochondrial Genes*, edited by P. SLONIMSKI, P. BORST and G. ATTARDI. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

ANDERSON, S., M. H. L. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER and I. G. YOUNG, 1982b The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156:** 683–717.

BARRELL, B. G., A. T. BANKIER and J. DROUIN, 1979 A different genetic code in human mitochondria. Nature **282:** 189–194.

BARRELL, B. G., S. ANDERSON, A. T. BANKIER, M. H. L. DE BRUIJN, E. CHEN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1980 Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc. Natl. Acad. Sci. USA **77:** 3164–3166.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BOYCE, T. M., M. E. ZWICK and C. F. AQUADRO, 1989 Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123:** 825–836.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71–94.

BRUTLAG, D. L., J. CLAYTON, P. FRIELAND and L. H. KEDES, 1982 SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res. **10:** 279–304.

CANTATORE, P., M. ROBERTI, G. RAINALDI, M. N. GADALETA and C. SACCONE, 1989 The complete nucleotide sequence, gene organization, and genetic code of mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264:** 10965–10975.

CHANG, D. D., T. W. WONG, J. E. HIXON and D. A. CLAYTON, 1985 Regulatory sequences for mammalian mitochondrial transcription and replication, pp. 135–144 in *Achievements and Perspectives of Mitochondrial Research, Vol. II. Biogenesis*, edited by E. QUAGLIARIELLO, E. C. SLATOR, F. PALMIERI, C. SACCONE and A. M. KROON. Elsevier, Amsterdam.

CHOMYN, A., and G. ATTARDI, 1987 Mitochondrial gene products. Curr. Top. Bioenerg. **15:** 295–329.

CLARY, D. O., and D. R. WOLSTENHOLME, 1983a Nucleotide sequence of a segment of *Drosophila* mitochondrial DNA that contains the genes for cytochrome *c* oxidase subunits II and III and ATPase subunit 6. Nucleic Acids Res. **11:** 4211–4227.

CLARY, D. O., and D. R. WOLSTENHOLME, 1983b Genes for cytochrome *c* oxidase subunit I, URF2 and three tRNAs in *Drosophila* mitochondrial DNA. Nucleic Acids Res. **11:** 6859–6872.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985a The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22:** 252–271.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985b The ribosomal genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res. **113:** 4029–4045.

CLARY, D. O., and D. R. WOLSTENHOLME, 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.

CLARY, D. O., J. M. GODDARD, S. C. MARTIN, C. M. R. FAURON and D. R. WOLSTENHOLME, 1982 *Drosophila* mitochondrial DNA: a novel gene order. Nucleic Acids Res. **10:** 6619–6637.

CLAYTON, D. A., 1982 Replication of animal mitochondrial DNA. Cell **28:** 693–705.

CLAYTON, D. A., 1984 Transcription of the mammalian mitochondrial genome. Annu. Rev. Biochem. **53:** 573–594.

DAHLBERG, A. E., 1989 The functional role of ribosomal RNA in protein synthesis. Cell **57:** 525–529.

DALE, R. M. K., B. MCCLURE and J. P. HOUCHINS, 1985 A rapid single-stranded cloning strategy for producing overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13:** 31–40.

DE BRUIJN, M. H. L., 1983 *Drosophila melanogaster* mitochondrial DNA: a novel organization and genetic code. Nature **304:** 234–241.

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

DIRHEIMER, G., G. KEITH, A.-P. SIBLER and R. P. MARTIN, 1979 The primary structure of tRNAs and their rare nucleosides, pp. 19–41 in *Transfer RNA: Structure, Properties and Recognition*, edited by P. R. SCHIMMEL, D. SOLL and J. N. ABELSON. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

DOOLITTLE, R. F., 1981 Similar amino acid sequences: chance or common ancestry. Science **214:** 149–159.

FREIER, S. M., R. KIERZEK, J. A. JAEGER, N. SUGIMOTO, M. H. CARUTHERS, T. NIELSON and D. H. TURNER, 1986 Improved free energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83:** 9373–9377.

GADALETA, G., G. PEPE, G. DECANDIA, C. QUAGLIARIELLO, E. SBISA and C. SACCONE, 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.

GARESSE, R., 1988 *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics **118:** 649–663.

GAREY, J. R., and D. R. WOLSTENHOLME, 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., C. ZWEIB and R. BRIMACOMBE, 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, M. W., D. SANKOFF and R. J. CEDERGREN, 1984 On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res. **12:** 5837–5852.

GUTELL, R., and G. E. FOX, 1988 A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res. **16:** r175-r269.

GUTELL, R., B. WEISER, C. R. WOESE and H. F. NOLLER, 1985 Comparative anatomy of 16-S-like ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol. **32:** 155–216.

HIMENO, H., H. MASAKI, T. KAWAI, T. OHTA, I. KUMAGI, I. MIURA and K. WATANABE, 1987 Unusual genetic codes and a novel gene structure for $tRNA^{ser}AGY$ in starfish mitochondrial DNA. Gene **56:** 219–230.

HONG, G. F., 1982 A systematic DNA sequencing strategy. J. Mol. Biol. **158:** 539–549.

JACOBS, H. T., D. J. ELLIOT, V. B. MATH and A. FARQUHARSON, 1988 Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202:** 185–217.

JUE, R. A., N. W. WOODBURY and R. F. DOOLITTLE, 1980 Sequence homologies among *E. coli* ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J. Mol. Evol. **15:** 129–148.

KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 333–348.

MAGGENTI, A. R., 1983 Nematode higher classification as influenced by species and family concepts, pp. 25–40 in *Concepts in*

Nematode Systematics, edited by A. R. STONE, H. M. PLATT and L. F. KHALIL. Academic Press, London.

MOAZED, D., and H. F. NOLLER, 1989a Interaction of tRNA with 23S rRNA in ribosomal A, P and E sites. Cell **57:** 585–597.

MONNEROT, M., M. SOLIGNAC and D. R. WOLSTENHOLME, 1990 Discrepancy in divergence of the mitochondrial and nuclear genes of *Drosophila teissieri* and *Drosophila yakuba*. J. Mol. Evol. **30:** 500–508.

MONTOYA, J., T. CHRISTIANSON, D. LEVENS, M. RABINOWITZ and G. ATTARDI, 1982 Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **79:** 7195–7199.

MORITZ, C., T. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–292.

NOLLER, H. F., M. ASIRE, A. BARTA, S. DOUTHWAITE, T. GOLDSTEIN, R. R. GUTELL, D. MOAZED, J. NORMANLEY, J. B. PRINCE, S. STERN, K. TRIMAN, S. TURNER, B. VAN STOLK, V. WHEATON, B. WEISER and C. R. WOESE, 1986 Studies on the structure and function of ribosomal RNAs, pp. 141–163 in *Structure, Function and Genetics of Ribosomes*, edited by B. HARDESTY and G. KRAMER (Springer Series in Molecular Biology, edited by A. RICH). Springer-Verlag, New York.

OJALA, D., J. MONTOYA and G. ATTARDI, 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature **290:** 470–474.

OJALA, D., C. MERKEL, R. GELFAND and G. ATTARDI, 1980 The tRNA genes puncuate the reading of genetic information in human mitochondrial DNA. Cell **22:** 393–403.

OKIMOTO., R., and D. R. WOLSTENHOLME, 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., J. L. MACFARLANE and D. R. WOLSTENHOLME, 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., H. M. CHAMBERLIN, J. L. MACFARLANE and D. R. WOLSTENHOLME, 1991 Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res. **19:** 1619–1626.

POINAR, G. O., JR., 1983 *The Natural History of Nematodes*. Prentice-Hall, Englewood Cliffs, N.J.

ROE, B. A., D. P. MA, R. K. WILSON and J. F.-H. WONG, 1985 The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SINGHAL, R. P., and P. A. M. FALLIS, 1979 Structure, function, and evolution of transfer RNAs. Prog. Nucleic Acids Res. Mol. Biol. **23:** 227–290.

SNYDER, M., A. R. FRASER, J. LAROCHE, K. E. GARTNER-KEPKAY and E. ZOUROS, 1987 Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc. Natl. Acad. Sci. USA **84:** 7595–7599.

SPRINZL, M., T. HARTMANN, J. WEBER, J. BLANK and R. ZEIDLER, 1989 Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **17:** r1–r172.

STADEN, R., 1982 Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10:** 4731–4751.

SULSTON, J. E., and S. BRENNER, 1974 The DNA of *Caenorhabditis elegans*. Genetics **77:** 95–104.

WAHLEITHNER, J. A., and D. R. WOLSTENHOLME, 1987 Mitochondrial plasmid DNAs of broad bean: nucleotide sequences, complex secondary structures, and transcription. Curr. Genet. **12:** 55–67.

WARRIOR, R., and J. GALL, 1985 The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. Arch. Sci. Geneva **38:** 439–445.

WILSON, A. C., S. S. CARLSON and T. J. WHITE, 1977 Biochemical evolution. Annu. Rev. Biochem. **46:** 573–639.

WILSON, E. O., T. EISNER, W. R. BRIGGS, R. E. DICKERSON, R. L. METZENBERG, R. D. O'BRIEN, M. SUSMAN and W. E. BOGGS, 1978 *Life on Earth*, Ed. 2. Sinauer, Sunderland, Mass.

WOLSTENHOLME, D. R., and D. O. CLARY, 1985 Sequence evolution of *Drosophila* mitochondrial DNA. Genetics **109:** 725–744.

WOLSTENHOLME, D. R., and C. M. R. FAURON, 1976 A partial map of the circular mitochondrial genome of *Drosophila melanogaster*: location of *Eco*RI-sensitive sites and the adenine-thymine-rich region. J. Cell Biol. **71:** 434–448.

WOLSTENHOLME, D. R., J. L. MACFARLANE, R. OKIMOTO, D. O. CLARY and J. A. WAHLEITHNER, 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA **84:** 1324–1328.

WOLSTENHOLME, D. R., R. OKIMOTO, J. L. MACFARLANE, G. A. PONT, H. M. CHAMBERLIN, J. R. GAREY and N. A. OKADA, 1989 Unusual features of lower metazoan mitochondrial genomes, pp. 103–116, in *Structure, Function and Biogenesis of Energy Transfer Systems*, edited by E. QUAGLIARIELLO, S. PAPA, F. PALMIERI and C. SACCONE. Elsevier, Amsterdam.

ZWEIB, C., C. GLOTZ and R. BRIMACOMBE, 1981 Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res. **9:** 3621–3640.

Communicating editor: M. R. HANSON