GENE ORGANIZATION AND EVOLUTION OF MITOCHONDRIAL GENOMES FROM TWO INVERTEBRATE PHYLA: VESTIMENTIFERA AND CHAETOGNATHA

BY

YUEWANG WEI

B.Sc. Peking University 1970

M.Sc. Peking University 1983

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Biological Sciences

Simon Fraser University

December 1992

APPROVAL

Name: **YUEWANG WEI**

Degree: **Doctor of Philosophy**

Title of Thesis:

GENE ORGANIZATION AND EVOLUTION OF MITOCHONDRIAL GENOMES FROM TWO INVERTEBRATE PHYLA: VESTIMENTIFERA AND CHAETOGNATHA

Examining Committee:

Chair: Dr. J. Borden, Professor

Dr. A.T. Beckenbach, Associate Professor,
Senior Supervisor, Department of Biological Sciences, SFU

Dr. M.J. Smith, Professor,
Department of Biological Sciences, SFU

Dr. D.L. Baillie, Professor,
Department of Biological Sciences, SFU

Dr. D. Sen, Associate Professor,
Department of Chemistry, SFU
Public Examiner

Dr. S.R. Palumbi, Associate Professor,
Department of Zoology, University of Hawaii,
Honolulu, Hawaii
External Examiner

Date Approved 14 December 1992

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Gene Organization and Evolution of Mitochondrial Genomes from Two Invertebrate Phyla: Vestimentifera and Chaetognatha

Author: _____
(signature)

Yuewang Wei
(name)

Dec. 14, 1992
(date)

ABSTRACT

Information concerning DNA sequence comparisons and gene organization of animal mitochondrial DNA has proven extremely valuable in determining relationships of animal groups. Within the vertebrates, only two gene rearrangements are known. A single major rearrangement separates the vertebrates from the echinoderms, while three major rearrangements distinguish these groups from the insect gene organization. In contrast, nematodes and the mollusc *Mytilis* have extensive differences from arrangements of other animal groups.

In this thesis, I have used the polymerase chain reaction and direct sequencing to obtain partial sequence and gene organization for members of two enigmatic phyla: Vestimentifera (deep sea vent tube-worms) and Chaetognatha (arrow worms). Approximately one-third of the mitochondrial genome was determined for both groups.

The organization of the vestimentiferan genome bears some similarities to those of vertebrates and insects. In particular, the subunits of cytochrome oxidase (COI-COIII) and ATPase appear in the same order: COI, COII, ATPase 8, COIII. In addition, the two ribosomal RNA genes are transcribed in the same direction and are separated by tRNA valine gene, as occurs in vertebrates and insects. One major difference, however, is the removal of the ATPase 6 gene from its position between ATPase 8 and COIII. This change clearly separates them from those phyla.

The chaetognath gene organization bears no similarities to that of any other known phylum, and suggests that this group separated from other groups very early in the history of the animal kingdom. The large ribosomal RNA gene has been separated from the small RNA gene, and is located adjacent to the structural gene for COI. The cytochrome b gene

has been rearranged to the region upstream the 5' end of the large ribosomal RNA subunit. In addition, the NADH dehydrogenase 2 gene is contiguous with the COIII gene in the chaetognath mitochondrial genome. These features have not been observed in any other animal group.

ACKNOWLEDGEMENTS

I would first like to thank my fellow graduate students and associates for their support and helpful discussions throughout my time at Simon Fraser University. I would especially like to thank Karen Beckenbach, Kelly Thomas and Eric Cabot for their invaluable help during my reseach. I also wish to thank the members of my committee for their helful discussions and comments on this thesis.

I am most indebted to, and greatefully acknowledge the continous support of my supervisor Andy Beckenbach.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

INTRODUCTION

Mitochondrial DNA (mtDNA) has been characterized from a number of multi-cellular animals, and from fungi and protozoa over the past decades (Attardi, 1988). Among the vertebrates, the complete mitochondrial DNA sequences have been published for human (Anderson et al. 1981), mouse (Bibb et al. 1981), cow (Anderson et al. 1982), rat (Gadaleta et al. 1989), fin whale (Arnason et al. 1991), harbor seal (Arnason and Johsson 1992), *Xenopus* (Roe et al. 1985) and chicken and quail (Desjardins and Morais 1990; 1991). Among the invertebrate deuterostomes there are complete DNA sequences for the mitochondrial genomes of two sea urchin species *Strongylocentrotus purpuratus* (Jacobs et al. 1988) and *Paracentrotus lividus* (Cantatore et al. 1989), and partial sequences are available for several sea star species (Himeno et al. 1987; Jacobs et al. 1989; Smith et al. 1989; 1990). Among protostome invertebrates, the complete nucleotide sequence of a fly, *Drosophila yakuba* (Clary and Wolstenholme 1985) and a nearly complete sequence of a bivalve mollusk, *Mytilus edulis* (Hoffmann et al., 1992) have been determined. In addition, complete sequences are available for two nematode worms (Okimoto et al. 1992) and partial sequences have been obtained from a number of other vertebrates and invertebrates.

The animal mitochondrial genomes show a strikingly conserved organization. The gene content and genomic organization has remained stable since the divergence of the mammalian and amphibian lineages, approximately 350 million years ago (Brown, 1983). The mitochondrial genome of animals consists of a very compact circular DNA molecule approximately 15-16 kilobases in length, which encode the same genes for 12 or 13 polypeptides, two ribosomal RNAs (large or 16S RNA and small or 12S RNA), and 22 transfer RNAs (tRNAs), as well as a control region containing the initiation sites for

mitochondrial DNA replication and transcription. 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_0 ATPase complex (ATPase 6 and ATPase 8), and subunits 1-6 and 4L of the respiratory chain NADH dehydrogenase (ND1-6 and 4L) (Chomyn and Attardi, 1987).

In contrast with the near constant organization of vertebrate mitochondrial genomes, sequence data from invertebrates such as *Drosophila* (Clary and Wolstenholme 1985), sea urchin (Jacobs et al. 1988; Cantatore et al. 1989), starfish (Himeno et al. 1987; Jacobs et al. 1989; Smith et al. 1989; Smith et al. 1990), and nematode (Okimoto et al. 1992) mtDNAs reveal that many rearrangements, including inversions and transpositions, have occurred between these phyla. Gene rearrangements have also been observed by Desjardins and Morais (1990) in the chicken mitochondrial genome and by Pääbo et al., (1991) in marsupials.

Mitochondrial DNA has become a very useful tool for studies of evolutionary genetics because of its small size, high copy number, mostly unisexual mode of inheritance, and evolutionary behavior (Wilson et al. 1985). Studies of the organization of mtDNA in different animals are of interest for two reasons. First, the conserved and unconserved features can give insights into the mechanisms of mitochondrial gene expression and its regulation. Second, the infrequency of rearrangements provides a possible route to the elucidation of phylogenetic relationships between higher-order taxa (Brown 1983; Jacobs et al., 1988; Wolstenholme et al. 1985) for which other markers may prove unreliable.

The analysis of mtDNA genomic sequence is essential if we are to ascertain the sites and potential mechanism whereby major rearrangements occur. The phylogeny and evolution

of organisms can be inferred from DNA analyses at both the macro and micro level. The determination of the pattern of major gene rearrangements can lead to better understanding of the branch points of macrophylogenetic events. Detailed studies of the base changes in the structure genes in the mtDNA can give insights into the phylogeny of organisms at the species level.

Because of the simplicity of the molecule, lack of recombination, and high mutation rate, mitochondrial DNA has been, by far, the most popular genetic material for use in molecular evolutionary studies. The rate of animal mtDNA evolution varies among lineages, among genes, and within genes. The slow but perceptible rate of rearrangement and the very large number of arrangements possible suggest that mitochondrial gene order may provide useful information about the phylogenetic relationships of different organisms. Although present data provide encourgement for this approach, the methods for analyzing the evolutionary relationships among gene rearrangements are still being developed (Sankoff et al., 1990; Beckenbach and Smith, in prep.).

Although cloning is very powerful approach and is of popular use in molecular biology, this method usually requires significant amounts of material and highly purified mtDNA. The initial cloning and identification methods can often be very time consuming. When only very small organisms are available, obtaining enough mitochondrial DNA can be a limitation. A new method, the polymerase chain reaction (PCR), has been developed for amplifying selected nucleic acids (DNA or RNA) sequences *in vitro* (Mullis and Faloona 1987). The method consists of repetitive cycles of DNA denaturation, primer annealing, and extension by DNA polymerase (Mullis 1990). Two oligonucleotide primers are chosen to flank the DNA segment to be amplified. The template DNA is repeatedly heat denatured,

cooled to allow binding of primers, and replicated with DNA polymerase. The two primers hybridize to opposite strands of the target sequence, such that synthesis proceeds across the region between the primers, replicating that DNA segment. The product of each PCR cycle is complementary to and capable of binding the primers, and so the amount of DNA synthesized is doubled in each successive cycle.

PCR has made possible the rapid isolation and amplification of specific DNA segments, which can then be used for a wide range of applications. Nucleic acid amplification has added a new and revolutionary dimension to molecular biology. The detection and characterization of specific DNA sequence employing PCR are a simple, less expensive and less time consuming alternative to the cloning of specific genes. Highly conserved regions of mtDNA, such as 12S rRNA, 16S rRNA, cytochrome B, and other loci, are routinely being amplified via PCR technology to generate fragments of DNA. The original template DNA can be in a pure form and as discrete molecules or it can be a very small part of a complex mixture of biological substances. It can be a tissue specimen, a human hair, dried blood, mummified brain tissue, or tissue from a 40,000-year-old woolly mammoth frozen in a glacier (Mullis 1990).

In most work to date, PCR has been used to amplify short regions of DNA, usually a few hundred base pairs long (Saiki et al., 1988). Since Taq polymerase is capable of efficiently amplifying DNA of up to 2 kb long (Saiki et al., 1988), it therefore appeared possible that PCR could be extended to the amplification of longer fragments of mitochondrial DNA.

Although many sequences of mitochondrial genomes from various phyla are available, information about mtDNAs from two invertebrate phyla, Vestimentifera and Chaetognatha,

has never been reported. Eucoelomates are generally divided into two main branches: Deuterostomes and Protostomes. The former covers Echinoderms, Chordates (including vertebrates) and a minor phylum, Hemichordata. The protostomes include Arthropods, Annelids, Molluscs, and a few minor phyla. There are other phyla which do not fall cleanly into either group. Vestimentifera is a newly defined phylum (Jones 1985), encompassing tube worms from deep ocean vents. This phylum is closely related to Phylum Pogonophora and is included by some within the Phylum Pogonophora. Phylum Pogonophora is a group of deep sea tube worms discovered this century based on partial specimens dredged from the continental shelf (deep sea sediments). A remarkable feature of members of these phyla is the complete absence of a mouth and digestive tract. Prior to 1970, when the existence of the opisthosoma was unknown, Pogonophora were variously placed among deuterostomes, protostomes, or even as a class of Annelid worms. With the discovery of the opisthsoma (the posterior portion of adult) and various developmental stages (Southward 1988), placement among deuterostomes became untenable. In 1985, members of this group from deep sea vents were raised to equal status with Pogonophora. The phylogenetic affinities of these groups are at present, unknown.

The Chaetognatha is another phylum of about 100 species, called arrowworms because of their shape. Despite the similarities of adult chaetognaths to aschelminths, the embryogeny of the phylum appears to be deuterostome in nature. Phylum Chaetognatha is variously placed with deuterostomes, protostomes or even pseudocoelomate groups (Willmer 1990). There are, however, some peculiarities. For example, the coelom is enterocoelic in origin as in deuterostomes but does not arise by a direct outpocketing of the archenteron. Furthermore, only two pairs of coelomic pockets are formed instead of

three in deuterostomes. Moreover, there is no larval stage comparable to that of the echinoderms and the hemichordates. It is now unlikely that the characteristics of each of major group of Chaetognatha show any affinity with deuterostomes. Thus, the Chaetognatha cannot be allied with any specific deuterostome phylum. If chaetognaths are really deuterostomes, the phylum must have departed very early from the base of the deuterostome line and is only remotely related to the other deuterostome groups. The phylogenetic relationships of both phyla, Vestimentifera and Chaetognatha, to other animal phyla are still not clear.

In this thesis, I have investigated the gene organization and evolution of mitochondrial genomes from members of two invertebrate enigmatic phyla: Vestimentifera and Chaetognatha. Using the polymerase chain reaction, a total length of 9,297 nucleotides was amplified and sequenced. The sequence obtained from both each phylum represents about one third of the mitochondrial genome. Data from both phyla revealed some significant features different from those of other animal species reported to date.

This thesis is divided into four sections. In section I, I use the polymerase chain reaction to study the Phylum Vestimentifera mitochondrial genome. I have developed a suite of PCR primers which were based on analysis of sequences of various mitochondrial genes from different animal species. Using these primers, several fragments coding different functional genes, representing one third of the genome were amplified with PCR and sequenced. I have analyzed all these sequences and demonstrated novel features of this mitochondrial genome. Also, I have identified a number of complementary sequences between ribosomal RNA and protein genes. In section II, I use information from comparison of gene organization and analysis of various functional genes to examine

evolutionary features of Phylum Vestimentifera mitochondrial DNA. I have found that gene rearrangements have occurred in this genome and the rate of mtDNA evolution varies among and within gene regions and is related to functional constraints. In section III, I use PCR to study the mitochondrial genome of a representative of Phylum Chaetognatha. Analysis of these sequences reveals that the organization of the mitochondrial genome of the Phylum Chaetognatha is unlike that of any other known group. In last section, I compare the gene organization and sequence information from different functional genes, with those of other species to examine the evolutionary features of Phylum Chaetognatha mitochondrial DNA.

MATERIALS AND METHODS

Biological Materials

Adults of *Ridgeia phaephiale* (Phylum Vestimentifera) were supplied by Dr. Verena Tunnicliffe. They were collected by John Barross on Endeavour Segment of Juan de Fuca Ridge in September of 1988 by ALVIN. The tissue of was stored frozen at -70°C prior to DNA extraction. Adults of *Sagitta elegans* (Phylum Chaetognatha) were collected by myself using a plankton net near Bamfield Marine Station, Vancouver Island, British Columbia, Canada. The living adults were used as source for DNA extraction and mtDNA preparation.

Mitochondrial DNA Preparation for Cloning

Mitochondrial DNA was extracted from *Ridgeia* and *Sagitta* using modifications of the alkaline lysis technique of Tamura and Aotsuka (1988).

Briefly, 0.1 g of tissue was ground in 400 μl of MSB+Ca++ (0.21 M mannitol, 0.07 M sucrose, 0.05 M Tris HCl, pH 7.5, 3mM $CaCl_2$, 0.5 mg/ml of protein K). Eighty μl 0.5 M EDTA and 100 μl of solution I (50 mM glucose, 10 mM EDTA, 25mM Tris, pH 8.0) were added and the mixture was then heated to 65^0 C for 3 minutes. The strands were separated by adding two volumes of 0.1% SDS plus 0.2 N NaOH and chilled on ice for 5 minutes. Closed circular DNA was allowed to re-anneal by adding 1.5 volumes of 3 M potassium, 5 M acetate, pH 4.8. The supernatant was then extracted with phenol/sevag (chloroform/isoamylalcohol 24:1), and ethanol precipitated at room temperature.

DNA Extraction for PCR

A small amount of tissue of *Ridgeia* and a single adult of *Sagitta* were ground in microcentrifuge tubes containing 50 μl protease solution (0.1M Tris-OH, pH 8.0, 0.05M EDTA, 0.2M NaCl, 1% SDS, with 0.4mg/ml protease K) respectively, and immediately placed at 65°C for 3-5 minutes. The mixture was extracted with phenol (saturated with 10mM Tris, 1mM EDTA, pH 8.0), the aqueous phase removed and extracted with chloroform:isoamyl alcohol (24:1). The aqueous phase was again removed and 2 volumes of 95% ethanol (at -20°C) added to precipitate the DNA. The DNA was pelleted, washed twice with 70% ethanol (-20°C), dried under vacuum, and redissolved in 50 μl to 100 μl sterile distilled H_2O.

Cloning

Mitochondrial DNA from *Ridgeia* was cloned into the vector PVZ-1 (Henicoff and Eghtedarzadeh 1987). 1 μg of *Ridgeia* mtDNA sample was mixed with 0.2 μg of PVZ-1 and digested to completion with HindIII in a total volume of 10 μl for 2 hr at 37°C. The restriction enzyme was inactivated by heating at 65°C for 3-5 minutes. The digestion products were ligated by adding 6 μl of ligation mix (20 μl 10X ligation Buffer (0.66 M TRIS pH 7.5, 0.66 M $MgCl_2$); 20 μl 0.1 M dithiothreitol (DTT); 20 μl BSA (bovine serum albunin), 1mg/ml; 20 μl 10 mM ATP (adenosine-triphosphate); and 2 units of T4 ligase). This mixture was incubated at 14°C for 1 hr. The diluted ligation mixture was added to the ligation reaction to a final volume of 100 μl, mixed and incubated at 14°C overnight. 25 μl of ligation reaction was mixed with 200 μl of *E. coli* JM 83 competent cells, left on ice for 45 minutes to 1 hr and heated in bath block at 42°C for 5 minutes. Five μl of mixture was

Table 1. Primers used in amplification and sequencing of *Ridgeia*, vestimentiferan and *Sagitta*, chaetognath mitochondrial genes

Name	Sequence
COIF	5'-CCACGACGTTACTCAGACTA-3'
COIIR	5'-TCAGTATCATTGATGACC-3'
COIIR1**	5'-GGTAAAACTACTCGATTATCAAC-3'
COIIIR	5'-ATCTACAAAATGTCAGTATC-3'
12SF*	5'-AAACTAGGATTAGATACCC-3'
12SF1	5'-TGCCAGCCACCGCGGTTATAC-3'
16SF*	5'-ACGTGATCTGAGTTCAGACCGG-3'
16SF4**	5'-TAGGGATAACAGCGCAAT-3'
16SR*	5'-CCGGTCTGAACTCAGATCACGT-3'
CYTBR**	5'-TAAGAAATACCATTCAGG-3'

* Primers 12SF, 16SF, and 16SR from John Boom.

** Primers only for chaetgnath mitochondrial DNA.

added in 5 ml of L-broth medium and cultured in a shaking incubator at 37°C for 15-30 minutes. 50-200 μl of the transformation mixture was plated on nutrient plates containing 50 μg/ml Ampicillin, 40 μg/ml X-gal (5-bromo-4-chloro-3- indolyl-β-D-g, and 160 μg/ml IPTG (isoprophltio-β-D-galactoside). The white colonies were picked from each clone, replated on nutrient plates with only Ampicillin and cultured at 37°C overnight.

Plasmid DNA was prepared from each white colony using the alkaline lysis mini-preparation method (Birnboim and Doly 1979). The resulting plasmid DNA was digested with HindIII.

Polymerase Chain Reaction

Primer sequences for PCR amplification and sequencing of vestimentiferan and chaetognath mtDNA are given in Table 1. These sequences were based on examination of the organization and sequences of mitochondrial genomes from different species, including human (Anderson et al.,1981), mouse (Bibb et al.,1981), cow (Anderson et al.,1982), frog (Roe et al.,1985), *Drosophila yakuba* (Clary and Wolstenhome 1985), sea urchin (Jacobs et al.,1988), sea star (Smith et al.,1990), rat (Gadaleta et al.,1989), and chicken (Desjardins and Morais 1990). Conserved regions were chosen as the basis for PCR primer sequences.

The primer oligonucleotides were prepared in crude form by both Tom Atkinson (University of British Columbia) and the Institute of Molecular Biology and Biochemistry (Simon Fraser University) and were purified as in Atkinson and Smith (1985) using the optional Sep-Pack C18 cartridge (Waters).

Double strand PCR amplification was used for obtaining various fragments. PCR

amplification was carried out in 25 μl reactions, using 1 μl of extracted DNA samples, and 1:1 ratio of two end primers. We followed the protocol provided with the GeneAmp PCR kit (Perkin Elmer Cetus). Thermal cycling was accomplished in an Ericomp Programmable Cyclic Reactor, for 35 to 40 cycles. Each cycle consisted of 94°C denaturation (1 minute), 48° to 52°C annealing (30 to 60 sec) and 72°C extension (3 minutes) steps.

Sequencing

Templates from clones selected for sequencing were prepared from 5 ml overnight cultures. Plasmids containing insert fragments were prepared by the minipreparation procedure (Birnbom and Doly 1979). Supercoiled plasmid DNA was isolated from contaminating RNA, genomic DNA and nicked plasmid by electrophoresis on low melting point agarose. The DNA band containing supercoiled plasmid was cut out of the gel and the DNA was extracted using the phenol method described by Maniatis et al. (1982). The purified template was redissolved with 20 μl sterile distilled H_2O. The double strand product was denatured with 2 μl of 2N NaOH at room temperature for 5 minutes, 8 μl of 5 M NH_4OAc was added for neutralization and 3 volumes of 95% ethanol (at -20°C) was added to precipitate the DNA. The DNA was pelleted, washed with 70% ethanol (at -20°C), dried under vacuum, and finally redissolved in 20 μl sterile distilled H_2O. Templates were sequenced using the dideoxy sequencing protocols with the universal M13 sequencing primers (Pharmacia).

PCR amplified double strand products were purified using the phenol method as described above. The DNA was redissolved in 30 μl sterile distilled H_2O and 7 μl were

used in each sequencing reaction. Both primers used for amplification were also used as sequencing primers. Templates were sequenced using the Sequenase II (United States Biochemical Corporation) dideoxy sequencing protocol. The mixture containing double strands product, primer and sequenase buffer was denatured by heating to 95°C for two minutes, immediately quenched in dry-ice/ethanol for five minutes, then put into a dry-bath block which had been chilled to 0°C. The block was then allowed to come to room temperature. The sequencing reaction followed the manufacturer instructions. In all cases of double strand sequencing, Mn^{++} was added in order to read those regions near the primers.

Sequence Data Analyses

The sequences were read from autoradiographs using a Beckman Sonic Digitizing System (Gel Mate 1000). DNA sequences were analyzed using the Delaney (Delaney Software Ltd.), and ESEE sequence alignment programs (Cabot and Beckenbach 1989). Gene identity was determined by sequence comparison with corresponding genes from several animal species. Sequence alignments presented in the figures were done with the ESEE program (Cabot and Beckenbach 1989). Transfer RNA genes were identified by eye, within sequences lying between protein genes and between ribosomal RNA and protein genes, from the ability of these sequences to fold into secondary structures characteristic of tRNAs, and from the trinucleotide in the anticodon position of the folded sequence.

The trees were constructed by the neighbor-joining method (Saitou and Nei 1987), using the DISTANCE and NEIGHBOR programs of version 3.4 of PHYLIP (Felsenstein 1989).

SECTION I

GENE ORGANIZATION OF MITOCHONDRIAL GENOME FROM *RIDGEIA* (PHYLUM VESTIMENTIFERA)

INTRODUCTION

The mitochondrial genome of animals consists of very compact circular DNA molecule approximately 15-16 kilobases in length, which contains the genes for 22 tRNAs, small and large subunit ribosomal RNAs, and 12 or 13 proteins. Complete nucleotide sequences and gene content have been published for nine vertebrates, two sea urchin, an insect and two nematodes. In addition, most of the sequence of a bivalve mollusc, *Mytilis*, (Hoffmann et al. 1992), and partial mtDNA sequences from a number of other vertebrates and invertebrates have been obtained.

Data from invertebrate mitochondrial genomes reveal that many rearrangements, including inversions and transpositions, have occurred between those phyla. A gene rearrangement has been found by Desjardins and Morais (1990) in the chicken mitochondrial genome relative to other vertebrates. In *Drosophila* mtDNA, the gene arrangement differs from that found in most vertebrates. The urchin genome is characterized by a major rearrangement displacing the 16S RNA some distance from the 12S RNA gene with the ND1 and ND2 genes found between the ribosomal genes. In this arrangement, transcriptional polarity is uniform for both the ribosomal and ND1 and ND2 genes. In addition, this genome has many of its tRNAs clustered just 3' to the 12S RNA gene. Smith et al. (1989) found a major inversion event in sea star mtDNA. In this genome, a segment including the tRNA cluster, ND1, ND2, and 16S RNA genes, is inverted in relation to the sea urchin genome, and the transcriptional polarity of ND1, ND2, and 16S RNA genes are opposite to that of the 12S and COI genes. A recent report (Okimoto et al., 1992) shows that between mtDNAs of nematodes and *Drosophila*, and nematodes and mammals, extensive rearrangements involve almost all tRNA, rRNA, and protein coding

genes.

The genetic codes used by metazoan mitochondrial protein genes contain various modifications (Barrell et al. 1979; Barrell et al. 1980). In all metazoan mtDNAs, TGA specifies tryptophan rather than being a stop codon. In vertebrate mitochondrial protein genes, AGA and AGG are absent (Bibb et al. 1981) or are used as rare stop codons (Anderson et al., 1981; 1982; Roe et al., 1985). However, in *Drosophila 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; Okimoto et al., 1992).

Both translation initiation and termination of metazoan mitochondrial protein genes have unusual features. Some mitochondrial 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 et al., 1981).

The polymerase chain reaction (PCR) is an *in vitro* method for amplifying DNA without the use of microorganisms (Saiki et al. 1985; Mullis et al. 1986). In five years since the development of PCR, this technology has been modified for many uses (Innes et al. 1990) and has essentially revolutionized molecular biology (Guyer and Koshland 1990). PCR allows the rapid selection, isolation and amplification of DNA regions of interest from small amounts of tissue and can be used to help prepare DNA for sequencing. Highly conserved regions of mtDNA, such as rRNAs, cytochrome b and other loci, are routinely being amplified via PCR technology to generate fragments of DNA.

Although many sequences of mitochondrial genomes from various vertebrates and invertebrates have been obtained, the information about mtDNA from members of the Phylum Vestimentifera has never been reported. Vestimentifera, and its sister taxon, Phylum Pogonophora, are almost exclusively deepwater animals. The first specimen of a pogonophoran was dredged from Indonesian waters in 1900. Since that time more than 100 species have been described and more are being discovered. In most texts dated before 1964, pogonophorans treated as oligomerous deuterostomes. However, the segmented setiferous rear part of the body, the similarity of the setae to those of annelids, and the segmentation of the mesoderm indicate a protostome position of the pogonophorans (and therefore, the vestimentiferans) and suggest that they are closely related to the Annelida. Siewing (1975) and Cutler (1975) both argue that pogonophorans should more properly be considered as an intermediate between deuterostomes and protostomes or to be on a separate line of evolution. Jones (1985) argues that the pogonophora and vestimentifera are not particularly close to annelids in respect to coelomic organisation. The phylogenetic relationships of these animals is still not clear.

Based on investigation of sequences for various mitochondrial functional genes from several animal species, I have developed a suite of primers for studying the gene organization of mitochondrial genome from *Ridgeia*, a genus found deep sea vents off the Pacific Coast of North America. Using these primers, several fragments coding three different functional classes of genes were amplified and sequenced. Although the complete sequence of the mitochondrial genome from this phylum has not been determined, partial sequence reveals that the gene organization, lengths of some genes, genetic code and the overall codon usage of this genome show some differences from those of published animal

mitochondrial genomes. This genome exhibits a novel gene order, but bears certain similarities to those of *Drosophila* and of vertebrates.

RESULTS AND DISCUSSION

The Polymerase Chain Reaction

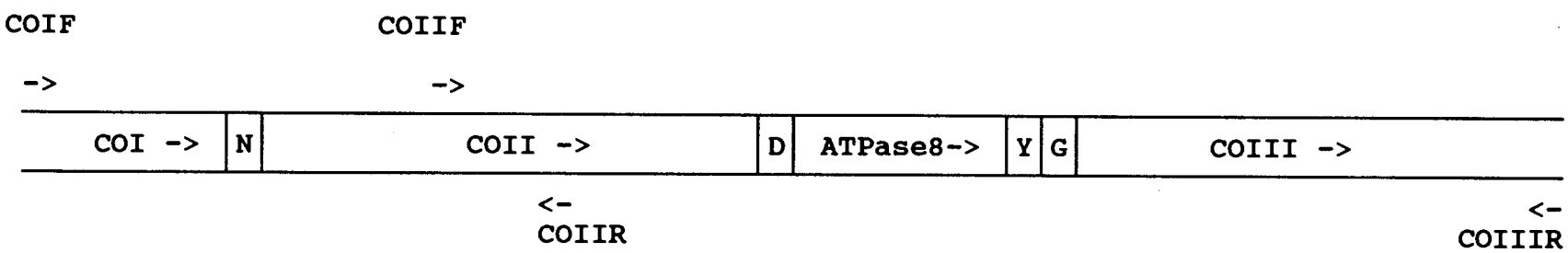
Several pairs of primers were used for PCR amplification of *Ridgeia* mitochondrial DNA. The sequences for these primers are listed in Table 1. The location of the genes and PCR amplification and sequencing strategy for *Ridgeia* mitochondrial DNA are shown in Figure 1. Knowledge of conserved gene regions has one more very important utility, it allows us to choose so-called "universal" oligonucleotide primers for amplification via PCR. This is of critical importance in the study of species which have not yet been sequenced. Indeed, there are some very conserved regions in the sequences from different phyla, making it possible to base primers on these sequences. The first pair of primers, COIF and COIIIR, was used for amplification of a fragment of 1985 base pairs, which contains COI (part), COII, ATPase 8, COIII (most) and four tRNA genes. The second pair of primers, 12SF and 16SR, was used for amplification of a segment of 1600 base pairs, across 12S rRNA and 16S rRNA genes. The third pair of primers, COIF and COIIR, was used for amplification of a small segment across the COI and COII gene junctions. When the sequence of each end of these fragments was determined, new primers was designed for amplification and sequencing of remainder of the fragments.

In most work to date, PCR has been used to amplify short regions of DNA, usually a few hundred base pairs long. Since Taq polymerase is capable of efficiently amplifying regions of DNA up to 2 kb (Saiki et al., 1988, Higuchi et al., 1988, Li et al., 1988), I tried to amplify some longer fragments. The PCR amplification products for these fragments are shown in Figure 2. In each case, double strand mtDNA was amplified using PCR. As shown in

FIGURE 1.

Location of the genes and PCR amplification and sequencing strategy for *Ridgeia* mtDNA. Abbreviations of protein names are as used elsewhere in text. The transcriptional polarity is shown by arrows. The tRNAs are designated with the single-letter amino-acid code: N, asparagine; D, aspartic acid; Y, tyrosine; G, glycine; and V, valine. The names and sequences of PCR primers are listed in Table 1. PCR conditions are described in text. A. A fragment obtained by using a pair of primers COIF and COIIIR; B. A segment amplified by using a pair of primers 12SF and 16SR.

A.


COIF
COIIF
->
->
COI ->
N
COII ->
D
ATPase8->
Y
G
COIII ->
<-
COIIR
<-
COIIIR

B.


12SF1
12SF
->
->
16SF
->
12S rRNA ->
V
16S rRNA ->
<-
16SR

FIGURE 2.

PCR products of the vestimentiferan mtDNA. Amplification products of mtDNA were analyzed by electrophoresis in a 1% agarose gel with ethidium bromide and photographed. Molecular weight markers (1 kilobase pairs, kb ladder) are as indicated (lane M). Lane A, B, C, D, E, and F are products amplified by using pairs of primers COIF and COIIR, COIF and COIIIR, COIIF and COIIIR, COIIIF and COIIIR, 12SF and 16SR, and CYTBF and CYTBR respectively. In all cases, double strand DNA was amplified.

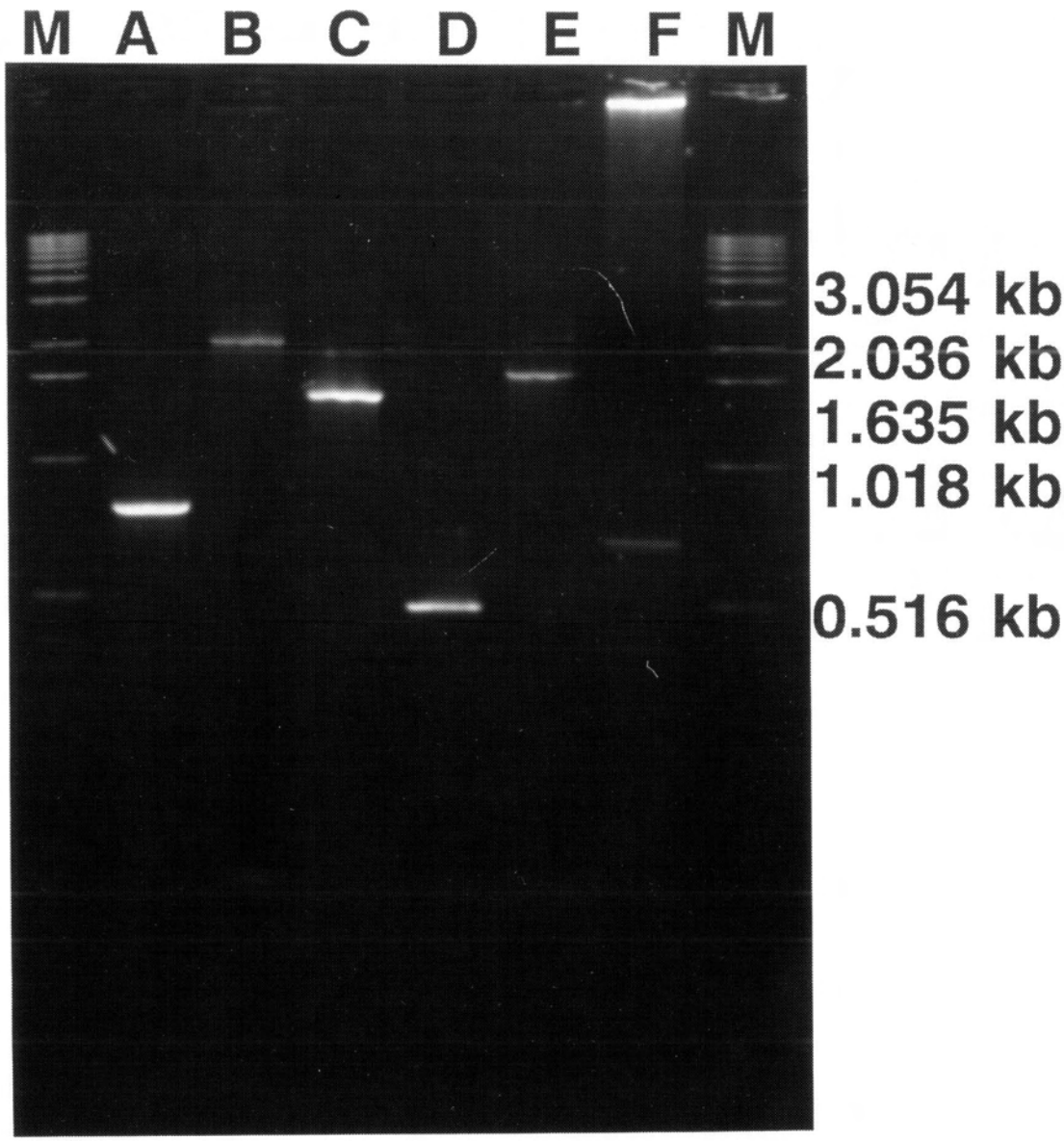

M A B C D E F M
3.054 kb
2.036 kb
1.635 kb
1.018 kb
0.516 kb

Figure 2, the sizes of PCR products range from 0.5 kilobase (kb) to 2 kb in length. The longest fragment with PCR amplification is about 3.2 kb.

Recently Ponce and Micol (1992) developed PCR conditions allowing the efficient amplification of DNA segments of up to 6 kilobases. They changed the PCR buffer content and found the absence of KCl to be optimal for the amplification of DNA molecules in the range of 3-6 kilobases. In fact, using standard PCR buffer, I have amplified fragments of more than 3 kilobases of mitochondrial DNA. If this buffer system and appropriate primers are used, it may simplify study of the whole mitochondrial genome.

Sequence and Location of Structural Genes

A total of 4,114 base pairs from *Ridgeia* mitochondrial genome was determined. Figure 3 shows the complete nucleotide sequence of three fragments. Two of them were amplified by PCR. Another was a clone. As shown in Figure 3A, the fragment of 1,985 base pairs contains four open reading frames that, from comparisons of inferred amino acid sequences of mitochondrial genes from several species, were identified as the genes for COI (part), COII, ATPase 8, and COIII (most). Those sequences lying between the COI and COII genes, between the COII and ATPase 8 genes, and between the ATPase 8 and COIII genes can be folded into the characteristic secondary structures of mitochondrial tRNA genes containing anticodons expected for $tRNA^{asn}$, $tRNA^{asp}$, $tRNA^{tyr}$, and $tRNA^{gly}$. All of the four protein genes and four tRNA genes would be transcribed in the same direction. The second fragment of 1,732 base pairs, shown in Figure 3B, codes for most of both ribosomal RNA genes (12S and 16S rRNAs) and $tRNA^{val}$ gene. The sequence between the two rRNA genes can also be folded into a characteristic secondary structure of mitochondrial tRNA gene.

FIGURE 3.

Nucleotide sequence of three coding regions from *Ridgeia* mitochondrial genome. The PCR amplification and sequencing strategy for these fragments is shown in Figure 1. The transcriptional polarity of individual genes is indicated by the arrows (->). Translation of the mitochondrial protein genes is indicated below the sequence using the single letter amino acid designations. Abbreviations of protein names are as used elsewhere in text. Stop codons are designated by asterisks. The limits of the tRNA genes are enclosed with brackets, and the anticodons are underlined. A. A fragment containing COI (part), COII, ATPase 8, COIII (part) and four tRNAs. B. A segment including part of the 12S rRNA, 16S rRNA and $tRNA^{Met}$. C. A clone of HindIII fragment coding cytochrome B gene (part).

A:

COI-->

```
CTCCTCAATCGGAGCATTCATTTCTTTCTCCTCGCTGCTATTCTTCATTTTCCTTATGTG 60
  S  S  I  G  A  F  I  S  F  S  S  L  L  F  F  I  F  L  M  W

AGAAGCCTTAGCTTCTCAACGAGGTGTATTAGCTTCACCCCATATGCCAACAGCCTTAGA 120
  E  A  L  A  S  Q  R  G  V  L  A  S  P  H  M  P  T  A  L  E

ATGACAAGAAACTCTTCCTCTAGACTACCACATATTCCAAGAAACAGGTCTAATTACTTC 180
  W  Q  E  T  L  P  L  D  Y  H  I  F  Q  E  T  G  L  I  T  S

                                  (   tRNA^asn-->
CCCTTCATTCTCAGCATCTTCTCTATATAAGCAGAAGCCAATTTTGGCATCTAACTGTTA 240
  P
                                     }COII-->
ATTAGAAGCTAGTCTTCCAGACCTGCTTAGATGGCCCACTGAGGACAATTAATATTTCAA 300
                                     M  A  H  W  G  Q  L  I  F  Q

GACGCTGCCTCACCTATCATAATTCAATTAGTAGCTCTTCACGACCACGCACTTACTATT 360
 D  A  A  S  P  I  I  I  Q  L  V  A  L  H  D  H  A  L  T  I

ATAATCATAGTTGTATCTTTAGTTCTCTATATACTTTATAGAATCTTAACTAACAAGTTT 420
 I  I  I  V  V  S  L  V  L  Y  I  L  Y  R  I  L  T  N  K  F

ACATGTCGAACACTCCTAGAAGCACAAGAAATTGAAACCATCTGAACAGTTCTTCCTGCC 480
 T  C  R  T  L  L  E  A  Q  E  I  E  T  I  W  T  V  L  P  A

ACAATTCTAGTCGTCCTTGCCCTCCCTTCCCTTCGCCTTCTCTACCTAATAGACGAGATC 540
 T  I  L  V  V  L  A  L  P  S  L  R  L  L  Y  L  I  D  E  I

TCTCAGCCCACCCTTACAGTAAAAACAATTGGGCATCAATGATATTGAAGATATGAATAC 600
 S  Q  P  T  L  T  V  K  T  I  G  H  Q  W  Y  W  R  Y  E  Y

TCCGACTTCTTAAACTTAGAATTTGATTCTTACATGCTCCCTACCGAAGAACTTCAAGAT 660
 S  D  F  L  N  L  E  F  D  S  Y  M  L  P  T  E  E  L  Q  D

GGAGAATTCCGCTTATTAGAAGTAGACCATCGAATAGTAATTCCCATGCAAACAGAAGTC 720
 G  E  F  R  L  L  E  V  D  H  R  I  V  I  P  M  Q  T  E  V

CGACTTCTAGTAACCGCTGCAGACGTAATTCACTCATGATGTGTACCTAGCCTAGGAATC 780
 R  L  L  V  T  A  A  D  V  I  H  S  W  C  V  P  S  L  G  I

AAACTAGACGGAATTCCGGGCCGCCTAAACCAAACAACCCTTTCTATTAACCGGCCAGGA 840
 K  L  D  G  I  P  G  R  L  N  Q  T  T  L  S  I  N  R  P  G

ATTTTCTATGGCCAATGCTCAGAAATATGCGGAGCCAACCACTCATTTATACCAATTGCC 900
 I  F  Y  G  Q  C  S  E  I  C  G  A  N  H  S  F  I  P  I  A
                                                          (
CTAGAAGTAATTGACCATCCCTCCTTCACCCAATGAGTAATAACATTTAGAGAATAGAAT 960
 L  E  V  I  D  H  P  S  F  T  Q  W  V  I  T  F  R  E  *
```

```
tRNA^asp-->                                                          )
TCTAGTTAAATAATAATATAGGACTGTCAGCCCTAAGTTACTAAATATAGTGAATTCTGA 1020

ATPase8-->
ATGCCTCATCTAGCCCCTCTAAATTGAATCCTCCTCCCTCTTTTCTTTTTATTCTCCCTA 1080
 M  P  H  L  A  P  L  N  W  I  L  L  P  L  F  F  L  F  S  L

CTTTTACTCGCTTCAATTACCTGATGAAACCAATTAATTTCTGTCCCTCAACTCAAATCT 1140
 L  L  L  A  S  I  T  W  W  N  Q  L  I  S  V  P  Q  L  K  S
                                                 {  tRNA^tyr-->
AAATCTAAACAGGCCCATTCTATATCCCCTTGAAAATGAAACTAAAAAAGATGGCCGAGTT 1200
 K  S  K  Q  A  H  S  I  S  P  W  K  W  N  *
                                                 {  }
ATAGGCAGAAGATTGTAATCCTTCCCACGGGCTTTCCCTCTTTTTACTTTCTCAGTATAA 1260

tRNA^gly-->                                                    }COIII-->
ATTTGTACAATTGCCTTCCAAGCAATAAGTTTGACATTCAAAGAAAGTAATGATCCGCCA 1320
                                                          M  I  R  Q

ACCTTTCCATGTATTAGAATATAGACCATGACCATTTCTAGTCGCCGTCGGTGTTTTAGC 1380
  P  F  H  V  L  E  Y  R  P  W  P  F  L  V  A  V  G  V  L  A

CATTACATGCGGTGCCGCAGCATGATTTCACAATCACGGTGCCCTATGCCTAATTATTGG 1440
  I  T  C  G  A  A  A  W  F  H  N  H  G  A  L  C  L  I  I  G

TTTAACACTTACCACTTTAACTTCAATTATCTGATGACGAGATGTAATTCGTGAAGGAAC 1500
  L  T  L  T  T  L  T  S  I  I  W  W  R  D  V  I  R  E  G  T

TTATCTAGGCTTCCACAGATCTGTAGTATCTAGAGGGCTACGCTGAGCAATAATTCAATT 1560
  Y  L  G  F  H  R  S  V  V  S  R  G  L  R  W  A  I  I  Q  F

TATTCTTTCAGAAGTACTCTTTTTCGCAGCTTTCTTCTGAGGCTTCTTTCACAGAAGTCT 1620
  I  L  S  E  V  L  F  F  A  A  F  F  W  G  F  F  H  R  S  L

GGCCCCTACTCCAGAAATTGGGTGCACTTGACCCCCAACCGGAATTAACCCTATCAACCC 1680
  A  P  T  P  E  I  G  C  T  W  P  P  T  G  I  N  P  I  N  P

ATTCTCCATCCCCCTGCTAAACACAGCCGTTCTTCTAGCATCAGGAGTCACAGTAACTTG 1740
  F  S  I  P  L  L  N  T  A  V  L  L  A  S  G  V  T  V  T  W

AGCCCACCATAGGGTAATAAACAAATCTCGAACAGAAACTCTTCAAGCCCTTTCCCTTAC 1800
  A  H  H  R  V  I  N  K  S  R  T  E  T  L  Q  A  L  S  L  T

AGTAATCCTAGGAGTTTACTTTACTTTCCTTCAAGCAGGAGAATATATAGAAGCCCCTTT 1860
  V  I  L  G  V  Y  F  T  L  L  Q  A  G  E  Y  I  E  A  P  F

TACTATCGCCGATAGAGCCTATGGTACCCTATTTTATGTCTGTACAGGCTTCCACGGCAT 1920
  T  I  A  D  R  A  Y  G  T  L  F  Y  V  C  T  G  F  H  G  I

ACATGTCCTAGTCGGAACTATATTTTTATCTATATACTTAATTCGAACATTTCTATATCA 1980
  H  V  L  V  G  T  I  F  L  S  I  Y  L  I  R  T  F  L  Y  H
```

TTTCT 1985
F

B.

12S rRNA-->
TCTCGGAAATTATGGCTTCAGATAATAACCTTTTCAAAAGAAGTCTAATTCTATCAATAT 60

TTCAAATCCAAACATATCTTTATAACAGTCCCATGAAAGCCTAAATAAAATCAAGGATTA 120

GATACCCTTTTATATCTAGGCCCAAAATTATCCAGGGCACTACAACCACAGGTTTAAAAC 180

CACAAAGAAATTGGCGGTACCTAAATCCAATCAGGGGAACCTGTCCTTTAACTCGAAAAT 240

TACCTCTTTGAATATACCTCAGCTTGTATACTGCCGTCGTCAGCCCACCATAAAAATGTA 300

AGTGAGCTAACAGATTTTTATCTTCACGTCAGGTCAAAGTGCAGCCTATGAGAAGGAAGA 360

GATGGGTTACAATTTAAATTCTAAATATGAATTATTAATGAAAATAAATATAAAAGTGGA 420

CTTGAAAGTAATTAAAAGTAACATACTTTTATGAATATGGCAACCGAAGGTGTGCACACA 480

TCGCCCGTCACTCTCGCCGAAAGGGGAGATAAGTCGTAACATAGCAGGTGTAATGGAAAT 540
{ tRNA^{val}-->
TGTACCTTCAAAATACAGCATCTAAAGAATGCCTTTCACTTACACTGAAAAGAGAATTTT 600
} 16S rRNA-->
TAAAATTTATTTTGAAATATCCTAAACCCTCTATTTTTCTTTAAACTTGAATAAAAATCT 660

ATCTACAACTCTATCATCCCAAATCTCCGCTATCTTAGTACTGCAAAGGAAATTATAAAT 720

TATTAATAAGTAAAAATAAAACCTTATACCTTCGTGCATTATGGCTTAGCAAGCCAACTC 780

TAATTCTAGCCTATCCCGAAATCTTCACGAGCTGATAAACATTTGTATAAGAACTCACTA 840

CCGCATGTTTCAAAATGCCTAGAAAAATTTTATCAGAAGCTACATACCTACCGCGCAGAC 900

TATAGCTGGTTCCCTAAAAGCTTCACATTAGTGAAACAAGATATATTCTCTGATAAATAA 960

TAAAGGAAAAGCTCTATTATAAAAGCTAATTTCGAATCCTACCCCCTAAAGTAAGCTTAG 1020

AAACTGCTAACTTTTAAATAACGTTATAGTATAAAAACTCTTTTTACAAAATCATAAACA 1080

AATTGCTACAACCTATTAGGGATTATTTATCAACTTTCCTAAATAAAAATTCTGCTAAGA 1140

TTAGTATTCATTCCCTCCTACAACACCTATAAAACCCCCTTTCCCCCAAAAATGAAATTC 1200

CTTATTATAAGGAACTCGGCAAATACAAGCTCCGACTGTTTAACAAAAACATTGCCTCTT 1260

GATTTATAAATAAGAGGTTCATCCTGCCCAATGACTTTAGTTCAATGGCCGCGGTACCCT 1320

GACCGTGCAAAGGTAGCATAATCACTTGCCCCTTAATTAGGGGCTGGCATGAATGGACAC 1380

```
ACGAAAGCTTAACTGTCTCATAATAACTAATAAAAATTAATCTTTAAGTGAAAAAGCTTA 1440

AATTCCATTGCAGGACAAGAAGACCCCGTTGAGCTTTATTCTCTATAGACCACCACTAAT 1500

ATTATACTATATCATAAACCTAAAAAAGAATTTAGTTGGGGTGACTAAGGAACATCAAAA 1560

TCTTCCTTTTATTAACCAGGGCTATTACCCAAAACAATTGACCCATAACCATGAACAAAA 1620

AAATAAGCTACCACAGGGATAACAGGCTAATCTTTCTCAAGAGCCCAAATTGTCAGAAAG 1680

GATTGGCACCTCGATGTTGGCTTAGGGGCCCCTAATAGTGCAGAAGCTATTA          1732
```

C.

```
HindIII                Cytochrome B-->
AAGCTTGCAACAATGGTTGCTCGGCGGTTATTCCGTGGGTAATCCAACGTTGCAACGCTT 60
  S  L  Q  Q  W  L  L  G  G  Y  S  V  G  N  P  T  L  Q  R  F 20

CTTCTCCCTGCATTACCTCTTGCCATTTATCATTGCGGCCTTGGTTGGTTTGCATGTGTG 120
  F  S  L  H  Y  L  L  P  F  I  I  A  A  L  V  G  L  H  V  W 40

GGCTGTGCATCATGTGGGACAAAACAACCCGATTGGCATCGACGTGAAAACCAAGGAAGA 180
  A  V  H  H  V  G  Q  N  N  P  I  G  I  D  V  K  T  K  E  D 60

CACATTATCCTTCCATCCTTATTACACCATGAAAGACGGGTTTGCGATTGCCGTGTTCTT 240
  T  L  S  F  H  P  Y  Y  T  M  K  D  G  F  A  I  A  V  F  F 80

TATGATTTTCGCGCTGTTTGTGTTCTATCTCCCAGACGCGCTTGGACATGCAGACAATTA 300
  M  I  F  A  L  F  V  F  Y  L  P  D  A  L  G  H  A  D  N  Y 100

TCAGCGGGCTGATCCTATGAAAACACCGCCGCATATTGTCCCTGAATGGTATTTCTTGCC 360
  Q  R  A  D  P  M  K  T  P  P  H  I  V  P  E  W  Y  F  L  P 120
                                HindIII
GTTCTACGCGATTTTACGAGCTGTGCCTGATAAGCTT                         397
  F  Y  A  I  L  R  A  V  P  D  K  L                          132
```

According to its anticodon, it is apparently $tRNA^{val}$. The 12S and 16S rRNA genes and the $tRNA^{val}$ gene are transcribed in the same direction. The third sequence (397 base pairs) was from a clone containing a HindIII fragment. Comparison of this sequence with published sequences from other species show that the sequence codes for cytochrome B (centre part). The base composition of the protein and both rRNA genes from *Ridgeia* mitochondrial DNA is shown in Table 2. In all cases except for the cytochrome b gene, a low G content is found.

The majority of the protein genes and the rRNA genes are punctuated by at least one tRNA gene. These tRNAs are thought to serve as a recognition signal(s) for mitochondrial RNA processing (Ojala et al., 1981). Short intergenic sequences of variable length, from one to ten nucleotides, are found between $tRNA^{asp}$ and ATPase 8, and between the COI and $tRNA^{ser}$ genes. Some genes are contiguous, however, such as $tRNA^{ser}$ and COII, 12S rRNA and $tRNA^{val}$, and $tRNA^{val}$ and 16S rRNA genes, while other genes apparently overlap. An overlap of 1-3 nucleotides is located between the stop codon of the ATPase 8 and $tRNA^{tyr}$. An apparent four-nucleotide overlap is found between two tRNA genes, $tRNA^{tyr}$ and $tRNA^{gly}$. A one-nucleotide overlap between two tRNA genes, such as $tRNA^{cys}$ and $tRNA^{tyr}$, that are both L-strand encoded has been reported in human and amphibian mitochondria but is not observed in the bovine or mouse sequences. The two-nucleotide overlap is observed between $tRNA^{gly}$ and the start codon of the COIII gene. In all cases, these short overlaps occur on the same strand. These observations indicate that, in some mitochondria, a single precursor RNA molecule could not, in principle, be processed to yield both of these tRNAs in their mature forms. Small overlaps also occur between some genes in vertebrate and invertebrate mtDNAs, the largest of which is a 40 to 46 nucleotide overlap of the 3' end

Table 2. Base composition in *Ridgeia* (Phylum Vestimentifera) mitochondrial genes

	Base composition % of total			
Gene	Guanine	Adenine	Thymine	Cytosine
COI	12.8	25.5	32.6	29.0
COII	14.6	31.6	27.9	25.9
COIII	16.1	26.8	31.5	25.6
ATPase8	6.8	27.2	35.2	30.9
CytB	22.9	21.4	31.0	24.7
12S rRNA	16.9	36.8	27.1	19.2
16S rRNA	13.2	36.6	28.6	21.6

of the ATPase 8 gene and the 5' end of the ATPase 6 gene in vertebrates and most invertebrate genomes (Anderson et al., 1981; Smith et al., 1990).

Genomic Organization

Comparison of the fragment containing COI, COII, ATPase 8, COIII, and four tRNA genes with corresponding regions from different species reveals that the organization of *Ridgeia* mitochondrial DNA is different from those of vertebrates and other invertebrates (Figure 4). In particular, the two protein genes, ATPase 6 and ATPase 8 are not adjacent in *Ridgeia*. As shown in Figure 4, in both vertebrates and invertebrates where the ATPase 8 gene is present, these two genes are always contiguous and partially overlapping. These two genes are translated in different reading frames on the same strand. In the vestimentiferan, the ATPase 6 gene was not found between the ATPase 8 and COIII genes, and as a replacement, the two tRNA genes, $tRNA^{tyr}$ and $tRNA^{gly}$ are located between the ATPase 8 and COIII genes. The ATPase 8 gene in *Ridgeia* mitochondria has an independent and complete reading frame and a set of orthodox mitochondrial initiation and termination codons, although all three nucleotides of the stop codon overlap with $tRNA^{tyr}$ gene. The ATPase 8 gene was not located in either of two nematode mtDNAs (Okimoto et al., 1992), nor in the mollusc, *Mytilus* (Hoffmann et al., 1992).

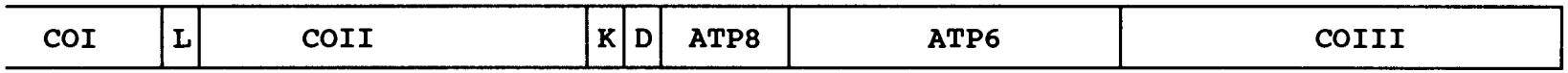
Further analysis of gene organization reveals that the kind and number of tRNA genes lying between protein genes are quite different. For example, between the COI and COII genes, in the vestimentiferan and *Drosophila yakuba* there is one tRNA gene, but in the vestimentiferan it is $tRNA^{asn}$, while *Drosophila* has $tRNA^{lys}$. In vertebrates, there are two tRNA genes, having opposite direction of transcription. In both sea urchin and sea star, this

FIGURE 4.

A comparison of gene order in linear mitochondrial molecules containing COI to COIII genes. Abbreviations of protein names are as used elsewhere in text. The tRNAs are designated with the single-letter amino-acid code. The transcriptional polarity of tRNAs from Phylum Vestimentifera are shown in Fig. 3. A. Phylum Vestimentifera; B. *Drosophila yakuba*; C. sea urchin and sea star; D. vertebrate.

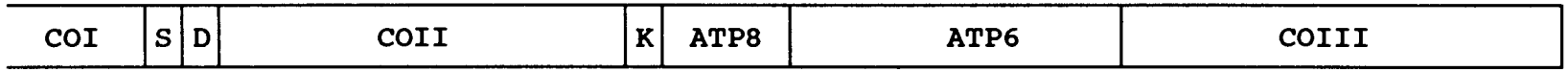
A. VESTIMENTIFERA

B. DROSOPHILA



C. SEA URCHIN AND SEA STAR

D. VERTEBRATES



region includes one tRNA gene and the ND4L gene, but the ND4L and COII genes are contiguous. In the region between the COII and ATPase 8 genes, *Drosophila* consists of both a $tRNA^{lys}$ and $tRNA^{asp}$ genes. Vertebrates, sea urchin, sea stars, have only $tRNA^{lys}$. In the vestimentiferan, the $tRNA^{lys}$ was not found, only the $tRNA^{asp}$ is located in this region.

The sequence of a fragment across the small and large rRNA genes and $tRVA^{val}$ gene was determined (Figure 3B). The gene order in the vestimentiferan is small rRNA-$tRNA^{val}$ -large rRNA, all transcribed in the same direction, as in vertebrate (Roe et al., 1985; Desjardins and Morais 1990) and *Drosophila* (Clary and Wolstenholme 1985) mitochondria genomes. This gene order is different from that of sea urchin (Jacobs et al., 1988; Cantatore et al., 1989). The two ribosomal genes 12S and 16S are separated by the ND1, ND2, and a cluster of 15 tRNA genes (Jacobs et al., 1988; Cantatore et al., 1989). In the sea star, both ribosomal RNA genes are punctuated by two tRNA genes. The transcriptional polarity of the rRNA genes is opposite (Smith et al., 1989).

Protein-Coding Genes

Similarities of the nucleotide sequences and predicted amino acid sequences of the five vestimentiferan mitochondrial protein genes and the corresponding genes of mouse, frog, *Drosophila yakuba*, sea urchin and *C. elegans* are summarized in Table 3. The degree of similarity of the different protein coding genes with their vertebrate and invertebrate homologues shows a broadly similar pattern of relative conservation. The degree of conservation for the nucleotide sequences and amino acid sequences of these protein genes is COII > COIII > Cyt B > COI > ATPase 8. It should be noted that the degree of similarity of the COI gene with its counterparts is comparatively low. The probable reason may be

Table 3. Similarities between mitochondrial protein genes in different taxa

	Protein genes of *Ridgeia* (Phylum Vestimentifera)									
	COI		COII		COIII		CytB		ATPase 8	
Species	N%	A%	N%	A%	N%	A%	N%	A%	N%	A%
YAKUBA	51.9	41.5	62.4	60.1	63.5	59.1	52.9	48.5	49.7	33.3
MOUSE	56.5	44.6	63.5	59.6	60.2	53.8	55.5	52.3	48.8	29.6
FROG	47.7	30.8	63.0	59.2	62.7	60.0	52.4	48.5	44.6	31.5
URCHIN	52.9	43.1	61.1	57.5	58.6	52.9	51.7	43.2	43.6	20.4
NEMATODE	36.9	23.1	47.6	42.1	49.1	41.8	45.0	40.2	-	-
MEAN	49.2	36.6	59.5	55.7	58.8	53.5	51.5	46.5	46.7	28.7

Notes: N%: Nucleotide sequence similarity %.

A%: Amino acid sequence similarity %.

that only the 3' end of this gene was obtained and compared. The 3' end region of the COI gene is the most variable region within this gene (see below). It is also noted that the relative degree of similarity of the nucleotide sequence for ATPase 8 gene with its counterparts is 46.7%, which is the highest similarity reported for this gene for members of different phyla, although the amino acid sequence of this gene shows a lower similarity. It is surprising that when the nucleotide sequence of the ATPase 8 gene is compared with that of *Drosophila yakuba*, 49.7% similarity is observed.

Each of the polypeptides is encoded in the same economical fashion as the corresponding mitochondrial genes of other animals, exhibiting a complete absence of non-coding nucleotides within genes. In at least one gene (COI), the stop codon is absent and is presumably supplied by post-transcriptional polyadenylation at a 3' terminal uridine residue.

Alignment of the vestimentiferan nucleotide sequences and inferred amino acid sequences with those present in *Drosophila yakuba*, mouse, frog, sea urchin and nematode mtDNAs is shown in Figures 5 and 6.

Alignment was carried out by first conducting all pairwise alignments of the inferred amino acid sequences using the Align program (Pearson and Lipman, personal communication), then comparing the pairwise alignments among all six sequences to minimize the total number of insertion/deletion events required. For most sequences, little adjustment of the optinal pairwise alignments was required. For the most divergent sequences, those of the nematode, *C. elegans*, the alignments provided by Okimoto et al. (1992) with *Drosophila* and mouse were used. Finally, the amino acid sequence alignments (Figure 6) were transferred to the nucleotide sequences (Figure 5). The COII gene of *Ridgeia* differs in size by the addition of three nucleotides (one amino acid codon) compared

FIGURE 5.

A comparison of the nucleotide sequences of vestimentiferan and other animal mitochondria: A. COI; B. COII; C. ATPase 8; D. COIII; and E. Cyt B. The comparisons include *Ridgeia* (Rp), *Drosophila yakuba* (Dy), mouse (Mm), frog (Xl), sea urchin (Sp) and nematode (Ce). All nucleotide sequences from the vestimentiferan were aligned so as to maximize the alignment of sequences conserved in the other genomes. Nucleotide identities to vestimentiferan are indicated by dots. Gaps are indicated by dashes.

A. COI

```
Rr  CTCCTCAATC GGAGCATTCA TTTCTTTCTC CTCGCTGCTA TTCTTCATTT TCCTTATGTG  60
Dy  G..TA.T..T ..GT..ACT. ....A..A.T AGGAA.TT.. ..T..TT.C. ATA....T..
Mm  ...T..T..A ...T....T. ....AC.AA. AG.TG.T..C A..A.G..C. .TA.A..T..
Xl  ...A..T... ..GT.C..A. .....C.TGT TG.CG.AA.T A.GA.A..A. ..A....C..
Sp  .........T ...T..AC.. .C..CG.AGT GG.TA..... ..T...C.C. ..T.A..C..
Ce  TG....TTAT ..TT.TA.T. ..AGAACTG. AGGA..AT.C ..A..T.... ATG.AT.A.T

Rp  AGAAGCCTTA GCTTCTCAAC GAGGTGTATT AGCTTCACCC CATATGCCAA CAGCCTTAGA  120
Dy  ....AGT... .TG....... ..CAA...A. TTA.C..ATT ..AT.AAATT ..T.TA.T..
Mm  ...G.....T .....AA... ...AA...A. .T.AGT.T.G T..GCTT... ..AAT.....
Xl  ......A..T ..AG..A... ...AA..TAC CA...ACGAA TTA.CAT... .CATA..G..
Sp  G..G.....C .......... .G.AA.G.A. CA.CC..GAG TTCTCA.ACG .CT.AC....
Ce  ...GT.T..C TT.AG.T.T. .TTTA...A. TAGAGATTAT T..TCTAAT. G.AGACCT..

Rp  ATGA---CAA GAAACTCTTC CTCTAGACTA CCACATATTC CAAGAAACAG GTCTAATTAC  177
Dy  ....---T.T C...A.ACA. .C.C..CTG. A..T.G..AT TCT...TT.C CA..TT.A..
Mm  .....---.TT C.TGGCTGC. ...C.CCA.. T....C.... G.G...C..A CCTATG.A.A
Xl  G.....---.TT C..GGCTGC. .CACTCCT.. ...T.CC..G A.GACC.GCC TCG.TCAA.T
Sp  G...CAAT.C ACCT.CT... .C.CTTCTC. .....CC... G.T.....C CCTCT.CC.T
Ce  G.AT---TGT ATG.G.AA.T A.G..TTTGG T....GT.AT ..GTCTGAGA T.TATT...G

Rp  TTCCCCTTCA TTCTCAGCA                                              196
Dy  AAAT------ ---------
Mm  AGTAAAA--- ---------
Xl  CAA..A.CA. A.AATTAA.
Sp  AATTAT.GT. AAG------
Ce  AA.TA..AG. ..AAA.AAT
```

B. COII

```
Rp  ATGGCCCACT GAGGACAA-- -TTAATATTT CAAGACGCTG CCTCACCTAT CATAATTCAA  57
Dy  ...T.TACA. ...CTA.T-- -...GGT..A .....TAGA. .T..T...T. A..GGAA...
Mm  ......T..C C.TTC....- -C.TGGTC.A ........CA .A..C..... T...GAAG.G
Xl  .....A...C C.TC.....- -...GGT... ........A. ....T..A.. T...GAAG..
Sp  ....GAACT. ...C...G-- -..TGGTC.A .....T..AT ....C...C. T..GGAGG.G
Ce  ..TAATA.T. TTTTT...GG A.AT.AT..A .T.TTTCAGC ATAGTTTAT. TGCT.G.T.T

Rp  TTAGTAGCTC TTCACGACCA CGCACTTACT ATTATAATCA TAGTTGTATC TTTAGTTCTC  117
Dy  ...A.TTT.T ....T..T.. T...T.ATTA ...T..G.A. ..A..AC.GT A.....AGGA
Mm  C..A..AA.T .C..T..T.. .A....A.TA ...G.TT..C ..A..AGC.. C.....C...
Xl  ...C.TCACT .C........ TA.C..C.TA GCCG.TT.TC .TA..AGTA. GC.......T
Sp  C.CAC.TACT .C.....TT. T...T.A.T. G.AC.T.C.C .CA..AC.AT AC.....T.T
Ce  A...ATTGAT ....TAGGTT TAATTG..G. T.AT.GT.GG G....T..GT G..T...ACT
```

```
Rp  TATATACTTT AT---AGAAT CTTAACTAAC AAGTTTACAT GTCGAACACT CCTAGAAGCA  174
Dy  ...T..A.G. T.---.T.T. A..TTT...T ..T.A.GT.A A....TTT.. TT..C.T.G.
Mm  .....CA.C. CG---CT... A.....A.CA ..AC.A...C A.AC..GCAC AA....T...
Xl  ..C..TA..A C.---.TT.. AA......CT ..AC.A..TA A.AC..AC.. AA.G..C...
Sp  ...GGGT.AG T.---TCCT. GC.TGTATC. TCTAA...TA AC...TTTT. .T.T..G.G.
Ce  .TAT.GT..G GATATTT... T..TGG..CT TTT.A.TTTA AAA...A.AA AA.T..GTAT

Rp  CAAGAAATTG AAACCATCTG AACAGTTCTT CCTGCCACAA TTCTAGTCGT CCTTGCCCTC  234
Dy  ...CTT.... ...TA..T.. ...TA....C ..A..T.TT. ..T..T.AT. TA....T..T
Mm  ......G... .......T.. ...TA....A ..A..TGT.. .C..TA..A. AA....T...
Xl  .....G..C. ...TAG.G.. ...TA..A.A ..A..T.TT. GC..CA..A. AA.......T
Sp  .....GT.A. ....A..T.. ......GA.. .....TCT.. .CT..A..T. AA.......T
Ce  ..GTTTGG.. ..TTAT.G.. T.GTA..T.. ..AA.T.TT. ..T..T.AA. A.AAATAG.T

Rp  CCTTCCCTTC GCCTTCTCTA CCTAATAGAC GAGATCTCT- --CAGCCCAC CCTTACAGTA  291
Dy  .....AT.A. .AT.A..T.. TT..T....T ..A..TAA.- --G.A..AT. AG.A..TT..
Mm  ..C..T..A. ..A....A.. TA........ ..A...AAC- --A.C...GT AT.A..C..T
Xl  ..A.....  .TA.C..A.. TT.......T ..AG.TAA.- --G.T..ACA .T.A...A.T
Sp  ........C. AA..C..T.. ......G... ...G.TAAA- --G.C...TT .T.G..TA.T
Ce  .....A...A .G...T.A.. TTATTAC.GA TTA..AAA.T TAG.TAGA.A TT.A.....T

Rp  AAAACAATTG GGCATCAATG ATATTGAAGA TATGAATACT CCGACTTCTT AAACTTAGAA  351
Dy  ....GT.... .T........ ...C.....T ........T. .A..T..TAA T..TA.T...
Mm  .....C..A. ....C..... ...C.....C ..C.....TA .T....ATGA .G..C..TGC
Xl  ...G....C. .C..C..... ...C.....C ..C.....TA .TA...ATGA GG.TC.CTC.
Sp  ..GG.GT.C. .T.....G.. ...C...... ..C..G...A .G......AA .G..C.T...
Ce  ..GGTT.CA. .A.....G.. .......... ........TA GA..TA.TCC TGGG......

Rp  TTTGATTCTT ACATGCTCCC TACCGAAGAA CTTCAAGATG GAGAATTCCG CTTATTAGAA  411
Dy  ........A. .T..AA.T.. ...AA.T... T.AGC.AT.. AT.G...T.. A........C
Mm  ........A. .T..AA.... A..AA.C..C ..AA..CC.. .T...C.A.. AC.GC.....
Xl  .....C.... .T..AA.T.. A..TA.T..C ...ACCCC.. ..C....... GC.GC.....
Sp  ..C..C.... .T...G.A.. ....TC...C G..TCCTT.. .TA.CCC... ..........
Ce  .....C.... .T..AAAAT. ACTA..TC.. T.AAGTTTA. .T...CCA.. T.........

Rp  GTAGACCATC GAATAGTAAT TCCCATGCAA ACAGAAGTCC GACTTCTAGT AACCGCTGCA  471
Dy  ..T..TA... ..G..A.TT. A..A..AA.T T..C..A.T. ..A..T.... ...A..C...
Mm  ..T..TA.C. ..G.C..TC. G..A..AG.. CTTCC.A... .TA.AT..A. TT.AT...A.
Xl  ..T..TA... ........G. C..A..AG.. T.TCC.AC.. .....T.... T..A..C.A.
Sp  ..G...A.C. ..T.G..CC. .......... .ACCCCA.A. ..G....... GT..T.....
Ce  ..T..TA... .TTGT..T.. ...TTGTG.T ..TA.CA.T. .TT..TGTA. T..AT....T

Rp  GACGTAATTC ACTCATGATG TGTACCTAGC CTAGGAATCA AACTAGACGG AATTCCGGGC  531
Dy  ..T.....  .T..T...AC A..C..AGCT T.....G.A. .GG.T..... ..C...T..A
Mm  .....CC.C. ........GC A..C..CTC. ......C.T. ..ACT..T.C C..C..A...
Xl  .....CC.C. ....G...GC ......CTC. T.G..TG... ..AC...T.C ...C..A..A
Sp  ..T...C.A. ....C...GC ...T..CTC. ..T....CT. .GA.G..T.C .G.C..A..A
Ce  ..T..T.... .TG.T..GGC AT..AA.TCT T..TCTG.A. ..T....T.C T..AAGA..A
```

```
Rp  CGCCTAAACC AAACAACCCT TTCTATTAAC CGGCCAGGAA TTTTCTATGG CCAATGCTCA  591
Dy  ..AT....T. ....T.ATT. ..T..... ..A.....GT .A..T..... T.....T...
Mm  ..A.....T. ..G....AG. AA.ATCA... ..A.....GT .A..... .........T
Xl  ..A..TC.T. .....T.AT. .AT.GC..CT ..T..G...G .A..T..C.. A.....T...
Sp  .....C.... .G..C..AT. C.T.GCAGCT ..CA.....G .G..... ...G.....C
Ce  ATTT...GAA C.TTT.GTTA CAGGT..CCT ATAGTG...G ....T..... T.....T...

Rp  GAAATATGCG GAGCCAACCA CTCATTTATA CCAATTGCCC TAGAAGTAAT TGACCATCCC  651
Dy  .....T.... .G..T..T.. TAGT.....G .......TAA .T...AGTG. .CCTGTAAAT
Mm  .....T..T. ..T.T..... TAGC.....G ..C....T.. .....A.GG. .CCA.TAAAA
Xl  .....T.... ....A..... .AGC..... .......TAG .T....C.G. ACCG.TAA..
Sp  .....T.... .G..T..... TAG...C..G .....A.TTA ....GTCTG. GCCATT.AAT
Ce  ..G..T..T. ....A..T.. TAGT..... ........TT ....G....C ATTATTGGAT

Rp  TCCTTCACCC AATGAGTAAT AACATTTAGA GAA---                          684
Dy  AAT..T.TTA .....A.TTC T.G.AA..AT TCT---
Mm  .AT...GAAA .C...TCTGC TT..A.A.TT ------
Xl  GA...TGAAA .C...TCTTC .T..A.ACT. ...GCA
Sp  A....TGAAA .C.....T.C TCA..ACTT. ...GAA
Ce  AAT..T.AAA G....TGTT. TGGTAC..T. ...---
```

C. ATPase 8

```
Rp  ATGCCTCATC TAGCCCCTCT AAATTGAATC CTCCTCCCTC TTTTCTTTTT ATTCTCCCTA  60
Dy  ..T..A..AA ....A..AA. T.GA...T.A T.A..ATT.A ..G.T....C TA.TA.AT.T
Mm  .....A..A. ...ATA.ATC ..CA...T.T A..ACAAT.A .C.CA.CAA. .A.TA.....
Xl  .....A..GT ..AA...AGG CCCA...T.. ..AA...TAA .C..T.CC.G .C.TGT...T
Sp  G....A..A. ...AATT.GC TTGA...... G.AAA.TT.T CCC..A...G .GCT...G..

Rp  CTTTTACTCG CTTCAATTAC CTGATGAAAC CAATTAATTT CTGTC---CC TCAACTCAAA  117
Dy  A.....T.TT G...T....A T.AT.ATTCA T.TA..CCAA ..TCA---- .A..TCT..T
Mm  T..A.CT.AT T.CA.C.A.A AGTC.C.TCA ...AC.T.CC .AC.GGCA.. .TC..CA...
Xl  T.AAC.T.TA TCC..CCA.A AGTT.T...A ..CAA.GCA. T.AATGAA.. AACTACAC..
Sp  T.AA..G..A T...CT.ACT A.T.AAT.G. TTTCC.CC.A ACAG.GCGGG ....TC.TCT

Rp  TCTAAATCTA AACAGGCC-- -CATTCTATA TCCCCTTGA- -----AAATG AAAC------  162
Dy  GAATT.AAA. .TATTAATTT AA........ ---AAC...- ---------- .---------
Mm  ..ACT.A.A. CCATAAAAGT AA.A------ A........G AATTA..... ..CGAAAATC
Xl  A.C.C.GAA. ..TCTAAACC TA.C------ ---.....A ACTGACC... ..C.------
Sp  ...TCC.TA. CTTTAAAA-- -A.GA.C.C. A.TAA....- -----C.... .CTA------

Rp  ---------- ---------- -------
Dy  ---------- ---------- -------
Mm  TATTTGCCTC ATTCATTACC CCAACAA
Xl  ---------- ---------- -------
Sp  ---------- ---------- -------
```

D. COIII

```
Rp  ATG---ATCC GCCAA----- -CCTTTCCAT GTATTAGAAT ATAGACCATG ACCATTTCTA  51
Dy  ...TCT.CA. A.TC.AATCA C.....T... T..G.T..T. ....C..... ...T..AAC.
Mm  ...---.C.. A....ACTCA TG.A.AT..C A..G.TA.TC CA..T..... ......AACT
Xl  ...---GCA. A....GCACA CG.C.A...C A..G.C..CC CC..C..T.. ....C.AACG
Sp  ...---GCTA TT...---CA C..A.AT... T..G....CC .A..C..... ...C..AGAC
Ce  ---------- --AT.TTTCA TAA...T... A.T...AG.C T.TC.AGG.A TG...A.AAT

Rp  GTCGCCGTCG GTGTTTTAGC CATTACATGC GGTGCCGCAG CATGATTTCA CAATCACGGT  111
Dy  .GT..TA.T. .A.C.A..A. A.C.GT..CA ...ATA.T.A A......... TC.AT.T.A.
Mm  .GA...T.TT CA.CCC.CCT TC.A....CA ...CTA.T.A T......... .T..A.TTCA
Xl  .GA..T..A. CA.C.C..CT .C......CA ..CTTA..TA T......... .TT.GGATCA
Sp  .GA..AT.TA .A.GC...AT G..G..T.CA ..CAAT.TCC T...G..C.. T.CC..AAAG
Ce  T.ATTTT.T. CCTCAGCC.G A..GTT.A.A TC.TTA.T.A T..TT...A. ATT.GGACTA

Rp  GCCCTATGCC TAATTATTGG TTTAACACTT ACCACTTTAA CTTCAATTAT CTGATGACGA  171
Dy  ATTTC..TAT .TT.AT.A.. .AAT.TTA.. ..T.T..... .AGTTTA.CA A.........
Mm  ATTAC.CTAT ...CCC.... CC..CTCACC .AT.TCC.C. .AAT.TA.CA A.........
Xl  ATAA.TCTTT ...CCC.A.. CC...TTAC. .TAGTAC... ..AT....CA A.........
Sp  A.TAAT.TAA CTT.AG.A.. C..TTT.T.A TTA.TAAC.. AAATGG...A ...G.....C
Ce  TA.GA..TAT .T...T..AC AC..TTTTC. GTGTTA..T. T...TT..GC T...G.TAAG

Rp  GATGTAATTC GTGAAGGAAC TTATCTAGGC TTCCACAGAT CTGTAGTATC TAGAGGGCTA  231
Dy  .....TTCA. .A........ ...C.A...A ..A..T.CT. AC.C....A. ..TT..TT..
Mm  ..C....... .......... C..C.A.... CA.....CTC ..A.T...CA A.A...A...
Xl  ..C....... .A........ A.TC.A...A CAT....CTC .ACCC..TCA A.A...AT..
Sp  ...A...... .AA.G.CC.A C.T..AG... AGA....CTG ..A.T...AA A.AG..AA.G
Ce  ...A.TGC.A TA.....TTT AAGA---..T .A...T.AT. T.T.C..TAT AGAC...T.T

Rp  CGCTGAGCAA TAATTCAATT TATTCTTTCA GAAGTACTCT TTTTCGCAGC TTTCTTCTGA  291
Dy  ..A....G.. .....TT... ....T.A... .....TT.A. ....T.TTAG A..T..T...
Mm  ..A.AT.GT. ......T... C..CG.C..G ......T.T. .C..T....G A.........
Xl  ..A.AT.G.. ....C.T... ....ACA... ......T... .C..TATT.G A.........
Sp  ..A.AT.GC. .G..C.T... ...AACC... ..G..TTG.. ....TTTC.. C..T..T..G
Ce  AAA.TC.G.G .....TT... .G.GT..AGG ..GT.TA.A. .C...TTTTG .A.T..T...

Rp  GGCTTCTTTC ACAGAAGTCT GGCCCCTACT CCAGAAATTG GGTGCACTTG ACCCCCAACC  351
Dy  .CA..T.... .T......T. AT.T..AG.A ATT...T.A. .AGCTT.A.. ...T..T.TG
Mm  .CG....A.. .TTCT..C.. C.TA..A..A .AT..TC.A. .AG..TGC.. ...T.....A
Xl  .CA..T.ACA ..TC...CT. A........A TAT...T.A. ..GAATGC.. G..A.....A
Sp  .C.......C. .T.....AT. A.....CT.C GTT.....A. ..GTAG.A.. ......G.GA
Ce  ACA..T...G .TGCTGC... A.TA..AGTA .AC..GT.G. .AGAG..... .T.A..TTTT

Rp  GGAATTAACC CTATCAACCC ATTCTCCATC CCCCTGCTAA ACACAGCCGT TCTTCTAGCA  411
Dy  .......TTT .AT.T..T.. ...TCAA..T ..TT.AT... .T.....TA. ....T....T
Mm  ......TCA. .AC.T..... TC.AGAAG.. ..A..A..T. .T..TT.A.. A.........
Xl  .......C.. .AT.A..... ...TGAAG.T ..A..TT... .......A.. A.........
Sp  .....A.C.. .CC.T..... T...CTAG.T ..T..AT... .A....G... .......T.T
Ce  .....GC..T TAG.T..T.. ...TGGTG.A ..GT.AT... .T...ATTA. .T.AT..AGG
```

```
Rp  TCAGGAGTCA CAGTAACTTG AGCCCACCAT AGGGTAATAA ACAAATCTCG AACAGAAACT  471
Dy  ........T. .......... ...T..T... ..AT.....G .A.G.AA..A TT..C.....
Mm  .....T..TT ..A.T..A.. ...T..T... ..CC.T...G .AGGTAAA.. ..ACC.C.TA
Xl  ........T. .T..C..A.. ...T...... ..CA.C..GC .TGGCGA... ..A....G.A
Sp  ........T. .TT...GA.. GT.......C ..AA.TC..G CAGGGAA... ...T...T..
Ce  AGT..T..T. .......... ...A..T..C ..AT..C.T- --.G.AA.AA ..G.---TG.

Rp  CTTCAAGCCC TTTCCCTTAC AGTAATCCTA GGAGTTTACT TTACTTTCCT TCAAGCAGGA  531
Dy  AC.....GAT .A.TTT.... ...TT.A..T ..GA....T. .C..AA.TT. A.....TTAT
Mm  AA.....    .ACTAA.... CA.T..A... ...C...... .C..CA.... C.....TTC.
Xl  A.....T.A. .AA.TT.A.. CA.TC.T..T ...C....T. ....AGC... ......CAT.
Sp  A.......A. .A.TT..G.. ...GGCT..C ..TAGG..T. ....CGCG.. ...G..GT..
Ce  AC.A.TAGTA .AATTT.A.. .TGTT.AT.G .C..C...T. ....AGGAA. ....TT.AT.

Rp  GAATATATAG AAGCCCCTTT TACTATCGCC GATAGAGCCT ATGGTACCCT ATTTTATGTC  591
Dy  ........T. ....T..A.. ......T..T ...TC..TT. .....T.AAC T......A.G
Mm  .....CT.T. ..A.AT.A.. CT.C..TT.A ...G.TAT.. .....T.TAC ...C.TCA.G
Xl  ......TAC. .......A.. ...A..T..A ...G...TG. .C..AT.AAC .....T...A
Sp  ........T. .C.....A.. ...C..T... .....T.TT. ....CT..AC C..C.T...T
Ce  ..G.....   ....TAGA.. .T....T..A ..CG...TA. T...A.GGA. T......T.A

Rp  TGTACAGGCT TCCACGGCAT ACATGTCCTA GTCGGAACTA TATTTTTATC TATATACTTA  651
Dy  GCC..T..A. ....T..AG. T.....T... A.T.....A. CT..C....T AG...GT...
Mm  GC...T..A. ....T..AC. C.....AA.T A.T...T.A. C...CC.TAT .G.T.G.C..
Xl  GCA..T..T. .......TC. T......A.T A.T..CT.AT ....CC.... .G.T.GTC.T
Sp  GC......A. .T..T..TC. C..G..AA.T A.A.....A. CT..CC.CAT GG...G.C..
Ce  .C...G..A. .T.....A.. T......T.G TGT..TGG.T .G..C...G. AT.TA.T..T

Rp  ATTCGAACAT TTCTATATCA TTTC                                         675
Dy  T.A..TCAT. .AAATA.... ...T
Mm  C.A...CA.C .AAA..T... C...
Xl  C.....CA.A ...A...... C...
Sp  T....G..TG CAGGCCGC.. C...
Ce  T.A..TCTTC .AAA.A.... ...T
```

E. Cyt B

```
Rp  AAGCTTGCAA CAATGGTTGC TCGGCGGTTA TTCCGTGGGT AATCCAACGT TGCAACGCTT  60
Dy  .GA...AGT. .....A..AT GA..A..A.T .G.T..A.A. ...G.T..T. .AACT..A..
Mm  ..C.C.AGTC G....AA.TT GA..G..C.T C..A..A.AC ..AG.C..C. ..ACC..A..
Xl  CGTAC.AGTC .....AAGTT .A..A..A.T C..T..A.A. ..CG.C..T. .AACC..A..
Sp  T.TTA.AGTT ..G..A..AT GA..G..A.T C.....C.AC ..AG.C..CC .TACC..A..
Ce  ..CTA.TGTT ACT..AA.TT GAA.A....T .GGT..TACA GGGG....A. .AA..TT...

Rp  CTTCTCCCTG CATTACCTCT TGCCATTTAT CATTGCGGCC TTGGTTGGTT TGCATGTGTG  120
Dy  T...A.AT.T ....TTA.T. .A..T..... TG..CTT..T A.AAC.ATAA .T...C.ACT
Mm  ....G.TT.C ..C.T.A... .A........ T..C...... C.A.CAATCG .T..CC.CCT
Xl  ....G.AT.T ..C.T....C .T..T..... T.....C.GA GCTAGCAT.C .C...C.T.T
Sp  T..TC..T.T ..C.T...T. .C..C..... A..A..A... ..A.CG.T.A .A..CT.AGT
Ce  T..TGTAT.A ....TTT.A. ....G.GAGC T...CTA.TT A.T..ATTGG G....T.AAT
```

```
Rp  GGCTGTGCAT CATGTGGGAC AAAACAACCC GATTGGCATC GACGTGAAAA CCAAGGAAGA  180
Dy  ATT.T.A... ..AACA...T CT..T..... T.....TT.A A.TTCT..T. TTG.TA..AT
Mm  CTTCC.C..C G.AACA...T C......... A.CA..AT.A A..TCAG.TG .AG.TA..AT
Xl  ATT.C.C..C G.AACT...T C..CA..... A.C...AT.A A..TCAG.CC .AG.TA...T
Sp  ATTCC.C..C A.CAGA...G CC........ TT...C.T.. A.AAGC..CT ATG.CA.G.C
Ce  TTT.T.A..C AGAACT..TA G..CATCTAG .T.ATATTG. C...GTG.TT ATG.TA...T

Rp  CACATTATCC TTCCATCCTT ATTACACCAT GAAAGACGGG TTTGCGATTG CCGTGTTCTT  240
Dy  TC.T------ ..T..C..A. .C.T...AT. T..G..TATT G.A.GAT..A TT..AA.AA.
Mm  TC..------ ..T..C..C. .C..T..A.. C.....TATC C.A.GT..CC TAA.CA.A..
Xl  AC.T------ .....C..A. .C.T.T.TTA C......CTT ..A.GCT.CC TTA.TA.AC.
Sp  .C.T------ .....CAT.. .C.T...G.C C..G...ACA G.C.G.T..A TTC.T..GG.
Ce  TTGT------ ..TAGA...G .G...TTAGG T.....T.CT .A.AAT---A TT..TA.T.G

Rp  TATGATTTTC GCGCTGTTTG TGTTCTATCT CCCAGACGCG CTTGGACATG CAGACAATTA  300
Dy  .T.T...C.A ATTTCA..A. .T..AAT.AG A...A.TTTA T.G...G.CC ....T..C.T
Mm  CT.A...C.. ATAACCC.A. .A..A.T.T. .......ATA ..A...G.CC .......C..
Xl  ..CAGCAC.T A.T..CC.A. CCA.A.T.TC ....A..CTT T.A...G.CC .........T
Sp  AGCCGCAC.A TTTAGC..A. CCC..CTAT. T..T.G...C ..AAA.G.CC ....G..A.T
Ce  AT.AT.A..T ATTG....AA G...GAT.TA ...GTTTAAT T.A..TG... ....G.TG.T

Rp  TCAGCGGGCT GATCCTATGA AAACACCGCC GCATATTGTC CCTGAATGGT ATTTCTTGCC  360
Dy  .ATT.CT... A.....T.AG T......AG. T..C...CAA ..A.....A. ....T..ATT
Mm  CATA.CA... A....AC.A. .C..C..A.. C......AAA ..C.....A. .....C.ATT
Xl  .ACC.CA... A.....C.A. TC..C..T.. A......AAA ..A.....A. .C...C.ATT
Sp  CATT.CT..A A.C..AC..G TG..T..C.. A..C...CAG ..A..... .C.....ATT
Ce  .ATTGAA... ..C.....A. TG.GG..AGT T........T ..A..G..A. ....T..ATT

Rp  GTTCTACGCG ATTTTACGAG CTGTGCCTGA TAAGCTT                          397
Dy  TGCT.....A ...C.T..TT .AA.T...A. ...AT.A
Mm  TGCA.....C ...C....CT .AA.C..CA. ...A..A
Xl  CGC......T ..CC.T...T .CA.A---A. C..A..A
Sp  CGC......C ...C.G...T ..A.C..CA. C...T.A
Ce  TGCT..T..A .....G..T. ..A.T..AA. ...AG.C
```

FIGURE 6.

Inferred amino acid sequences of the vestimentiferan mitochondrial genes, aligned with their counterparts from other species. Gaps are indicated by dashes, residues identical to those of the vestimentiferan by dots. Published sequences used for comparison are *Drosophila yakuba* (Dy), mouse (Mm), frog (Xl), sea urchin (Sp) and nematode (Ce). Alignments are shown for A. COI; B. COII; C. ATPase 8; D. COIII; and E. Cyt B. Amino acid sequences are depicted using the standard 1-letter code.

A. COI

```
Rp  SSIGAFISFS SLLFFIFLMW EALASQRGVL ASPHMPTALE W-QETLPLDY HMFQETGLIT  59
Dy  .T..ST..LL GI...FYII. .S.V...Q.I YPIQLNSSI. .-YQNT.PAE .SYS.LP.L.
Mm  ..M.S...LT AV.IM..MI. ..F..K.E.M SVSYAS.N.. .-LHGC.PP. .T.E.PTYVK
Xl  ....SL..LV AVIMMM.II. ..F.AK.E.T TYELTS.M.. .-LQGC.TP. .TLKTSLVQI
Sp  ....ST..VV AM...L..I. ..F....EGI TPEFSHAS.. .QYTSF.PSH .T.D..PSTM
Ce  A.Y.SI..TA G.FL..YVLL .SFF.Y.L.I SDYYSNSSP. Y-CMSNYVFG .SY.SEIYFS

Rp  SPSFSA                                                            65
Dy  N-----
Mm  VK----
Xl  NHQMIK
Sp  IIVK--
Ce  TT.LKN
```

B. COII

```
Rp  MAHWGQ-LMF QDAASPIMIQ LVALHDHALT IMIMVVSLVL YMLY-SILTN KFTCRTLLEA  58
Dy  .ST.AN-.GL ..S...L.E. .IFF.....L .LV.ITV..G .LMF-MLFF. NYVN.F..HG
Mm  ..YPF.-.GL ...T....EE .MNF...T.M .VFLIS.... .IIS-LM..T .L.HTSTMD.
Xl  ...PS.-.G. ........EE .LHF...T.M AVFLIST... .IIT-IMM.T .L.NTN.MD.
Sp  .GT.A.-FGL ...S..L.EE .TYF..Y..I VLTLITM..F .G.V-.L.VS SN.N.FFF.G
Ce  INNFF.GYNL LFQH.LFASY MDWF.SFNCS LLLG.LVF.T LLFGYL.FGT FYFKSKKI.Y

Rp  QEIETIWTVL PATILVVLAL PSLRLLYLMD EIS-QPTLTV KTIGHQWYWS YEYSDFLNLE 117
Dy  .L..M...I. ..I..LFI.. ........L. ..N-E.SV.L .S........ ......N.I.
Mm  ..V.....I. ..V..IMI.. ....I..M.. ..N-N.V... ..M....... ...T.YED.C
Xl  ....MV..IM ..IS.IMI.. ....I..... .VN-D.H..I .A........ ...TNYED.S
Sp  ..L......I ..L..ILI.. ...Q..... .VK-D.F..I .AF....... ...T..KD..
Ce  .FG.LLCSIF .TI..LMQMV ...S...YYG LMNLDSN... .VT....... .....IPG..

Rp  FDSYMLPTEE LQDGEFRLLE VDHRMVIPMQ TEVRLLVTAA DVIHSWCVPS LGIKLDGIPG 177
Dy  .....I..N. .AIDG....D ..N.VIL..N SQI.I..... ......T..A ..V.V..T..
Mm  .....I..ND .KP..L.... ..N.V.L..E LPI.M.ISSE ..L...A... ..L.T.A...
Xl  .....I..ND .TP.Q..... ..N...V..E SPT......E ..L...A... ..V.T.A...
Sp  .....V..SD VSF.NP.... ..N.L.L... NPM.V..SS. ..L...A... ..T.M.AV..
Ce  .....KSLDQ .SL..P.... ..N.C...CD .NI.FCI.S. ....A.ALN. .SV...AMS.

Rp  RLNQTTLSIN RPGIFYGQCS EMCGANHSFM PIALEVIDHP SFTQWVMTFS E          228
Dy  .....NFF.. ...L...... .I........ ..VI.SVPVN N.IK.ISSNN S
Mm  ....A.VTS. ...L...... .I..S..... ..V..MVPLK Y.EN.SASMI -
Xl  ..H..SFIAT ...V...... .I........ ..VV.AVPLT D.EN.SSSML .
Sp  ......FFAA .T.V...... .I........ .MVM.SVPFN T.EN..TQYL .
Ce  I.STFSY.FP MV.V...... .I........ ......TLLD N.KS.CFGTM .
```

C. ATPase 8

```
Rp  MPHLAPLNWI LLPLFFLFSL LLLASITWWN QLISV-PQLK SKSKQA-HSM SPW--KWN--  54
Dy  I.QM..IS.L ..FIV.SITF I.FC..NYYS YMPTS-.KSN ELKNINLN.. -N.--..---
Mm  ..Q.DTST.F ITIISSMIT. FI.FQLKVSS .TFPLA.SP. .LTTMKVK-- T..EL..TKI
Xl  ..Q.N.GP.F .ILI.SWLV. .TFIPPKVLK HKAFNE.TTQ TTE.SKPN-- -..NWP.T--
Sp  V.Q.EFAW.. VNFSLIWA.V .MVI.LLLNS FPPNSAG.SS .SLTLK-KTT TN.--Q.L--

Rp  ---------
Dy  ---------
Mm  YLPHSLPQQ
Xl  ---------
Sp  ---------
```

D. COIII

```
Rp  M-IRQ--PFH VLEYSPWPFL VAVGVLAITC GAAAWFHNHG ALCLIIGLTL TTLTSIIWWR  57
Dy  .STHSNH... LVD.....LT G.I.AMTTVS .MVK...QYD ISLFLL.NII .I..VYQ...
Mm  .-TH.THAY. MVNP....LT G.FSA.LL.S .LVM...YNS ITL.TL..LT NI..MYQ...
Xl  .-AH.AHAY. MVDP....LT G..AA.LL.S .L.M...FGS MIL.TL..IT MV..M.Q...
Sp  .-AI.-H.Y. LVDQ....LD G.FSG.MM.S .NVL...TQK TNLTLV.FL. LMTKMVN...
Ce  ----MFHN.. I.SL.SYAYN LFFASAGMLS SLVMF.KFGL YELF.FT.FS VLFI.FA.GK

Rp  DVIREGTYLG FHSSVVSSGL RWAMIQFILS EVLFFAAFFW GFFHSSLAPT PEIGCTWPPT  117
Dy  ..S.....Q. L.TYA.TI.. ..G..L.... .....VS... A......S.A I.L.AS...M
Mm  ........Q. H.TPI.QK.. .YG..L..V. ..F...G... A.Y....V.. HDL.GC....
Xl  .......FQ. H.TPP.QK.. .YG..L..T. ..F..IG... A.YN...... Y.L.EC....
Sp  .M..KANFQ. S.TAI.KK.M .YG..L.MT. ..C..F.... A........S V.M.VA...S
Ce  .IAM..LS-. Y.NFF.MD.F KFGV.L.VF. .FM..FCI.. T..DAA.V.V H.L.E..S.F

Rp  GINPINPFSI PLLNTAVLLA SGVTVTWAHH SVMNKSRTET LQALSLTVIL GVYFTFLQAG  177
Dy  ..ISF...Q. ......I... .......... .L.ESNHSQ. T.G.FF..L. .I...I...Y
Mm  ..S.L..LEV .....S.... ...SI..... .L.EGK.NHM N...LI.IM. .L...I...S
Xl  ..T.L...EV .......... .......... .I.HGD.K.A I.S.T..IL. .L...A...M
Sp  .MT.L...LV ...K.G...S ....LS.S.. .ILAGN...S I...F...A. .S...A...W
Ce  .MHLV...GV .....II..S .......... .LL-SNKS-C TNSMI..CL. AA...GI.LM

Rp  EYMEAPFTIA DSAYGTLFYV CTGFHGMHVL VGTMFLSMYL IRTFLYHF              225
Dy  ..I....... ..V..ST..M A.....V... I..T..LVC. L.HLNN..
Mm  ..F.TS.S.S .GI..ST.FM A.....L..I I.ST..IVC. L.QLKF..
Xl  ..Y....... .GV..ST.F. A.....L..I I.SL...VC. L.QIQ...
Sp  ..ID...... ..V..ST.F. A.....LQ.I M..T..MVC. F..AGR..
Ce  .....S.S.. .GVF.SI..L S.....I... C.GL..AFNF L.LLKN..
```

E. Cyt B

```
Rp  SLQQWLLGGY SVGNPTLQRF FSLHYLLPFI IAALVGLHVW AVHHVGQNNP IGIDVKTKED  60
Dy  D.V...W..F A.D.A..T.. .TF.FI.... VL.MTMI.LL FL.QT.S... ..LNSNIDKI
Mm  T.VE.IW..F ..DKA..T.. .AF.FI.... ....AIV.LL FL.ET.S... T.LNSDADKI
Xl  V.V..S...F ..D.A..T.. .AF.F..... ..GASI..LL FL.ET.ST.. T.LNSDPDKV
Sp  IMV...W..F ..DKA..T.. .PF.F.F..M M...AVM.LV FL.NS.A... FAFKSNYDKA
Ce  TIVT.IWS.F G.TGA..KF. .V..F...WA .LVI.LG.LI FL.ST.STSS LYCHGDYDKV

Rp  TLSFHPYYTM KDGFAIAVFF MIFALFVFYL PDALGHADNY QRADPMKTPP HIVPEWYFLP  120
Dy  P--....F.F ..IVGFI.MI F.LISL.LIS .NL..DP..F IP.N.LV..A ..Q......F
Mm  P--......I ..ILG.LIM. L.LMTL.LFF ..M..DP... MP.N.LN... ..K......F
Xl  P--....FSY ..LLGFLIML TALT.LAMFS .NL..DP..F TP.N.LI... ..K......F
Sp  P--..I.F.T ..TVGFILLV AALFSLALLF .G..KDPEKF IP.N.LV... ..Q......F
Ce  C--.S.E.LG ..AYN-I.IW LL.IVLSLIY .FN..D.EMF IE....MS.V .........F

Rp  FYAILRAVPD KL                                                     132
Dy  A.....SI.N ..
Mm  A.....SI.N ..
Xl  A.....SM-N ..
Sp  A.....SI.N ..
Ce  A......I.N .V
```

to the corresponding mouse gene and by the loss of three nucleotides (one amino acid codon) from the corresponding *Xenopus* and sea urchin genes. Both of these changes have occurred at or near the 3' end of the gene. The length of the COII gene of the vestimentiferan is equal to that of *Drosophila yakuba*. The ATPase 8 gene of the vestimentiferan is 20% smaller than its mouse counterpart, lacking a segment of 11 codons at its 3' end (Figure 6).

In contrast to the mitochondrial genomes of other animal species, which occasionally use ATG, ATT or ATC codons as translational initiators, the three protein genes for which the 5' end was obtained (COII, ATPase 8, and COIII) in *Ridgeia* begin with the orthodox ATG codon. The 3' end was sequenced for three of the protein coding genes, ATPase 8, COI and COII. The ATPase 8 gene of *Ridgeia* appears to end in a complete TAA termination codon (Figure 3A). The COI gene apparently does not encode a complete stop codon but rather ends with T or TA (Figure 3A). The reading frame of the COI gene in *Ridgeia* extends 31-33 nucleotides into the $tRNA^{asn}$ to the first in frame TAA. If the gene actually extends to that point, it would be lengthened, adding 15 amino acids and overlapping 33 nucleotides with the $tRNA^{asn}$ gene. The exact termination position of this gene cannot be determined from the DNA sequence alone. I hypothesize that the transcript is processed at either nucleotide 185 or 197 (Figure 3A), and that the termination codon is completed by polyadenylation. The former position would make the *Ridgeia* COI gene equal to that of mammals, the latter would make this gene four amino acids longer than that of mammals, but almost equal to that of the frog and sea urchin. Here I prefer the latter. It has been reported (Clary and Wolstenholme 1985) that the presence or absence of complete termination codons at 3' end of the mRNA genes is poorly conserved in mammals or flies.

However, even here some rules seem to be followed. Because of the absence of a stop codon, a functional stop codon is presumably supplied by post-transcriptional polyadenylation at 3' terminal uridine residue, as in mammalian or amphibian mitochondria (Ojala et al., 1981).

The 3' end of the COII gene presents a similar problem. Whether or not the stop codon is coded in the DNA depends on the transcript processing. If the transcript is processed at the U (position 955 in Figure 3A) prior to the start of the $tRNA^{asp}$ gene, then the TAA stop codon must be completed by polyadenylation. If the transcript is not processed at that point, a UAG stop codon is encoded in the DNA sequence, overlapping the $tRNA^{asp}$ gene by two nucleotides.

A number of studies have suggested functional roles for the existence of stable secondary structure in mtDNA sequences. Such structures may be important for two reasons. First, since the transcripts are normally single stranded, secondary structures can form and may play a role in processing. Second, since the DNA may be single stranded for significant periods during replication (Clayton 1982), secondary structures can form during replication and may have an effect on replication.

Analysis of these protein genes reveals a number of potential secondary structures within each protein gene or between these protein genes which may be functional in the mitochondrial system. In the COI gene, I have identified a potential secondary structure (Figure 7C) located at 3' end of this gene. In the COII gene, a stem loop structure was found at two EcoRI restriction sites (Figure 7D). In the ATPase 8 gene, a stem loop structure was located at 22nd of nucleotide downstream from 5' end of this gene (Figure 7F). In the COIII gene, I have identified a stem loop structure which has a eight base stem

and a 97 base loop (Figure 7G). A potential secondary structure was found in the cytochrome b sequence 103rd nucleotide of the 5' end of this gene, which has a 10 base stem including one unpaired base pair and a 28 base loop (Figure 7H). Two stem loop structures were found between the COI and $tRNA^{asn}$ genes and between the COII and ATPase 8 genes (Figure 7B and Figure 7E).

Another interesting stem loop in the mitochondrial sequence of *Ridgeia* occurs near the COI/$tRNA^{asn}$ junction. The alternative structure shown in Figure 7A has a 10 base stem and a 88 base loop.

Genetic Code

The codon AGA, which specifies rare termination codons in some vertebrate mtDNAs, human and bovine, and serine in *Drosophila* and sea urchin mtDNAs, is found frequently in vestimentiferan mitochondrial genes (Table 4). The triplet AGA specifies arginine in the universal genetic code. In the vestimentiferan mtDNA, internal AGA codons (a total of 8) are present in two of five protein genes, the COII and COIII genes. It is clear that these AGA codons do not specify arginine. None of the AGA codons in either protein gene corresponds in position to arginine-specifying codons (CGN) in the equivalent positions of mtDNAs of five species including both vertebrates (mouse and frog) and invertebrates (*Drosophila*, sea urchin and nematode).

Further analysis of those positions where AGA occurs in the vestimentiferan sequence supports the view that AGA specifies serine. Of the eight AGA codons found in the vestimentiferan COII and COIII genes, which have the greatest similarity to the equivalent genes from other species (Table 3), two correspond in position to serine-specifying codons

FIGURE 7.

Secondary structures found within each protein gene or across gene junctions. Standard base pairs are designated by dashes. Asterisk denotes mispairing bases. The number in the centre of the loop is the nucleotide length of loop.

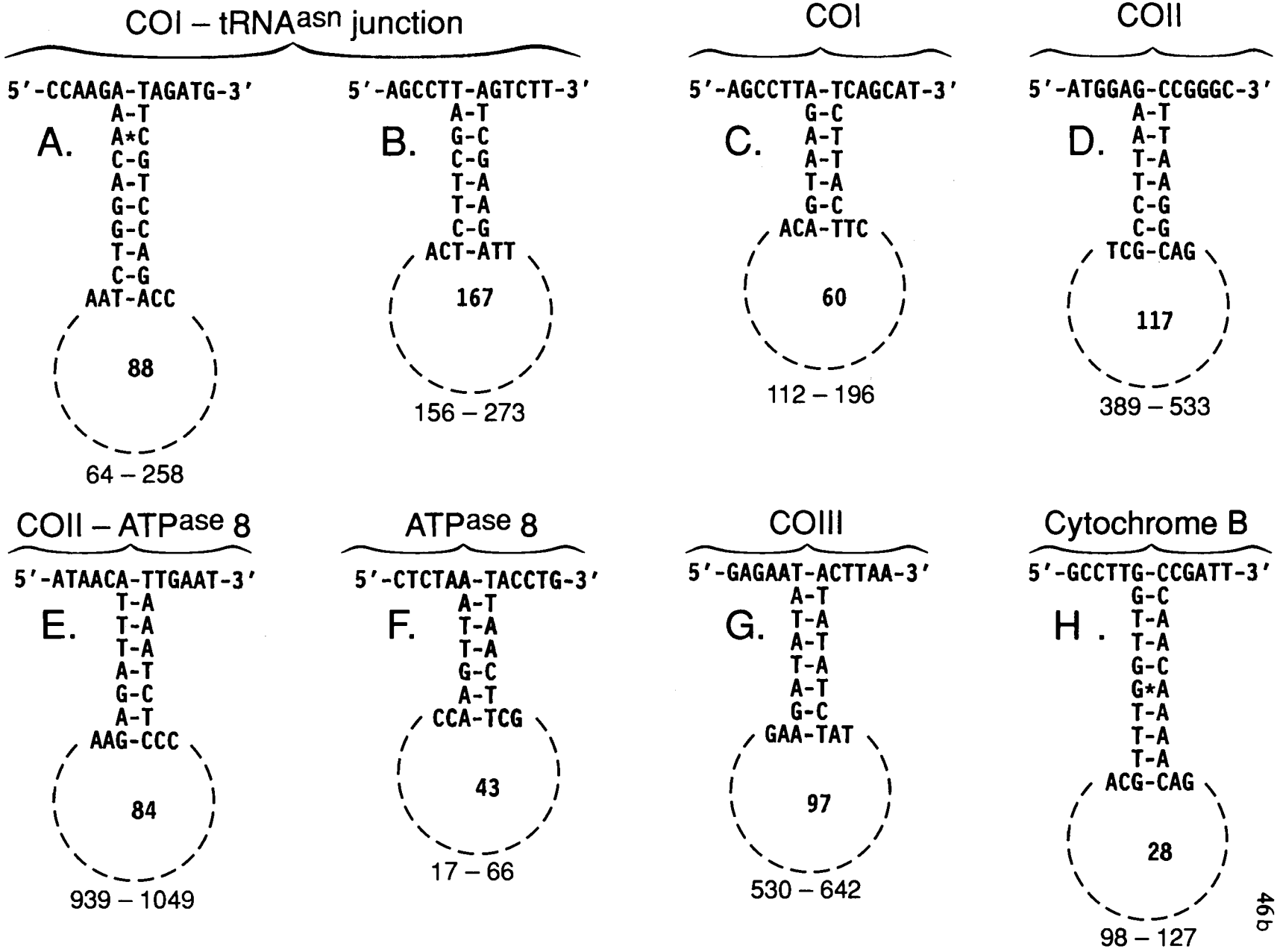

COI – tRNA^asn junction
5'-CCAAGA-TAGATG-3'
A-T
A. A*C
C-G
A-T
G-C
G-C
T-A
C-G
AAT-ACC
88
64 – 258
5'-AGCCTT-AGTCTT-3'
A-T
B. G-C
C-G
T-A
T-A
C-G
ACT-ATT
167
156 – 273
COI
5'-AGCCTTA-TCAGCAT-3'
G-C
C. A-T
A-T
T-A
G-C
ACA-TTC
60
112 – 196
COII
5'-ATGGAG-CCGGGC-3'
A-T
D. A-T
T-A
T-A
C-G
C-G
TCG-CAG
117
389 – 533
COII – ATPase 8
5'-ATAACA-TTGAAT-3'
T-A
E. T-A
T-A
A-T
G-C
A-T
AAG-CCC
84
939 – 1049
ATPase 8
5'-CTCTAA-TACCTG-3'
A-T
F. T-A
T-A
G-C
A-T
CCA-TCG
43
17 – 66
COIII
5'-GAGAAT-ACTTAA-3'
A-T
G. T-A
A-T
T-A
A-T
G-C
GAA-TAT
97
530 – 642
Cytochrome B
5'-GCCTTG-CCGATT-3'
G-C
H. T-A
T-A
G-C
G*A
T-A
T-A
T-A
ACG-CAG
28
98 – 127

Table 4. Codon usage in Phylum Vestimentifera mitochondrial DNA

Phe	TTT	20	Ser	TCT	14	Tyr	TAT	16	Cys	TGT	3
(F)	TTC	31	(S)	TCC	13	(Y)	TAC	9	(C)	TGC	5
Leu	TTA	22		TCA	12	TER	TAA	1	Trp	TGA	21
(L)	TTG	7		TCG	1		TAG	1	(W)	TGG	3
Leu	CTT	24	Pro	CCT	20	His	CAT	17	Arg	CGT	1
(L)	CTC	13	(P)	CCC	6	(H)	CAC	12	(R)	CGC	6
	CTA	25		CCA	11	Gln	CAA	23		CGA	8
	CTG	5		CCG	5	(Q)	CAG	3		CGG	2
Ile	ATT	33	Thr	ACT	14	Asn	AAT	4	Ser	AGT	1
(I)	ATC	16	(T)	ACC	12	(N)	AAC	13	(S)	AGC	2
Met	ATA	18		ACA	21	Lys	AAA	10		AGA	8
(M)	ATG	10		ACG	1	(K)	AAG	3		AGG	1
Val	GTT	7	Ala	GCT	10	Asp	GAT	6	Gly	GGT	10
(V)	GTC	11	(A)	GCC	23	(D)	GAC	14	(G)	GGC	8
	GTA	19		GCA	12	Glu	GAA	26		GGA	15
	GTG	8		GCG	5	(E)	GAG	1		GGG	4

in the genes from all five other species and two others specify serine in at least two of the other species. Other AGA codons in the COII and COIII genes correspond in position to codons specifying different amino acids including leucine, isoleucine, threonine, asparagine, lysine, glycine, aspartic acid, alanine, methionine and serine in the equivalent genes from the other five species. In view of this observation and the finding that AGA specifies serine in other invertebrate mitochondrial genetic codes (Wolstenholme and Clary 1985; Himeno et al. 1987; Wolstenholme et al. 1987; Garey and Wolstenholme 1989), it seems reasonable to conclude that, in vestimentiferan mtDNA, AGA specifies serine, as do AGC and AGT.

One codon, AGG, which does not occur in *Drosophila* or vertebrate mtDNAs, is found in *Ridgeia*, although much less frequently than AGA (Table 4). Only serine is found at the corresponding position in equivalent gene from the other five species. This observation suggests that codon AGG also specifies serine. Wolstenholme et al. (1985) drew similar conclusions for the assignment of the AGG codon in *Ascaris suum* mtDNA.

Internal TGA codons are found in four protein genes (all except for Cyt B gene) of the vestimentiferan mitochondrial DNA. Of a total of 21 TGA codons, 17 correspond in position to tryptophan-specifying codon (TGA or TGG) in equivalent genes from the other five species, indicating that in vestimentiferan mtDNA, as in mammalian, fungal, and *Drosophila* mtDNAs (Barrell et al. 1979, 1980; Fox 1979; Bonitz et al. 1980; Heckman et al.1980; Clary and Wolstenholme 1985), TGA specifies tryptophan.

Codon Usage

The codon usage among the *Ridgeia* mitochondrial protein genes is summarized in Table 4. All codons were used in the five protein genes. In a total of 705 codons, the frequencies

of first positions with A, T, C, and G are almost equal. The frequency of codons ending in A, T, C, and G is 36.6%, 28.4%, 27.5%, and 8.5% respectively. The overall codon usage for four of five protein-encoded genes has a strong bias against the use of codons ending in G (Table 5). This infrequent use of G at the third position has been reported for other animal mtDNAs.

The pattern of codon bias is markedly different from that found in other taxa. In *Drosophila* an enormous preponderance of A or T residues in the third-base position has been reported. In vertebrates, C appears to be a preferred wobble base in human mtDNA, but A is more prevalent in the mouse. In sea urchin, the third base is 60% (A+T). In the vestimentiferan, the third base is 64% (A+T), which is almost equal to that of sea urchin, but in other respects no obvious rationale can be advanced for the particular pattern of bias observed. For example, GAA and GAG are used with quite different frequencies to specify glutamate. The former is used 26 times more frequently than the latter.

The frequency of leucine is high and remarkably constant, in the range 15.0 to 16.9%, among mitochondrial proteins of different metazoa. In *Ridgeia*, leucine accounts for 13.4% of all amino acids in five mitochondrial DNA-encoded proteins. This value is a slightly lower than that of other metazoa.

Table 5. Base composition at 3rd codon positions in vestimentiferan mitochondrial protein genes

	Base composition % of total			
Gene	Guanine	Adenine	Thymine	Cytosine
COI	6.2	43.1	24.6	26.2
COII	4.0	42.5	25.0	28.5
COIII	2.7	38.7	31.1	27.6
ATPase8	3.7	35.2	33.3	27.8
CytB	28.8	15.2	29.6	26.5

Transfer RNA Genes

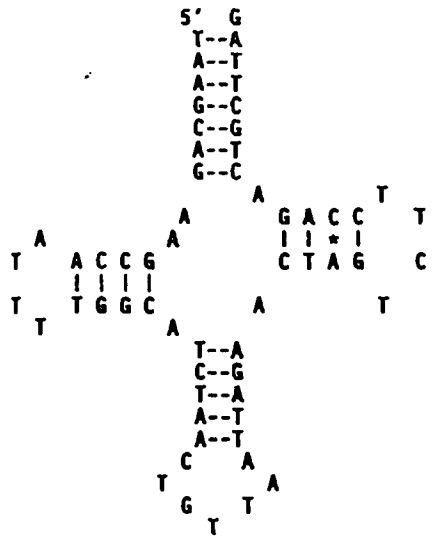
Each of the tRNA^{asp}, tRNA^{ser}, tRNA^{tyr}, tRNA^{phy}, and tRNA^{val} genes identified in the vestimentiferan mitochondrial DNA sequence can be folded into the four-arm cloverleaf structure characteristic of almost all known tRNAs. The major structural features of these five vestimentiferan tRNA genes resemble those of tRNA genes from vertebrates and other invertebrates (Figure 8). The vestimentiferan mitochondrial tRNA genes range in size from 63 nucleotides for tRNA^{ser} to 71 nucleotides for tRNA^{val}. They are generally smaller than their vertebrate and invertebrate counterparts. The sequence similarity to the corresponding mitochondrial tRNA genes from sea urchin, mouse, frog, and *Drosophila yakuba*, are summarized in Table 6. For all five tRNA genes, the highest mean similarity is between the vestimentiferan and sea urchin. The highest value of similarity between sea urchin and the vestimentiferan is 67.1%, in the tRNA^{asp} gene.

The predicted structures of five tRNAs (Figure 8) are conventional clover-leafs. The aminoacyl stems contain seven nucleotide pairs for tRNA^{ser} and tRNA^{asp}, and eight pairs for tRNA^{tyr} and tRNA^{phy}. The nine nucleotide stem in tRNA^{val} is the longest of them. Complete pairing is observed in the aminoacyl stems of all five tRNAs. The dihydro-uridine (DHU) stems have three to five nucleotides and the DHU loops four to eight nucleotides. The anticodon stems of five tRNAs vary from five nucleotides to seven nucleotides in length, and the anticodon loops are of a constant seven nucleotides in all cases. For tRNA^{asp}, there is the potential for an additional base pairing interaction in the anticodon stem, which would reduce the anticodon loop to five nucleotides, as is the case with a small number of tRNAs encoded by vertebrate or invertebrate mitochondrial genomes. The extra loops are generally variable regions and range from one nucleotide for tRNA^{ser} to four for tRNA^{asp}

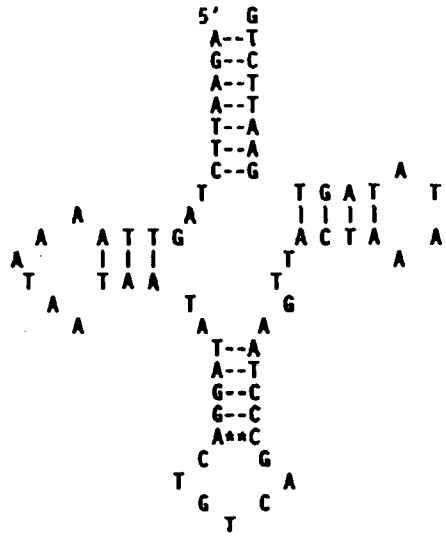
FIGURE 8.

Sequence of five *Ridgeia* mitochondrial tRNA genes represented in the cloverleaf form. Standard base-pairings(G-C or A-T) are indicated by dashes. Non-standard base-pairings are shown by asterisks. A. $tRNA^{ser}$; B. $tRNA^{asp}$; C. $tRNA^{tyr}$; D. $tRNA^{gly}$; and E. $tRNA^{val}$.

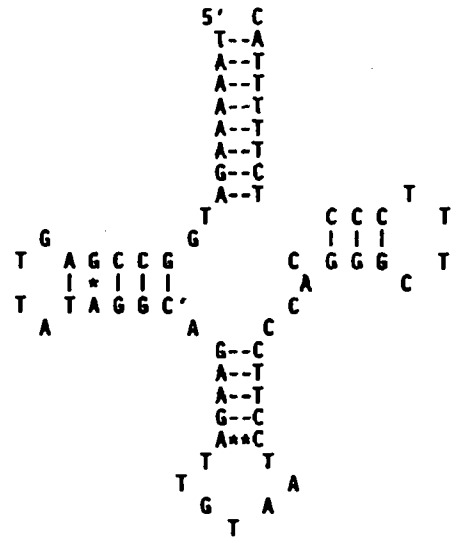
A. tRNA^{asn}



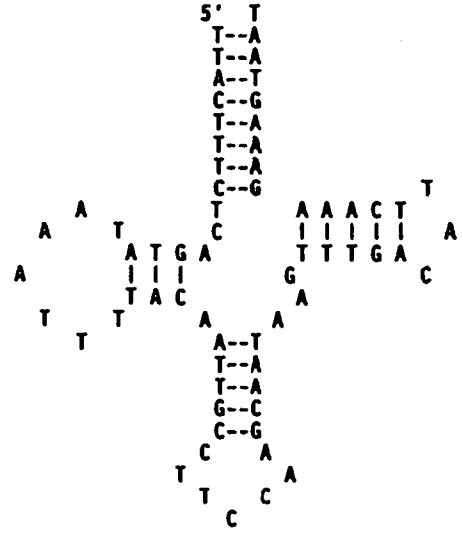
B. tRNA^{asp}



C. tRNA^{tyr}



D. tRNA^{gly}



E. tRNA^{val}

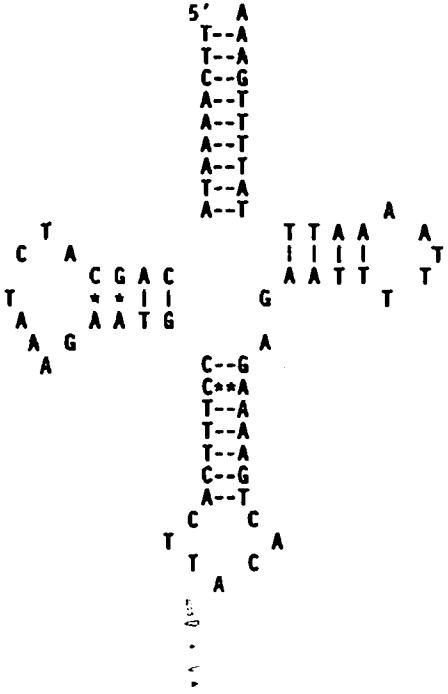


Table 6. Similarities between vestimentiferan tRNA genes to those from different taxa

	% Nucleotide similarity					
Species	$tRNA^{asn}$	$tRNA^{asp}$	$tRNA^{tyr}$	$tRNA^{gly}$	$tRNA^{val}$	MEAN
URCHIN	52.8	67.1	63.0	53.9	60.0	59.4
MOUSE	54.2	50.7	56.2	54.1	62.3	55.5
FROG	48.6	46.6	54.8	54.4	42.3	49.3
YAKUBA	43.1	49.3	47.9	50.0	60.0	50.1
MEAN	50.0	53.4	55.5	53.1	56.2	53.6

and $tRNA^{Tyr}$. The ribothymidine pseudouridine cytosine (TΨC) stems are of three to five nucleotide pairs and TΨC loops of three to five nucleotides. Unorthodox base-pairings are found in all of five tRNAs and occur in all three stems. These include a total of five A-C pairs, two G-A pairs.

I have carried out sequence analysis by comparing similar tRNA genes in five different species. A comparison of five tRNAs with their counterparts from sea urchin, frog, mouse and *Drosophila yakuba* is presented in Figure 9. Anticodon loops are highly conserved compared with those of other species. For example, the anticodon loop of $tRNA^{Gly}$ is identical to that in all four of the other taxa. In $tRNA^{Ser}$, the vestimentiferan anticodon loop is identical to three of the other taxa, with only one nucleotide change from that of sea urchin. For the remaining tRNAs, they have at least 80% similarity to each other. The highly conserved nature of anticodon loop sequences allows us to unambiguously identify the homologous set of tRNAs in every kind of species studied. The best conserved regions besides the anticodon loop are DHU stems. This conclusion is consistent with analysis by Gadaleta et al (1989). In comparisons of mitochondrial tRNAs from five vertebrate species (rat, mouse, human, cow, and *Xenopus*), Gadaleta et al. (1989) showed that the degree of conservation of different functional regions within tRNAs varied, with the anticodon loop being the most conserved region, followed by the dihydrouridine (DHU) stem. The DHU loop, the TΨC loop and stem are the most variable regions. In general the degree of conservation of the 5' half is higher than that of the 3' half, suggesting that the 5' region contains stronger functional constraints. In closely related salmonid species, for $tRNA^{Arg}$ and $tRNA^{Gly}$, variation is concentrated in the DHU loop (Thomas and Beckenbach 1989). In nematodes, *Ascaris suum* and *Caenorhabditis elegans* (Wolstenholme et al., 1987), the TΨC

FIGURE 9.

Sequence comparison of vestimentiferan mitochondrial tRNA genes with those from sea urchin (Jacobs et al., 1988), mouse (Bibb et al., 1981), frog (Roe et al., 1985), and *Drosophila yakuba* (Clary and Wolstenholme 1985). On the top the different functional areas are indicated. In each case, the vestimentiferan sequence is shown in full. Nucleotides identical to those of *Ridgeia* are indicated by dots, insertions and deletions by dashes.

	AA STEM	DHU STEM	DHU LOOP	DHU STEM	AC STEM	AC LOOP	AC STEM	VAR LOOP	TψC STEM	TψC LOOP	TψC STEM	AA STEM	
tRNA^{val}													
VEST	TTCAAAATA	CAGC	ATC-TAAAG	AATG	CCTTTCA	CTTACAC	TGAAAAG	AG	AATT	TTTA------	AAAT	TTATTTTGAAA	71
URCHIN	.....GGC.	T...	TAG..T.GT	..-.	TT.C..C	T.....	G..G.CC	.C	.C.C	G.GCAATT	CG.G	..G.....G	
FROG	--....G..	T...	T--....CC	..A.	.....G	.....	C....CA	.T	.TC.G	.TAAACC	CGG..	.-.C.....	
MOUSE	--..T.G.G	T...	T.AA..T-T	..A.	.A.C.GG	.C.....	CC.G...	.T	TTCA	.G----ACCA	.TGAA	C.C.C...	
YAKUBA	--...TT..	A...	T.AT...GT	..A.	.A.....	T.....T	.....	.T	TT..	G.GCAAATC	..TA	.A.A.--...	
tRNA^{asn}													
VEST	TAAGCAGAA	GCCA	ATTT-----	TGGC	ATCTAA	CTGTTAA	TT-AGAA	G	CTAG	TCTT	CCAGA	CCTGCTTAG	63
URCHIN	.GG..T.T.	..T	..A.G.AAA	-A..GC.TGG		.C.....	CCA....	ATA	A...	GA.AAATA	..TAT	.T.C.C-..	
MOUSE	..GATT...	....	G.AA....-	.A.G	G.A.TTAG	.....	C.A.ATT	TTCG	...	GT..AAT-T	..T.	..AATC...	
FROG	..GAAT...	..TC	G..GGA	TT.AG	.T..G	.....	C.A.A	.T	.TT GCG.G	A.CGAG.-	..C.T	..TTC...	
YAKUBA	.T.ATT...	....	.AAAGAG.	-..G	.A.C.	.....	.GATAT.	ATTG	AG	.A.A	AACT	..AATTA..	
tRNA^{asp}													
VEST	-AGAATTCTAGTTA	AA-TAA-		TAAT	ATAGGA	CTGTCAG	CCCTAA	GTT	ACT--	---AAA-T	ATAGT	GAATTCTG	64
URCHIN	....GGC.....	..C.T..		...C	T.T..G	T.....	G..A..	A..	G..GG	.TT...C.	CC..C	.GC.....	
MOUSE	A...TA.-....A.	..TC...		.T.C	...ACT	T.....A	AGT...	A..	.TAGA	.TC..TAA	TCTA.	AT..CT.A	
FROG	G...TG.-....A.	..CA.T.		..GC	.CGCC-	T.....A	GG.G..	A.A	G..GG	.TT.G.C.	CCG.C	AC..CTCA	
YAKUBA	A.A..A.-.....	..T..TA		...C	..TA.T	A.....A	A.TA..	A..	.T.AA	ATT.TTAA	T----	ATT..T.A	
tRNA^{tyr}													
VEST	TAAAAAGA	TGGCC	GAGTT-ATA	GGC	AGAAGA	TTGTAAT	CCTTCCC	AC	GGG	CTTT	CCC	TCTTTTTAC	65
URCHIN	-.GT...G	....T	...GGA...	A..	G.TG..	.....A	T..A.	AT..	AA...	T.AGAC	T..TT	.TC..ACTA	
MOUSE	-GGT..A.	....T	....-.--.	A..	.TT...	C.....A	T.TAA	AC.G	AGA..	T..AAA	TC..C.	.T....C.	
FROG	-GGT...G	.....	....AAT-.	...	G.CG..	.....GC	T.CGT	GT..	AGA..	T.CAAG	C.T.T	....C.ATCA	
YAKUBA	-G.TT.AG	....T	..AG.TT-.	...	GAT...	.....A	T..ATT	TAT	AA.AA	T.A.TC	TT.T	CT.AA.C.AT	
tRNA^{gly}													
VEST	TTACTTTCTC	AGTAT	AAAT	TTGTACA	ATTGC	CTTCCAA	GCAAT	AAG	T-TTG	ACATT	CAAA-	GAAAGTAA	67
URCHIN	--G....A.G	.A..C	..CA	--...TT	T...A	.....	T..GA	C..TCT.G.		TG.AAAT	TT..G	AT..AGC.	
MOUSE	--...CC..T	.....	..T.	A-A..T.	.C..A	.....	TT.G.	-..	AT.CT	GA..AAA	.CC.GA	AG.GAGT.	
FROG	--.....T	.....	T..C	CA.....	CG..A	.....	T..CA	...	.C..A	GTTAGAAT	.T..G	AG..AGT.	
YAKUBA	--.TC.ATAT	.....	...-	-A...T.	T...A	.....	T..TA	.G.	.C.AT	TA..	...T	AGT.TAG.TA	

loop is absent.

Mitochondrial tRNAs have been found to lack most of the conserved features of other tRNAs, especially those nucleotides in the DHU and TΨC stems which are believed to function as transcriptional control signals in nuclear DNA (Hall et al., 1982). In the vestimentiferan, the only universally conserved nucleotides are Py32 (C or T), Py33 (T in all cases) and Pu37 (A in all cases), the nucleotides flanking the anticodon (Figure 8). T8 occurs in three of five tRNAs, the $tRNA^{Met}$, $tRNA^{Trp}$ and $tRNA^{Asp}$. The Py11-Pu24 base-pair in the DHU stem which is conserved in *Drosophila* and sea urchin, is also conserved in all these tRNAs except $tRNA^{Ser1}$. The conventional sequence T54-Ψ-C-Pu-A is not found in the TΨC loop of all tRNAs.

The trinucleotide sequence CCA, which occurs at the 3' end of prokaryotic and eukaryotic nuclear-encoded tRNA genes, is absent in this position from the vestimentiferan mitochondrial tRNA genes, as it is from mammalian and *Drosophila* mitochondrial tRNA genes.

Ribosomal RNA Genes

In the mitochondrial genome there are two ribosomal subunits, the small or 12S subunit and the large or 16S subunit (Attardi and Schatz 1988). Each subunit is divided into three or four sections or domains.

Sequences of length of 457 and 1,117 base pairs coding both 12S and 16S rRNA genes have been obtained (Figure 3B). Comparison with corresponding regions of their vertebrate and invertebrate counterparts shows that both 12S and 16S rRNA genes in *Ridgeia* appear to be shorter than those of vertebrates and may be equal to those of *Drosophila yakuba*,

although the complete sequences of both ribosomal RNA were not obtained.

The 5' termini of the large ribosomal RNA and the 3' termini of the small ribosomal RNA have not been mapped directly by primer extension. Since $tRNA^{Val}$ gene is located upstream of the large ribosomal RNA gene, the 5' termination of the large rRNA gene should follow the $tRNA^{Val}$ assuming an absence of non-coding intergenic nucleotides and no overlap between the $tRNA^{Val}$ and the large ribosomal RNA genes. Similarly, assuming the 12S RNA and $tRNA^{Val}$ genes are butt-joined, the terminus of the 12S rRNA gene should be located immediately upstream from the 5' end of the $tRNA^{Val}$ gene. In order to locate both termini, the 5' end of 16S rRNA gene and the 3' end of 12S rRNA gene, I compared both the 5' end of the large ribosomal RNA gene and 3' end of the small rRNA gene of *Ridgeia* with their counterparts from other taxa. Figures 10 and 11 show these results. As shown in Figure 10, for the 12S rRNA gene, sequence similarity with both vertebrate and some invertebrate 12S sequences continues at least up to the second T residue upstream from the 5' end of the $tRNA^{Val}$ gene. Considering that the sequence similarity with vertebrate 12S rRNA sequences continues up to the second C residue upstream from the 5' end of the 12S rRNA gene, and the sequence similarity with *Drosophila yakuba* 12S sequence continues up to the first C residue upstream from the 5' end of the $tRNA^{Val}$ gene immediately preceding the gene for $tRNA^{Val}$, it is reasonable that the 3' terminus of the 12S rRNA gene of *Ridgeia* is located at the first residue upstream from the 5' end of the $tRNA^{Val}$ gene. It is therefore likely that the 3' end of the 12S rRNA is generated by RNA processing event that concomitantly defines the terminus of the adjacent tRNA.

In contrast to the 3' end of the 12S rRNA gene, the region of the 5' end of the 16S rRNA gene is poorly conserved (Figure 11). The sequence similarity with vertebrate 16S

```
                12S rRNA->                                          tRNA^val->
VEST        AAGTCGTAACATAGCAGGT-----GTAATGGAAATTGTACC TTCAAAAT
            ************ * * **      *** ****** ** **
HUMAN       AAGTCGTAACATGGTAAGT-----GTACTGGAAAGTGCACTTGGACGAAC
            **********  * * *        ** ****** *** *
MOUSE       AAGTCGTAACAAGGTAAGC-----ATACTGGAAAGTGTGCTTGGAATAAT
            **********  * * *        ** ****** *** *
BOVINE      AAGTCGTAACAAGGTAAGC-----ATACTGGAAAGTGTGCTTGGATAAAT
            ************** ** *      *** ****** **** *
YAKUBA      AAGTCGTAACATAGTAGAT-----GTACTGGAAAGTGTATCTAGAATGA
            ********** *  ***       * *   **** *
URCHIN      AAGTCGTAACACAATAGGCACACCGGACGCGAAAAT
```

FIGURE 10. Sequence alignment of the 3' end of the 12S rRNA gene from *Ridgeia* (Figure 3B), human (Anderson et al., 1981), mouse (Bibb et al., 1981), cow (Anderson et al., 1982), *Drosophila yakuba* (Clary and Wolstenholme 1985), and sea urchin (Jacobs et al., 1988). The sequence for the $tRNA^{val}$ gene that follows is underlined. Gaps are denoted by dashes, nucleotides identical to those of *Ridgeia* by asterisks, above the sequence to which they refer. The names and direction of tRNA and rRNA are included at the top of the sequences.

```
        tRNA^val->  16S rRNA->
VEST    ATTTTGAAA TATCCTAAACCCTCTATTTTTCTTTAAACTTGAATAAAAATCTATCTA
                   *******
HUMAN                GCTAAACCTAGCCCCAAACCCACTCCACCTTACTACCAGACAACC
                    *****  **
MOUSE          ACTAATCCTAGCCCTAGCCCTACACAAATATAATTATACTATTATATAAAT
                    ****  **
COW            ACTAGACCTAGCCCAAAGATACCCTCTCGACTAAACAACCAAGATAGAATA
                    *** *** *
FROG                TCCAAAAACCTAGCATTCCAATTATAATAACAATAACCTCATATTC
```

FIGURE 11. Sequence alignment of the 5' end of the 16S rRNA gene from *Ridgeia* (Fig. 2), human (Anderson et al., 1981), mouse (Bibb et al., 1981), cow (Anderson et al., 1982), and frog (Roe et al., 1985). The locations and direction of the mitochondrial tRNA and rRNA genes are indicated above the sequence. The nucleotides identical to those of *Ridgeia* are indicated by asterisks (only those nearest to the 5' end are shown). The sequence for the $tRNA^{val}$ gene that is upstream from 5' end of 16S rRNA gene is underlined.

rRNA sequences is only found in a few of bases downstream from the 3' end of the $tRNA^{Val}$. Localizing the 5' terminus of the 16S rRNA gene to the first residue downstream from the 3' end of the $tRNA^{Val}$ is reasonable.

The vestimentiferan small and large rRNA genes are higher in G+C content (36.1% for 12S rRNA gene and 34.8% for 16S rRNA gene respectively) than is observed in *Drosophila yakuba* mtDNA . The G+C contents for 12S and 16S rRNA genes are approximately those of sea urchin (44% for 12S and 38% for 16S).

In spite of the small size of the vestimentiferan mitochondrial rRNA genes, segments throughout the entire lengths of both of these rRNA genes may be folded into secondary structures analogous to those proposed for mouse (Zweib et al., 1981, Glotz et al., 1981), *Drosophila* (Clary and Wolstenholme 1985b), and *Xenopus* (Roe et al., 1985) mitochondrial rRNAs. According to Simon et al. (1990), the stems that hold those sections together are particularly well conserved, and some of the short range stems are poorly conserved, some of them are not. I have observed a number of very conserved regions in both ribosomal RNA genes among several animal species (Figure 12 and Figure 13). In both ribosomal RNA genes, these structurally conserved sequence blocks may be important for maintaining the characteristic secondary and tertiary structures of the rRNA molecules.

Differences in size between corresponding the vestimentiferan and vertebrate and invertebrate mitochondrial rRNA genes may result from the absence of specific blocks in the vestimentiferan rRNA genes. Indeed, I have found a number of deletion events in both sequences of the ribosomal RNA genes (Figure 12 and Figure 13). In the small ribosomal RNA gene, at least six deletion events have been observed among four species (Figure 12). The longest deletion segment consists of 45 nucleotides in the small ribosomal RNA gene

FIGURE 12.

Partial nucleotide sequence of the *Ridgeia* mitochondrial 12S rRNA (546 base pairs), aligned with its counterparts from other species: human (Anderson et al., 1981), frog (Roe et al., 1985) and sea urchin (Jacobs et al., 1988). Sequences are: *Ridgeia* (top); human (second); frog (third) and urchin (bottom). Gaps are indicated by dashes, residues identical to those of *Ridgeia* by dots. The positions for deletions are shown by asterisks, above the sequence. The positions for more than three bases of deletions are indicated by using DEL+numbers.

```
 12S rRNA ->
TCTCGGAAATTATGGCTTCAGATAATAACCTTTTCAAAAGAAGTCTAATTCTATCAATAT 60
C...CCC...A.A.CTAAA.CTC.CCTGAG..G.A....ACTCCAGTTGA.ACAA....G
A..A.AGTCAA.CTC.AA.CA.GCTGTCG.ACGCTTTCGTTG..T.G.AGAACA.TCACG
.G.TA....AGT.CAT.GGTTG..GCTTTGGGC.AGCGGT.TAAT.TTAATCCCA-CACG

                        ***                 *       *
TTCAAATCCAAACATATCTTT---ATAACAGTCCCATGAA-AGCCTAAA-TAAAATCAAG 115
ACT.---.G...G-.GG....AAC...T.T.AA.ACAC..T...TA.G.CCC...-.TG.
AAAGT.A.TCT..CC..A..ACAC--TTG.A-.T..C..CC-..TAGG.AAC...-.TG.
A.ACCC.T-.CC.CC.AA..AAGC..TTG.AG.-..C...ATCTAAGGCC-T...C.-..

                                      * <- DEL.1              ->*
GATTAGATACCCTTTTATATCTAGGCCCAAAA-------------------------TT 149
............CAC...GCT...-...T...CCTCAACAGTTAAATCAACAAAACTGC.
............CAC...GC....C.AT...CTTTGACTACTTA....CGCAAAAATCC-
............G.....CT...-----------------------------------
                                    *                     *
ATCCAGGGCACTACAACCACAGGTTT-AAAACCACAAAGAAATTGGCGGTACCTAA-ATC 207
CG....AA......G.G.CACA.C..-.....-T.....G.CC.......G.T.C.T...
-G....--A.....G.G.CT.A.C..-......-.....G.C........G.TCC.A.C.
--AAGT.TA..-..C--.CT.AAG..C....T-T.....G.T........TTTCC.A.C.

                                        *DEL.2 *          **
CAATCAGGGGAACCTGTCCTTTAACTCGAAAATT--------ACCTCTTTG--AATATAC 257
.CTCT..A...G.....T..G...-....T..ACCCCGATCA.....ACCA..CC.C.TG
..CCT..A...G.....T..G...TCGAT.CCCCTCGCTA-A.....ACCACTTC.TGC.
TCCCTG.A.AGTTTGCCAT.G-..TCGAT..CCCACGATC-C.....ACCAATTT.TGTA

**                                      *DEL.3*
--CTCAGCTTGTATACTGCCGTCGTCAGCCCACC-------ATAAAAATGTAAGTGAGCT 308
--......C.A.....C...A..T.....AA...CTGAT--GA.GGCTACA....A...G
AAAC.C..C.A.....CA......C.........TCGTGAG--.G.TT-C.T...AG...
ACAC............CAT......A..T.T..TTCTTGAG.A.GTTGAC.TTA-AG.GA

                  *                                   *
AACAGATTTTTATCTT-CACGTCAGGTCAAAGTGCAGCCTATGAGAAGG-AAGAGATGGG 366
C.AGT.CCCACG.AAAG-....T.......G...T....C.....GT..C....A.....
T.AT.......CATCAA.............G...T...A.....AGT..G....A.....
G.ACCC.GG---------.......A....G...........A..TT..GG.T..G..A.

          DEL.4       **    *            **           *     **
TTACAATTT-----AAATTC--TAAA-TATGAATTAT--TA-ATGAAAA-TAAAT--ATA 412
C....T...CTACCCC.--GA--...C..C..TAGCCC-.T.......C..T..GGGTCG
C....T...CTATACCT.AGAA....CG.AAG..CTC--.-........-CC.G..CG.G.
C......G.TT--------GAAC...CC.GTTGGA.GGAGGG......TACCCC.CG-G.
```

```
     *                                  **                ***
AAAGT-GGACTTGAAAGTAATTAA------AAG--TAACATACTTTTA---TGAATATGG 460
..G..-...T..AGC.....ACT-------...AG..GAG.G...AGT---....C.G..
....GC...T..---..C.G.A--------...AGA....AGAGAG.TCCTCTTTA.AAC
.--T.T...T.C---..C.G.AGGCCCACT...AGA.TGGGG..GAA.AGAGCTC.GGAA

     *    *                                    *<-DEL.5
CAACC-GAAG-GTGTGCACACATCGCCCGTCACTCTCGCCG------------------- 499
.CCT--....C.C..A.....--.........C...CT.AAGTATACTTCAAAGGACAT
GGC..T.G..C.C.--......C.........C...TT.TAC-AAAAATCAACCAACGT
TGCGT-----------.....................TAGTTAAGATTACAC-----

                           ->*                  *
--------------------------AAAGGGGAGATAAGTCGTA-ACATAGCA------ 526
TTAACTAAAACCCCTACGCATTTAT..T..A.....C........-....G.T.------
CTA..TAAACACACAATTAACACAAAG...A..CC--........G..CACTG.GATCGG
-------------------------.........C........-...C.AT.-----G

DEL.6        ->*
---------------GGTGTAATGGAAATTGTACC                          546
---------------A.....C......G..C..TTGGACGAAC
GATCACCAGCCGGAA.....GC.T.G..-.
GCACACC----GGAC-----GC--....A.
```

FIGURE 13.

Partial nucleotide sequence of the *Ridgeia* mitochondrial 16S rRNA (1115 base pairs), aligned with those from frog (Roe et al., 1985) and sea urchin (Jacobs et al., 1988). The sequences are: *Ridgeia* (top); frog (middle) and urchin (bottom). Gaps are indicated by dashes, nucleotides identical to those of *Ridgeia* by dots. The positions for insertions and deletions are shown by letter X and asterisks respectively. The positions for more than three bases of insertion or deletion are indicated by using IN+numbers or DEL+numbers.

```
XX                                IN.1                                    IN.
TATCCTAAACCCTCTATTTTTCTTTAAACTTGAATAAAAATCTATCTACAACTCTATCAT 60
--...A...A...AGCA..CCAA...T.------....C...AACCTC.T.TTCTC..---
---AAAGTTATGA..C.AGG.GC----------...G...ATCCAAACA.T---------

2                     *      <- DEL.1 ->
CCCAAATCTCCGCTATCTT---------------------------------------- 79
---....T..TAA.TAAACCATTCTAAAATTTTAGTATAGGCGATAGAACAATCATAATA
---....T..TATA.AA.AGAAAACACTCTTACTCTCTAGTAAATTCATTTGAAATCTTA

                  *                                          *<-DEL.2
-----------------AGTACTGCAAAGGAAATTATAAATTATTAATAAG--------- 113
GCTATAGAAAA------.....C.T..G.A..-GATG....AGAA.TG..ATAATTAACT
TTTTAACCAAAAGAAGC.....C....G.....GATG....ACCC.TA.TTAAACAAACC

->  *                                                DEL.3
-----TAAAAATAAAACCTTATACCTTCGTGCATTATGGCTTAGCA---AGCCAACTCTA 165
AAGCAAC....AGC.GAGAAC.TA.C....A.C..T..CA.A.TGGTCT.....GTCA..
-----......AGG..AGACTA.A.C T...ACC.G..TA.A.TGGATT.A.G.GAAA..

             *<-DEL.4->      *              X                DEL.
ATTCTAGCCTAT-----------------CCCGAAATCTTCACGAGCTGATAAACA---- 204
..CAAGCAAA.CGAATTTCAGTTTGACTA.......CTAAG-...T..AC.CCGAGACAG
TA--AGAAAA-CTAGTCTAAT--------.......CTGGG-......A..CTT.CCTC-

5*                              IN.3
--TTTGTATAAGAACTCACTACCGCATGTTTCAAAATGCCTAGAAAAATTTTATCAGAAG 262
CT...TAGAGCA...C.GTCT.T.TGG----.....GAGTGG...G.TC.CCGAGTAGG.
CT...TAGA.G..TAC.C.C..T.TTG----...T.GTG---......AAGGGAAGATTA

 **            *           *
C--TACATACCTAC-CGCGCAGACT-ATAGCTGGTTCCCTAAAAGCTTCACATTAGTGAA 318
GTG....G...A.A-..A..CT.G.G..........G.TC.GG.AA.GA.T..A...TCT
GATGTG.A.T...AC.....CC.GAG..........T..C....AA..AGTT.G..CT..

                                                 *
ACAAGATATATTCTCTGATAAATAATAAAGGAAAAGCTCTA-TTATAAAAGCTAATTTCG 377
..CCT.A....AGAT.TT...CA.T.....T....AG....C...GG.TTTA.TCAA..A
G.CTC---...AAG------------...AAT...A...CT-...AT.TTATA..AAGAA

     *<-DEL.6->        *                          **
AATCC-----------------TACCCCCTAAAGTAAGCTTAGA--AACTGCTAACTTT 417
GGG--------------------...AG...G.TTG..A.AGGATAC...CTA...TAC.
TC.TTTAATTTTTAAGCAGAGGT..GG....T...G.TAAGCT-TAA.G.CCAC..TGGA

          *          *  <-- DEL.7
TAAATAACG-TTATAGTA----------------------------------------- 434
GGG-...A.A.....A.CTTCAAGGAAAGTTGAGTCAGTG-------------------
A....CCAACA.TG.TAGTTAAAGACAACAAGCCCCAAAATAGCTATTCTCGAAGGAATA
```

```
   DEL.7
-------------------------------------------------------------
GGCCTAAAAGCAGCCACCTGTAAAGACAGCGTCAAAGCTCACTCAATCATTTAACCCTTT
GGCCTAGAAGCAGCCACTTAACAAGAAAGCGTTAAAGCTCAATTGTTCTTGCTAGCTAAA

   DEL.7
-------------------------------------------------------------
AATTAGTATAACTAATTCTAAACCCCCAAACAATACTGAGCTATTCTATAAACTATAGAA
AATTTTTGGGCAATCATCTCTAACACTACAAATTATTGGGACATTCTGTAATCAGAAGAG

   DEL.7
-------------------------------------------------------------
GCACTTATGCTAGAACTAGTAATGTGATACA---CGATTCTCC-----------------
ACAATGTTAATATAAGTAAGATAACTATACCTAGCGTTTTATACGCTTTTAACCATGGAA

   DEL.7-->    *       **         XXX*        IN.4
---------------TAAAAA--CTCTTTTTACA-AAATCA-------TAAACAAATTGC 469
---------------....TG---.AAG.G.----......GATCGAA....TC.C. .A
GAAAAACATTGAAAC......TC........---A..G..GTCTTCC--..CTC.GGA.A


TACAACCTATTAGGGATTATTTATCAACTTTCCTAAATAAAAATTCTGCTAAGATTAGTA 529
..ATTAACG.CCTCCC.G.GA.CCTTG.AA.AAC...AC..G.AAACCA.GCAC...T..
A-------------------------------------------------------------

                                                      IN.5 ->
TTCATTCCCTCCTACAACACCTATAAAACCCCCTTTCCCCCAAAAATGAAATTCCTTATT 589
CCGT.AAT..AAC....G.A.AT.T.C.GGAAAGA.TAAAAG.CGCA.------------
-----------------------------GAAAAA.AAAAA..GTGG...----------

 X                 *                                  *
ATAAGGAACTCGGCAAA-TACAAGCTCCGACTGTTTAACAAAAACATT-GCCTCTTGATT 647
--..............C..TG.A.C...C.......C.........C........C..
--..............-..AGGTT.-..C.......C.........-C..TC.CCA.AA

           DEL.8                                  ***
TATAAATAAG-------AGGTTCATCCTGCCCAATGACTTTA---GTTCAATGGCCGCGG 697
A.A...C.TTGTATAAG....C..G.........GTGAC-GTATATG....C........
.T....GCCTG----GGGA..C.TG.........-GTGAC.AGAGGT.A-..C......T.

                                       XXX                 ***
TACCCTGACCGTGCAAAGGTAGCATAATCACTTGCCCCTTAATTAGGGGCTGGC---ATG 754
..T-...................G........GT---......A..AA-....---T...
..T.T.......CGG.G-............T.GT---...CC.A.TTA.AGACTAGT...

                                                        **
AATGGACACACGAAAGCTTAACTGTCTCATAATAACTAATAAAAATTAATCTTTAA--GT 812
..C..C...--...G.T.C.........C.GCATC.A.TCC.TT.AACTGACC.CC--..
.....CA.G.G.G..ATAA.TGACAT.TT.TC..T--.GCCCTT.A..C.ACC.CCCC..
```

```
GAAAAAGCTTAAATTCCATTGCAGGACAAGAAGACCCCGTTGAGCTTTATTCTCTATAGA 872
.C.G.G..GGGG..AGA.CCAT..A..G.....-...TA.G........AA--...A...
...G.G..GGGG..AAA.CG.TTA...G.........T..G.......TAAGCGG.AGTT

                                                  * <- DEL.9
CCACCACTAATATTATACTATATCATAAACCTAAAAAAGAATTTA--------------- 917
A.TG.CAAGT.GAACCTAACC--.....GG--...T..C....A.ACAAGCAGAAACTGA
AA.TTTTA.CACACT.--ACCT.GTG.CTAACT..T.TACC-.ATCCAACAGT-------

     DEL.9 ->                    *                        *
---------------------------------GTTGGGGTGACTAAGGAACATCAAAA- 943
CCTAAAGTTTTC.....................G.......C...C.C...GA..A....A
----AAGTTTTAAACATTTTAGCAAAAGCTTTG.......CA..G-C...GT.AG..G.-

                          *         * <- DEL.10 ->             *
TCTTCCTTTTATTAACCAGGGC-TATTACCC----------------------------- 973
..C.....GA.GA.T--.....C..CC....TTTCACCAAGAACCACCATTCTAAGTAAC
C.C...GC.A..A.GA----------..TTACTATAAAAAGAATTACGGTTCTACAAT-C

     * DEL.11 *                  *
AAAAC----------AATTGACCCATAACCA-TGAACAAAAAAATAAGCTACCACAGGGA 1022
....TTTATGACTAT......T...GTC.T.C...T...CG..CC...T....CT.....
....TGAAAG-----...-..T...CT.AGG-...T....G...C...T....G.. ...


TAACAGGCTAATCTTTCTCAAGAGCCCAAATTGTCAGAAAGGATTGGCACCTCGATGTTG 1082
......CGC....CA.T....A..TT.CT..C.A..A.TG..T..ACG............
......CG.T......TCTG....TT..C....A.GA.....T...CG............

          **     X
GCTTAGGGGC--CCCTAATAGTGCAGAAGCTATTA                          1115
.AA.CA....AT...AG-.G......CC....C..
.A.CG..ACATC.TAA--GG...........T...
```

(Figure 12). In the large ribosomal RNA gene, at least 11 deletion events can be observed among three species, the longest deletion consisting of 237 nucleotides (Figure 13). A number of deletion events in the ribosomal RNA genes result in sizes of both rRNAs much shorter than those of vertebrates and some invertebrates. These observations imply that varying sizes of ribosomal RNA genes can serve the same function in different animal species. How long a sequence is necessary for the function of both ribosomal RNA genes? Is there a novel mechanism for this function in the mitochondrial genomes?

Sequence analysis of both 12S rRNA and 16S rRNA and protein genes in *Ridgeia* revealed a number of examples of complementarity between ribosomal RNA and protein genes. In the 12S rRNA gene I identified a number of interesting sites, some of which are illustrated in Figure 14. Some of them are very conserved regions, some of them are not located in the conserved regions of the small ribosomal gene. At position 175 from the 5' end, a sequence that is conserved almost perfectly in organisms ranging from vertebrate to invertebrate mitochondria is present. It is found complementary to a region in the Cytochrome b protein gene. At position 322 from the 5' end, a sequence which is highly conserved in all small ribosomal RNA genes, is complementary to a sequence from the COIII protein gene. Another sequence, from 360 to 375 of the 5' end of this small RNA gene, which is also located in the very conserved regions in almost all animal small rRNA genes, is complementary to a region in the COII protein gene. In rat mitochondrial genome, a similar case was found (Gadaleta et al. 1989).

In the 16S rRNA gene I also identified a number of interesting sites, some of which are illustrated in Figure 15. Like those observed for the 12S rRNA, some of them are conserved regions while some are not, in the 16S rRNA gene. At position 587 from the 5'

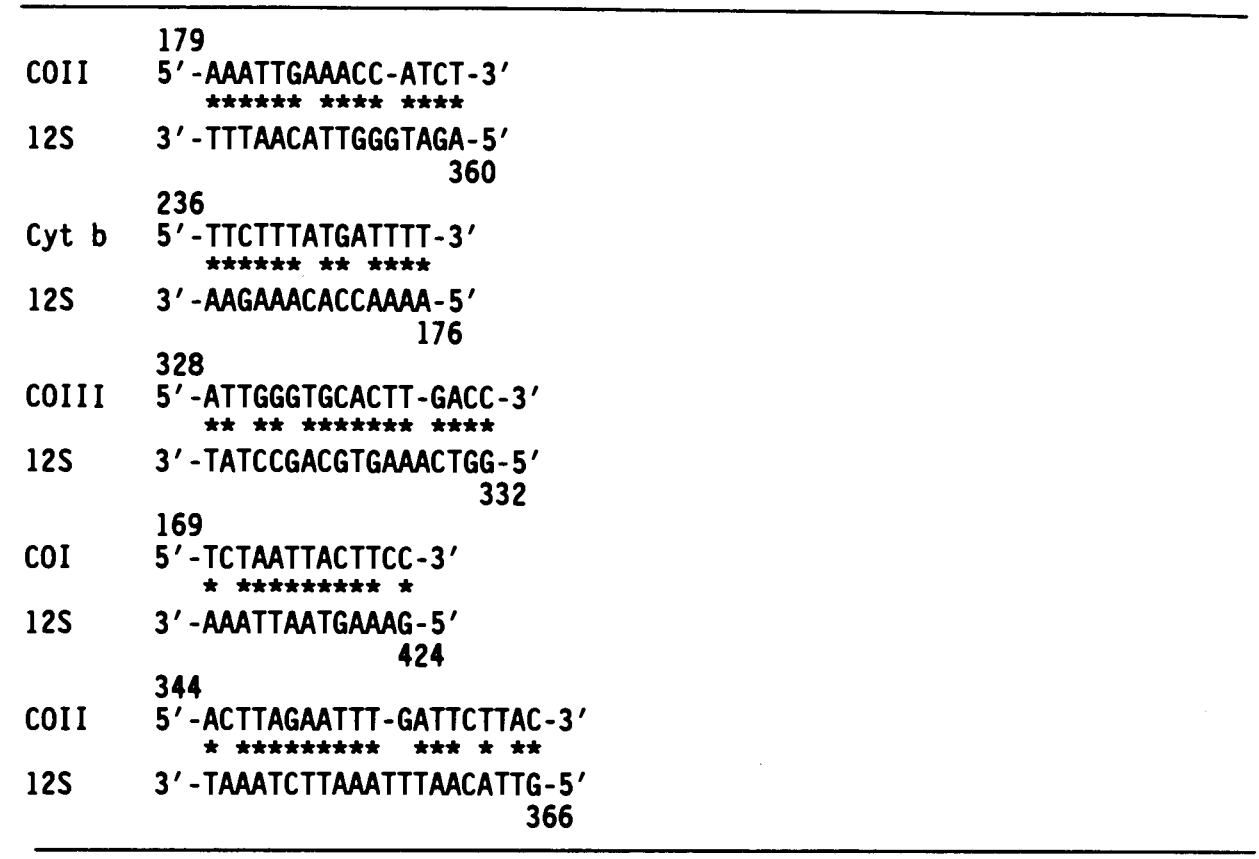


FIGURE 14. Complementary sequences found between the small ribosomal RNA and protein genes. Standard base pairs are designated by asterisks.

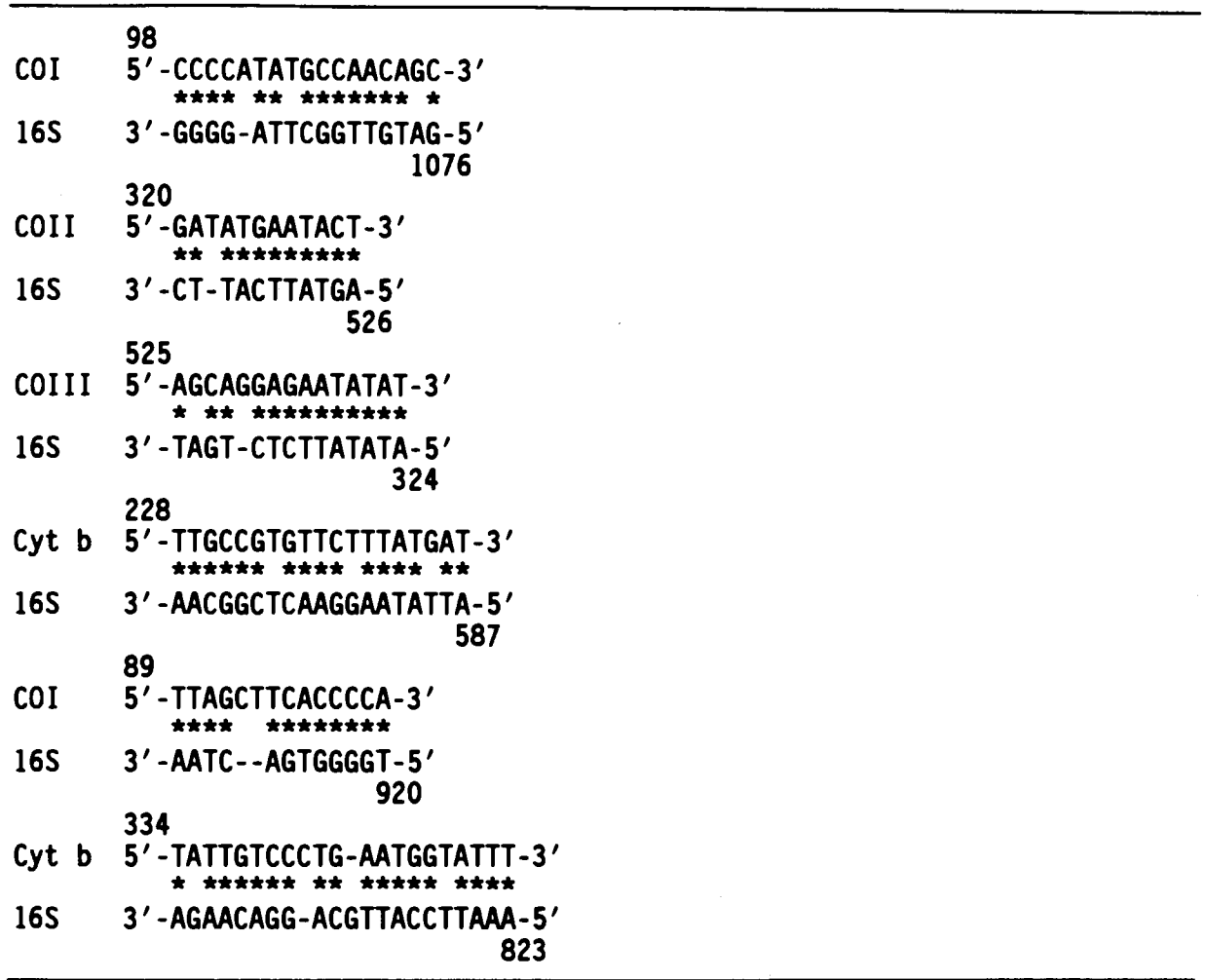


FIGURE 15. Complementary sequences found between the large ribosomal RNA and protein genes. Standard base pairs are designated by asterisks.

end of the 16S rRNA gene, a sequence that is very conserved in comparison to that of frog and sea urchin, is found to be complementary to a region in the Cytochrome b gene. Another sequence, located from 1076 to 1090 from the 5' end of the 16S rRNA gene, is very conserved in almost all large ribosomal RNA genes, and is complementary to part of the COI protein gene. The degree of complementarity of these sequences varies from gene to gene. The evolutionary conservation and these interactions strongly suggest an important function in the regulation of the translation.

Summary

Prior to this study, there was no information available concerning the gene organization or sequences of genes from any member of this phylum. Comparisons of sequences of funtional genes from a variety of animal phyla revealed regions of sequence conservation. Using primers based on these sequences, I have amplified and sequenced several fragments coding various functional classes of genes from *Ridgeia*, a deep sea vent representative of the Phylum Vestimentifera.

Analysis of sequence of about one third of the genome from *Ridgeia* mitochondrial DNA reveals that the mitochondrial genome of this species has a novel organization. In the vestimentiferan, the ATPase 6 gene does not occur between the ATPase 8 and COIII genes. This position is occupied by two tRNA genes. Further analysis of this gene order shows that the kind and number of those tRNA genes lying among protein genes are quite different. In *Ridgeia*, AGA and AGG specify serine and TGA specifies tryptophan.

Many insertion or deletion events ranging from a single amino acid codon to a 12 amino acid domain can be observed in these five protein genes of the vestimentiferan. The COI

gene of the vestimentiferan mtDNA does not encode a complete stop codon but rather ends with T or TA. A number of potential secondary structures within each protein gene or across gene junctions have been observed. They may be functional in the mitochondrial system.

Five tRNA genes were identified in the vestimentiferan mtDNA. Each of them can be folded into the four-arm cloverleaf structure characteristic of almost all known tRNAs. Sequence similarities of the tRNA genes with those of members of other animal phyla ranged from 49% with *Xenopus* to over 59% with sea urchin.

A number of deletion events in the ribosomal RNA genes of the vestimentiferan mtDNA result in the reduction of sizes of both rRNAs compared to those of vertebrates and some invertebrates. Sequence analysis revealed a number of examples of complementarity between ribosomal RNA and protein genes. This suggests that they might play an important role in the regulation of mitochondrial translation and transcription mechanisms.

SECTION II

EVOLUTION OF MITOCHONDRIAL DNA FROM *RIDGEIA* (PHYLUM VESTIMENTIFERA)

INTRODUCTION

A paradigm of molecular evolution is that rate of genetic change is negatively correlated with degree of functional constraint. Gene or nucleotide positions whose physiological functions are least disrupted by genetic alterations are those that typically evolve rapidly (Nei 1987). MtDNA is generally considered to provide an extreme example of genetic economy arising through functional constraint (Attardi 1985). Before 1980, it was widely assumed that functional constraint on mtDNA would be so severe that most genetic changes would likely be lethal (Brown 1985). Therefore, the first report (Brown et al. 1979) that mtDNA evolves rapidly (about 1-2% sequence divergence per lineage per million years, or perhaps 5-10 times faster than typical single-copy nuclear DNA) came as a great surprise. A rapid pace for mtDNA nucleotide substitution was subsequently confirmed in many animal groups, although particular rate estimates vary considerably among different portions of the molecule and perhaps among taxa.

Two general explanations have been advanced to account for rapid evolution of mtDNA (Wilson et al. 1985): (a) relaxation of functional constraint; and (b) inefficiency in the repair mechanisms of mtDNA. The former suggestion arises in part because mtDNA does not code for the proteins involved directly in its own replication, transcription, or translation, and because a molecule of mtDNA that produces only 13 kinds of polypeptide chain might tolerate less accuracy in the translation machinery. It is also consistent with inferred "drift" in the mtDNA code, which implies a role for inaccuracy during shifts from one code to another (Wilson et al. 1985). This explanation may apply with particular force to mitochondrial tRNA and rRNA genes, where rates of evolution are accelerated as much as 100-fold relative their nuclear counterparts (Brown 1985; Cann et al. 1984).

The second class of explanation arises because of an enhanced mutation rate in mtDNA. The silent nucleotide substitution rate in mammalian mtDNA has been estimated at $4.7x10^{-8}$ per site per year (Brown and Simpson 1982), or about 10 times higher than in nuclear pseudogenes (Nei 1987). A likely reason is that mitochondria may lack a mismatch repair system that could otherwise repair the transitions characteristic of much mtDNA evolution (Wilson et al. 1985). Unfortunately, very little is known of how any potential repair enzymes actually operate or function in mitochondria, including γ-DNA polymerase, the only eukaryotic DNA polymerase known to be present in mitochondria (Lewin 1990). As Brown (1985) noted, "Lack of information about the enzymology of mtDNA replication and gene expression is presently the single greatest barrier to better understanding of the evolution and molecular biology of mtDNA".

In all likelihood, both selection and mutation are involved in rapid evolution of mtDNA. An enhanced evolutionary rate due solely to inefficient DNA repair should result in uniform divergence at all mtDNA nucleotide positions, but observed genetic changes are concentrated disproportionately where relaxed selection seems likely on functional grounds, e.g., at synonymous positions in protein-coding genes, in the noncoding D-loop, and in certain regions of the rRNA and tRNA genes.

Variable selective constraints can be imposed by functional considerations such as amino acid codon specificity, active sites of proteins, tRNA attachment sites in rRNA, or structural considerations such as secondary and tertiary structure in rRNA, tRNA and proteins and can affect the rate of evolution of the region under consideration. The degree of constraint varies among genes. A clear demonstration of variability of constraint among different mitochondrial protein coding regions can be seen by comparing rates of evolution calculated

for the "slowly evolving" COII gene versus the "rapidly evolving" ATPase 8 gene (Pumo et al. 1992).

The 22 mitochondrial tRNA genes are less constrained by structure and function than their highly conserved nuclear counterparts; they evolve at a higher rate (Wilson et al. 1985). Nevertheless, mitochondrial tRNAs evolve more slowly than mitochondrial protein coding genes (Wolstenholme and Clary 1985), an indication that they are structurally or functionally more constrained than these protein genes. An excellent demonstration of the effect of structure and function on the rate of evolution of mitochondrial tRNAs is the study of Cantatore et al. (1987) and Thomas et al. (1989) who showed that in sea urchin, the tRNA leucine CUN gene appears to have lost its function, having been incorporated into the ND5 protein subunit gene, it is now diverging at the typically higher rate of the ND5 gene.

Using information from comparing gene organization and sequences of various functional genes, I examine the evolutionary features of the *Ridgeia* (Phylum Vestimentifera) mitochondrial DNA. For evolutionary change in the mitochondrial genome, I have found that the ATPase 6 gene of the vestimentiferan mtDNA has been independently translocated during evolution; also, several tRNA genes appear to have been translocated as a single segment. For the ribosomal RNA genes, the evolution involves mainly a number of deletion events and substitutions. Different protein genes have different rates of evolution.

RESULTS AND DISCUSSION

Gene Rearrangement

The analysis of sequence from *Ridgeia* (Phylum Vestimentifera) mtDNA revealed that gene rearrangement has occurred. The ATPase 6 gene that follows the ATPase 8 gene and overlaps with it in all vertebrates and some invertebrates (*Drosophila yakuba*, sea urchin, and sea star) mtDNAs was not found between the COII and COIII genes. Two tRNAs, $tRNA^{Tyr}$ and $tRNA^{Gly}$, were located in this region of *Ridgeia* (Phylum Vestimentifera) mitochondrial genome. The relative arrangement of tRNA and protein genes in the sequenced segment of the *Ridgeia* mtDNA molecule suggests that there is a considerable difference in relative gene order between the mtDNAs of vestimentiferan, vertebrates, and some invertebrates. In contrast with the relatively conserved organization of the mitochondrial genome within the vertebrates, sequence data from invertebrates such as *Drosophila* (Clary and Wolstenholme 1985), sea urchin (Jacobs et al., 1988; Cantatore et al., 1989), sea star (Himeno et al., 1987; Jacobs et al. 1989; Smith et al., 1989; Smith et al., 1990), the mollusc, *Mytilis* (Hoffmann et al., 1992) and *Ascaris* (Wolstenholme et al., 1987) mtDNAs reveal that many rearrangements, including inversions and transpositions, have occurred during the evolution of phyla. Gene rearrangement has also been observed by Desjardins and Morais (1990) in chicken mtDNA and by Pääbo et al., (1991) in marsupials. These are the first indication that mtDNA gene transpositions have occurred in higher vertebrate lineages. In bird mtDNA, the contiguous $tRNA^{Glu}$ and ND6 genes which are encoded on the light strand (L-strand), are located immediately adjacent to the D-loop region; the Cytochrome b and ND5 genes, which border the $tRNA^{Glu}$, and the ND6 genes

in other vertebrate mtDNAs are contiguous and separated only by a few nucleotides. Between vertebrate and *Drosophila* mtDNAs differences in organization of protein coding genes and rRNA genes can be accounted for by three inversion events. In addition, 11 of the 22 tRNAs occupy different relative positions. In contrast, extensive reorganization involving all functional classes of genes is evident in comparisons of vertebrates, mollusc and nematode sequences.

Mitochondrial DNA exhibits a seemingly paradoxical combination of rapid evolutionary change and extreme conservatism: its high rate of sequence evolution (Ferris et al., 1983; Brown, 1983) contrasts with the relatively slow rate of genome reorganization. Since animal mtDNA is a highly economical genetic system in which there is almost no non-coding information, random break-points for rearrangements are almost certain to interrupt an essential function. Data from comparable genomes, such as plant and fungal mtDNAs and chloroplast DNAs, suggest that the degree of organizational plasticity of organelle genomes is largely a function of their compactness, supporting the view that selection has maintained the structure of animal mtDNA. Transpositions could also occur by some recombination events. Comparison of the vestimentiferan gene order with those of other animals (Figure 4) indicates that the ATPase 6 gene has been independently translocated during evolution. Also, the $tRNA^{Ala}$ gene, $tRNA^{Arg}$ gene, and both of $tRNA^{Tyr}$ and $tRNA^{Met}$ genes appear to have been translocated as a single segment. However, there is no evidence for intermolecular recombination in animal mtDNA; hence the modes of rearrangement are restricted.

Protein-Coding Genes, Mode and Tempo of Evolution

Partial sequence coding for five protein genes (COI, COII, COIII, ATPase 8, and Cyt B) from *Ridgeia* mtDNA was obtained (Figure 3). Comparisons of DNA and predicted amino acid sequences for these genes with those from *Drosophila*, mouse, frog, sea urchin and nematodes are presented in Figures 5 and 6. Although some of these genes only have partial sequence, the comparisons show that both insertions and deletions have occurred. In the COIII gene, a two codon deletion is evident near the 5' end of the gene; in the cytochrome b gene, a two codon insertion is evident. The *Ridgeia* COII gene is three nucleotides (one codon) longer the 3' end compared to the mammalian sequence, and three nucleotides (one codon) shorter compared to those of frog and sea urchin. The COII gene is 687 nucleotides (including stop codon) in length in *Ridgeia* which is equal to that of *Drosophila*. The ATPase 8 gene is 165 nucleotides in length in *Ridgeia*, compared with 204 nucleotides in mouse, 168 nucleotides in sea urchin and frog, and 162 nucleotides in *Drosophila yakuba*. The ATPase 8 gene is 20% smaller than its mammalian counterpart, lacking a segment of about 11 codons at its 3' end. In this gene, there are several 1-2 codon insertion/deletion events evident in addition to those at the 3' end (Figure 6C). The coding strand of *Ridgeia*, like other sequenced vertebrates and some invertebrates, exhibits a strong nucleotide bias against G except for cytochrome b (Table 3).

Many amino acid replacements in these protein genes in the vestimentiferan mtDNA are non-conservative. For example, single amino acid mutations are found in many sites of these protein genes, and blocks containing multi-amino acid mutations were also observed in these protein genes. In the COII gene, at position 18 downstream from the 5' end of this gene, there is a isoleucine in the vestimentiferan, a glutamic acid is found in the

corresponding sites of the COII gene of most other species (Figure 6B). At position 23 downstream from the 5' end of this gene, a leucine occurs in the vestimentiferan where a phenylalanine is found in the corresponding sites of the COII gene of the other species (Figure 6B). In a conserved region of this gene, from the 134th to 141st codon from the 5' end, there is a histidine at position 140 in the vestimentiferan where a asparagine is found in the corresponding sites of the other species (Figure 6B).

In the COIII gene, a block containing three amino acid differences was located at a position 16 codons from the 5' end (Figure 6D). In this conserved domain, there are usually three amino acids, leucine, threonine (aspartic acid for sea urchin), and glycine. These three amino acids were replaced with phenylalanine, leucine, and valine in the vestimentiferan (Figure 6D). There is a leucine at codon position 66 from the 5' end of this gene where a glutamine is found at the corresponding sites in most other species. Only the nematode carries a leucine at this position. There is usually a leucine at codon position 83 from the 5' end of this gene where glutamine is found in the vestimentiferan. Similarly, there is an alanine at codon position 80 where a glycine occurs in the corresponding site of the other species. Only a few of these kinds of examples are found. It seems likely that these changes represent amino acid replacements rather than a genetic code change between the mtDNAs of the vestimentiferan and other animal species.

In the cytochrome b gene, several multi-amino acid changes were found in addition to single amino acid replacements. The longest altered amino acid domain was located at codon position 39, and includes four codons (Figure 6E). At these sites, the amino acids, valine, tryptophan, alanine, and valine are inferred in the vestimentiferan, the four amino acids, leucine, leucine (valine for sea urchin), phenylalanine, and leucine were found in the

corresponding sites of the other species. Another two-codon-change domain was found in the most conserved region of the cytochrome b gene, two amino acids, phenylalanine and alanine located in the other species mtDNAs, were replaced with amino acids proline and phenylalanine in the vestimentiferan mtDNA (codon position 120-121; Figure 6E). These observations suggest that some of the amino acids which are completely conserved in the protein genes of vertebrate and some invertebrate mtDNAs, have been changed in the vestimentiferan during evolution.

Table 7 shows the numbers and frequencies of specific base substitutions for the five protein coding genes included in this study. The divergences between *Ridgeia* and representives of other taxa are extensive. Corrected divergence estimates (Jukes and Cantor 1969) range from about 50% to complete saturation. For divergences as great as these, estimates are highly dependent on the accuracy of the alignment. Where alignment is uncertain, part of the divergence estimates may be due to comparisons at non-homologous sites. This problem is particularly acute at the 3' ends of genes, where alignment is often difficult. Divergences of the COI gene appear quite high, from 64% to saturated (corrected divergences- Table 7), but may just reflect the fact that only 195 residues of the 3' end are included for comparison.

Since complete sequences were obtained for the COII and ATPase 8 genes, I will focus on analysis of these two protein genes. As shown in Table 4, the degree of similarity to most species is different for the two genes. For the COII gene, there is a mean of 60% nucleotide similarity, and a mean of 56% amino acid similarity compared to those of five species. In contrast, the ATPase 8 gene has a lower similarity with four species compared, having a mean of 47% for nucleotide similarity and a mean of 29% for amino acid

Table 7. Nature and frequency of specific base substitutions in five protein coding genes between *Ridgeia*, a vestimentiferan and *Drosophila*, mouse, frog, sea urchin and nematode worm

A. COI

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	5	8	12	9	17
C--T	26	20	22	32	24
Total	31	28	34	41	41
Transversions					
A--C	14	14	28	16	17
A--T	26	19	22	14	39
G--C	7	8	12	8	11
G--T	9	10	6	9	16
Total	56	51	68	47	83
Indels	1	1	0	2	0
Number of sites	181	184	195	191	195
Total changes	88	80	102	90	124
% changes	48.6	43.5	52.3	47.1	63.6

Table 7(2). Nature and frequency of specific base substitutions in five protein coding genes between *Ridgeia*, a vestimentiferan and *Drosophila*, mouse, frog, sea urchin and nematode worm

B. COII

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	33	38	37	45	42
C--T	71	59	68	72	81
Total	104	97	105	117	123
Transversions					
A--C	41	67	59	55	56
A--T	79	56	54	56	109
G--C	18	14	21	19	29
G--T	15	14	14	19	40
Total	153	151	148	149	234
Indels	0	1	0	0	3
Number of sites	684	682	684	684	687
Total changes	257	249	253	266	360
% changes	37.6	36.5	37.0	38.9	52.4

Table 7(3). Nature and frequency of specific base substitutions in five protein coding genes between *Ridgeia*, a vestimentiferan and *Drosophila*, mouse, frog, sea urchin and nematode worm

C. ATPase 8

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	4	2	5	8	-
C--T	18	19	24	26	-
Total	22	21	29	34	-
Transversions					
A--C	22	28	26	21	-
A--T	26	22	18	25	-
G--C	1	4	1	6	-
G--T	6	3	9	5	-
Total	55	57	54	57	-
Indels	3	4	4	1	-
Number of sites	159	160	157	163	-
Total changes	80	82	87	92	-
% changes	50.3	51.2	55.4	56.4	-

Table 7(4). Nature and frequency of specific base substitutions in five protein coding genes between *Ridgeia*, a vestimentiferan and *Drosophila*, mouse, frog, sea urchin and nematode worm

D. COIII

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	26	25	27	47	50
C--T	85	81	71	76	79
Total	111	106	98	123	129
Transversions					
A--C	34	55	58	51	50
A--T	76	73	65	62	86
G--C	12	22	17	15	32
G--T	12	12	13	28	35
Total	134	162	153	156	203
Indels	2	1	1	1	5
Number of sites	677	676	676	676	662
Total changes	247	269	252	280	337
% changes	36.5	39.8	37.3	41.4	50.9

Table 7(5). Nature and frequency of specific base substitutions in five protein coding genes between *Ridgeia*, a vestimentiferan and *Drosophila*, mouse, frog, sea urchin and nematode worm

E. Cyt B

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	41	36	32	34	43
C--T	45	49	45	46	43
Total	86	85	77	80	86
Transversions					
A--C	18	26	29	21	17
A--T	38	22	29	32	45
G--C	10	25	22	32	18
G--T	31	15	26	23	44
Total	97	88	106	108	126
Indels	1	1	2	1	2
Number of sites	391	391	389	391	389
Total changes	184	174	185	189	214
% changes	47.1	44.5	47.6	48.3	55.0

similarity.

Further comparison of both COII and ATPase 8 genes with their counterparts revealed the nature and frequency of specific base substitutions (Table 7). The data presented in Table 7 suggest that the two genes exhibit strikingly different patterns, or modes of evolution. For example, in comparisons of the COII gene, more than 60% of the detectable base pair differences are transversions, and more than 75% of these are A-C or A-T transversions. In comparisons of ATPase 8 genes about 68% of the detectable base pair differences are transversions, and more than 84% of these are A-C or A-T transversions. In the COII gene, the content of A-G transitions is about 37% of total transitions; thus the frequency of C-T transitions is about 1.8 times that of A-G transitions. In contrast, in the ATPase 8 gene, A-G transitions comprise only 18% of total transitions; the frequency of C-T transitions is 5 times that of A-G transitions. Although the number of G-C transversions from both COII and ATPase 8 genes is low, the number of G-C transversions from the ATPase 8 gene is lower and is less than half that observed in the COII gene.

Base substitutions in all five protein coding genes, relative to five taxa are summarized in Table 8. Assuming the alignments are reasonably accurate, about 60% of all substitutions are transversions. The most common type of substitution is C-T transitions while the most common transversions are A-T. This latter result probably reflects the general high A+T content in these mtDNAs.

Aquadro et al. (1984) modeled the dynamics of the substitution process. They showed that the decrease in the proportion of transitions observed as divergence increases is consequence of transversions obscuring transitions. They further suggest that although a portion of the mtDNA molecule evolves at an extremely rapid rate, a significant portion

Table 8. Summary of mutation types for protein coding genes, between *Ridgeia* and members of five other taxa

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	109	109	113	143	152
C--T	245	228	230	252	227
Total	354	337	343	395	379
Transversions					
A--C	129	190	200	164	142
A--T	245	192	188	189	279
G--C	48	73	73	80	90
G--T	73	54	68	84	135
Total	495	509	529	517	646
Indels	7	8	7	5	10
Number of sites	2092	2093	2101	2105	1933
Total changes	856	854	879	917	1035
% changes	40.9	40.8	41.8	43.6	53.5
Jukes-Cantor Divergence	58.7	58.4	60.8	64.9	93.0

is under strong selective constraints. Variable selective constraints can be imposed by functional considerations or structural considerations and can affect the rate of evolution of the region under consideration. The degree of constraint varies among genes as can be seen by comparing rates of evolution calculated for the "slowly evolving" COII gene versus the "rapidly evolving" ATPase 8 gene. The extreme differences in substitution rates for the COII and ATPase 8 genes was previously noted by Clary and Wolstenholme (1985).

Transfer RNA Genes

Five potential tRNA genes were identified. All of these sequences can be folded into the cloverleaf structure characteristic of almost all known tRNAs (Figure 8). Sequence comparison indicated that these genes are homologous to their sea urchin, mouse, frog, and *Drosophila yakuba* counterparts (Figure 9) with similarity ranging from 50% for $tRNA^{ser}$ up to 56.2% for $tRNA^{val}$ (Table 6). They are generally smaller than their vertebrate and invertebrate counterparts in size.

Comparisons of mitochondrial tRNAs from five vertebrate and invertebrate species (mouse, frog, *Drosophila yakuba*, sea urchin, and *Ridgeia*) show that the degree of conservation of different functional regions within tRNAs varied. The most conserved region is anticodon loop, there are 100% identity for $tRNA^{asp}$ among five taxa. The other tRNA genes are more than 80% similar in anticodon loop. The most conserved regions besides the anticodon loop are the dihydrouridine (DHU) stems. The DHU loop, the ribothymidine pseudouridine cytosine (TΨC) loop and TΨC stem were the most variable regions. The degree of conservation of the 5' half was higher than that of the 3' half.

The differences between the mitochondrial tRNA genes of *Ridgeia* and those of other

taxa involve mainly substitutions, although small insertion/deletion events of 1-5 nucleotides in length also occur in the different functional regions (Figure 9). Detailed comparison of these five tRNA genes with their counterparts shows the nature and frequency of specific base substitutions (Table 9). The data presented in Table 9 suggest that these tRNA genes exhibit strikingly different pattern, or modes of evolution. Of all five tRNA genes, three ($tRNA^{Val}$, $tRNA^{Asn}$, and $tRNA^{Gly}$) have similar numbers (39, 42, and 41 respectively) and frequencies (40%, 43%, and 42% respectively) of transitions. In the remaining two tRNA genes, more than half (52% and 60%) of the detectable base pair differences are transitions. For all five tRNA genes, the number of A-G and C-T transitions is almost same. In all five tRNA genes, the number and frequency of A-C transversions are much higher than those of A-T, G-C, and G-T transversions, ranging from 45% for $tRNA^{Val}$ up to 74% for $tRNA^{Tyr}$ of total transversions. The number and frequency of A-T and G-T transversions is approximately same, the lowest is that of G-C transversions, ranging from 7% for $tRNA^{Asp}$ to 1% for $tRNA^{Tyr}$ of total substitutions. These results suggest that the rate of evolution varies among as well as within mitochondrial tRNA genes. In the rat, mouse, cow and frog comparison, Gadaleta et al. (1989) demonstrated that the tRNAs which coded for the most commonly used amino acids were most conserved. Comparisons among humans showed that tRNA genes that recognize four-fold degenerate codons are generally more variable than tRNA genes that recognize two-fold degenerate codons (Cann et al. 1984).

In the attempt to identify common features, Gadaleta et al. (1989) found that many tRNAs possess the triplet TAG at positions 8, 9, and 10. A second TAG is present also in 11 genes at a fixed distance (20-27 bases) from the 3' terminus. TAG at positions 8, 9, and 10 is also found at high frequency in prokaryotic and eukaryotic tRNAs (ref. in Sprinzl et

Table 9. Nature and frequency of specific base substitutions in tRNAs in mitochondrial DNA of vestimentiferan, human, mouse, yakuba, frog, and sea urchin

	tRNA genes				
Type of changes	tRNA[Val]	tRNA[Ala]	tRNA[Asp]	tRNA[Tyr]	tRNA[Gly]
Transitions					
A--G	18	22	26	25	22
C--T	21	20	25	31	19
Total	39	42	51	56	41
Transversions					
A--C	26	27	31	28	31
A--T	13	15	7	4	11
G--C	6	3	7	1	2
G--T	13	11	3	5	13
Total	58	56	48	38	57
Total substitutions	97	98	99	94	98

al. 1987), although only T at position 8 is considered invariant. In the $tRNA^{Asp}$ gene of the vestimentiferan, the triplet TAG is also present at positions 8, 9, and 10, the remaining two triplets TAG were found at positions 24, 25, and 26 and positions 53, 54, and 55. The position of second triplet TAG is not located at a fixed distance (20-27 bases) from the 3' terminus, as described by Gadaleta et al. (1989). In the $tRNA^{Ser}$ gene, three sets of triplet TAG were found. However, all positions of three TAG are quite different from that of rat, the first triplet TAG is found at positions 36, 37, and 38 downstream from 5' end, the second TAG is located at positions 19, 20 and 21 upstream from the 3' end, and last TAG occurs at the 3' end of this gene. In the $tRNA^{Asp}$ gene, the sequence TAGTTAAA occurs, starting from the position 8 downstream from 5' end of this gene. This sequence is at the start of the A block of the Pol III enzyme and is probably important for the evolution of the tRNAs in general. The same sequence has not been found in the protein, ribosomal RNA genes, and intergenic sequences.

Ribosomal RNA Genes

Although complete sequences for both ribosomal RNA genes were not determined, partial sequences (Figure 3B) of both small and large ribosomal RNA genes show some features of these two genes. Alignment of the vestimentiferan small and large ribosomal RNA genes with the corresponding parts from vertebrates and invertebrates is presented in Figures 12 and 13. Both small and large ribosomal RNA genes in the vestimentiferan are shorter than those of vertebrates and some invertebrates (sea urchin).

The differences in size between corresponding vestimentiferan and vertebrate and invertebrate mitochondrial ribosomal RNA genes may result from the absence in the

vestimentiferan rRNA genes of specific blocks. Indeed, as shown in Figure 12, there are at least six major deletion events in addition to those of one to two nucleotides in the small ribosomal RNA gene. Similarly, there are at least 11 major deletion events and three insertion events besides those of one to two nucleotides in the large ribosomal RNA gene (Figure 13). It is surprising that a region of 237 nucleotides, located at position 434 from the 5' end of this gene, accounts for more than half of the total nucleotide deletions in vestimentiferan large ribosomal RNA gene (Figure 13). In contrast, the longest deletion in the vestimentiferan small ribosomal RNA gene is a region of 45 nucleotides in length (Figure 12). These addition and deletion blocks are variable in length, and range from single, up to 237 nucleotides. Although a small number of insertions are evident in both small and large ribosomal RNA genes, the majority are deletions. The differences result in the variation in size of some loops or stems and in the addition or deletion of some domains. This situation agrees with the view of Gray et al. (1984) who consider addition or deletion of rRNA pieces a common feature of rRNA evolution.

Detailed comparison of the vestimentiferan small RNA primary sequence with those reported from human (Anderson et al. 1981), frog (Roe et al. 1985), and sea urchin (Jacobs et al. 1988) mtDNAs reveals the presence of structural domains that are conserved among these species (Figure 12). Similarly, detailed comparison of the vestimentiferan large rRNA primary sequence with those reported from frog (Roe et al. 1985) and sea urchin (Jacobs et al. 1988) mtDNAa shows the presence of structural domains that are conserved among these species (Figure 13). It is found that in both small and large rRNA genes, the regions near the 3' end of these genes are more highly conserved than those near the 5' end of these genes. Although the sequences of both small and large rRNA have not been folded into

the secondary structures, detailed comparisons of these secondary structures with those from various vertebrates and invertebrates have also not been finished, the data from these primary sequences of two rRNA genes reveal that the presence of primary structural domains that are conserved among all organelle and even nuclear rRNAs and define a universal core structure presumably essential for ribosomal functions.

Further comparison of both small and large rRNA genes with those from other taxa revealed the nature and frequency of specific base substitutions (Table 10 and Table 11). In the small ribosomal RNA gene, the number of transitions in comparisons to human, frog and sea urchin are very similar, comprising 42% of total base substitutions. The numbers of A-G and C-T transitions are also equal. More than 57% of base substitutions are transversions. The frequencies of A-C and A-T transversions are high 66% (sea urchin), 75% (human), and 82% (frog) of observed transversions. The number of G-C transversions is less than 5% of total base substitutions.

In the large ribosomal RNA gene, transitions comprise 34-37% of total base substitutions with the frequencies of A-G and C-T transitions being almost equal. More than 63-66% of the detectable base pair differences are transversions, more than 71-73% of transversions occurred in A-C and A-T transversions. Like that of small ribosomal RNA gene, the number of G-C transversions is less than 6-7.6% of total base substitutions. The ratios of transversion to transition are 1.7 and 1.9 respectively for frog and sea urchin.

Table 10. Nature and frequency of specific base substitutions in 12S rRNAs in mitochondrial DNA of vestimentiferan, human, frog, and sea urchin

	Species compared		
Type of changes	HUMAN	FROG	SEA URCHIN
Transitions			
A--G	37	39	44
C--T	46	44	49
Total	83	83	93
Transversions			
A--C	38	37	36
A--T	45	54	47
G--C	13	10	12
G--T	15	16	30
Total	111	117	125
Insertions	15	28	64
Deletions	115	114	75
Total substitutions	194	200	218

Table 11. Nature and frequency of specific base substitutions in 16S rRNAs in mitochondrial DNA of *Ridgeia* compared to frog and sea urchin

	Species compared	
Type of changes	FROG	SEA URCHIN
Transitions		
A--G	72	64
C--T	80	76
Total	152	140
Transversions		
A--C	88	76
A--T	104	114
G--C	26	31
G--T	46	46
Total	264	267
Insertions	66	181
Deletions	363	442
Total substitutions	416	407

Summary

The relative arrangement of tRNA and protein genes in the sequenced segment of *Ridgeia* (Phylum Vestimentifera) mtDNA molecule suggests that there is a considerable different in relative gene order between the mtDNAs of *Ridgeia*, vertebrates, and some invertebrates. This study has demonstrated that the ATPase 6 gene has been independently translocated during evolution. Also, the $tRNA^{ser}$ gene, $tRNA^{asp}$ gene, and both of $tRNA^{tyr}$ and $tRNA^{gly}$ genes appear to have been translocated as a single segment.

In addition to a few insertions and deletions, a number of single and multi-codon mutations were observed in the protein coding genes of *Ridgeia* mtDNA. This result suggests that those amino acids which are completely conserved in the corresponding protein genes of vertebrate and some invertebrate mtDNAs, have accepted non-conservative replacements in the vestimentiferan evolution. The data from comparison of both COII and ATPase 8 genes with those from other taxa reveals that the two genes exhibit strikingly different patterns, or modes of evolution. A clear demonstration of variability of constraint among different mitochondrial protein coding regions has been seen by comparing rates of evolution calculated for the "slowly evolving" COII gene versus the "rapidly evolving" ATPase 8 gene.

The differences between vestimentiferan mitochondrial tRNA genes and those of other taxa involve mainly substitutions, although small insertion/deletion events also occur in the different functional regions. The most conserved region is anticodon loop. The degree of conservation of the 5' half was higher than that of the 3' half. The DHU loop and TΨC loop and TΨC stem were the most variable regions. Data on the nature and frequency of specific base substitutions of these tRNA genes suggest that the rate of evolution varies

among as well as within mitochondrial tRNA genes.

A number of deletion events in both ribosomal RNA genes result in both rRNA genes much shorter than those of vertebrates and some invertebrates. These differences reflect mainly the addition or deletion of entire domains, a common feature of rRNA evolution. Besides a number of deletion events, the evolution of both ribosomal RNA genes involves specific base substitutions. These results show that two ribosomal genes exhibit different patterns.

SECTION III

GENE ORGANIZATION IN THE MITOCHONDRIAL GENOME OF *SAGITTA ELEGANS*, (PHYLUM CHAETOGNATHA)

INTRODUCTION

Complete nucleotide sequences of the mitochondrial DNA genome are available for several animal species. These genomes, each representing different metazoan taxa, encode the same genes for 12 or 13 polypeptides, two ribosomal RNAs and 22 transfer RNAs, as well as a control region containing the initiation sites for mtDNA replication and transcription. These and other similarities support the general view that the mtDNAs of animals are highly conserved in terms of their gene content, structure, and function.

Vertebrate mitochondrial genomes have a relatively constant organization. The gene arrangement of eutherian mammals and the frog, *Xenopus* is identical. Only two gene rearrangements are known among the vertebrates. A transposition has been demonstrated in the chicken sequence, relative to that of mammals (Desjardins and Morais 1990), and a transposition of two tRNAs has been observed in marsupials, compared to the eutherian arrangement (Pääbo et al., 1991).

In contrast with the relatively constant gene organization within the Phylum Chordata, sequence data from various invertebrate groups reveal that many rearrangements, including inversions and transpositions, have occurred between these phyla. Between vertebrate and *Drosophila* mtDNAs there is a difference in gene order for 11 tRNA genes, the 2 ribosomal RNA genes, and 5 protein genes. In sea urchin, there are novel gene arrangements as compared to the pattern found in the vertebrates. The order of genes in sea urchin mtDNA shows two major differences from that of vertebrates: the genes for 16S rRNA and for ND4L are in new positions, though they remain encoded on the same strand. The 16S rRNA gene is located between the genes for ND2 and COI, rather than between the 12S rRNA gene and the ND1 gene; the ND4L gene is between the genes for COI and COII,

rather between those for ND3 and ND4. In contrast to the two rearrangements involving the protein-coding and ribosomal genes, since the divergence of the vertebrate and echinoid lines up to 15 independent transpositions and inversions of tRNA genes have taken place. A major inversion in the mtDNA of the sea star versus the sea urchin has been observed by Smith et al. (1989). Between mtDNAs of nematodes and *Drosophila*, and nematodes and mammals extensive rearrangements involving almost all tRNA, rRNA, and protein genes have occurred.

Protein genes of metazoan mtDNAs use genetic codes that are modified (Barrell et al. 1979, 1980; Bibb et al. 1981). In all metazoan mtDNAs, TGA specifies tryptophan rather than being a stop codon. ATA has been interpreted as specifying methionine rather isoleucine in all metazoan mitochondrial genetic codes except those of echinoderms (Jacobs et al. 1988; Cantatore et al. 1989). In vertebrate mitochondrial protein genes, AGA and AGG are absent (Bibb et al. 1981) or are used as rare stop codons (Anderson et al. 1981, 1982; Roe et al. 1985). However, in *Drosophila yakuba*, 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; Okimoto et al. 1992).

Both translation initiation and translation termination of metazoan mitochondrial protein genes have unusual features. Among many of these protein genes, triplets other than ATG (AUG) are used as translation initiation codons. These include ATA, ATT (except sea urchin), ATC (except *Drosophila yakuba*), GTG, GTT and TTG. Also, ATAA has been suggested as the translation initiation codon of the *Drosophila* COI gene.

Some intergenic sequences have been found within the mitochondrial genomes from various animal species. Some of them have been assigned functions related to replication and transcription (Bibb et al. 1981; Roe et al. 1985). However, in most cases other than vertebrates the function of the intergenic sequence is poorly understood as is the mechanism of replication and transcription in these organelles.

PCR has recently been devoloped as a standard laboratory technique, with applications in all areas of molecular biology research. Highly conserved regions of mtDNA can be chosen as primers for PCR amplification.

Although many sequences of mtDNA from vertebrates and invertebrates are available, information about mtDNA from the Phylum Chaetognatha has never been reported. Members of the phylum, known as arrow worms, are common animals found in marine plankton. About 65 species have been described. The adults possess none of the features common to the other deuterostome phyla, and they are like aschelminths in many respects. Only the embryogeny of arrowworms would suggest a deuterostome position for these animals. Chaetognatha are an ancient and conservative phylum. Some of their fossils are five hundred million years old, indicating that arrowworms had already evolved by Cambrian times. Chaetognatha are a very enigmatic group, and nobody has ever felt confident of their status; de Beauchamp described them as 'possibly the most isolated group in the animal kingdom', and they have at various times been allied with molluscs, nematodes, spiders, or brachiopods. The chaetognatha were in fact considered as pseudocoelomic; but the muscles are very different in structure to those of nematodes, and the cuticle is not moulted. Their nervous system is described as more like that of protostomes than deuterostomes (Rehkamper & Welsch 1985). In some respects the chaetognatha are quite unique.

Independent evolution of the peculiar chaetognatha design from a very early stage, probably no further advanced than the earlist eucoelomate form, seems the most plausible explanation for this phylum's origins. The phylogenetic relationships of chaetognatha to other animal phyla are still not certain. In order to gain the information on gene organization and evolution of this mitochondrial genome, I chose *Sagitta* (Phylum Chaetognatha) for analysis. Using PCR, two major segments coding three functional classes of genes as well as some intergenic sequences were amplified and sequenced. Sequence analysis of this mitochondrial genome shows substantial differences from those of all vertebrates and invertebrates.

RESULTS AND DISCUSSION

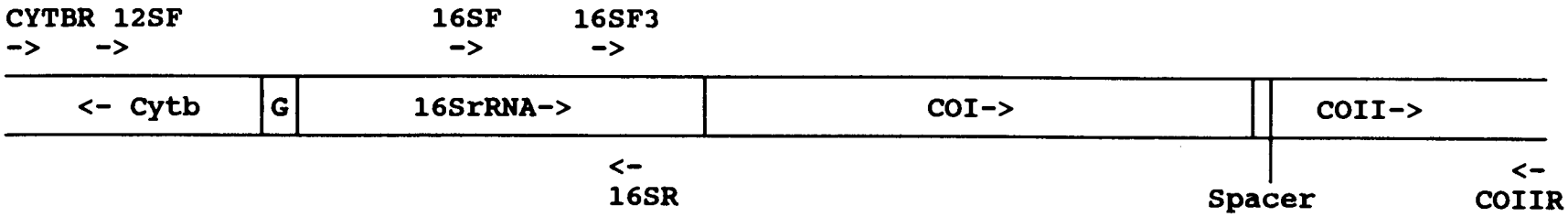
The Polymerase Chain Reaction

Several pairs of primers were used for PCR amplification of *Sagitta* mitochondrial DNA. Their sequences are listed in Table 1. The location of the genes and PCR amplification and sequencing strategy for *Sagitta* mtDNA are shown in Figure 16. Two of same pairs of primers used for the vestimentiferan, *Ridgeia*, also amplified fragments in the chaetognath. No information on gene organization from the chaetognath mitochondrial DNA was known, and it turned out to be radically different. The first pair of primers, COIF and COIIIR, amplified a fragment of 1,397 base pairs, containing partial ND2 and COIII genes of the chaetognath mitochondrial DNA (Figure 16). Using the same pair of primers, a fragment of 1,985 base pairs long was obtained from the vestimentiferan, *Ridgeia*, encoding COI, COII, ATPase 8, COIII, and four tRNA genes. It can be inferred that COIF primer had sufficient similarity to a site in ND2, to allow amplification and that COI is either located too far from COIII to efficiently amplify a fragment or its transcriptional orientation is reversed relative to COIII. The sequence shows that the COI gene is not located between ND2 and the 5' end of COIII. The second pair of primers, 12SF and 16SR, amplified a segment of the chaetognath mitochondrial DNA, encoding cytochrome B, the large ribosomal RNA, and a tRNA gene. As described above, using this pair of primers, a segment across both small and large ribosomal RNA genes and $tRNA^{Val}$ gene was obtained in case of the vestimentiferan mitochondria. It can be inferred that 12S rRNA has moved from position near 5' end of 16S rRNA, where it occurs in vertebrates, insects and vestimentiferans. These results demonstrate that specificity imposed by a single primer is

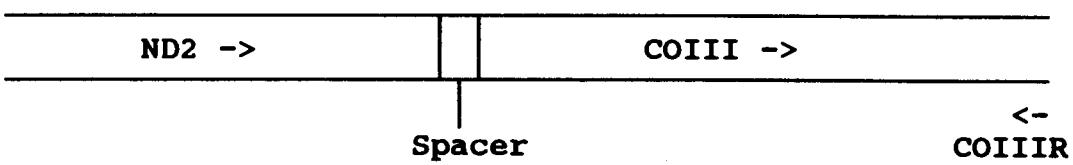
FIGURE 16.

Location of the genes and PCR amplification and sequencing strategy for *Sagitta* (Phylum Chaetognatha). Abbreviations of protein names are as used elsewhere in text. The transcriptional polarity is shown by arrows. The $tRNA^{Gly}$ is designated with the single-letter amino-acid code G. The names and sequences of PCR primers are listed in Table 1. PCR conditions are described in text. A. A fragment containing cytochrome B, $tRNA^{Gly}$, 16S rRNA, COI, and COII genes; B. A segment coding both ND2 and COIII genes amplified by using a pair of primers COIF and COIIIR.

A.


CYTBR 12SF
-> ->
16SF
->
16SF3
->
<- Cytb
G
16SrRNA->
COI->
COII->
<-
16SR
Spacer
<-
COIIR

B.


COIF
->
ND2 ->
COIII ->
Spacer
<-
COIIIR

often sufficient to allow amplification of the mitochondrial gene for which it was designed, if a secondary priming site for the paired primer is present. The third pair of primers, COIF and COIIR (used only for the chaetognath), was used for amplification of another fragment from *Sagitta*. When the sequence of all these fragments was determined completely, a set of new primers was designed for finding the relationships between these fragments. The fourth pair of primers, 16SF3 and COIIR, was used for amplification of the longest fragment of about 3 kilobase long, which joins the two fragments together. Therefore, a large fragment was obtained, which codes for cytochrome B, $tRNA^{asp}$, large ribosomal RNA, COI, and COII genes.

In this study, using adults of the chaetognath as source material, preparation of total cellular DNA is sufficient for PCR amplification. It was not necessary to purify mitochondrial DNA prior to amplification. In all cases, double strands of mitochondrial DNA were amplified. The longest fragment amplified was about 3 kilobases in the chaetognath using the normal buffer system under general conditions.

Sequence and Location of Structure Genes

A total of 5,183 base pairs from the *Sagitta* mitochondrial genome was determined. Figure 17 shows the complete nucleotide sequence of two fragments amplified by PCR. The first fragment of 3,786 base pairs in length contains three open reading frames that, from comparisons of inferred amino acid sequences to those of previously identified mitochondrial genes of human, mouse, cow, *Drosophila yakuba*, frog, and sea urchin, were identified as the genes for cytochrome B (most), COI, COII (most). This fragment also encodes the entire large ribosomal RNA gene. The second segment of 1,397 base pairs ,

FIGURE 17.

Nucleotide sequence of two coding regions from *Sagitta* mitochondrial genome. The PCR amplification and sequencing strategy for these two fragments is shown in Figure 16. The transcriptional polarity of individual genes is indicated by the arrows (->). Translation of the mitochondrial protein genes is indicated below or above the sequence using the 1-letter amino acid code. Abbreviations of protein names are as used elsewhere in text. Stop codons are designated by asterisks. The limits of the tRNA genes are enclosed with brackets. The tRNA anticodon is underlined. The non-coding sequences between genes are indicated as "spacer". A. A fragment containing cytochrome B, $tRNA^{Gly}$, COI and COII genes. Both strands of this fragment are shown. B. A segment coding ND2 and COIII genes.

A.

```
5'
TTGGCGGCGATAAAGTTATCCGGTTCTCCCAAGAAATTCGGAGCAAAAATAACGAAAACA 60
AACCGCCGCTATTTCAATAGGCCAAGAGGGTTCTTTAAGCCTCGTTTTTATTGCTTTTGT
3'P  P  S  L  T  I  R  N  E  W  S  I  R  L  L  F  L  S  F  L
  <- cytochrome B

GAAGCGTAAACATTACAACAAAACCTAACATATCTTTTACACTGTAATATCAATGGAAAG 120
CTTCGCATTTGTAATGTTGTTTTGGATTGTATAGAAAATGTGACATTATAGTTACCTTTC
  L  T  F  M  V  V  F  G  L  M  D  K  V  S  Y  Y  W  H  F  P

GAATCTTATCTGAAGTAGAGTTAACACCCAATGGATTGTTGCTACCTGTCTCATGTAAGA 180
CTTAGAATAGACTTCATCTCAATTGTGGGTTACCTAACAACGATGGACAGAGTACATTCT
  I  K  D  S  T  S  N  V  G  L  P  N  N  S  G  T  E  H  L  L

GAAAGATGTGAAGTCCAGCTATTGCTGAAGCGATAAAAGGTACTAAGAAGTGGAAAGTAA 240
CTTTCTACACTTCAGGTCGATAACGACTTCGCTATTTTCCATGATTCTTCACCTTTCATT
  F  I  H  L  G  A  I  A  S  A  I  F  P  V  L  F  H  F  T  F

AGAAGCGAGTAAGAGTAGCGTTCTCCACAGCGAACCCTCCCCACAACCAGTGCACTAGTC 300
TCTTCGCTCATTCTCATCGCAAGAGGTGTCGCTTGGGAGGGGTGTTGGTCACGTGATCAG
  F  R  T  L  T  A  N  E  V  A  F  G  G  W  L  W  H  V  L  R

TGGGTCCCACATAAGGGAATGCAGAAAATAAATTTGTAATAACAGTAGCCCCTCAGAACC 360
ACCCAGGGTGTATTCCCTTACGTCTTTTATTTAAACATTATTGTCATCGGGGAGTCTTGG
  P  G  V  Y  P  F  A  S  F  L  N  T  I  V  T  A  G  W  F  R

TTATTTGGCCTCAAGGCAACACGTAGCCAAGAAAAGCGGCCGCCATTACTAAAATTAACA 420
AATAAACCGGAGTTCCGTTGTGCATCGGTTCTTTTCGCCGGCGGTAATGATTTTAATTGT
  I  Q  G  W  P  L  V  Y  G  L  F  A  A  A  M  V  L  I  L  L

AAGCTAAAACCCTAAACCATGTCTTAATAAACATATATCTGCCATAATAAATACCTCGTC 480
TTCGATTTTGGGATTTGGTACAGAATTATTTGTATATAGACGGTATTATTTATGGAGCAG
  A  L  V  R  F  W  T  K  I  F  M  Y  R  G  Y  Y  I  G  R  G

CGATGTGACAGTAAAGACATATAAAGAAAAAAGAAGCCCCGTTTGCGTGTACTCTGCGGA 540
GCTACACTGTCATTTCTGTATATTTCTTTTTTCTTCGGGGCAAACGCACATGAGACGCCT
  I  H  C  Y  L  C  I  F  F  F  S  A  G  N  A  H  V  R  R  L

GGAACCATCCTCTATTTACGTCCCGGAAAATATGATCCACAGAAGAAAAAGCCATATTCA 600
CCTTGGTAGGAGATAAATGCAGGGCCTTTTATACTAGGTGTCTTCTTTTTCGGTATAAGT
  F  W  G  R  N  V  D  R  F  I  H  D  V  S  S  F  A  M  N  V

CATCACATGTGTAATGTATAGCTAAAAAAAGACCTGTTGCAAGTTGGATTACCAAGACTA 660
GTAGTGTACACATTACATATCGATTTTTTTCTGGACAACGTTCAACCTAATGGTTCTGAT
  D  C  T  Y  H  I  A  L  F  L  G  T  A  L  Q  I  V  L  V  L

AACCTAATAGACTACCAAAATTTCACCATACTCTAATATTGATTGGTGAAGGCAAATCAA 720
TTGGATTATCTGATGGTTTTAAAGTGGTATGAGATTATAACTAACCACTTCCGTTTAGTT
  G  L  L  S  G  F  N  W  W  V  R  I  N  I  P  S  P  L  D  I
```

```
TCAGAGTATTATTAATAATTTTCAAGCTAGGCTCGCTCAGCCGTGTTCGTTTTTGTGTAA 780
AGTCTCATAATAATTATTAAAAGTTCGATCCGAGCGAGTCGGCACAAGCAAAAACACATT
  L  T  N  N  I  I  K  L  S  P  E  S  L  R  T  R  K  Q  T
                                                <- cytochrome B <-

AGTCCTATAGTAGTATGTTTGTAAATATATTCTGATGCTGCTTTTCGTCTGGAATAAGAA 840
TCAGGATATCATCATACAAACATTTATATAAGACTACGACGAAAAGCAGACCTTATTCTT
   Spacer          ->(                  <- tRNA^gly

                      16S rRNA ->
TAGTTTATTTGAACATATTTTTGAAAAAGATAAAGACCTGTTAAGTGATTGTAAATCTAG 900
ATCAAATAAACTTGTATAAAAACTTTTTCTATTTCTGGACAATTCACTAACATTTAGATC
                     )

TTTAACGTCAAACATGTTTTTTACACAGACATAAACTTCAACCGGTGCATTCTATCCTTT 960
AAATTGCAGTTTGTACAAAAAATGTGTCTGTATTTGAAGTTGGCCACGTAAGATAGGAAA

TGTATAATGAAAGTAGGATAGAATTTGTGGGATCTCTGTTTAAAGGATACATTAAATATT 1020
ACATATTACTTTCATCCTATCTTAAACACCCTAGAGACAAATTTCCTATGTAATTTATAA

ATAATGAACAGGTAGGACAAAAATTGTATATCGACCTCAACAAACTAAATATACAAACAA 1080
TATTACTTGTCCATCCTGTTTTTAACATATAGCTGGAGTTGTTTGATTTATATGTTTGTT

GAAAGAAAAGCTTAATCTTAAAAACCTAGGTAAAACGATCTAGACCTTTAAGATATACTT 1140
CTTTCTTTTCGAATTAGAATTTTTGGATCCATTTTGCTAGATCTGGAAATTCTATATGAA

ATTTATTTAACCTTCATATTAGGTTTATAACATTAACCATTTACAATATAGGATTTAATA 1200
TAAATAAATTGGAAGTATAATCCAAATATTGTAATTGGTAAATGTTATATCCTAAATTAT

TAATTAATAGTTTCGATATAGCATCTACTTTGAAATAAAAATTTAAAAGTAAAATAAAAT 1260
ATTAATTATCAAAGCTATATCGTAGATGAAACTTTATTTTTAAATTTTCATTTTATTTTA

GGACTTCCGAAAGTTTCACAAAAACCTTTCCTTATAATATACAAGGTACACCCTGCTAAA 1320
CCTGAAGGCTTTCAAAGTGTTTTTGGAAAGGAATATTATATGTTCCATGTGGGACGATTT

GAAAAGATGTAGCATAATTAATAGCCACTTTATTGGTGGAGAGTCTGAAAGGGCACTCGA 1380
CTTTTCTACATCGTATTAATTATCGGTGAAATAACCACCTCTCAGACTTTCCCGTGAGCT

GAGGTTTCAGTATTACGTATAAAAAAAAATCTATATATATTTATATCTAAAAAAAATATT 1440
CTCCAAAGTCATAATGCATATTTTTTTTTAGATATATATAAATATAGATTTTTTTATAA

GAAGGCGAAAATACCTCTACAATGAATAAAGGCGATAAGACCCTAAAAGCTCTATCAATA 1500
CTTCCGCTTTTATGGAGATGTTACTTATTTCCGCTATTCTGGGATTTTCGAGATAGTTAT

GTTGTTGGGGCAACGAGAAATATAACATTTCTTTCTTCTAGCTATTTATACTAGTTACTT 1560
CAACAACCCCGTTGCTCTTTATATTGTAAAGAAAGAAGATCGATAAATATGATCAATGAA

TAGGGATAACAGCGCAATATCCATTAACAAGATCATATTTATAATGGAGAGTGTGACCTC 1620
ATCCCTATTGTCGCGTTATAGGTAATTGTTCTAGTATAAATATTACCTCTCACACTGGAG
```

```
GATGTTGAACTAAAGGAAAACTAGAGTATACACTAGTGTCTATCGTCTGTTCGCCGATTT 1680
CTACAACTTGATTTCCTTTTGATCTCATATGTGATCACAGATAGCAGACAAGCGGCTAAA

TACCTTTACGTGATTTGAGTTCAAACCGGTGTAAACCAGGTTGGTTTCAACCTTCTTTAA 1740
ATGGAAATGCACTAAACTCAAGTTTGGCCACATTTGGTCCAACCAAAGTTGGAAGAAATT

TATTATTATTTGGATAGTACGAAAGGAAATTCGGTTAGTTACTGATTTAAGTTATTTTTA 1800
ATAATAATAAACCTATCATGCTTTCCTTTAAGCCAATCAATGACTAAATTCAATAAAAAT

ATATTTTCTTGTCGAGAAAAAGAACTTCTAATTCTGCGTATATAGTTATCAAAATATTTT 1860
TATAAAAGAACAGCTCTTTTTCTTGAAGATTAAGACGCATATATCAATAGTTTTATAAAA

                    COI ->
        16S rRNA -> M  T  R  W  L  F  S  T  N  H  K  D  I  G  14
TCTTGTAAAAAAAAAGCTAATGACCCGTTGGTTATTTTCCACTAATCATAAAGACATCGGT 1920
AGAACATTTTTTTTCGATTACTGGGCAACCAATAAAAGGTGATTAGTATTTCTGTAGCCA

 T  L  Y  F  V  L  G  I  W  S  A  F  L  G  T  A  L  S  A  L  34
ACTCTTTATTTCGTATTGGGTATCTGATCTGCATTCCTGGGCACCGCTCTCTCAGCACTA 1980
TGAGAAATAAAGCATAACCCATAGACTAGACGTAAGGACCCGTGGCGAGAGAGTCGTGAT

 I  R  L  E  L  G  N  A  G  S  L  L  G  D  D  Q  L  Y  N  V  54
ATTCGACTAGAGTTAGGTAATGCAGGATCGTTGTTGGGTGATGATCAGCTATACAACGTG 2040
TAAGCTGATCTCAATCCATTACGTCCTAGCAACAACCCACTACTAGTCGATATGTTGCAC

 I  V  T  A  H  A  F  I  M  I  F  F  F  V  M  P  T  M  I  G  74
ATTGTCACTGCACATGCATTTATTATGATTTTTTTCTTTGTTATGCCGACCATGATAGGA 2100
TAACAGTGACGTGTACGTAAATAATACTAAAAAAAGAAACAATACGGCTGGTACTATCCT

 G  F  G  N  W  L  V  P  L  I  V  N  A  P  D  I  A  F  P  R  94
GGTTTCGGAAATTGGTTGGTTCCGCTAATAGTTAACGCTCCCGACATAGCTTTTCCTCGT 2160
CCAAAGCCTTTAACCAACCAAGGCGATTATCAATTGCGAGGGCTGTATCGAAAAGGAGCA

 L  N  N  I  R  F  W  L  L  P  P  A  L  M  L  L  L  L  S  G  114
TTGAACAATATAAGATTTTGGCTTTTGCCACCTGCTCTAATGTTACTATTACTATCGGGT 2220
AACTTGTTATATTCTAAAACCGAAAACGGTGGACGAGATTACAATGATAATGATAGCCCA

 M  V  E  S  G  V  G  T  G  W  T  V  Y  P  P  L  S  T  V  G  134
ATGGTTGAAAGCGGTGTGGGCACTGGTTGGACTGTCTACCCCCCTTTGAGCACAGTAGGC 2280
TACCAACTTTCGCCACACCCGTGACCAACCTGACAGATGGGGGGAAACTCGTGTCATCCG

 H  T  G  G  A  V  D  L  G  I  F  S  L  H  L  A  G  V  R  S  154
CATACAGGAGGCGCTGTTGATTTGGGTATCTTTTCTCTACACTTGGCGGGGGTCAGGAGT 2340
GTATGTCCTCCGCGACAACTAAACCCATAGAAAAGAGATGTGAACCGCCCCCAGTCCTCA

 I  L  G  S  A  N  F  I  T  T  I  V  N  I  K  G  E  G  I  T  174
ATTCTTGGCAGTGCTAATTTTATTACTACGATTGTGAATATAAAAGGAGAGGGTATAACA 2400
TAAGAACCGTCACGATTAAAATAATGATGCTAACACTTATATTTTCCTCTCCCATATTGT

 I  E  L  M  S  L  F  V  W  S  V  L  L  T  A  I  L  L  L  L  194
ATAGAACTTATGTCCCTGTTTGTATGATCGGTTTTGTTAACGGCCATTTTGCTATTGTTG 2460
TATCTTGAATACAGGGACAAACATACTAGCCAAAACAATTGCCGGTAAAACGATAACAAC
```

```
 S  L  P  V  L  A  G  A  I  T  M  L  L  T  D  R  N  F  N  T  214
TCACTCCCGGTATTGGCTGGTGCAATTACTATGCTTCTAACTGACCGCAATTTCAATACA 2520
AGTGAGGGCCATAACCGACCACGTTAATGATACGAAGATTGACTGGCGTTAAAGTTATGT

 S  F  F  D  P  A  G  G  G  D  P  I  L  Y  Q  H  L  F  W  F  234
TCCTTCTTTGATCCGGCTGGGGGCGGGGATCCAATTTTATATCAACATCTTTTTTGGTTT 2580
AGGAAGAAACTAGGCCGACCCCCGCCCCTAGGTTAAAATATAGTTGTAGAAAAAACCAAA

 F  G  H  P  E  V  Y  I  L  I  L  P  G  F  G  I  V  S  Q  I  254
TTTGGACATCCTGAAGTTTATATTCTCATTCTTCCGGGTTTCGGTATAGTGTCGCAGATT 2640
AAACCTGTAGGACTTCAAATATAAGAGTAAGAAGGCCCAAAGCCATATCACAGCGTCTAA

 I  N  H  Y  S  A  K  G  N  R  F  G  A  L  G  I  I  Y  A  M  274
ATTAATCACTATTCCGCTAAAGGGAATAGGTTTGGAGCTTTAGGCATAATTTATGCCATG 2700
TAATTAGTGATAAGGCGATTTCCCTTATCCAAACCTCGAAATCCGTATTAAATACGGTAC

 R  S  I  A  L  L  G  F  V  V  W  A  H  H  M  F  T  V  G  I  294
AGGTCCATTGCGTTGTTAGGTTTCGTTGTGTGGGCCCACCATATGTTCACGGTGGGTATA 2760
TCCAGGTAACGCAACAATCCAAAGCAACACACCCGGGTGGTATACAAGTGCCACCCATAT

 D  V  D  T  R  A  Y  F  T  S  A  T  I  I  I  A  V  P  T  G  314
GACGTTGATACCCGGGCTTATTTTACATCTGCTACTATAATTATTGCAGTACCTACTGGC 2820
CTGCAACTATGGGCCCGAATAAAATGTAGACGATGATATTAATAACGTCATGGATGACCG

 I  K  I  F  S  W  L  A  T  L  H  G  T  P  S  L  L  E  T  P  334
ATTAAGATTTTCTCTTGATTAGCCACGCTACATGGTACTCCCTCTTTGCTCGAGACTCCT 2880
TAATTCTAAAAGAGAACTAATCGGTGCGATGTACCATGAGGGAGAAACGAGCTCTGAGGA

 L  M  W  V  L  G  F  L  F  L  F  T  V  G  G  L  T  G  V  V  354
TTGATGTGAGTTTTGGGCTTTCTGTTCCTTTTTACAGTTGGCGGTCTCACTGGCGTAGTC 2940
AACTACACTCAAAACCCGAAAGACAAGGAAAAATGTCAACCGCCAGAGTGACCGCATCAG

 L  A  N  S  S  L  D  I  R  L  H  D  T  Y  Y  V  V  A  H  F  374
TTAGCCAATAGTTCTTTAGATATTAGACTGCATGACACTTATTATGTAGTAGCCCATTTC 3000
AATCGGTTATCAAGAAATCTATAATCTGACGTACTGTGAATAATACATCATCGGGTAAAG

 H  Y  V  L  S  I  G  A  V  F  A  I  F  A  G  V  T  F  W  Y  394
CACTACGTATTGTCAATAGGTGCAGTATTCGCCATTTTCGCGGGGGTCACTTTTTGGTAT 3060
GTGATGCATAACAGTTATCCACGTCATAAGCGGTAAAAGCGCCCCCAGTGAAAAACCATA

 P  V  I  S  G  M  T  M  S  A  R  G  T  Q  V  Q  F  A  I  I  414
CCTGTGATTAGTGGGATGACTATGTCCGCTCGTGGTACTCAAGTACAATTCGCGATTATA 3120
GGACACTAATCACCCTACTGATACAGGCGAGCACCATGAGTTCATGTTAAGCGCTAATAT

 F  I  G  V  N  L  T  F  F  P  Q  H  F  L  G  L  Q  G  I  P  434
TTTATTGGTGTAAATCTGACTTTTTTTCCACAACACTTTTTAGGTTTACAAGGAATACCA 3180
AAATAACCACATTTAGACTGAAAAAAAGGTGTTGTGAAAAATCCAAATGTTCCTTATGGT

 R  R  Y  S  D  Y  P  D  S  F  S  T  W  N  V  V  S  S  S  G  454
CGACGGTATTCGGACTACCCTGACTCGTTCTCTACTTGGAACGTAGTGAGTTCATCAGGT 3240
GCTGCCATAAGCCTGATGGGACTGAGCAAGAGATGAACCTTGCATCACTCAAGTAGTCCA
```

```
 R  L  V  R  I  V  G  V  F  M  F  V  G  V  I  W  K  S  L  S  474
AGACTAGTTAGGATTGTTGGAGTTTTTATGTTTGTAGGTGTAATATGAAAATCTTTATCG 3300
TCTGATCAATCCTAACAACCTCAAAAATACAAACATCCACATTATACTTTTAGAAATAGC

 A  I  N  A  R  G  D  V  D  N  L  S  V  E  F  R  P  R  L  P  494
GCTATTAATGCCAGGGGAGACGTAGACAACTTGTCTGTAGAGTTTAGACCTCGTCTACCT 3360
CGATAATTACGGTCCCCTCTGCATCTGTTGAACAGACATCTCAAATCTGGAGCAGATGGA

                                    COI ->    <- Spacer
 V  S  W  H  S  F  N  E  S  T  V  L  G  *                    507
GTCTCTTGACACTCTTTTAACGAGAGCACAGTATTAGGTTAGAGAGGTGTATAGCACATT 3420
CAGAGAACTGTGAGAAAATTGCTCTCGTGTCATAATCCAATCTCTCCACATATCGTGTAA

Spacer          -> COII ->
                   M  S  L  N  F  Q  N  R  N  S  P  L  M  E
ATTTTTTGGTAGTAAGGGATGTCTTTAAACTTCCAGAACAGGAATTCGCCCTTAATGGAA 3480
TAAAAAACCATCATTCCCTACAGAAATTTGAAGGTCTTGTCCTTAAGCGGGAATTACCTT

 Q  L  I  F  F  H  D  W  V  I  V  F  V  S  R  I  T  V  G  Y
CAGTTAATTTTTTTTCATGATTGGGTTATAGTTTTTGTTAGTAGAATCACTGTGGGATAC 3540
GTCAATTAAAAAAAAAGTACTAACCCAATATCAAAAACAATCATCTTAGTGACACCCTATG

 L  I  L  I  V  S  N  K  P  T  H  R  V  L  L  E  S  Q  G  V
CTGATTCTCATTGTTTCAAACAAACCTACTCATCGTGTTTTGTTAGAGTCTCAAGGTGTG 3600
GACTAAGAGTAACAAAGTTTGTTTGGATGAGTAGCACAAAACAATCTCAGAGTTCCACAC

 E  F  A  W  T  A  L  P  C  L  V  L  V  A  I  A  L  P  S  L
GAGTTCGCCTGGACAGCCCTCCCTTGTCTCGTTCTGGTCGCTATTGCTCTTCCATCTCTA 3660
CTCAAGCGGACCTGTCGGGAGGGAACAGAGCAAGACCAGCGATAACGAGAAGGTAGAGAT

 R  L  L  Y  S  M  D  E  I  I  D  P  S  L  T  I  K  A  I  G
CGATTATTGTATTCAATGGATGAAATTATTGACCCTTCTTTAACTATTAAAGCAATAGGT 3720
GCTAATAACATAAGTTACCTACTTTAATAACTGGGAAGAAATTGATAATTTCGTTATCCA

 H  Q  W  Y  W  S  Y  E  Y  S  D  V  D  E  E  S  I  E  F  D
CATCAATGATACTGATCTTATGAATATTCTGATGTAGATGAAGAATCCATAGAATTCGAC 3780
GTAGTTACTATGACTAGAATACTTATAAGACTACATCTACTTCTTAGGTATCTTAAGCTG

COII->
 S  Y 3'
TCTTAT                                                       3786
AGAATA
      5'
```

B.

```
ND2->
TTGTATCTCAATCTTGCGGTTCTCTAATGATTTTGTTGGGTGGTATAGTAGCCGATTCTA 60
   V  S  Q  S  C  G  S  L  M  I  L  L  G  G  I  V  A  D  S

GTTTTTTTTCTGTGTCTTTACTTTTGCATGGCGTGGTTTTTAAGATGGGCCTTATGCCTC 120
S  F  F  S  V  S  L  L  L  H  G  V  V  F  K  M  G  L  M  P

TTCATTTTTGGGTGCCTTGTGTAGTTATAAATTTAACTCGTTTCAACTTATATCTACTCA 180
L  H  F  W  V  P  C  V  V  I  N  L  T  R  F  N  L  Y  L  L

TATCTTGGCAAAAGATTGGTCCTATTGTCATTGTCATGACCGCTTCGGTGGGGTATACTG 240
I  S  W  Q  K  I  G  P  I  V  I  V  M  T  A  S  V  G  Y  T

TGTTGTGTCTAGTCAATGCTGTTGGCGGGTCGTTAGCTATGAGCGGTGTTACCGTGTTGC 300
V  L  C  L  V  N  A  V  G  G  S  L  A  M  S  G  V  T  V  L

CGTTGCTTCTTATTTTTAGAGGAATAGTTCAGATAGGTTGAGTGTTCATAACCACCGGGG 360
P  L  L  L  I  F  R  G  I  V  Q  I  G  W  V  F  I  T  T  G

TTTTCACCTTCTATTATCTTTTTGTTTACTACATTGTCTTAAGTGCTGTTGTCCTATACA 420
V  F  T  F  Y  Y  L  F  V  Y  Y  I  V  L  S  A  V  V  L  Y

GGGCAGTCGCTTCGGTTCAATTCGGTTGGGCACTGCTCAACGCAGGGGGGTTACCTCCTT 480
R  A  V  A  S  V  Q  F  G  W  A  L  L  N  A  G  G  L  P  P

TCTCTGGTTTTATAATCAAACTAAAAGCGATTTTACATATTAAGGGGAGTATAGTGGTAC 540
F  S  G  F  I  I  K  L  K  A  I  L  H  I  K  G  S  I  V  V

TGTTGGTTGGTGCAAGAGGATTGGCTCTAACTTCTTATATCCGCTTACTGTTGAATACAC 600
L  L  V  G  A  R  G  L  A  L  T  S  Y  I  R  L  L  L  N  T

GTTTGAAGTCGGGCCCCAGCTCTGGTTTTCTTGTAGCCACTATGGTTGCTGGTAGAGTTT 660
R  L  K  S  G  P  S  S  G  F  L  V  A  T  M  V  A  G  R  V
                                                    COIII->
AAATACCTTTGATTTCCAATCTATTGTATCTGATTCTTTGTGGCTGAATATATGAACAAA 720
*                                                    M  N  K

CACCCGTTCCATATCGTAGACGTAAGACCTTGACCACTGATAGGTTCTGTGGGGAGTTTG 780
 H  P  F  H  I  V  D  V  R  P  W  P  L  I  G  S  V  G  S  L

TGTCTGGTCGGAGGGCTGGTGACGACCATGCACCGTTACGGTAGCTCTCTTTTCTGGTTG 840
 C  L  V  G  G  L  V  T  T  M  H  R  Y  G  S  S  L  F  W  L

GGTATTGCTCTCATTCTTGCCACTATGTTCCAGTGGTGGCGTGATGTTACACGTGAAGCT 900
 G  I  A  L  I  L  A  T  M  F  Q  W  W  R  D  V  T  R  E  A

ACGTTTCAAGGCAAACACACTGCGAAGGTAGAAAGAGGCATGCGATATGGTATACTACTA 960
 T  F  Q  G  K  H  T  A  K  V  E  R  G  M  R  Y  G  I  L  L
```

```
TTCATCAGTTCCGAGGTTTTTTTCTTTTTAGCTTTTTTTTGGGCATTCTTCCACTCAGCA 1020
 F  I  S  S  E  V  F  F  F  L  A  F  F  W  A  F  F  H  S  A

TTGAGGCCTAACGTGGAGGTCGGGTCTGTTTGGCCTCCGTTAGGTATTTCAGCTATTAAC 1080
 L  R  P  N  V  E  V  G  S  V  W  P  P  L  G  I  S  A  I  N

CCTTTTGATGTCCCTTTACTCAATACATCTATCCTATTGTCTTCCGGTGCCACTATTACA 1140
 P  F  D  V  P  L  L  N  T  S  I  L  L  S  S  G  A  T  I  T

TGAGCACACAGAGCCTTACTAGAAAACCGGTGATTGGAGTCTCAACTAAGCTTGATTATC 1200
 W  A  H  R  A  L  L  E  N  R  W  L  E  S  Q  L  S  L  I  I

ACTGTCGTTCTCGGTTTTTATTTTTCTATATTGCAAGGCTTAGAGTATGTGTGAGCTGGT 1260
 T  V  V  L  G  F  Y  F  S  I  L  Q  G  L  E  Y  V  W  A  G

TTTTCGCTTTCAGACGGTATTTACGGTAGTACTTTTTACGTTGCTACTGGTTTTCACGGT 1320
 F  S  L  S  D  G  I  Y  G  S  T  F  Y  V  A  T  G  F  H  G

TTACATGTTCTCATTGGCACCCTCTTTATTGCGGTGATGGCTTACCGAAACTTATATCAC 1380
 L  H  V  L  I  G  T  L  F  I  A  V  M  A  Y  R  N  L  Y  H

CATTTCAGATGCAGGCA                                             1397
 H  F  R  C  R
```

shown in Figure 17B, contains two open reading frames, that from comparisons of inferred amino acid sequences were identified as the protein genes, ND2 and COIII. The sequence lying between the cytochrome B and large ribosomal RNA genes can be folded into the characteristic secondary structure of mitochondrial tRNA gene that contains anticodon expected for $tRNA^{Met}$. The large ribosomal RNA, COI, and COII genes are transcribed in the same direction, while the transcriptional polarty of the cytochrome B and $tRNA^{Met}$ genes is reversed (Figure 17A). The ND2 and COIII genes share the same transcriptional polarity (Figure 17B). The base composition of these protein genes and large ribosomal RNA gene from the *Sagitta* mitochondrial DNA is shown in Table 12. In all cases, a high percent of T residue is found.

Unlike those of the vestimentiferan (*Ridgeia*), the majority of the protein genes and the ribosomal gene are not punctuated by tRNA genes. The secondary structure provided by tRNA genes is thought to act as a recognition signal(s) for mitochondrial RNA processing (Ojala et al., 1981). Intergenic spaces of variable length from 23 to 49 nucleotides are found between the $rRNA^{Met}$ and cytochrome B, COI and COII, and ND2 and COIII genes. The 16S rRNA and COI genes are butt-joined (Figure 17A). No gene overlap is observed in these sequences.

Genomic Organization

Analysis of the sequence from these two regions reveals that there is a novel gene organization in chaetognath mitochondria. As shown in Figure 17A, the gene order of the first fragment is the cytochrome B, $tRNA^{Met}$, large ribosomal RNA, COI and COII genes; the ND2 and COIII genes located in second fragment are adjacent to one another. This order

Table 12. Base composition in *Sagitta* mitochondrial genes

	Base composition % of total			
Gene	Guanine	Adenine	Thymine	Cytosine
ND2	24.0	19.0	39.3	17.7
COI	22.5	22.2	36.0	19.3
COII	17.8	26.2	37.4	18.7
COIII	22.7	21.0	34.8	21.4
CytB	20.7	24.2	36.8	18.4
16S rRNA	15.2	38.0	32.8	14.0

of genes in the chaetognath mitochondrial DNA shows no similarity to the organization observed in other animal groups.

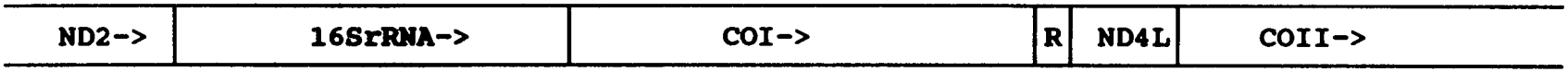
Comparison of this region with corresponding regions from other animal species is shown in Figure 18. In all vertebrates and in *Drosophila*, the order of genes for this region is similar: ND2, tRNA(s), COI, tRNA(s), and COII. In sea stars, this order of genes has been changed. The first change is observed at the 5' end of the COI gene, where the ND2 and tRNA genes are replaced by ND1 and a cluster of 13 tRNA genes. The second change is located in the region between COI and COII genes, where the ND4L gene is inserted. The order of genes in the corresponding sea urchin mitochondrial genome also shows two major differences from that of vertebrates: the genes for large ribosomal RNA and ND4L are in new positions. The large ribosomal RNA gene is found between the ND2 and COI genes, rather than between the small ribosomal RNA gene and the ND1 gene; the ND4L gene is added between the COI and COII genes. In two nematodes, *Caenorhabditis elegans* and *Ascaris suum*, the order of genes for corresponding region is ND4, COI, tRNAs, COII, and large ribosomal RNA (Okimoto et al. 1992). Further analysis of these orders of genes shows that in all vertebrates and some invertebrates including sea urchin, *Drosophila*, and nematodes, these corresponding genes are transcribed in same direction. In sea star, a gene inversion has been observed. In the chaetognath, the cytochrome B gene and large ribosomal RNA genes are adjacent. This is the first finding this order of genes. Another major difference of the order of genes in chaetognath is the finding that the ND2 and COIII genes are contiguous, which has never been observed in all previously reported mitochondrial genomes.

FIGURE 18.

A comparison of gene order in linear mitochondrial molecules containing cytochrome b, 16S rRNA, COI, and COII genes. Abbreviations of protein names are as used elsewhere in text. The transcriptional polarity is shown by arrows. The tRNAs are designated with the single-letter amino-acid code. The single-letter amino-acid codes were not shown in D, because the number and kind of those tRNAs were different from each other. The transcriptional polarity of tRNA from chaetognath is shown in Figure 17A. A. Phylum Chaetognatha; B. sea urchin; C. sea star; D. vertebrates or *Drosophila yakuba*; E. nematodes.

A. PHYLUM CHAETOGNATHA

B. SEA URCHIN



C. SEA STAR

D. VERTEBRATES OR <u>DROSOPHILA</u>

E. NEMATODES

Protein-Coding Genes

The sequence of one entire protein gene (COI) and partial sequence of four other protein genes (COII, COIII, Cyt B, and ND2) are given in Figure 17.

Similarities of the nucleotide sequences and predicted amino acid sequences of the five chaetognath mitochondrial protein genes and the corresponding genes of human, mouse, cow, frog, *Drosophila yakuba*, and sea urchin are summarized in Table 13. The degree of similarity of the different protein coding genes with their vertebrate and invertebrate homologues shows a broadly similar pattern of relative conservation. The order of similarity for the nucleotide sequences of these protein genes is COI=Cyt B>COII>=COIII>ND2. When inferred amino acid sequences are compared, COI is the most highly conserved: COI>Cyt B>COII=COIII>ND2. The highest similarity for nucleotide sequences is 66% of the COI gene from *Drosophila yakuba*; the lowest is 37% of the ND2 gene from the nematode. The highest similarity for amino acid sequences is 72% of COI gene from *Drosophila yakuba*; the lowest is only 30% of the ND2 gene from mouse.

Alignment of the chaetognath nucleotide sequences and protein sequences with those present in *Drosophila*, mouse, frog, sea urchin and nematode mtDNAs is shown in Figures 19 and 20. The COI gene is 1,524 nucleotides in length in *Sagitta*, compared with 1,540 nucleotides in *Drosophila*, 1,545 nucleotides in mouse, 1,554 nucleotides in sea urchin, and 1,555 nucleotides in frog. The COI gene of the chaetognath mitochondria is the shortest among these species. It is surprising that two nematodes, *Caenorhabditis elegans* and *Ascaris suum*, have the longest sequence of 525 amino acids for the COI gene. Lengths of other protein genes are shorter in nematodes than in all vertebrates and some invertebrates (Okimoto et al. 1992). This difference reflects mainly deletion of codons for amino acids

Table 13. Similarities between mitochondrial protein genes in different taxa

	Protein genes of *Sagitta*									
	COI		COII		COIII		CytB		ND2*	
Species	N%	A%	N%	A%	N%	A%	N%	A%	N%	A%
MOUSE	61.1	68.3	56.9	55.2	55.5	54.8	62.2	64.0	38.7	30.1
YAKUBA	65.6	72.0	64.4	58.6	56.3	56.1	67.8	69.4	40.8	34.2
FROG	60.9	67.7	57.5	53.4	55.8	54.4	59.0	63.6	38.9	32.4
URCHIN	60.7	64.2	55.5	51.7	54.7	52.6	57.6	57.4	40.9	37.0
C. elegans	60.3	61.2	45.8	32.8	47.6	43.3	52.6	45.9	37.4	40.6

Notes: N%: Nucleotide sequence similarity %.

A%: Amino acid sequence similarity %.

* Parts of the gene could not be reliably aligned.

FIGURE 19.

A comparison of the nucleotide sequences of chaetognath (Se) and other animal mitochondria: A. COI; B. ND2; C. COIII; D. COII; and E. Cyt B. The comparisons include *Drosophila* (Dy), mouse (Mm), frog (Xl), sea urchin (Sp), and nematode (Ce). All nucleotide sequences from the chaetognath were aligned so as to maximize the similarity of sequences conserved in the other genomes. Nucleotide identities to chaetognath are indicated by (.). Gaps are indicated by dashes.

A. COI

```
Se  ---------- ---------- ----ATGACC ------CGTT GGTTATTTTC CACTAATCAT  30
Dy  ---------- ---------- -----..A--- TCGCGA.AA. .......... T..A......
Mm  ---------- ---------- ----...TT. ATTAAT.... .A.....C.. A..C.....C
Xl  ---------- ---------- ----...G.A ATTACT.... .A.....C.. A..A.....C
Sp  ---------- ---------- ----...CAA CTAAGA..A. .A........ T.....C..C
Ce  ATTAATCTTT ATAAAAAATA TCAAGGAGGA TTGGCAGT.. .A...GAGAG AT........

Se  AAAGACATCG GTACTCTTTA TTTCGTATTG GGTATCTGAT CTGCATTCCT GGGCACCGCT  90
Dy  .....T..T. .A...T.A.. ....A.T..T ..AGCT...G .C.G.A.AG. A..A..AT..
Mm  .....T.... .A..C..C.. .C.AC....C ..AGC....G .G.G.A.AG. ...T..T..A
Xl  ........T. .C..C..... C..A..T..T ...GCT...G .A.GGC..G. C..A......
Sp  ..G....... .A..A..... ...AA.T..T ..GGC....G ...GCA.GG. A.....A...
Ce  .....T.... .A........ ...TA.T..T ..AC.T.... ...GTA.GG. T..T..TAGA

Se  CTCTCAGCAC TAATTCGACT AGAGTTAGGT AATGCAGGAT CGTTGTTGGG TGATGATCAG  150
Dy  T.AAG.ATTT ........GC ...A...... C..C.....G .A..AA.T.. A........A
Mm  ..AAGTATTT ........GC ...A...... C.AC....TG .AC.T..A.. A.....C..A
Xl  ..TAGCTT.T ........GC T..AC.TA.C C.GC.C...A .AC.AC.T.. A.....C..A
Sp  A.GAGT.TGA .T..C..TGC C.....G.CA C.AC.T..T. ..C..C.AAA A.....C...
Ce  T.T..TTT.T .......TT. ...A....C. ..AC....T. TT..TC.TA. GA...GA...

Se  CTATACAACG TGATTGTCAC TGCACATGCA TTTATTATGA TTTTTTTCTT TGTTATGCCG  210
Dy  A.T..T..T. .A.....T.. .........T ........A. .......TA. A..A..A..T
Mm  A.T.....T. .T..C..A.. ...C.....T ...G....A. ....C...A. A..A..A..A
Xl  A.T..T..T. .T..C..T.. A........T ........A. ....C...A. A..G.....T
Sp  A.......A. ..G.C..T.. C.....AT.G C.AG.C.... ....C...A. G..A.....A
Ce  T.G..T..TT CAG..A.T.. A..T.....T A..T.A..A. .......TA. G..A..A..T

Se  ACCATGATAG GAGGTTTCGG AAATTGGTTG GTTCCGCTAA TAGTTAACGC TCCCGACATA  270
Dy  .TT..A..T. .G..G..T.. ......A..A ..G..TT... ..T.AGGA.. ...T......
Mm  .TA..A..T. ....C..T.. ...C..AC.T ..C..A.... ..A.CGGA.. C..A..T...
Xl  .TT..A..C. .T..A..T.. G..C..A..A .....AT... ..A..GGA.. C..A..T...
Sp  .TA.....T. .T..A..T.. G.....AC.C A....A.... .GA.CGGT.. G..A..T..G
Ce  ..T..A..C. .T.....T.. T..C..A..A T.A..A..T. .GT.AGGA.. A..T..T...

Se  GCTTTTCCTC GTTTGAACAA TATAAGATTT TGGCTTTTGC CACCTGCTCT AATGTTACTA  330
Dy  ..A..C..A. .AA.A..T.. .......... ..AT.AC.A. .T........ TTCT...T..
Mm  ..A..C..A. .AA.A..T.. ......T... ..A..CC.A. ....AT.AT. TC.CC.T..C
Xl  ..A.....G. .AA.A..T.. ......C... ..A...C.T. .C..AT.AT. TC.T...T..
Sp  ..C..C..C. .CA....A.. ...G..T... ..A...A.T. .C...T..T. T..A.....T
Ce  AGA.....A. ....A..T.. .T........ ...T.A..A. .TA.AT..A. .T.A...A.T

Se  TTACTATCGG GTATGGTTGA AAGCGGTGTG GGCACTGGTT GGACTGTCTA CCCCCCTTTG  390
Dy  ...G..AGAA .A..A..... ..A...A.CT ..T..A.... .A.....T.. ...T.....A
Mm  C..GC...AT CA..A..A.. .GCA..A.CA ..A..A..A. .A..A..... ...A...C.A
Xl  C..GC...AT C.GG...... .GCA..A.CC .....A.... .A.....G.. ...G.....A
Sp  ...GCG..C. CAGGA..A.. ..AA..A.CA ..A.....C. .A...A.... ...T...C.C
Ce  ...GATG.TT ..T.T..A.. T.TA...TGT ..G...A.G. .A..A..... ...A.....A
```

```
Se  AGCACA---G TAGGCCATAC AGGAGGCGCT GTTGATTTGG GTATCTTTTC TCTACACTTG  447
Dy  TCTT..GGTA .C.CT...GG T....CTT.. ..A.....A. C...T..... ...T..T..A
Mm  GC.GG.AATC C..T....G. .....CAT.A ..A..CC.AA CA..T..C.. C..T..T..A
Xl  GCTGG.AACC ...CA...G. T....CAT.A .....CC.AA CA..T..C.. C..T.....A
Sp  TCT.GTAAAA ..ACA..CG. C..TA.GT.C ........A. CA.....C.. C..T...C..
Ce  ..A.....-A .G..G..CC. T...A.TAGA ..A.....A. C...T...AG .T....TGCA

Se  GCGGGGGTCA GGAGTATTCT TGGCAGTGCT AATTTTATTA CTACGATTGT GAATATAAAA  507
Dy  ..T..AA.TT CTTCA...T. A..AGC..TA .......... .G..TG.AA. T......CG.
Mm  ..T..A..GT CATC....T. A..TGCAAT. .......... .C..T...A. C..C..G...
Xl  ..T..TA.TT CATC....T. A..AGCAAT. ..C..C..C. .A..A.CAA. T..C......
Sp  ..C..T.C.T CTTCC..CT. G...CTCAT. ..A.....A. .A..A...A. T.....GCGG
Ce  ..A...T.A. .ATC...CT. A..TG..AT. ........GT G...T.C.AA A...T..CGT

Se  GGAGAGGGTA TAACAATAGA ACTTATGTCC CTGTTTGTAT GATCGGTTTT GTTAACGGCC  567
Dy  TC.ACT..A. .T...T.... C.GA..AC.T T.A....... ....A..AG. TA.T..T..T
Mm  CCCCCA.CC. .....CAGT. T.AA.CTC.A ..A.....C. ....C..AC. TA.T..A...
Xl  CC.CCA.C.. ..T.TCA.T. C.AA.CCC.A ..A.....T. ....A..A.. AA.C..A..T
Sp  AC.CC...G. .GT.TT.G.. T.G.C.TC.T T.A..C..C. ....C..C.. TG.C..T...
Ce  A..AGTTC.. .TT..T.... ..A...AA.T T.A.....T. ..A.T..A.. TG....A.TG

Se  ATTTTGCTAT TGTTGTCACT CCCGGTATTG GCTGGTGCAA TTACTATGCT TCTAACTGAC  627
Dy  T.A..A..T. .AC.A..TT. A..A..TC.T ..C..A..T. .......AT. AT....A...
Mm  G.AC....CC .A..A..... A..A..GC.A ..C.CA.GC. .......A.. A.....A...
Xl  G.AC.TT..C .TC.T..T.. T..T..C..A ..C.CA.G.. .C..A...T. AT....A..T
Sp  T.C.....CC .CC.T..T.. T..A.....A ..A..A.... ....A..... ...C..A..T
Ce  T....A..GG .T..A..T.. A.....T..A ..A..G..T. ........T. GT.......T

Se  CGCAATTTCA ATACATCCTT CTTTGATCCG GCTGGGGGCG GGGATCCAAT TTTATATCAA  687
Dy  ..A.....A. ....T..T.. T........A .....A..A. .A.....T.. ...G..C...
Mm  .....CC.A. .C...A.T.. .........C .....A..A. ....C..... .C.C..C..G
Xl  ..T...C.G. .....A.T.. ......C..T ..C..A..A. .T..C...G. AC.T..C...
Sp  ..T..AA.A. .C...A.T.. ......C..T ..A..A..G. .A........ .C...T....
Ce  ..T.....A. ....T..A.. T........A AGAACT..A. .TA....TC. .A.T......

Se  CATCTTTTTT GGTTTTTTGG ACATCCTGAA GTTTATATTC TCATTCTTCC GGGTTTCGGT  747
Dy  ...T.A.... .A........ T..C...... .........T .A...T.A.. ...A..T..A
Mm  .....G..C. .A..C..... G..C..A... .......... .T..C..C.. A..A..T..A
Xl  ..C..G..C. .A..C..... G..C..A... ..G..C.... .T..CT.A.. A..G..T..C
Sp  ..C..A..C. ..C....... ...C..C..G ..G....... .T..CT.A.. ...A..T...
Ce  ...T.G.... .A........ T......... ..A......T .G...T.A.. A.C...T...

Se  ATAGTGTCGC AGATTATTAA TCACTATTCC GCTAAAGGGA ATAGGTTTGG AGCTTTAGGC  807
Dy  ...A.T..T. .T.......G A..AG.A..T .G....AA.G .A.CT..C.. TT.......A
Mm  ..TA.T..A. .TG.AG...C .T....C... .GA...AAAG .ACCT..C.. CTA.A....A
Xl  ..GA.C..C. .T..CG.A.C .T.T..C..A .GA...AAAG .ACCT..C.. CTA.A....A
Sp  ..GA.C..A. .CG....AGC ......C..T .G...GC.AG .GCCT..C.. .TACC.G..G
Ce  ..T..CAGA. .ATC..CACT .T.T.TAA.A .GA...AAAG .AGTT..... T.....G..T
```

```
Se  ATAATTTATG CCATGAGGTC CATTGCGTTG TTAGGTTTCG TTGTGTGGGC CCACCATATG  867
Dy  .....C.... .T..ACTTG. T....GA..A .....A..TA ....T..A.. T..T.....A
Mm  ...G.A.GA. .A..A.T... T....GC..T C....C..TA ....A..A.. ......C..A
Xl  ...G.C.GG. .A..A.TA.. A....GAC.T C....C..TA ....C..A.. ...T..C..A
Sp  T.GG...... ......TTG. A..A.GAG.T .....A...C ....C..... ..........
Ce  ...G...... .A..TTTAAG A....GT..A A.T....GT. .A..A..A.. T.........

Se  TTCACGGTGG GTATAGACGT TGATACCCGG GCTTATTTTA CATCTGCTAC TATAATTATT  927
Dy  ..T..A..T. .A........ ......A..A .........T .T........ ..........
Mm  .....A..A. .AT....T.. A..C..A..A ....GC.... ....A..C.. .........C
Xl  ..T.....T. A.C..A.... A.....T..A .....C..C. ....A..A.. A.....C..C
Sp  ..T..A..A. .G..G..T.. ......A..A ..A..C..C. .TG.C..C.. A..G......
Ce  .AT..A..A. .......TT. G...T.A..T .........T .GG....... ....G.....

Se  GCAGTACCTA CTGGCATTAA GATTTTCTCT TGATTAGCCA CGCTACATGG TACTCCCTCT  987
Dy  ..G..T.... .A..A..... A.....TAGA ........T. .TT....... A....AACT.
Mm  ...A.T.... .C..TG.C.. AG.A..TAGC ...C.T..A. .C.....C.. AGG.AATAT.
Xl  ...A.T.... .A..TG.... AG.A..TAGC ........T. .AA....C.. .GGGA.AAT.
Sp  ..T..C..A. .A..AT.A.. .G.....AGA ...A.G..A. A...C..A.. GT..AATCTA
Ce  .....G..A. .A..TG.... AG.G..TAGA .....G..T. .AT..TT... ..TAAAAATG

Se  TTGCTCGAGA CTCCTTTGAT GTGAGTTTTG GGCTTTCTGT TCCTTTTTAC AGTTGGCGGT 1047
Dy  .CTTATTCTC .AG..A.TT. A....C...A ..A...G.T. .TT.A..C.. ...A..A..A
Mm  AAATGATCTC .AG..A.AC. A....CC..A ......A.T. ..T.A..... ......T...
Xl  AAATGA..CG .C..AA.AC. T....CC..A .....CA.T. ..T.G..... T..A..A...
Sp  CAATGAAGAC TC.....AT. A...ACC... ..GA..G.A. .TT.A..C.. .T.A..A..A
Ce  G.AT.TA.TC .A.T...AT. ......A... ..T...A.T. .TT.G..... TT.A..T..G

Se  CTCACTGGCG TAGTCTTAGC CAATAGTTCT TTAGATATTA GACTGCATGA CACTTATTAT 1107
Dy  T.A..A..A. .T..A..... T...TCA..A G.T....... TTT.A..... T.........
Mm  ..A..C..AA .T..T...T. ...CTCA..C C.T..C..CG TG..T..C.. T..A..C...
Xl  T.A..A..TA .T..TC.T.. ...CTCA..A C.T....... T...A..C.. T..C..C...
Sp  .....A..TA .T..TC.T.. ....TCC..C A.T..CT..G TT..T..... T..C..C..C
Ce  T.G..A..T. .T..A...T. T...TCAAGA ..G....... TTT.A..... T.........

Se  GTAGTAGCCC ATTTCCACTA CGTATTGTCA ATAGGTGCAG TATTCGCCAT TTTCGCGGGG 1167
Dy  ........T. .......... ...T..A... .....A..T. ....T..T.. .A.A..A..T
Mm  .......... .......... T..TC.A... ..G..A.... .G..T..T.. CA.A..A..A
Xl  .......... .......T.. T...C.T..T .....A..T. ....T..G.. CA.G.GA...
Sp  ..G.....T. .C..T..... ...TC.T... ..G..G..T. ....T..A.. C.....T..T
Ce  .....TAGA. ....T..T.. T..T..AAGT T....A..T. .T..T.GG.. ....A....T

Se  GTCACTTTTT GGTATCCTGT GATTAGTGGG ATGACTATGT CCGCTCGTGG TACTCAAGTA 1227
Dy  T.T.T.CAC. .A..C..AT. AT...C...A T....AT.AA ATAA.AAAT. GTTAA..AGT
Mm  T.TGT.CAC. .A.TC..AT. AT..TCA..C T.C..CC.AG AT.ACACAT. AG.AA...CC
Xl  T...T.CAC. .A.TC..GT. AT...C...T TAT..AC.AC AT.AAACAT. AG.AA..A.C
Sp  T.....CAC. ...T...CC. CT.CTC...T TAT.GCC.AC A.C.ATTAT. AGGAA.G..T
Ce  ..T..AC.A. .A.GAAGAT. T....CA... TATGTGT.AG ATAAA.T.AT G.TATCT.C.
```

```
Se  CAATTCGCGA TTATATTTAT TGGTGTAAAT CTGACTTTTT TTCCACAACA CTTTTTAGGT 1287
Dy  .....TATT. ....G..... ...A...... T.A..A.... .C..C..... T........A
Mm  ..C.....C. .C.....CG. A..A.....C A.A..A..C. .C..T..... T..CC.G..C
Xl  ..T..T.GAG .A......GC ......T... T.A..C..C. .C..T..... T...C....C
Sp  ..C...TTC. .A..G...G. ...A..C..C T.A..C.... .C..T..... ...C......
Ce  GT...TATTT .AT....... ...G...... T.A..A.... .C..G.T... T...GC...A

Se  TTACAAGGAA TACCACGACG GTATTCGGAC TACCCTGACT CGTTCTCTAC TTGGAACGTA 1347
Dy  ...GC..... ....T..... T.....A..T ........TG .T.A.A.... A..A..T..T
Mm  C.TTC..... .......... C..C..A... .....A..TG .T.A.A.C.. A..A...ACT
Xl  ...AGC.C.. ....T..... A..C..T... .....A...G .T.ATA.ATT A..A..TACC
Sp  C..GCC.... .G........ ...C..A... ..T..A...G .C.ATA.ACT ...A..TACT
Ce  C....C..GT .C.....TAA A....TA..T ........TG TT.AT..GGT A..A..TA.T

Se  GTGAGTTCAT CAGGTAGACT AGTTAGGATT GTTGGAGTTT TTATGTTTGT AGGTGTAATA 1407
Dy  ...TC.A.TA TT..GTC.AC TA..TCAT.A T.A...A... .AT.T...T. CTA.A.T..T
Mm  ..CTC...TA T...ATC.T. TA..TCAC.A ACA.CT...C .C..CA.GA. CTT.A....T
Xl  ..CTCA..TA TC..GTCCT. .A..TCTC.. ....CC..AA .....A.AA. .TTCA.T..C
Sp  A.CTCC...A TT..ATC.AC CA.CTCCG.A ..G.CTA.GC .AT.T..CC. CTTCT....C
Ce  A.TGCC..T. AT...TCTA. TA....A.C. .CA...C.A. .CT.A...A. TTA....T..

Se  TGAAAATCTT TATCGGCT-- ----ATTAAT GCCAGGGGAG ACGTAGACAA CTTGTCTGTA 1461
Dy  ...G..AG.. ..GT.T..CA ACGACAAGTA ATTTATCC.A TTCA.TTA.. T.CA...A.T
Mm  ...G.GG.C. .TG.TT.AAA ACGAGAAGTA ATATCA.T.T CGTAT.CTTC AACAAA.T..
Xl  ...G..G.A. .TG.A...AA ACGAGAAGT. A...CTTAC. .AT..ACATC AACCATAT.G
Sp  ..GG.GG.C. .CG.TT..CA ACGGGAAGGA AT..CCCC.. .GT.CTCAC. .GCC..AC..
Ce  .T.G.G.... .C.TTAG.TA TCGTT.AGTA ATT..A.ATT .TTATTCT.. TAGAAGACCT

Se  GAGTTTAGAC CTCGTCTACC TGTCTCTTGA CACTCTTTTA ACGAGAGCAC AGTATTAGGT 1521
Dy  ..A.GATAT. AAAA.AC... CCCAG..GA. ..TAGA.A.T CT..ATTAC. .C.T...ACA
Mm  ..A.GACTT. A.G.CTGC.. .CCAC.A.AT ...A.A..CG .G..ACCA.. CTATG..AAA
Xl  ....GACTT. AAG.CTGC.. CACTC...AC ..TA.C..G. .GACC...CT C..TCA.ATC
Sp  ....GACA.T ACACCTCCTT .CC.C...CT ...CACACCT T...TGAA.. .CCC.CTACC
Ce  ....A.T.TA TGA..AATTA ...A.T.G.T ...AG..A.C .GTCTGAG.T TTAT..TA.A

Se  TAG------- -------- 1524
Dy  A.T------- --------
Mm  GTAAAA---- --------
Xl  A.CCATCAAA TAATTAAA
Sp  ATAATTATTG TAAAG---
Ce  ACTACTAGAT TAAAAAAT
```

B. ND2

```
Se  TTGTATCTCA ATCTTGCGGT TCTCTAATGA TTTTGTTGGG TGGTATAGTA ----------  50
Dy  ..T..A.C.. .G...TG.CA ..AACTG.TT .A..A..TTC .TCA..TT.. CTT------A
Mm  .C..CA.A.. .G.AACA.CC ..AA....T. .CC.CC...C CATCG..C.C AACTATAAAC
Xl  ..T..A.A.. .G.AGCA.CA ...GC.C.TC .CC.C..CTC .A..T..AAT AATGCCTGAC
Sp  ..C.TGT... .G...CAA.G G..GC.C.TT .GC.AAAA.. C.CCC.T.G. CAAGCTTGGT
Ce  ....TAT... .GAG.CTTTA GG.T..T.AT ..C.TC.TT. .A.A------ ----------
```

```
Se  ---------- ---------- ---GCCGATT CTAGTTTTTT TTCTGTGTCT TTACTTTTGC  87
Dy  TATTGGCAAA TAATTTAAAT AAT.AAAT.A A.GAA.C... .A.ATCAATA A.TA..A.AT
Mm  AACTAGGAAC ATGAATATTT CAACAAC.AA .A.ACGG.C. .ATCC.TAAC A..ACA..AA
Xl  TCACCGGAGA ATGGTCAATT TTA.ATTTAA .A.ACCC.C. G..ATGCG.A ACTA.AACCA
Sp  TGACGGGATC ATGATCAATC CTA.ATCC.G T..AAGAGG. AA.CTCCAT. .GC..AAGAA
Ce  ---------- ---------- ---------- ---------- -GGA.GTCTA ....AA..TT

Se  ATGGCGTGGT TTTTAAGATG GGCCTTATGC CTCTTCATTT TTGGGTGCCT TGTGTAGTTA 147
Dy  CG.C.T.AT. A..A..A.GA ..AGCCGCT. ..T..... ...AT.T... AA.A..A.AG
Mm  TA.C.C.ATC CA.A..AC.A .....CGCC. .AT.C..C.. C..AT.A..A GAA...AC.C
Xl  T..CAA.CTG .A.A..AC.A ..A...GCA. ..T.C..C.. C..AT.A... GAA..CC..C
Sp  TA.C.C.T.C A........A ......GCC. ..G.C..C.. C..AT.T..A GA...TT.AC
Ce  T.ATTA.TT. A..G..A..T ..TG.AGCA. .G..C..C.. ...AA.TTT. AA....ACA.

Se  TAAATTTAAC TCGTTTCAAC TTATATCTA- --CTCATATC TTGGCAAAAG ATTGGTCCTA 204
Dy  A.GGA..... AT.AA.A..T GCT.TGA.-- --T.A...A. ...A.....A ....C...AT
Mm  A.GGGA.CC. A.TGCA..TA GG.CT.A.T- ---.TC.TA. A..A.....A ....C...CC
Xl  A.GGAC.T.G .TTAACA.CA GG..TAA.C- --..ATCCA. A..A.....A C.C.CC..A.
Sp  A.GGG..GC. CTT....C.A GGGCTAA..- --A.AGCCA. .......... ..A.CC...C
Ce  AT..CA.TTT .AA..AT--- GGGCTAA.GT GGT.TT..A. A.TT.....A T.ACCATT.T

Se  TTGTCATTGT C--------- ------ATGA CCGCTTCGGT GGGGTATACT GTGTTGTGTC 249
Dy  .AA.AT.AA. TTCTTAT--- ------T.A. ATAT.AAAAA TTTA.TATTA A.TAGTGTAA
Mm  .ATCA...T. AATTCAAATT TACCCGC.AC T.AAC..TAC TATCAT.TTA A.AC.AGCAA
Xl  .A.CT..CT. ATATCAAATC GCTCCA..AT TAAA.A.ACC ACTTCT.CTC ACTC.AG...
Sp  .AA.AC.AA. GTTTTATTTT AGCCAGT.AG GTTT...TTA CCTACT..TA ACACCTA.GT
Ce  .AACT...T. ATTACAAATT TTT---TG.T TAAGA..C.. .TATAT.TTG T.A..TG..T

Se  TAGTCAATGC TGTTGGCGGG TCGTTAGCTA TGAGCGGTGT TACCGTGTTG CCGTTGCTTC 309
Dy  .TT.ATCA.T .A..ATT..A G.AA.T.GAG GTTTAAACCA A..TTCAC.C .GAAAAT.AA
Mm  .TACTTC.AT .T.CATA... G.A.G..GAG GACTTAACCA A..ACAAA.A .GAAAAA..A
Xl  .CACATCAA. AC..AT...C GGA.G..GGG GACT.AA.CA A..TCAAC.A .GAAAAA.CT
Sp  ..A.TTC..T GC.GATA... GGC.GG.GAG G.CTAAA.CA G..GCAAG.. .GTAA.A..T
Ce  ..T.G.T.TG .TA..TTCAA ATT------- --TTT.TCA. A.AAAGT.AT AAAAATT.GT

Se  TTATTTTTAG AGGAATAGTT CAGATAGGTT GAGTGTTCAT AACC------ ---ACCGGGG 360
Dy  .AGCA...TC TTCT..TAA. ..TT....A. ..A.A..A.G .T.TTTAATG ATT.GA.AAT
Mm  .AGCC.A.TC .TC...T.CC ..C.....A. ..A.A..AGC ..TTCTTCCT TAC.A.CCAT
Xl  .AGC...CTC .TCT..T.CC ..TC.T.... ..A.AA.TTC T.TTCTCCCA TTCT.ACCCC
Sp  .AGCA..CTC TTC.....GA A.A..G..C. ..T.AG.... ...ATCAGCT TACT..TTTA
Ce  .A...A..TC .TCC.C..AG TCTT.TAA.. .GA.TG.TT. GGGAGTATTT TTTT.AAT.T

Se  TTTTCACCTT CTATTATCTT TTTGTTTACT ACATTGTCTT AAGTGCTGTT GTCCTATAC- 419
Dy  CAA.TTGA.. AAT....T.. A..T....T. CAT.CT.A.C TTT..TAT.A ACAT.TATAT
Mm  CCC....TC. ACTCA.C..C A.AA.C..TA TT...C.TAC .GCCC..A.A T..A..GCAC
Xl  AG..A.TAA. T.TAA.CT.A ACAA..... TA...A.GAC CTCCA.GA.A T....TGTAC
Sp  ACGCTG.GA. .AT.ATGT.A G..A..... TA...A.TAA C.C.T..T.G T.TT...TGT
Ce  ..AAT..A.. T...---T.A ...A....T. ..T....A.. ..TA.T.T.A T.AA.T.CTA
```

```
Se  ---------- -------AGG GCAGTCGCTT CGGTTCAATT CGGTTGG--- ----------  449
Dy  TTAATATTTT TAAATTATTT CATT.AAA.C AAT.ATTT.C TT.A.TTGTA AACAGAAAAA
Mm  TTATACTAAA TAACTCT.TA A.CA..AAC. .AA.CTC.C. TCTA..AAAT AAAACTCCAG
X1  TAAAAACTAT CTCATCC.CA AA.A.TT... .TT.AGCTAC .TCG..ATCT AAAACCCCAT
Sp  TTGACCACTT AAAGGTCTCC A..T.G.GAC ACT.AA..AC TAT..CTCAG CTTTCACCAA
Ce  AG-------- ---------- ---------- ---------- ----.TTTCT AAAACTAGGG

Se  ---------- ---------- ---GCACTGC TCAACGCAGG GGGGTTACCT CCTTTCTCTG  486
Dy  TTTTAAAATT TTCATTATTT ATAAATT.TT .ATCTTT... T..A..... ..A..T.TA.
Mm  CAATACTAAC TATAATCTCA CTGAT.T.A. .ATC.CT... A..CC.T..A ..AC.AA.A.
X1  CCACTACGGC ACTCTCACTT TTAA.T..T. .TTCTTT... T..CC.T..A ...C.T..A.
Sp  TTAGAGTTGC TCTTGTTCTC CTA.TGA... ..TCTCT... A..CC.T..C ..A..AA.C.
Ce  GTTATAATTT TATTAATTGA GAAA..ACAT .AGTATTTTT AAATA.T..A TT.AGAGT.T

Se  GTTTTATAAT CAAACTAAAA GCGATTTTAC ATATTAAGGG GAGT------ ------ATAG  534
Dy  .A...T..CC A...TG.TT. .TA...CA.. .AT.A.CAAT AT..AATCAA TATTTTT..T
Mm  .A..CT..CC A...TG..TT ATC..CAC.G .AC...TAAA A.ACAACTGT CTAATT....
X1  .G...G..CC A...TG.TTT ATT...CA.G .AT.G.CAA. CCAAAACACA ACTATTC...
Sp  .G.....CC. A..GT.T.CC T.CC...ATT TCT.GGTT.C C.AAAATTTT ATCATTT..T
Ce  CA...T.TG. A...A.TTTC T.AT.GAGGG .A...TTTAA ATA.GATAGA TTCTTT.CTC

Se  TGGTACTGTT GGTTGGTGCA AGAGGATTGG CTCTAACTTC TTATATCCGC TTACTGTTGA  594
Dy  .AAC.T.AA. AA.AATAT.. .CTTT.A.TA .AT..TT..T ....T.A..A A.TTGT.ACT
Mm  CAAC...CA. A.CAATAAT. GCTCT.C.AA AC...TTC.T ......T... C..A.T.ATT
X1  CCAC.ACAC. A.C.CTGT.. GC.CT.C.CA GC...TT..T C..CC.T... C..ACT.AT.
Sp  CTTCTA.TA. .A.AAT..G. .ATCTTCAA. A.TATTT..T ....C....A A.TTC...T.
Ce  .AT.TT..C. TT..ACAAT. TTTTT..CT. TAT.GG.A.T .AGAT.TT.A ...A.TAACT

Se  ATACACGTTT GAAGTCGGGC CCCAGCTCTG GTTTTCTTGT AGCCACTATG GTTGCTGGTA  654
Dy  CAG.TTT.A. ATTAAATTAT TT.GAAAA.A AC.GAA.CA. G.AA.TA.AT A.AAA.A...
Mm  CC..TTCAC. A.CAATATTT ..A.C.AACA A.AACTCAAA .ATA.TA.CT CACCAAACA.
X1  T.GTCACAC. A.CA..ATCA ..A.ATA.AT CAAACGCATC .TTA..ATGA CGACACCACT
Sp  .A..TA.C.. ATTTCT.TTT ...CAACACA T.A..AGATC C...T.ATG. CGAAA.A.G.
Ce  TG.G.ATGAA A..TAAT.AA GAA.CT..AA A.AA.AA.AA .ATA.A.TAT T..AT.AT.T

Se  GAGTT----- ---------- ---------- ---------- ---------- ----------  659
Dy  ATAA.ACTAA TTTATATTTA ATTATAACTT TTTTTTCAAT TTTCGGATTA TTTTTAATTT
Mm  A.AC.AAACC CAACCTAATA TTTTCCACCC TAGCTATCAT AAGCACAATA ACCCTACCCC
X1  CTAAACAACC ATCACTTTTA TTATCAATCG CATTAATCCT ATCCTCATTT ATTATTCCAA
Sp  C.A.AATTTC ACCTCTCGCC CCAAAGGCAT GATTAAGTTC GGTCTCCACT GTGTTGAGTA
Ce  TTCCGTTAAT AGTTATTTCT ATTATT

Se  ---------- ---------- ---------- ---------- -
Dy  CTTTATTTTT TTTTATACTT
Mm  TAGCCCCCCA ACTAATTACC
X1  TTTCACCATT AACTTTAACA
Sp  CTCTTGCAAT ACCCCTTACC CTCCCCTTAT ATATAATTAC A
```

C. COIII

```
Se  ATGAACAAA- --------CA CCCGTTCCAT ATCGTAGACG TAAGACCTTG ACCACTGATA  51
Dy  ...TCT.C.C ACTCAAAT.. ...T..T... T.A..T..TT AT..C..A.. ...TT.A.C.
Mm  ...---.CCC ACCAAACT.. TG.A.AT..C ..A..TA.TC C...T..A.. ....T.A.CT
Xl  ...---GC.C ACCAAGCA.. .G.C.A...C ..A..C...C CC..C..... ......A.CG
Sp  ...---GCTA TTCAA..... ...A.AT... T.A......C A...C..A.. ...CT.AGAC
Ce  ..A------- -----TTT.. TAAT..T... ..TT..AGAC .TTC.AGG.A TG..TAT.AT

Se  GGTTCTGTGG GGAGTTTGTG TCTGGTCGGA GGGCTGGTGA CGACCATGCA CCGTTACGGT  111
Dy  ...G..A.T. .AGC.A.AAC AACT..ATC. ..TA.A..A. AATGAT.T.. T.AA..T.A.
Mm  ..AG.CT.TT CAGCCC.CCT ...AACATC. ..T..A..A. TATGAT.T.. .TA.A.TTCA
Xl  ..AG....A. CAGC.C.ACT C..TACATC. ..CT.A.CT. TATGAT.T.. .TT.GGATCA
Sp  ..AG.AT.TA .AG.C..AAT GA..ACTTC. ..CAAT..CC TATGGT.C.. TACCC.AAAG
Ce  TTA.T.T.T. CCTCAGCCG. AA..T.AA.. TCTT.A..A. TATTTT.TA. ATT.GGACTA

Se  AGCTCTCTTT TCTGGTTGGG TATTGCTCTC ATTCTTGCCA CTATGTTCCA GTGGTGGCGT  171
Dy  .TT..AT.A. .T.TA..A.. ..A.AT.A.T .C.A..TTA. .AG.T.AT.. A..A..A..A
Mm  .TTA.A..A. .AACCC.T.. CC.ACTCAC. .A.A.CCT.. .A..A.AT.. A..A..A..A
Xl  .TAAT..... .AACCC.A.. CC.AAT.ACT ..AG.ACTA. ....AA.T.. A..A..A..A
Sp  .CTAA.T.AA CT.TAG.A.. CT..TTAT.A T.AA.AA.A. AA...G.TA. C.....A..C
Ce  TA.GAAT.A. .TATT..TAC AC.ATT.TCT G.GT.ATTT. T.TCT..TGC T..AG.TAAG

Se  GATGTTACAC GTGAAGCTAC GTTTCAAGGC AAACACACTG CGAAGGTAGA AAGAGGCATG  231
Dy  ......T... .A....GA.. T.AC.....A TT...T...T ACGCA...AC T.TT..TT.A
Mm  ..C..A.TT. ......GA.. C.AC..... C.C......C .T.TT...C. ..A...AC.A
Xl  ..C..A.TT. .A....GA.. A..C.....A C.T......C .ACCC..TC. ..A...AT.A
Sp  ...A.A.TT. .AA.G..C.A C.....G... .G........ .T.TT...A. ..AG..A...
Ce  ...A..G.TA TA....G.-- -..AAG...T T.C..T.A.T TTTTC..TAT .GAC..GT.T

Se  CGATATGGTA TACTACTATT CATCAGTTCC GAGGTTTTTT TCTTTTTAGC TTTTTTTTGG  291
Dy  ....GA..A. ..A.TT.... T..TTTA..A ..A.....A. .T...G.TAG A........A
Mm  .......... ..A.T..... ....GTC..G ..A..A.... .....GC..G A..C..C..A
Xl  ........A. ..A.C..... T..T.CA..A ..A..A..C. .....A.T.G A..C..C..A
Sp  ........C. .GA.C..... T..A.CC..A .......GC. .T.....C.. C.........
Ce  AA..TC..AG ..A.TT.... TG.GTT.AGG ...T..A.A. ....C..TTG .A.......A

Se  GCATTCTTCC ACTCAGCATT GAGGCCTAAC GTGGAGGTCG GGTCTGTTTG GCCTCCGTTA  351
Dy  .....T..T. .TAG.AGT.. ATCT..AGCA A.T..AT.A. .AG..TCA.. A.....TA.G
Mm  ..G....AT. .T..TAGCC. CGTA..A.CA CAT..TC.A. .AGGCTGC.. A.....AAC.
Xl  .....T.A.A .....AGC.. AGCC....CA TAT..AT.A. ..GAATGC.. ...A..AAC.
Sp  ..C..... .TAG.AG... AGCC..CTC. ..T..AA.A. ..GTA.CA.. A..C...AG.
Ce  A....T..TG .TG.T..TC. AGTA..AGTA CAC...T.G. .AGAGAC... AT.A..T..T

Se  GGTATTTCAG CTATTAACCC TTTTGATGTC CCTTTACTCA ATACATCTAT CCTATTGTCT  411
Dy  ..A...ATTT .AT....T.. A...C.AA.T ......T.A. .....G.... T..T..AG..
Mm  ..A......C .AC..... .C.A..A... ..AC....T. ....T..AG. A..TC.AG.A
Xl  ..A...A.CC .AT.A..... A.....A..T ..AC.TT.A. .C...G.AG. A..TC.AG.A
Sp  ..A..AA.CC .CC..... ...CCTA..T ...C..T.A. .A...GGCG. T..TC.A...
Ce  ..A..GCACT TAG....T.. A....G...A ..G...T.A. .....AT... TT....AAGG
```

```
Se  TCCGGTGCCA CTATTACATG AGCACACAGA GCCTTACTAG AAAACCGGTG ATTGGAGTCT  471
Dy  ..A..A.TT. .AG.A..T.. ...T..TCAT AGA...A... ...GAAATCA T.CAC.AA..
Mm  ..A....TTT .A..... ...T..TCAT AG.C.TA... ..GGTAAAC. .AACC.CATA
Xl  ..A..A.TT. ..G.C..... ...T...CAT AG.A.CA.GC .TGG.GATC. .AAA..AG.A
Sp  ..A..A.TT. ..T.A.G... GT.C...CAC AGAA.T.... C.GGGAATC. .ACT..A...
Ce  AGT....TT. .AG.A..T.. ......TCAC AGA.....TA G...TAAA-- ----AGA.G.

Se  CAACTAAGCT TGATTATCAC TGTCGTTCTC GGTTTTTATT TTTCTATATT GCAAGGCTTA  531
Dy  ACT.A.G.A. .AT..T.T.. A..TT.A..T ..GA..... .CA.A..T.. A....CT.AT
Mm  A.T.A.GC.C .AC.A..T.. CA.TA.A..A ..AC....C. .CA.C..CC. C....CT.C.
Xl  ATT.A.TCAC .A.C.T.A.. CA.TC....T ..AC..... ..A.AGCCC. T....C.A..
Sp  ATT.A.GCAC .AT..C.G.. A..G.C.... ...AGG.... ..A.CGCGC. T..G.CG.G.
Ce  ACTAAT..TA .A...T.A.. ATGTT.AT.G .CAGC..... ..A.AGG.A. T...TTAA..

Se  GAGTATGTGT GAGCTGGTTT TTCGCTTTCA GACGGTATTT ACGGTAGTAC TTTTTACGTT  591
Dy  ..A...A.TG A....CCA.. .A.TA..G.T ..TTCAG... .T...TCA.. ......TA.G
Mm  ..A..CT.TG A.A.ATCA.. C..CA..... ..T.....C. .T...TC... A..C.T.A.G
Xl  ..A...TACG A...CCCA.. .A.AA..G.. ..T..AG.G. ....ATCA.. A....TT..A
Sp  ..A...A.TG AC..CCCA.. .A.CA..G.C ..TA..G... .T..CTCC.. C..C.TT...
Ce  ......A.AG A....A.A.. ...TA..G.. .....AG.A. TT..A..G.T ......TT.A

Se  GCTACTGGTT TTCACGGTTT ACATGTTCTC ATTGGCACCC TCTTTATTGC GGTGATGGCT  651
Dy  ..C.....A. .C..T..AG. T........A .....A..AA CT..CT.ATT A..ATGTTTA
Mm  ........A. .C..T..AC. C.....AA.T .....AT.AA CA..CC..AT T..TTGCCTA
Xl  ..A..... .C......C. T.....CA.T ......T.AT .A..CC.AT. T..TTGTCT.
Sp  .....A..A. ....T...C. C..G..AA.T ..A..A..AA CT..CC.CAT ...ATGCCTA
Ce  T....G..A. .......AA. T.....CT.G TG...TGGTT .G..CT.A.. AT.T.ATTT.

Se  TACCGAAACT TATATCACCA TTTCAGATGC AGGCA                             686
Dy  .TA..TC.T. ..A..A.T.. ...TTC.AAA .AT..
Mm  CTA...C.AC ..A.ATTT.. C....C..CA .AA..
Xl  CTT...C.AA .TC.AT.T.. C....C..CA .AA..
Sp  .TT..G.CTG C.GGC.G... C...TC.AC. CAT..
Ce  .TA..TCTTC ..A.AA.T.. ...T.AT.AT .AT..
```

D. COII

```
Se  ATGTCT---- ---------- -------TTA AACTTCCAGA ACAGGAATTC GCCCTTAATG   39
Dy  ......ACAT GAGCTAAT-- --------... GGT..A..AG .T..AGC... T..T......
Mm  ...G.CTACC CATTCCAA-- -------C.T GGTC.A..AG ..GCC.CA.. C..TA.T..A
Xl  ...G.ACACC CATCACAA-- ---------. GGT..T..AG ..GCAGCC.. T..AA.T..A
Sp  ...GGAACTT GAGCACAG-- --------.T GGTC.A..AG .TGCATCC.. C..TC.T...
Ce  ..TAA.AATT TTTTTCAAGG ATATAAT... CTA..T...C .T..TTTA.T TG.TAGTTAT

Se  GAACAGTTAA TTTTTTTTCA TGATTGGGTT ---------- --ATAGTTTT TGTTAGTAGA   87
Dy  .....A.... .......... ....CAT.CA TTATTA---- ---.TT.AG. AA.A.T..C.
Mm  ...G..C... .AAA...C.. ....CACACA CTAATA---- ---.T..... CC.A.T...C
Xl  ...G.A...C ..CAC..C.. C..CCATACC CTCATA---- --GCC..... .C...T...T
Sp  ..GG..C.C. CA.AC..C.. C....AT.CA TTAATT---- --G..C..AC CC.C.T..C.
Ce  AT.G.T.G.T ..CA.AGGTT .A....TAG. TTATTGTTGG GAG.TT.AG. GT..GT..CT
```

```
Se  ATCACTGTGG GATACCTGAT T---CTCATT GTTTCAAACA AACCTACTCA TCGTGTTTTG  144
Dy  G.ATTA..A. ....TT.A.. GTTTA.AT.A T...TT..T. .TTA.GTAA. ...AT..C.T
Mm  TC.TTA..CC TC..TA.C.. CTCG..A..A T.AA...CA. ...TA..A.. .ACAAGCACA
Xl  .CGCTA..TC TT...A.T.. .ACTA.T..A A.AA.T.CT. ...TA...A. .ACAAACC.A
Sp  ..ACTA..TT TT..TGG.T. AGTTTC.T.G C..GT.TC.T CTAA....A. C..AT....C
Ce  T.ATTGT.T. ....TT.A.. .TTTGGT.C. T...ATTTT. ..AGA.AAA. AAT..AG.AT

Se  TTAGAGTCTC AAGGTGTGGA GTTCGCCTGG ACAGCCCTCC CTTGTCTCGT TCTGGTCGCT  204
Dy  ...C.TGGA. ..CT.A.T.. AA.AATT..A ..TATT.... .AGC.A.TA. .T.AT.ATT.
Mm  A....TG.A. ...AA..T.. AAC.ATT..A ..TATT..A. .AGC.G.AA. C..TA..ATA
Xl  A.G..CG.A. ...AGA.C.. AA.A.TG..A ..TATTA.A. .AGC.A.TAG C..CA..ATA
Sp  ..T...GGA. ...AGT.A.. AACAATT..A ....TGA.T. ..GC...AA. CT.AA..TTA
Ce  CAGTTTGG.G ..TTAT..TG TAGTATT.TT C..A.TA.TA T..TAT.AA. A.AAA.A.T.

Se  ATTGCTCTTC CATCTCTACG ATTATTGTAT TCAATGGATG AAATTATTGA CCCTTCTTTA  264
Dy  .......... .T..AT.... ....C.T... .T.T.A.... ......A... A..A..AG..
Mm  ........C. .C........ CA.TC.A... AT...A..C. ....C.ACA. ...CGTA...
Xl  .....C.... ....C..T.. TA.CC.A... .T...A.... ..G...A... T..ACAC...
Sp  .....C.... .T..C..C.A .C.CC.T..C CT......C. .GG...AA.. ...C.TC..G
Ce  CC.T.A...A GGCT.T..TA T.AT.ACGGA .T...AA..- -----T.A.. TAGAAA....

Se  ACTATTAAAG CAATAGGTCA TCAATGATAC TGATCTTATG AATATTCTGA TGTAGATGAA  324
Dy  ...T.A...A GT..T..... .......... ...AG..... .......A.. .T.TA..A.T
Mm  ..CG.....A .C.....G.. C.......... ...AGC..C. .....A.... CTAT..A..C
Xl  ..A.....   ....C..C.. C.......... ...AGC..C. .....A..A. CTAT..G..T
Sp  ........G. .GT.C..... ...G..... ...AGA..C. .G..CA.G.. CT.CA.A..C
Ce  ..AG....G. TT.C...A.. ...G.....T ...AGA.... .....AGA.. .A.TCC..GG

Se  GAATCCATAG AATTCGACTC TTAT                                         348
Dy  -------..T. ....T..T.. A...
Mm  ------C..T GC..T..T.. A...
Xl  ------C.CT C...T..... ....
Sp  ------C.T. .......... ....
Ce  ------T... ....T..... ....
```

E. CYTB

```
Se  ACACAAAAAC GAACACGGCT GAGCGAGCCT AGCTTGAAAA TTATTAATAA TACTCTGATT  60
Dy  .TG..T.... CTTT...AAA TTC.C.C... TTA..T.... ..GC..... .G..T.AG..
Mm  .TG---.C.A AC.T...AAA A.CAC.C..A TTA..T.... .......CC. CT.AT.C...
Xl  .TGGC.CCCA AC.TC..TAA ATCTC.T..A TTAA.T.... .......... .T..T.C...
Sp  .TGGC.GCT. C.TT...AAA .GAAC.T..A .TT..CCG.. ..C.G..A.G ...AT.CG..
Ce  ---------- --TTGAAAA. T.ATA.TAGA TTA..A..TT ..G.....GG G.TGT..G.G

Se  GATTTGCCTT CACCAATCAA TATTAGAGTA TGGTGAAATT TTGGTAGTCT ATTAGGTTTA  120
Dy  .....A..AG .T.....T.. ....TC.AG. ..A..... ....ATCAT. .C.T..A...
Mm  ..CC.A...G .C...TC... C...TC.TC. ..A.....C. ....GTCC.. TC....AG.C
Xl  ..CC.C..AA .C...TCA.. C...TC.TC. .TA.....C. .C..CTC... TC....GG.C
Sp  ..CC.C..CC TT..CTC... CC..TCCA.T ........C. CG..CTC... .C....GC.G
Ce  ACA.....A. .TAG..AA.C .T.A.C.T.. A.A..... ........A. ...G...A..
```

```
Se  GTCTTGGTAA TCCAACTTGC AACAGGTCTT TTTTTAGCTA TACATTACAC ATGTGATGTG  180
Dy  TGT..AA.T. .T...A..TT ...T..AT.A .......... ....C..... .GCA.....T
Mm  TG.C.AA..G .....A.CAT T......... ..C.....C. ....C..... ..CA...ACA
Xl  TGT..AA.TG C....A.CAT T.....AT.A ..C..... .......T.. .GCA..CACA
Sp  TG...A..TG .T...A.ATT G..T..AA.A ..CC....A. .G..C..... .GCC...A.T
Ce  ..T..AA.CT .T..GA..TT .......ACA ........AT .TT....T.. GCCC...AG.

Se  AATATGGCTT TTTCTTCTGT GGATCATATT TTCCGGGACG TAAATAGAGG ATGGTTCCTC  240
Dy  ..CT.A.... ...A.AG... TA........ .G...A..T. .....TAT.. T..A..AT.A
Mm  .TA.CA..C. ....A..A.. AACA..C... .GT..A.... .....TAC.. G..AC.AA..
Xl  TC...A..C. .C..A..A.. A.CC...... .GTTTT.... .T..CTAT.. ..TA..AA.T
Sp  .CCT.A..A. ....A..C.. TATG..C... ..G..A..T. .....TAT.. ...A..TT.A
Ce  TTA..A..A. ....AA.A.. .C.GT..... A.GTAT..G. .....TTT.. ...AG.AT.T

Se  CGCAGAGTAC ACGCAAACGG GGCTTCTTTT TTCTTTATAT GTCTTTACTG TCACATCGGA  300
Dy  ..A.CTT... ....T..... T..A..A... ..T.....T. ..A......T A..T..T..T
Mm  ..ATATA... .......... A..C..AA.A ..T.....T. .CT.A.T.CT ...TG.....
Xl  ....ATC.C. .T..C..T.. ACTC..A..C .....C..T. .CA.C...CT ..........
Sp  ..ATAT.... ....C..A.. C.TC...C.C ..T.....C. ..A.G..... C.....A...
Ce  ..A.TTT.T. .TTTT..T.. ...CAGG..A ..T.....T. T.T.G..T.T A..T..TTTT

Se  CGAGGTATTT ATTATGGCAG ATATATGTTT ATTAAGACAT GGTTTAGGGT TTTAGCTTTG  360
Dy  .....A.... ....C..ATC ....T.A... .CACCA..T. .A..AGTA.G AG..AT.A.T
Mm  .....CT.A. .......ATC .....CA... ..AG.A..C. .AAAC.TT.G AG..CT.C.A
Xl  .....GT.G. .C..C...TC T.TCT.A.A. .AAG.A.... .AAA..TT.G .G.GATCC.C
Sp  ..G..AC.A. .C..C..GTC T....AAAAG ...G....C. .AAAAGTT.G .G..ATCC.A
Ce  AA...GT.A. T..T.ATA.. ....CGT..A .AA..AGT.. .AA.GTCT.G ....A.AA.T

Se  TTAATTTTAG TAATGGCGGC CGCTTTTCTT GGCTACGTGT TGCCTTGAGG CCAAATAAGG  420
Dy  ...T...... ....A.GAA. A......A.A ..T..T..T. .A........ A......TCA
Mm  C.GT.CGC.. .C..A..CA. A..A...A.A ........CC .T..A..... A......TCA
Xl  C..T...... .T..A..TA. A..A...G.A ..A..T..TC .A..A..... A......TCT
Sp  ..TT.GG.TA CC..CCTAA. ...C...A.G .....T..C. .AGTC..G.. G.....GTCC
Ce  .ATT.A.... ....AATA.A A......A.A ..T..T..T. .AGT.....C T........A

Se  TTCTGAGGGG CTACTGTTAT TACAAATTTA TTTTCTGCAT TCCCTTATGT GGGACCCAGA  480
Dy  ..T.....A. .A.....A.. ...T...... ..G..A..TA ........T. A..TATAGAC
Mm  ........T. .C..A..... ......CC.C C.A..A..CA ....A...A. T...A.A.CC
Xl  ..T....... ....A..A.. ...T...C.T C.......TA AA..G..CA. C...AA.GT.
Sp  ..T....CT. .C..A..A.. C......... G.G......A .T..C..CA. A...A.T.TT
Ce  ..T....CA. .AGTA..... ...T.GAC.T ..AAGA.TTA .T..AAT.TG A..G..A.CT

Se  CTAGTGCACT GGTTGTGGGG AGGGTTCGCT GTGGAGAACG CTACTCTTAC TCGCTTCTTT  540
Dy  T....A..A. .A..A..A.. ...A..T... ..A..T..T. .....T.A.. ...A..T..C
Mm  .....CG.A. .AA.T..A.. G..C...T.A ..A..C..A. .C..CT.G.. C..A.....C
Xl  .....C..A. .AAGT.TA.. ...A...T.. ..A..T.... .C...T.A.. C..A.....C
Sp  A....T..G. .A..A..A.. G..A...T.C ..C..C..A. .C..C..... C..A..T...
Ce  A.T..TACT. .AA.T..AA. ...T..T.G. ..TACAGGG. .A..AT.A.A ATT...T...
```

```
Se  ACTTTCCACT TCTTAGTACC TTTTATCGCT TCAGCAATAG CTGGACTTCA CATCTTTCTC  600
Dy  ..A..T..T. .TA.TT.... ......T.T. CTT..T...A ..AT.A.... TC.AC.AT.T
Mm  G......... ..A.CT.... A.....TATC G.G..CC... .AATCG.... .C..C.CT..
Xl  G.A..T.... ..C.CC.T.. ......TAT. G.C.G.GCTA GCATT..C.. TC.T..AT.T
Sp  C.C..T.... ..C.TT.C.. C.....AATA G....CT... .G.TTA.A.. .T.AG.AT..
Ce  GTA..A..T. .T...T.G.. G.GAGCTAT. CT..TT..T. TATTGGGG.. TT.AA..T.T

Se  TTACATGAGA CAGGTAGCAA CAATCCATTG GGTGTTAACT CTACTTCAGA TAAGATTCCT  660
Dy  ......C.A. ....ATCT.. T..C..TA.T ...T.A..T. ...A.ATT.. ...A......
Mm  C.C..C..A. ....ATCA.. ...C...ACA ..AT.A.... .AGA.G.... ...A.....A
Xl  C.C..C..A. .T..ATCA.C A..C...ACT ..AT.A.... .AGACC.... ...AG.A...
Sp  C.C..CA.C. G...AGC... ...C..T..T .CCT.C..AA GC.AC.AT.. C...GCC...
Ce  .....CAGA. .T.....A.C ATC.AGG..A TA.TGCC..G G.GA..AT.. ...AG..TG.

Se  TTCCATTGAT ATTACAGTGT AAAAGATATG TTAGGTTTTG TTGTAATGTT TACGCTTCTG  720
Dy  ..T..CCC.. .C.T..CAT. T..G.....T G....A...A .......AA. .TTTA....A
Mm  ..T..CCCC. .C..T.CAA. C........C C.....A.CC .AA.C..A.. CTTAA....C
Xl  .....CCC.. .C.T.TC.TA C.....CC.T .....C..CC ..A.T..AC. ...AGCA..T
Sp  .....CATT. .C.T..CGAC C..G..C.CA G.C..G...A ..C.TT..G. AG.CGCA..A
Ce  ..TAGACCTG .G...TTA.G T......GCT .ATAA.---A ....T..T.G ATTAT.AT.T

Se  TTTTCGTTAT TTTTGCT--- -CCGAATTTC TTGGGAGAAC CGGATAACTT TATCGCCGCC  776
Dy  A....A...G ....AA.TAG A..A.....A ........C. .A........ ...TC.T..T
Mm  A.AA.CC..G .A..AT.TTT C..AG.CA.A C.A.....C. .A..C....A C..AC.A..T
Xl  AC.CTCC..G CCA.AT.TTC C..A..CC.T ..A.....C. .A..C..T.. ..C.C.A..T
Sp  ...AGC...G CCC.C..ATT T..TGGCGC. C.AAA...C. .A..G..A.. C..TC.T..A
Ce  A..GT....A GG...A.TTA C...TT.AAT ..A..T..TG .A..G.TG.. ...T.AA..T

Se  AA                                                                778
Dy  ..
Mm  ..
Xl  ..
Sp  ..
Ce  G.
```

FIGURE 20.

Amino acid sequences of the chaetognath (Se) mitochondrial genes, aligned with their counterparts from other species. Gaps are indicated by dashes, residues similarity with those of the chaetognath by dots. Published sequences used for comparison are *Drosophila* (Dy), mouse (Mm), frog (Xl), sea urchin (Sp), and nematode (Ce). Alignments are shown for A. COI; B. ND2; C. COIII; D. COII; E. Cyt B. Amino acid sequences are depicted using the standard 1-letter code.

A. COI

```
Se  --------MT --RWLFSTNH KDIGTLYFVL GIWSAFLGTA LSALIRLELG NAGSLLGDDQ  50
Dy  --------.- SRQ..... ........IF .A.AGMV..S ..I...A... HP.A.I....
Mm  --------.F IN..... .......LLF .A.AGMV... ..I...A... QP.A......
Xl  --------.A IT..... .......L.F .A.AGLV... ..L...A..S QP.T......
Sp  --------.Q LS..... .......LIF .A.AGMV... M.VI..A..A QP....K...
Ce  INLYKKYQGG LAV..E.S.. ........IF .L..GMV..S F.L......A KP.FF.SNG.

Se  LYNVIVTAHA FIMIFFFVMP TMMGGFGNWL VPLMVNAPDM AFPRLNNMSF WLLPPALMLL  110
Dy  I......... ......M... I.I..... ....LG.... ....M..... .......S..
Mm  I......... .V....M... M.I..... ....IG.... ....M..... .....SFL..
Xl  I......... ......M... I.I..... ....IG.... ....M..... .....SFL..
Sp  M.K.V...QS LV....M... M.I..... I...IG.... ....MK.... ..I..SF...
Ce  ...SVI.... IL....M... ..I..... L...LG.... S......L.. ....TSML.I

Se  LLSGMVESGV GTGWTVYPPL ST-VGHTGGA VDLGIFSLHL AGVSSILGSA NFITTIVNMK  169
Dy  .V.S...N.A .......... .SGIA.G.AS ...A...... ..I.....AV .....VI..R
Mm  .A.S...A.A .......... AGNPV.A.AS ...T...... ........AI ......I...
Xl  .A.SG..A.A .......... AGNLA.A.AS ...T...... ..I.....AI .....TI...
Sp  .A.AG..K.A .....I.... .SKMT.A.SS ...A...... ..A.....LI K.M...I..R
Ce  .DACF.DM.C ..S..... ..-M..P.SS ...A.....A ..L.....GI ..MC.TK.LR

Se  GEGMTMELMS LFVWSVLLTA ILLLLSLPVL AGAITMLLTD RNFNTSFFDP AGGGDPILYQ  229
Dy  ST.I.LDR.P ......VI.. L......... .......... ..L....... ..........
Mm  PPA..QYQTP .......I.. V......... .AG....... ..L..T.... ..........
Xl  PPA.SQYQTP .......I.. V......... .AG....... ..L..T.... ......V...
Sp  TP..SLDRLP ......FV.. F......... .......... .KM..T.... ........F.
Ce  SSSISL.H.T ....T.FV.V F..V...... .......... ..L....... ST..N.LI..

Se  HLFWFFGHPE VYILILPGFG MVSQIINHYS AKGNSFGALG MIYAMSSIAL LGFVVWAHHM  289
Dy  .......... .......... .I.H..SQE. G.KET..S.. .....LA.G. ...I......
Mm  .......... .......... II.HVVTY.. G.KEP..YM. .VW..M..GF ...I......
Xl  .......... .......... .I.H.VTY.. G.KEP..YM. .VW..M..G. ...I......
Sp  ....L..... .......... .I.HVMA... G.REP..Y.. LV...IAMGV ...L......
Ce  .......... .......A.. I...STLYLT G.KEV..... .V..IL..G. I.C.......

Se  FTVGMDVDTR AYFTSATMII AVPTGIKIFS WLATLHGTPS LLETPLMWVL GFLFLFTVGG  349
Dy  .......... ...S...... .......... ........QL SYSPAIL.A. ..V.......
Mm  ....L..... .C........ .I...V.V.. .......GNI KWSPAML.A. ..I.......
Xl  ...DLN.... .......... .I...V.V.. ....M..GTI KWDA.ML.A. ..I.......
Sp  .......... ....A..... .....L.V.. .M.K.Q.SNL QWSL..L.T. .IV....L..
Ce  Y.....L.S. ...SA...V. .....V.V.. .....F.MKM VFNPL.L... ..I....L..

Se  LTGVVLANSS LDISLHDTYY VVAHFHYVLS MGAVFAIFAG VTFWYPVISG MTMSARGTQV  409
Dy  .......... V..I...... .......... .......M.. FIH...LFT. L.LNNKWLKS
Mm  ...I..S... ...V...... .......... .......M.. FVH.F.LF.. F.LDDTWAKA
Xl  ...I...... ...M...... .......... .......MG. FIH.F.LFT. Y.LHETWAKI
Sp  ...I...... I.FV...... .......... .......... F.H.F.LF.. YSLHPLWGK.
Ce  ......S... ...I...... ..S....... L....G..T. ..L.WSF.T. YVLDKLMMSA
```

```
Se  QFAIMFIGVN LTFFPQHFLG LQGMPRRYSD YPDSFSTWNV VSSSGSLVSI VGVFMFVGVM  469
Dy  ..I.......  ..........  .A........  ...AYT....  ..TI..TI.L  L.ILF.FYII
Mm  H.....V...  M.........  .S........  ...AYT...T  ...M..FI.L  TA.LIMIFMI
Xl  H.GV..A...  ..........  .SA.......  ...AYTL..T  ...I...I.L  .A.I.MMFII
Sp  H.FM..V...  ..........  .A........  ...AYTL..T  I..I..TI.V  .AMLF.LFLI
Ce  V.ILL.....  .....L..A.  .H.F..K.L.  ...VY.V..I  IA.Y..II.T  A.L.L.IY.L

Se  WKSLSA--IN ASGDVDNLSV EFSPRLPVSW HSFNESTVLG ------                  507
Dy  .E..VSQRQV IYPIQL.S.I .WYQNT.PAE ..YS.LPL.T N-----
Mm  .EAFASKREV M.VSYASTNL .WLHGC.PPY .T.E.P.YVK VK----
Xl  .EAFA.KREV TTYELTSTML .WLQGC.TPY .TLKT.L.QI NHQMIK
Sp  .EAFASQREG ITPEFSHA.L .WQYTSFPPS .HTFDE.PST MIIVK-
Ce  LE.FFSYRLV I.DYYS.S.P .YCMSNY.FG ..YQSEIYFS TTSLKN
```

B. ND2

```
Se  VSQSCGSLMI LLGGMV---- -------ADS SFFSVSLLLH GVVFKMGLMP LHFWVPCVVM   49
Dy  LT.ALA.TVL .FSSILL--M LANNLNNEIN ES.TSMIIMS ALLL.S.AA. F...F.NMME
Mm  .T.ATA.MI. ..AIVLNYKQ LGTWMFQQQT NGLILNMT.M ALSM.L..A. F...L.E.TQ
Xl  LT.AAA.ALL .FSSLNNAWL TGEWSILDLT NPL.CATMTI AICM.L..A. F...L.E.LQ
Sp  LV.ASSAALL .K.ALGQAWL TGSWSILDPV KEVTSIC.SM ALA.....A. V...F.D.LQ
Ce  .I.ESLG.LF ..CS------ ---------- ---GGL.QFF IILL.I.VA. ....IFN.TN

Se  NLTRFNLYL- LMSWQKIGPI VIV-----MT ASVGYTVLCL VNAVGGSLAM SGVTVLPLLL  103
Dy  G..WM.ALM- ..T....A.L MLISY---LN IKNLLLISVI LSVII.AIGG LNQ.S.RK.M
Mm  GIPLHMGLI- .LT....A.L S.LIQIYPLL N.TIILM.AI TSIFM.AWGG LNQ.QMRKIM
Xl  G.SLTTGLI- .ST...LA.M A.LYQIAP.L NTPLLLT.G. TSTLI.GWGG LNQ.Q.RKI.
Sp  G.PF.QGLM- MAT...MA.L MLMFYFSQLG F.YLLMTPS. ISVLM.GWGG LNQ.QVRKI.
Ce  .IFNY-GLMW FLTF..LPFL T.LLQIF-WL S..YILLFG. LICYVQI--- FVMKSYKN..

Se  IFSGMVQMGW VFMT---TGV FTFYYLFVYY IVLSAVVLY- -----SAVAS VQFGW-----  149
Dy  A..SINHL.. MLSSLMISES IWLI.FIF.S FLSFVLTFMF NIFKLFHLNQ LFSWFVNSKI
Mm  AY.SIAH... MLAILPYNPS L.LLN.MI.I .LTAPMFMAL MLNNSMTIN. ISLL.NKTPA
Xl  A..SIAHL.. MISILPFSPQ LMILN.TI.L .MT.TMF.VL KTISSTKIS. LATS.SKTPS
Sp  A..S.GK... LV..SAYSFN AAIIM.VI.L .INTSLF.LF DHLKV.TLGH LKTISQLSPI
Ce  .I.STESFN. IVLGVFFSMF N...-..I.. F..MVLLISK ---------- ----FSKTSG

Se  -------ALL NAGGLPPFSG FMIKLKAILH IKGS----MV VLLVGASGLA LTSYIRLLLN  198
Dy  LKFSLFMNF. SL......L. .LP.WLV.QQ LTMCNQYFLL T.MMMSTLIT .FF.L.ICYS
Mm  MLTMISLM.. SL.....LT. .LP.WII.TE LMKNNCLI.A T.MAMMAL.N .FF....IYS
Xl  TTALSLLT.. SL.....L.. .VP.WFI.QE LTSQNTTILA TT.ALSAL.S .FF.L..TYI
Sp  SVALVLLVM. SL.....LT. .IL.FTSLYF LVAKNFIILS SIMMIGNLQD YFF.L.ISFK
Ce  YNFINWETT. VFLNI.FSVS .FV.IFSLSE .FKYDSFFTL F..FTMFLSV .AFSFW.INL

Se  TRLKSGPSSG FLVATMVAGS V                                             219
Dy  AFMLNYFENN WIMEMNMNSN NTNLYLIMTF FSIFGLFLIS LFFFML
Mm  .S.TMF.TNN NSKMMTHQTK TKPNLMFSTL AIMSTMTLPL APQLIT
Xl  VT.T.S.NTS NASL.WRHH. KQPSLLLSIA LILSSFIIPI SPLTLT
Sp  .S.FLF.QHI ISS.SWRNST MISPLAPKAW LSSVSTVLST LAMPLTLPLY MIT
Ce  SMKNNEET.N NNKMNYFIIF PLMVISII
```

C. COIII

```
Se  MNK---HPFH IVDVSPWPLM GSVGSLCLVG GLVTTMHRYG SSLFWLGIAL ILATMFQWWR  57
Dy  .STHSN.... L..Y.....T .AI.AMTT.S .M.KWF.Q.D I...L..NII TIL.VY....
Mm  .-THQT.AY. M.NP.....T .AFSA.L.TS ...MWF.YNS IT.LT..LLT NIL..Y....
Xl  .-AHQA.AY. M..P.....T .A.AA.L.TS ..AMWF.FGS MI.LT..LIT MVL..I....
Sp  .-AIQ...Y. L..Q.....D .AFSG.MMTS .N.LWF.TQK TN.TLV.FL. LMTK.VN...
Ce  .--..F.N.. .LSL.SYAYN LFFA.AGMLS S..MFFKFGL YE..IFTLFS V.FIS.A.GK

Se  DVTREATFQG KHTAKVESGM RYGMLLFISS EVFFFLAFFW AFFHSALSPN VEVGSVWPPL  117
Dy  ..S..G.Y.. L..YA.TI.L .W..I...L. ..L..VS... .....S...A I.L.AS...M
Mm  ..I..G.Y.. H..PI.QK.L ....I...V. .....AG... ..Y..S.V.T HDL.GC...T
Xl  ..I..G.... H..PP.QK.L ....I...T. .....IG... ..YN.S.A.T Y.L.EC...T
Sp  .MI.K.N... S...I.KK.. ....I..MT. ..C..F.... .....S.A.S ..M.VA...S
Ce  .IAM.G-LS. Y.NFF.MD.F KF.VI..VF. .FM..FCI.. T..DA..V.V H.L.ET.S.F

Se  GISAINPFDV PLLNTSILLS SGATITWAHS ALLENRWLES QLSLIITVVL GFYFSMLQGL  177
Dy  ..ISF...QI .....A...A ..V.V....H S.M.SNHSQT TQG.FF..L. .I..TI..AY
Mm  ...PL..LE. ......V..A ..VS.....H S.M.GKRNHM NQA.L..IM. .L..TI..AS
Xl  ..TPL...E. .....AV..A ..V.V....H SIMHGDRK.A IQ..TL.IL. .L..TA..AM
Sp  .MTPL...L. ...K.GV... ..V.LS.S.H SI.AGNRT.. IQA.FL..A. .S..TA..AW
Ce  .MHLV...G. .....I.... ..V.V....H S..S.K--SC TN.M.L.CL. AA..TGI.LM

Se  EYVWAGFSLS DGIYGSTFYV ATGFHGLHVL IGTLFIAVMA YRNLYHHFSC S           228
Dy  ..IE.P.TIA .SV......M ......V... ...T.LL.CL L.H.NN...K N
Mm  ..FETS..I. ........FM .........I ..ST.LI.CL L.Q.KF..TS K
Xl  ..YE.P.TIA ..V.....F. .........I ..S..LS.CL L.QIQY..TS K
Sp  ..ID.P.TIA .SV.....F. .......Q.I M..T.LM.CL F.TAGR...T H
Ce  ..ME.S..IA ..VF..I..L S.....I... C.G..L.FNF L.L.KN..NY N
```

D. COII

```
Se  MS-------L NFQNSNSPLM EQLIFFHDWV ----MVFVSS ITVGYLI-LI VSNKPTHRVL  48
Dy  ..TWAN---. GL.D.A.... ........HA LL--ILVMIT VL....MFML FF.NYVN.F.
Mm  .AYPFQ---. GL.DAT..I. .E.MN...HT LM--I..LI. SL.L.I.S.M LTT.L..TST
Xl  .AHPSQ---. G..DAA..I. .E.LH...HT LM--A..LI. TL.L.I.TIM MTT.L.NTN.
Sp  .GTWAQ---F GL.DAS.... .E.TY...YA LI--VLTLIT ML.F.GLVSL LVSSN.N.FF
Ce  INNFFQGYN. L..H.LFASY MDWFHSFNCS LLLGVLVFVT LLF....FGT FYF.SKKIEY

Se  LESQGVEFAW TALPCLVLVA IALPSLRLLY SMDEIIDPSL TIKAMGHQWY WSYEYSDVDE  108
Dy  .HG.LI.MI. .I..AII.LF .......... LL...NE..V .L.SI..... .......FNN
Mm  MDA.E..TI. .I..AVI.IM .......I.. M....NN.V. .V.T..... .....T.YED
Xl  MDA.EI.MV. .IM.AIS.IM .......I.. L...VN..H. ....I..... .....TNYED
Sp  F.G.EL.TI. .VI.A.I.IL ......Q... L...VK..F. ....F..... .....T.FKD
Ce  QFGELLCSIF PTIIL.MQMV PS.SL.YYYG L.N--L.SN. .V.VT..... .......IPG
```

```
Se  ESMEFDSY                                                    116
Dy  --I.....
Mm  --LC....
Xl  --LS....
Sp  --L.....
Ce  --L.....
```

E. CYTB

```
Se  TQKRTRLSEP SLKIINNTLI DLPSPINISV WWNFGSLLGL VLVIQLATGL FLAMHYTCDV  60
Dy  MH.PL.N.H. LF..A..A.V ...A.....S .......... C.I..IL... .......A..
Mm  M-TNM.KTH. LF....HSF. ...A.S...S .........V C.MV.II... .......S.T
Xl  MAPNI.K.H. LI.....SF. ...T.S...S L.........V C.IA.II... .......A.T
Sp  MAAPL.KEH. IFR.LKS.FV ...L.S.L.I ...S...... C..V.ML..M .......A.I
Ce  ----LKINNS L.NFV.GM.V T...SKTLTL S.....M..M ..IF.IL..T ...FY..P.S

Se  NMAFSSVDHI FRDVNSGWFL RSVHANGASF FFMCLYCHIG RGIYYGSYMF IKTWFSVLAL 120
Dy  .L..Y..N.. C....Y..L. .TL....... ..I.I.L... ........L. TP..LVGVII
Mm  MT.....T.. C....Y..LI .YM......M ..I..FL.V. ..L.....T. ME..NIGVL.
Xl  S......A.. CF...Y.LLI .NL....L.. ..I.I.L... ..L....FLY KE..NIGVI.
Sp  TL.....M.. L....Y.... .Y...K.V.L ..I.M...M. ..L.....KK .E..KVGVI.
Ce  L....T.QY. MYE..F..VF .IF.F....L ..IF..L..F K.LFFM..RL K.V.M.G.TI

Se  LILVMAAAFL GYVLPWGQMS FWGATVITNL FSAFPYVGPS LVHWLWGGFA VENATLTRFF 180
Dy  .F...GT..M .......... .......... L..I..L.MD ..Q..... .D........
Mm  .FA...T..M .......... .......... L..I..I.TT ..E.I....S .DK.......
Xl  .F....T..V .......... .......... L..K..I.NV ..Q.SL...S .D........
Sp  FLVTILT..M ....V..... ..A....... V..I..M.TI M.Q......S .DK.......
Ce  YL...ME..M ....V.A... ..A.V...S. L.VI.IW..T I.T.I.S..G .TG...KF..

Se  TFHFLVPFIA SAMAGLHIFL LHETGSNNPL GVNSTSDKIP FHWYYSVKDM LGFVVMFTLL 240
Dy  ....IL...V L..TMI.LLF ..Q......I .L..NI.... ..P.FTF..I V..I..IFI.
Mm  A...IL...I A.L.IV.LLF .........T .L..DA.... ..P..TI..I ..ILI..LI.
Xl  A....L...I AGASI..LLF ......T..T .L..DP..V. ..P.F.Y..L ...LI.L.A.
Sp  P....F..MM A.L.VM.LVF ..NS.A...F AFK.NY..A. ..I.FTT..T V..ILLVAA.
Ce  VL...L.WAI LVIVLG.LIF ..S...TSS. YCHGDY..VC .SPE.LG..A YN-I.IWL.F

Se  FSLFL--PNF LGEPDNFIAA                                       258
Dy  I..V.IS..L ..D.....P.
Mm  MT.V.FF.DM ..D...YMP.
Xl  TL.AMFS..L ..D....TP.
Sp  ...A.LF.GA .KD.EK..P.
Ce  IV.S.IY.FN ..DAEM..E.
```

(Figure 20A). A two codon deletion occurred at the 5' end of the COI gene, and a codon deletion was located within the gene. The remaining deletions were located at the 3' regions of this gene. Although the complete coding for COII, COIII, Cytochrome b, and ND2 genes sequence have not been determined, available data show that addition and deletion events have occurred in these protein coding genes. In the COII gene, at least one addition (two codons) and three deletions (involving from one to seven codons) are required to align the amino acid sequences (Figure 20D). In the COIII gene, one addition (one codon) and one deletion (three codons) have been found (Figure 20C). It is surprising to note that deletions have occurred at several separate domains of the ND2 gene (Figure 20B). The first domain is located near the 5' end of this gene and includes deletion of 9-11 codons relative to other eucoelomate species. The second domain is found at codon position 72 downstream from 5' end of this gene and includes deletions of 2-5 codons. Other domains including at least 28 codons are located in the middle and at 3' end of the ND2 gene (Figure 20B).

Unlike the gene organization of the vestimentiferan, in which all of four protein-coding sequences are punctuated by intervening tRNA genes, only one tRNA gene was found in *Sagitta*, at the 5' end of the cytochrome b gene. Many of the protein-coding sequences are contiguous with one another, rather than being flanked tRNA genes on the same strand. The ND2 and COIII, COI and COII gene junctions are examples. In most metazoan mtDNAs (mammals, amphibia, *Drosophila*) tRNA genes are located between most pairs of protein genes and the secondary structure of the tRNA is important for the precise cleavage of the mature protein gene transcript from primary multicistronic transcripts (Ojalaet al. 1980; Ojala et al. 1981). Consistent with this view is the finding that in these mtDNAs the

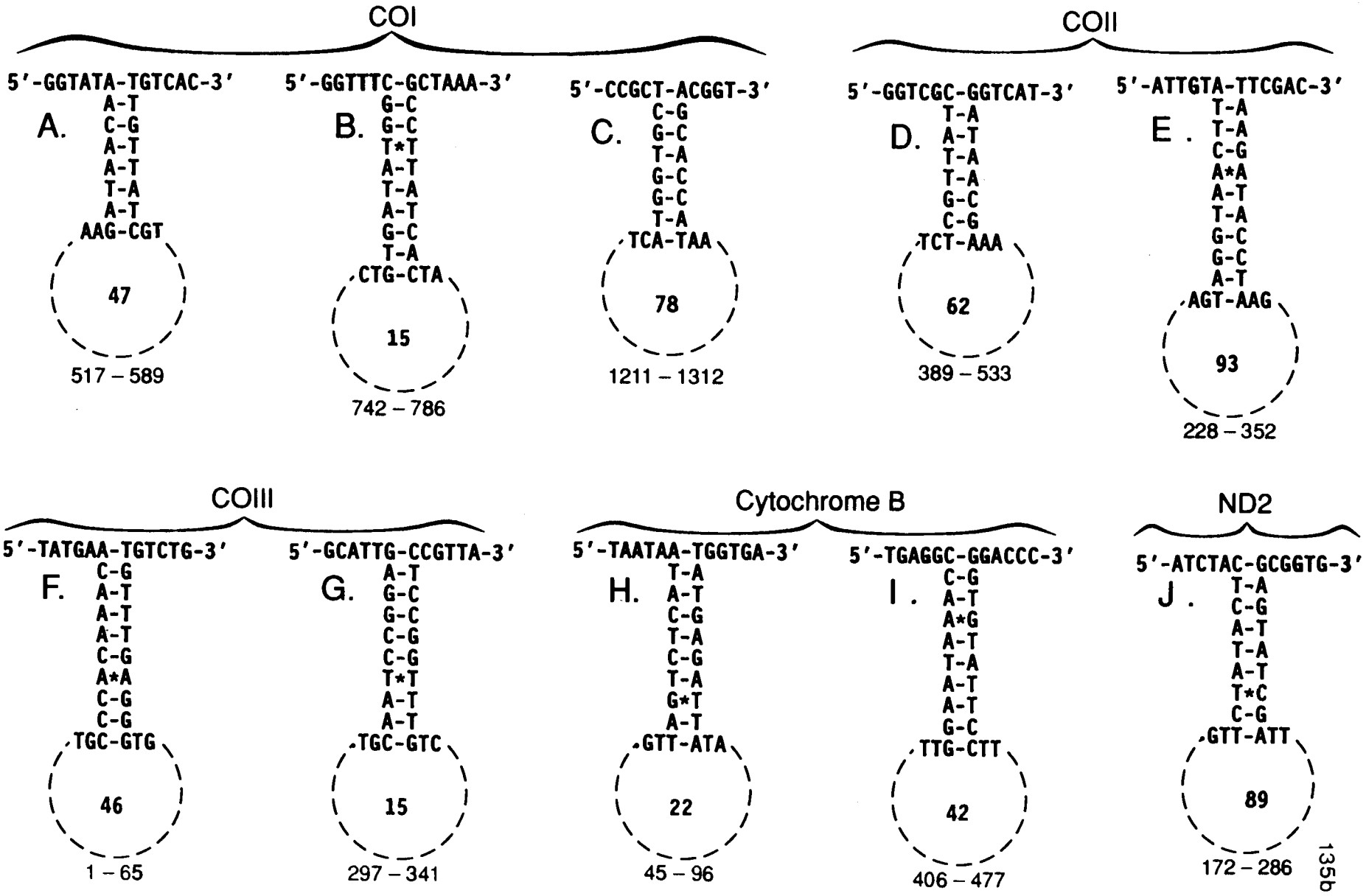
sequence adjacent to the 3'-terminus of protein genes, that are followed immediately by another protein gene (rather than a tRNA gene), may have the potential to form a hairpin structure (Bibb et al. 1981; Clary and Wolstenholme 1985). In the chaetognath, hairpin structures were not found in these regions, although some intergenic sequences do occur between these protein genes. Only one potential secondary structure was found, near the 5' end of the chaetognath COIII gene (Figure 21F). The intergenic sequences between protein genes lack stable secondary structure potential.

In contrast to the human, mouse, and cow mitochondrial genomes, which occasionally use ATG, ATT or ATC codons as translational initiators, all of three identified protein genes (COI, COII, and COIII) except for the cytochrome b gene in Phylum Chaetognatha begin with the orthodox ATG codon. Two identified protein genes end with either TAA termination codon (ND2) or TAG termination codon (COI).

Further sequence analysis of these protein genes revealed that there were a number of potential secondary structures within each gene which may be functional in the mitochondrial system. In the COI gene, I have identified a number of potential secondary structures, three of them are illustrated in Figure 21A-C. As shown in Figure 21, a stem loop structure was found at nucleotide position 522 downstream from 5' end of the COI gene, which consists of a 8 base stem and a 47 base loop. The second potential hairpin structure with a 10 base stem (one mispairing) and a 15 base loop was located at nucleotide position 744 downstream from 5' end of the COI gene. Another stem loop structure occurs at nucleotide position 1215 downstream from 5' end of the COI gene has a 8 base stem and a 78 base loop. In the COII gene, at least two stem loop structures were found, one was located at nucleotide 203 downstream from 5' end of this gene (Figure 21D) and the other

FIGURE 21.

Secondary structures found within each protein gene or across gene junctions. Standard base pairs are designated by dashes. Asterisk denotes mispairing bases. The number in the centre of the loop is the nucleotide length of loop.


COI
5'-GGTATA-TGTCAC-3'
A-T
A. C-G
A-T
A-T
T-A
A-T
AAG-CGT
47
517 – 589
5'-GGTTTC-GCTAAA-3'
G-C
B. G-C
T*T
A-T
T-A
A-T
G-C
T-A
CTG-CTA
15
742 – 786
5'-CCGCT-ACGGT-3'
C-G
C. G-C
T-A
G-C
G-C
T-A
TCA-TAA
78
1211 – 1312
COII
5'-GGTCGC-GGTCAT-3'
T-A
D. A-T
T-A
T-A
G-C
C-G
TCT-AAA
62
389 – 533
5'-ATTGTA-TTCGAC-3'
T-A
E. T-A
C-G
A*A
A-T
T-A
G-C
G-C
A-T
AGT-AAG
93
228 – 352
COIII
5'-TATGAA-TGTCTG-3'
C-G
F. A-T
A-T
A-T
C-G
A*A
C-G
C-G
TGC-GTG
46
1 – 65
5'-GCATTG-CCGTTA-3'
A-T
G. G-C
G-C
C-G
C-G
T*T
A-T
A-T
TGC-GTC
15
297 – 341
Cytochrome B
5'-TAATAA-TGGTGA-3'
T-A
H. A-T
C-G
T-A
C-G
T-A
G*T
A-T
GTT-ATA
22
45 – 96
5'-TGAGGC-GGACCC-3'
C-G
I. A-T
A*G
A-T
T-A
A-T
A-T
G-C
TTG-CTT
42
406 – 477
ND2
5'-ATCTAC-GCGGTG-3'
T-A
J. C-G
A-T
T-A
A-T
T*C
C-G
GTT-ATT
89
172 – 286

occurs at nucleotide 233 downstream from 5' end of same gene (Figure 21E). In the COIII gene, there are at least two potential secondary structures, a stem loop structure with a 10 base stem and 46 base loop at 5th nucleotide downstream from 5' end of this gene (Figure 21F), and a potential hairpin structure with a 10 base stem and a 15 base loop located at position 312 downstream from 5' end of this gene (Figure 21G). Two potential hairpin structures were found in the cytochrome b gene. The first one is located at nucleotide 50 downstream from 5' end of this gene and has a 10 base stem including one unpaired site and a 22 base loop (Figure 21H); the second has a 10 base stem (one unpairing) and 42 base loop and was found at nucleotide 411 downstream from 5' end from this gene (Figure 21I). In the ND2 gene, a potential secondary structure occurs at nucleotide 177 downstream from 5' end and has a 9 base stem and a 89 base loop (Figure 21J).

Genetic Code

The codon AGA, which specifies rare termination codons in some vertebrate mtDNAs, and serine in *Drosophila* and sea urchin mtDNAs, is found frequently in the chaetognath mitochondrial genes (Table 14). The triplet AGA specifies arginine in the standard genetic code. In chaetognath mtDNA, internal AGA and AGG codons are present in all five protein genes. None of the AGA and AGG codons in these protein genes correspond in position to arginine-specifying codons (CGN) in the equivalent genes of mitochondrial DNAs of five species including both vertebrates (mouse and frog) and invertebrates (*Drosophila*, sea urchin and nematode).

Further analysis of those positions where AGA and AGG occur in the chaetognath sequence supports the view that AGA and AGG specify serine. In conserved genes, COI,

Table 14. Codon usage in COI, COII, COIII, ND2 and Cytochrome B genes of *Sagitta* (Phylum Chaetognatha) mitochondrial DNA

Phe	TTT	65	Ser	TCT	41	Tyr	TAT	28	Cys	TGT	7
	TTC	43		TCC	9		TAC	17		TGC	2
Leu	TTA	48		TCA	13	TER	TAA	1	Trp	TGA	17
	TTG	47		TCG	15		TAG	1		TGG	24
Leu	CTT	26	Pro	CCT	29	His	CAT	22	Arg	CGT	10
	CTC	18		CCC	6		CAC	18		CGC	4
	CTA	27		CCA	9	Gln	CAA	16		CGA	7
	CTG	16		CCG	9		CAG	6		CGG	6
Ile	ATT	62	Thr	ACT	39	Asn	AAT	26	Ser	AGT	14
	ATC	15		ACC	11		AAC	21		AGC	10
Met	ATA	34		ACA	19	Lys	AAA	13		AGA	17
	ATG	33		ACG	7		AAG	8		AGG	11
Val	GTT	42	Ala	GCT	40	Asp	GAT	19	Gly	GGT	56
	GTC	18		GCC	16		GAC	14		GGC	22
	GTA	35		GCA	20	Glu	GAA	12		GGA	19
	GTG	29		GCG	8		GAG	15		GGG	18

COIII, and cytochrome b genes, which have the greatest similarity to the equivalent genes from all five species compared, about 50% of them correspond in position to serine-specifying codons in the other species. Other AGA and AGG codons in these protein genes correspond in position to codons specifying different amino acids. In view of these observations and noting that AGA and AGG specify serine in other invertebrate mitochondrial genetic codes (Wolstenholme and Clary 1985; Himeno et al. 1987; Wolstenholme et al. 1987; Garey and Wolstenholme 1989), it seems reasonable to conclude that, in chaetognath, like that in Phylum Vestimentifera, AGN codons all specify serine.

Internal TGA codons are also found in all these five protein genes of chaetognath mitochondrial DNA. Analysis of those positions where TGA occurs in the chaetognath sequence reveals that almost all correspond in position to tryptophan-specifying codon in equivalent genes from other species. This indicates that in chaetognath mtDNA, as in Phylum Vestimentifera (see above), in mammalian, fungal, and *Drosophila* mtDNAs (Barrel et al. 1979, 1980; Fox 1979; Bonitz et al. 1980; Heckman et al. 1980; Clary and Wolstenholme 1985), TGA specifies tryptophan.

It is interest to note that the triplet ACA may be used for translation initiation codon of the cytochrome b gene. As shown in Figure 17A, the cytochrome b gene lacks an ATG (ATN) traditional translation initiation codon, the first ATG codon is located at 25th codon position of upstream of 5' end from the ACA codon. Two termination codons, TAA and TAG, immediately follow the ATG codon. If one of the ATT or ATA triplets occurring in the region between the ATG and ACA codons were used to initiate translation of this gene, a TAG termination codon three positions upstream would stop this reading frame immediately. The exact initiation position of the cytochrome b gene has not been

determined yet. The possible sites are either the GAC codon which follows the TAG terminator or the ACA codon. The former would make the cytochrome b gene two amino acids longer than those of other species including all vertebrates and some invertebrates. Therefore, it has been reasoned that ACA may serve to function as the translation initiation codon of the chaetognath mitochondrial cytochrome b gene. In many of the protein genes of mtDNAs from other metazoan phyla, triplets other than ATG are used as translation initiation codons. Some or all ATN codons are used in this way among mammals, *Drosophila*, and sea urchin (Anderson et al., 1981, 1982; Bibb et al., 1981; Clary and Wolstenholme 1985; Jacobs et al., 1988; Cantatore et al., 1989). GTG and GTT have been reported as rare mitochondrial protein gene translation initiation codon in various metazoan (Bibb et al., 1981; Clary and Wolstenholme 1985; Jacobs et al., 1988; Gadaleta et al., 1989). The possible utilization of a GTG codon as translation initiator of the COI gene has been suggested in chicken (Desjardins and Morais 1990) and this unusual start codon is also found at the 5' end of the Japanese quail COI gene (Desjardins and Morais 1991). ATAA has been suggested as a translation initiation codon of *Drosophila* COI gene (de Bruijn 1983; Clary and Wolstenholme 1983). Recently, Okimoto et al. (1990; 1992) reported evidence obtained from alignment of the *Caenorhabditis elegans* and *Ascaris suum* nucleotide sequences indicating that in two nematode mitochondrial protein genes, TTG is used as the translation initiation codon. Also, GTT seems to be the translation initiation codon of the *Ascaris suum* COIII gene.

Codon Usage

The codon usage among the chaetognath mitochondrial protein genes is summarized in

Table 14. All sense codons are used but not with an equal frequency. In all of 1,330 codons, the frequency of first position with A, T, C,and G is 25.6%, 28.4%, 17.2%, and 28.8% respectively. The frequency of codon T group (second codon position) is 42%, the highest in all four groups; 21.9% for C group; and the frequency of both A and G groups is almost same, 17.8% for A group, 16.8% for G group. The frequency of codon usage of T group is 2 times more than that of A, C, and G group. The T group with G in third position favoured is almost equal to the sum of all A, C, and G group. The A group shows the greatest bias against G in the terminal position. The frequency of codons ending in T, A, C, and G is 39.5%, 23.1%, 18.3%, and 19% respectively.

The pattern of codon bias is markedly different from that found in other taxa. In *Drosophila yakuba*, the usage of A and T in the terminal position among 13 protein genes is 93.8% of all codons. In vertebrates, C appears to be a preferred wobble base in human mtDNA, but A is more prevalent in the mouse. In sea urchin, the the frequency of A and T in third position is 60%. In the Phylum Vestimentifera, data from the five protein genes show that the frequency of codons ending in A and T is 64.4%. In the chaetognath, analysis of codon usage in five protein genes reveals that the frequency of T in the third position is 39.5% of all codons, two times higher than that of A, C, and G. The frequency of A and T residues in the third-base position is 62.6% of all codons. This value is almost equal to that of both the Phylum Vestimentifera (64.4%) and sea urchin (60%), but in other respects no obvious rationale can be advanced for the particular pattern of bias observed. For example, GAA and GAG are used with almost equal frequency to specify glutamate, whereas tryptophan codons are 1.4 times more likely to be TGG than TGA.

The frequency of leucine is high and remarkably constant, in the range 15.0 to 16.9%, among mitochondrial proteins of different metazoa. In the chaetognath, the frequency of leucine is also high and is 13.4% of all codons. Among mammalian, insect (*Drosophila yakuba*), platyhelminth (*Fasciola hepatica*) and nematode 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 (Clary and Wolstenholme 1985; Garey and Wolstenholme 1987; Okitomo et al. 1992). Data from the chaetognath mitochondrial protein genes add to the generality of this correlation; the ratio of TTR:CTN codons is 1.1:1, and the corresponding ratio of T:C nucleotides in the third positions of codons is 2.2:1. However, this correlation is not found for sea urchin (Cantatore et al. 1989), frog (Roe et al. 1985), nor Phylum Vestimentifera (see above) mitochondrial protein genes, suggesting that in 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.

Transfer RNA Gene

Only one tRNA gene was identified in the sequences obtained of the chaetognath mitochondrial genome. This tRNA gene is located between the cytochrome b and large ribosomal RNA genes (Figure 17A) and was defined as $tRNA^{Ser}$ based on its predicted secondary structure and anticodon feature. In fact, an intergenic sequence of 23 nucleotides in length occurs between the cytochrome b and this tRNA gene (Figure 17A). The size of the $tRNA^{Ser}$ is only 59 nucleotides long, the shortest among several animal species (70, 68, 68, and 65 nucleotides for that of frog, mouse, sea urchin, and *Drosophila yakuba*).

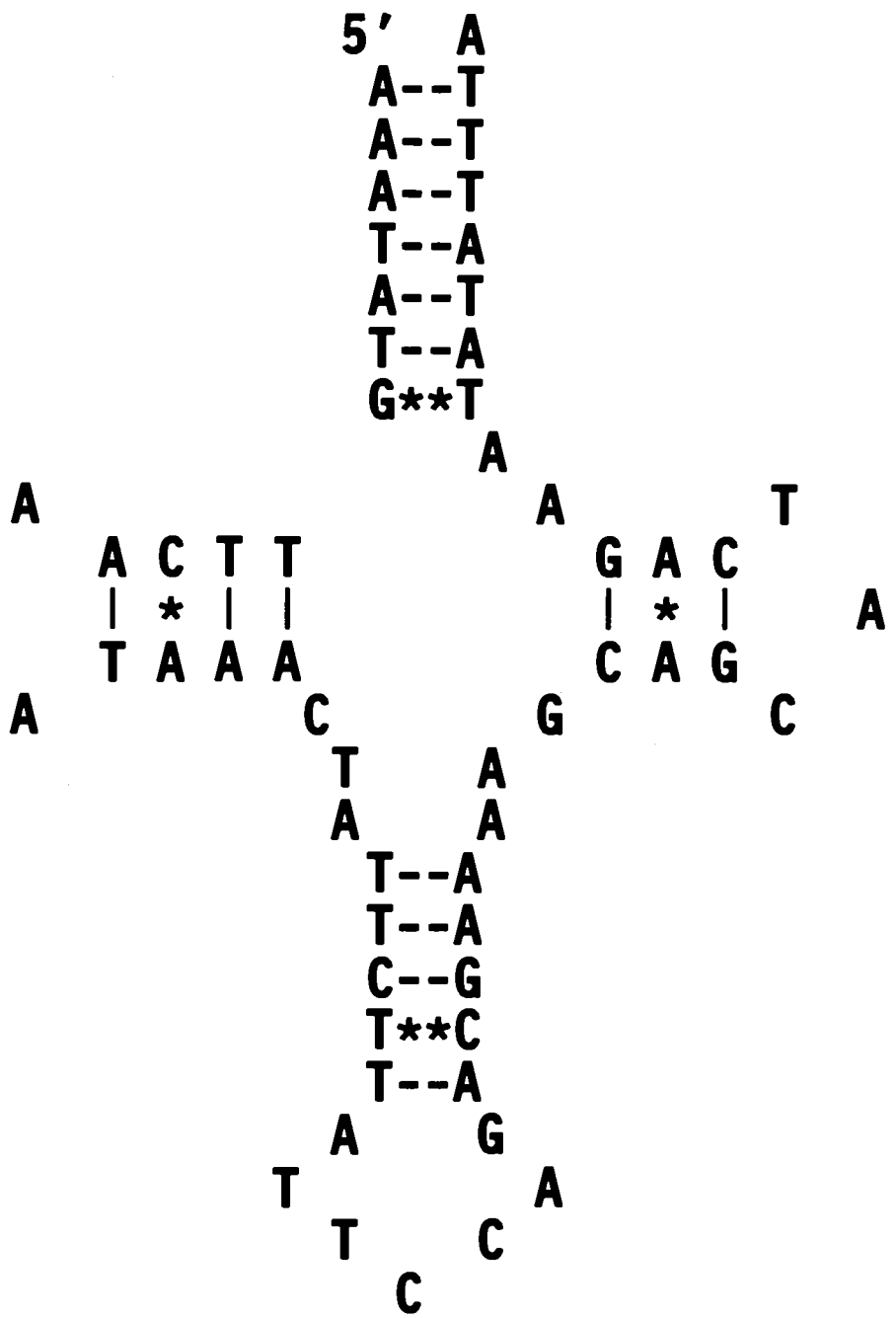
The predicted secondary structure of this tRNA gene is a conventional clover-leaf (Figure 22). As shown in Figure 21, this tRNA gene has an aminoacyl stem of 7 nucleotides, a DHU stem of 4 nucleotides, a DHU loop of only 2 nucleotides, an anticodon stem of 5 nucleotides, and an anticodon loop of 7 nucleotides. Unorthodox base-pairing is found in all four stems. These include one of G-T pair located on the aminoacyl stem, one of C-A pair in DHU stem, one of T-C pair found in the anticodon stem, and one of an apparent A-A pair observed in TΨC stem. For this tRNA gene, several major features can be observed. Firstly, this tRNA gene has the shortest length in comparison with the corresponding tRNA gene from other animal species. Secondly, the difference in size between the chaetognath and other animal species results from a deletion event in the DHU and TΨC loops. Thirdly, those bases lying between the aminoacyl stem and DHU stem of the corresponding tRNA gene from other animal species has been lost. It is interesting to note that there is a possible interaction between those nucleotides located between DHU stem and anticodon stem and variable loop (Figure 22).

The chaetognath mitochondrial tRNA gene is quite distinct in regard to various nucleotides that are usually conserved in prokaryotic and eukaryotic nuclear-encoded tRNAs. Only the conserved T33 and Pu37 nucleotides that lie immediately 5' and 3', respectively, to the anticodon are found in the $tRNA^{Ser}$ gene of the chaetognath. At the 8th position downstream from the 5' end of this gene there is a T residue. If this T is T8 like that of all tRNA genes, it would result in mispairing of all bases in the DHU stem. The structure shown in Figure 22 seems more likely. In this case, T8 is absent, but the DHU stem can be paired with three orthodox base pairs. The conserved nucleotide pair Py11-Pu24 in the DHU stem is found in this tRNA. As in the Phylum Vestimentifera, the

FIGURE 22.

Sequence of *Sagitta* (Phylum Chaetognatha) mitochondrial $tRNA^{Gly}$ gene represented in the cloverleaf form. Standard base-pairings(G-C or A-T) are indicated by dashes. Non-standard base-pairings are designated by asterisks.

$tRNA^{gly}$


5′ A
A--T
A--T
A--T
T--A
A--T
T--A
G**T
A
A A T
A C T T G A C
| * | | | * | A
T A A A C A G
A C G C
T A
A A
T--A
T--A
C--G
T**C
T--A
A G
T A
T C
C

conventional sequence T54-ψ-C-Pu-A is not found in the TψC loop of the $tRNA^{Pro}$ of the chaetognath.

Ribosomal RNA Gene

The sequence of length of 1,018 nucleotides coding the entire large ribosomal RNA gene of *Sagitta* (Phylum Chaetognatha) mitochondrial DNA was obtained (Figure 17A).

The chaetognath mitochondrial large ribosomal RNA gene was identified from similarities to the corresponding gene in vertebrate and invertebrate mitochondrial DNA. The relative position of the large ribosomal gene is different from that observed in other animals. In vertebrates, *Drosophila* and *Ridgeia* (vestimentiferan), the large ribosomal gene is preceded by the small ribosomal gene and a $tRNA^{Val}$. In the bivalve mollusc, *Mytilis*, seven tRNA genes separate the two ribosomal genes. Although the small ribosomal subunit has not been mapped in the chaetognath, it is evidently been separated from the large subunit by one or more protein genes. The separation of the two ribosomal genes by protein coding genes has been observed in sea urchin and nematode genomes.

Neither the 5' nor 3' termini of the large ribosomal RNA of the chaetognath mitochondria have been mapped directly. Since the $tRNA^{Pro}$ gene is located upstream of the large ribosomal RNA gene, the 5' termination of the large ribosomal RNA gene should follow the $tRNA^{Pro}$ supposing no non-coding intergenic nucleotides and no overlap between the $tRNA^{Pro}$ and the large ribosomal RNA genes. Similarly, assuming the large ribosomal RNA and the COI genes are butt-joined, the 3' termini of the large ribosomal RNA gene should end at a position upstream from the start codon ATG of the COI gene. Thus, the large ribosomal RNA gene appears to contain at least 1,018 nucleotides in chaetognath

and is smaller than those of most animal ribosomal RNA genes reported to date: 1,559 nucleotides in human, 1,525 in mouse, 1,571 in cow, 1,640 in frog, 1,326 in *Drosophila*, and 1,525 nucleotides in sea urchin. It is longer than those of two nematodes (953 and 960 nucleotides for *Caenorhabitis elegans* and *Ascaris suum* respectively). These differences reflect mainly the addition or deletion of entire domains, a common feature of ribosomal RNA evolution (Gray et al. 1984). The highly conserved regions are found throughout this gene and alternate with poorly conserved regions. The same pattern is found in all vertebrate and some invertebrate mtDNAs, suggesting that these constant regions are important for ribosomal RNA functions.

The base composition of the large ribosomal RNA gene in chaetognath is presented in Table 12. The chaetognath large ribosomal RNA gene shows a lower G+C content of 29%, compared with that of mammals (e.g. bovine, 39%), and some invertebrates (e.g. sea urchin, 38%, Phylum Vestimentifera, 35%), but higher than observed *Drosophila* (17%).

Analysis of the large ribosomal RNA and protein genes in the chaetognath revealed a number of significant examples of complementarity between ribosomal RNA and protein genes. In the large ribosomal RNA gene, I identified a number of interesting sites, some of them illustrated in Figure 39. At position 51 downstream from 5' end of the large rRNA gene, a sequence partially conserved compared to that of mouse is present, which is complementary to a segment of the cytochrome b protein gene. At position 235 downstream from 5' end of large ribosomal RNA gene, a 12 nucleotide sequence which is not located in a conserved region, is perfectly complementary to a sequence from the ND2 gene. Another sequence at 3' end of large ribosomal RNA gene, 12 nucleotides in length, is completely complementary to a region in the COIII protein gene. At position 4 downstream

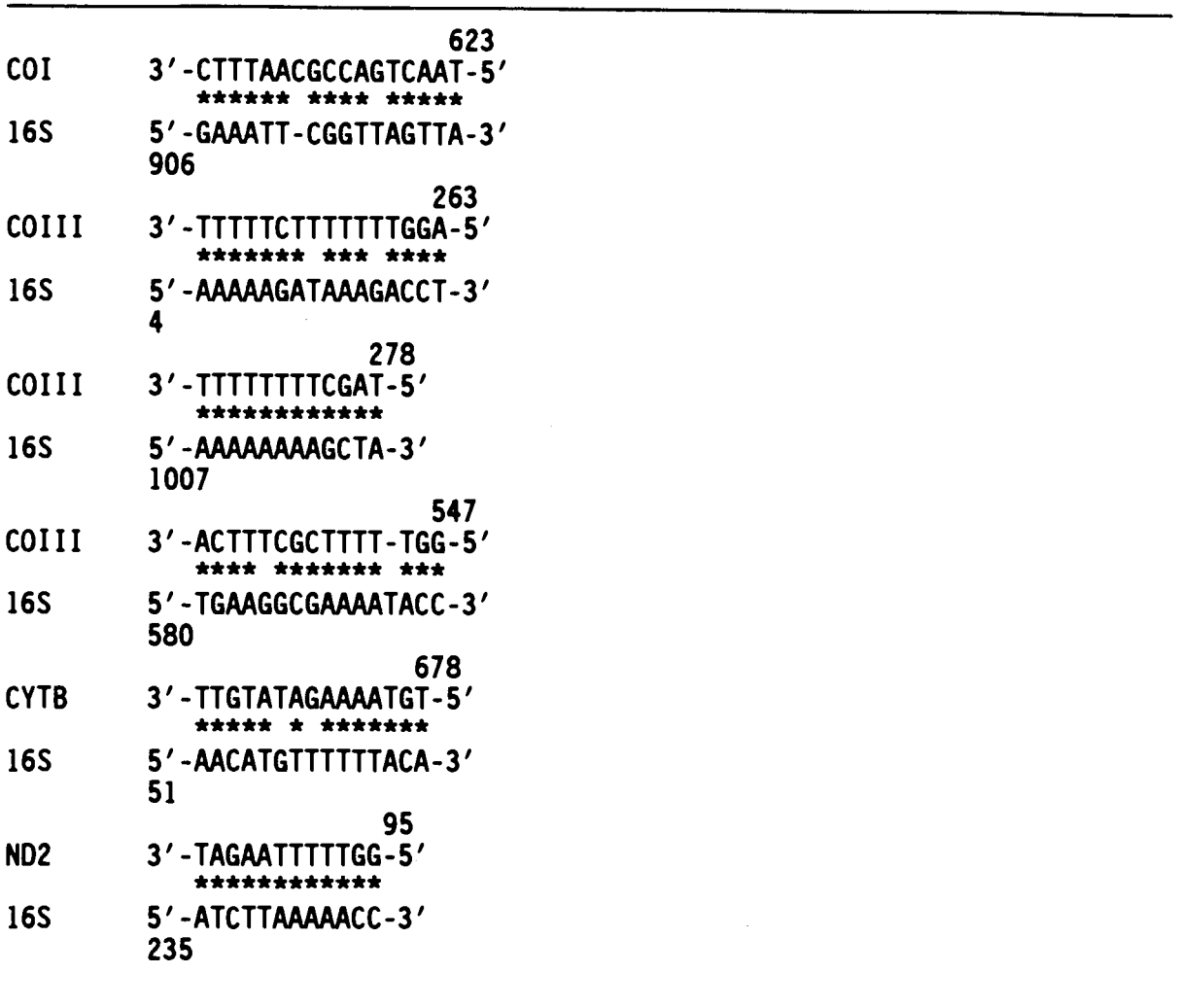


FIGURE 23. Complementary sequences found between the large ribosomal RNA and protein genes. Standard base pairs are designated by asterisks.

from the 5' end of large ribosomal RNA gene, 14 of a 16 base sequence is complementary to a sequence from the COIII protein gene. Another sequence complementary to the COI protein gene was found at position 906 downstream from 5' end of the large ribosomal RNA gene. As found in *Ridgeia* (Phylum Vestimentifera see above), some of these sequences complementary to protein genes are very conserved regions, while others are in less conserved regions in the large ribosomal RNA gene. The degree of complementarity of these sequences varies from gene to gene. The potential for interaction between ribosomal RNA and protein genes strongly suggests that they may serve important function in the regulation of the translation in the mitochondrial system.

Intergenic Sequences

Several intergenic sequences which are not assigned to structural genes have been found in chaetognath mitochondrial DNA (Figure 17). The length of these intergenic sequences varies from 23 to 49 base pairs. The first intergenic sequence, 23 base pairs long, is located between cytochrome B and $tRNA^{Asp}$. The second intergenic sequence, 48 nucleotides in length, lies between the COI and COII genes. The third intergenic sequence, 49 nucleotide long, is found between the ND2 and COIII genes. The intergenic sequence found between the ND2 and COIII genes contains a TATATAA sequence similar to the sequence TTATATATAA, which has been suggested as a bidirectional promoter (Jacobs et al., 1989). However, both ND2 and COIII genes are in the same transcriptional direction.

The analysis of base composition of the three intergenic sequences reveals that the frequency of T is much higher than that of A, G, and C. All three regions are AT-rich, more than 64%, while the content of C is the lowest.

Intergenic sequences have been observed in many mitochondrial genomes of invertebrates, such as *Drosophila* (Clary and Wolstenholme 1985; our unpublished data), sea urchin (Jacobs et al., 1988) and sea star (Smith et al., 1990). Some of these intergenic sequences have been assigned functions related to replication and transcription (Bibb et al. 1981; Roe et al. 1985). However, in most cases other than vertebrates the function of the intergenic sequences are poorly understood, as are the mechanisms of replication and transcription in these organelles.

The intergenic sequences have been searched for similarities with the conserved sequence blocks and other motifs of vertebrate mitochondrial D-loops. The only example found is a partial homology with D-loop of the frog. I have not identified any convincing potential stem-loop structure between these intergenic sequences and their adjacent regions nor in the intergenic sequences elsewhere in the genome. The ends of both of these intergenic sequences and the ends of those protein genes lack stable secondary structure potential. The significance of these intergenic sequences is not known yet.

Summary

Using PCR, two large fragments containing three functional classes of genes have been amplified and sequenced from *Sagitta elegans* (Phylum Chaetognatha).

Analysis of sequence of more than one third of the mitochondrial genome reveals that organization of the chaetognath mitochondrial genome is substantively different from those of other animals, including both vertebrates and invertebrates. The gene order of the first fragment is the cytochrome b, $tRNA^{Ser}$, large ribosomal RNA, COI and COII genes; the ND2 and COIII genes located in the second segment are contiguous. In the chaetognath, the

codons AGA and AGG specify serine, while TGA specifies tryptophan. The triplet ACA may be used for translation initiation codon of the cytochrome b gene from the chaetognath mtDNA.

Many insertion or deletion events ranging from a single codon to a 28 codon domain are observed in these five protein genes of the chaetognath. These protein genes are shorter than those of other animal species. The COI gene is the shortest among these species. Unlike those of most animals, many of the protein-coding sequences are contiguous with one another, rather than being flanked tRNA genes. Only one tRNA gene was found in more than one third of this mitochondrial genome. In all five protein genes, a number of potential secondary structures have been found. While they may be functional in the mitochondrial system, none appear to be associated with transcript processing.

The large ribosomal RNA gene of the chaetognath contains only 1,018 nucleotides, one third shorter than those of vertebrates and some invertebrates. A number of deletion events are appearent between the chaetognath and mouse. Sequence analysis revealed a number of complementary examples between the large ribosomal RNA and protein genes. This suggests that they might play an important role in the regulation of mitochondrial translation and transcription mechanisms.

Several intergenic sequences have been found in the chaetognath mitochondrial DNA. I have not identified any convincing potential stem-loop structure between these intergenic sequences and their adjacent regions nor in intergenic sequences elsewhere in the genome. The function of these intergenic sequences is not known yet.

SECTION IV

EVOLUTION OF MITOCHONDRIAL DNA FROM *SAGITTA* (PHYLUM CHAETOGNATHA)

INTRODUCTION

In recent years the mitochondrial DNA molecule has attracted great interest for determining evolutionary relationships at various levels. The advantages of using mtDNA include ease of extraction and manipulation, the simplicity of the molecule, lack of recombination, and high mutation rate.

The rate of animal mtDNA evolution varies among lineages, among genes, and within genes. The nucleotide sequences of mammalian mtDNA evolve much more rapidly than comparable regions encoded in the nuclear genome (Brown 1981). Estimates provided by Brown et al. (1981; 1982) suggest that the rate of base substitution (transitions and transversions) in mammalian mtDNA is 5-10 times greater than that of single-copy nuclear DNA. This rapid rate of base substitution quickly leads to the accumulation of parallel and back mutations, particularly among the more distantly related taxa (Brown et al. 1982). In contrast, Powell et al. (1986) and Solignac et al. (1986) found that in *Drosophila* the average rate of evolution of nuclear and mitochondrial DNAs were similar.

The observed rapid average rate of evolution of the mtDNA molecule is what gives these data the potential for use in species and population level studies, yet this view of mtDNA obscures probably the most useful aspect of the molecule. That is, it contains slowly evolving "highly conserved" regions as well as rapidly evolving "highly variable" regions (Cann et al. 1987; Carr et al. 1987). Conserved regions are useful for studies of distantly related taxa. It is only in conserved regions that distantly related taxa can be aligned. In protein coding regions, identification of amino acid codons, in which second positions are highly conserved, aid in alignment. In ribosomal RNA gene, structurally or functionally important conserved sequence blocks can serve as markers between which the more variable

regions can be aligned with differing degrees of confidence (Olsen 1988).

Sequence comparisons in protein coding regions have proven less useful for analysis of distant relationships. In vertebrates, Kocher et al. (1989) found evidence which suggests that there may be a five-fold higher rate of amino acid substitution in mammals and birds relative to fishes. Thomas and Beckenbach (1989) sequenced of more than 2000 base pairs in six salmonid fish species and found high amino acid similarity with the corresponding genes of the frog *Xenopus*. It is obvious that different genes are evolving at different rates. Bibb et al. (1981) compared mouse mtDNA to that of humans and Anderson et al. (1982) compared cow mtDNA to human mtDNA. In these examples, protein coding genes varied in sequence similarity between approximately 60% and 80%.

Variation in the rate of mtDNA evolution within genes is also striking. Brown et al. (1982) found transitions to outnumber transversions 9:1 in a comparison of an 896 base pairs segment containing genes for three tRNAs and parts of two proteins (ND4 and ND5) for human and chimps. In comparisons of mitochondrial DNA of the small ribosomal subunit in these same primates, Hixson and Brown (1986) found similar transition/transversion ratios. For the more distantly related pair *Drosophila melanogaster* and *Drosophila yakuba*, transitions accounted for only 18% of changes at the third codon positions (6 protein coding genes) (Wolstenholme and Clary 1985). In comparisons of nematode strains for the COII gene, transitions outnumbered transversions 12/1. For comparisons among species, the observed substitutional pattern showed transitions only slightly more common that transversions (1.3:1). In comparisons of two nematode genera, transversions outnumber transitions 1.5:1 (Thomas and Wilson 1991). Brown et al. (1982) suggested that the high proportion of transitions in their primate comparisons was due chiefly to a bias in the

mutation process, selection at the molecular level, rather than to selection at the level of gene products. Wilson et al. (1985) reiterated this point and suggested that transitions are selected for by tautomeric base pairing that allows purine-pyrimidine mismatches and that these mismatches are rarely if ever corrected due to deficient mtDNA repair system.

Using data from comparing gene organization and various functional genes, this study has demonstrated certain evolutionary features of the *Sagitta elegans* (Phylum Chaetognatha) mitochondrial DNA. Many rearrangements have occurred between the chaetognath and other phyla. The divergence from the vertebrate lineage has relocated the large ribosomal RNA gene adjacent to the structure gene for COI. The cytochrome b gene has been rearranged to the region upstream the 5' end of the large ribosomal RNA gene. Also, the ND2 and COIII genes are contiguous.

RESULTS AND DISCUSSION

Gene Rearrangements

Although the complete sequence of the *Sagitta* mitochondrial genome was not determined, analysis of about one third of this genome (5,183 nucleotides) revealed that the organization of this genome has some significant features different from those of vertebrates and other invertebrates. The first fragment contains the cytochrome b, $tRNA^{tyr}$, large ribosomal RNA, COI, and COII genes (Figure 17A). The second segment codes for the ND2 and COIII genes (Figure 17B). These features of gene organization in the chaetognath mitochondrial genome are different from those of all other animal mtDNAs reported to date (see above).

The various protein, ribosomal RNA and tRNA genes have identical arrangements in amphibian, bird and mammalian mtDNAs, except that in bird mtDNA, the segments of the molecule comprising the cytochrome 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 *Drosophila yakuba* (Clary and Wolstenholme 1985), and sea urchin (Jacobs et al. 1988; Cantatore et al. 1989) mtDNA molecules relative to vertebrate mtDNAs. Smith et al. (1989) demonstrated a major inversion in the mtDNA of the sea star versus the sea urchin. This indicates that the inversion event must have followed an earlier gene rearrangement that resulted in the distinction between vertebrate mtDNA gene order and that of the urchins. There is little similarity between the gene arrangements of nematodes and those found in vertebrate, *Drosophila yakuba*, sea urchin, and sea star mtDNAs. The

differences between vertebrate and *Drosophila* mtDNAs can be accounted for by three major inversions and transposition of 11 tRNA genes; but between mtDNAs of nematodes and *Drosophila*, and nematodes and mammals, extensive rearrangements involving almost all tRNA, rRNA, and protein genes have occurred. The most notable exception is that the COI and COII genes are adjacent to each other and transcribed in the same direction in two nematode mtDNAs, as they are in *Drosophila yakuba* and vertebrate mtDNAs. In the chaetognath extensive differences in gene rearrangement are evident. The two ribosomal RNA genes are separated by the protein genes, although the small ribosomal RNA gene has not been found. In all vertebrates and *Drosophila*, the gene order for the region containing the COI, COII genes and the 5' flanking area of the COI gene is similar, that is ND2, tRNA(s), COI, and COII (Figure 18D). Comparison of the first fragment with the gene order in the corresponding sea urchin mitochondrial genome shows two major differences. The first difference is that in the chaetognath, the COI and COII genes are adjacent, and in sea urchin, these two genes were separated by ND4L and one tRNA genes. The second difference is that in the chaetognath, the cytochrome b and one tRNA genes were located the flanking region of the 5' end of the large ribosomal RNA gene, in sea urchin the ND2 gene was found in the 5' flanking region of the large ribosomal RNA gene (Figure 18B). In sea star, the ND1 and a cluster of 13 tRNA genes were located the 5' flanking region of the COI gene. Detailed analysis of these orders of genes of corresponding region revealed that in all vertebrates and some invertebrates including sea urchin, *Drosophila*, and nematodes, the corresponding protein and ribosomal RNA genes are transcribed in the same direction (Figure 18). In the chaetognath, the cytochrome b gene is transcribed in a direction opposite to that of the other major genes on the fragment. Only one tRNA gene

($tRNA^{Ser}$) was found in more than one third of the genome from the chaetognath mitochondrial DNA and the majority of the protein genes and the ribosomal gene are adjacent. This finding implies that those tRNA genes must be concentrated in some region in the genome and may be clustered, and that they apparently do not serve as recognition signals for mitochondrial RNA processing in the chaetognath.

A novel location for the large ribosomal RNA gene and tRNA genes has potential implications for the transcriptional regulation of the genome. It has been proposed that, in vertebrates, the relative rates of synthesis of mitochondrial ribosomal and messenger RNAs are controlled by the selective use of a transcriptional attenuator located just beyond the ribosomal RNA genes (Montoya et al. 1983; Attardi et al. 1983). The separation of the ribosomal RNA genes in the chaetognath would require another, more complex mechanism to achieve this end.

Protein-Coding Genes

The comparisons of DNA and predicted amino acid sequences of the chaetognath COI gene and parts of COII, COIII, cytochrome b and ND2 genes for these genes with those of *Drosophila yakuba*, mouse, frog, sea urchin and nematode are presented in Figures 19 and 20. The comparisons show that a number of insertion and deletion events have occurred. In the COI gene, at least five codons have been lost within this gene in addition to deletion events occurring at the 3' end. The COI gene is shorter than those of all vertebrate and invertebrate mtDNAs reported to date. In the COII gene, a total of seven codons have been deleted and two codons inserted in a region that includes only about half the gene (Figure 20). In the COIII gene, a one codon insertion and three codon deletions have

occurred in a region including about 90% of the gene. In the cytochrome b gene, a one codon insertion at the second codon position relative to the sequences of this gene in mammals, occurs in the chaetognath and is present in *Drosophila yakuba*, frog, and sea urchin. The cytochrome b gene of *Sagitta* is particularly interesting. The gene ranges in length from 370 codons in *C. elegans* to about 381 in mammals. The sequence examined here for the chaetognath represents more than two-thirds of the gene, from the 5' end. The inferred amino acid sequence aligns with those of *D. yakuba*, *Xenopus* and sea urchin without introducing gaps, up to codon position 246 (nucleotide 737), with amino acid similarities of 69%, 63% and 58% respectively (Figure 20). All similarity appears to be lost in the last 13 codons. Introduction of a four-base gap at nucleotide position 737 (Figure 19E), shifting the reading frame, restores the alignment (Figures 19 and 20). There are several possible explanations for an observed shift of reading frame in an essential gene. There may have been a very recent (perhaps tandem) duplication of the gene, such that this copy is no longer functional; the reading frame may be corrected by some form of RNA editing; or the reading frame may be restored at a point more than 13 codons downstream from this point. Evidence currently available does not permit choosing between these alternatives. In contrast, at least 54 amino acids of deletions which were located in three domains have been observed in the partial sequence of the ND2 gene. These deletions make the *Sagitta* ND2 sequence much shorter than those of all six species which are 341 to 347 amino acids long. It is more likely that the length of the ND2 gene may like that of nematodes which is 281 or 282 amino acids in length (Okimoto et al. 1992).

The extremely infrequent use of G at third position has been reported for all vertebrate mtDNA. In some invertebrates, such as sea urchin, sea star and the vestimentiferan (see

above), this extremely low incidence of G at terminal position has also been observed. However, this situation has not been observed in the Phylum Chaetognatha. As shown in Table 15, in all five protein genes, the high frequency of T at the third position has been found. In contrast to the mtDNA of most other species, both C and G are relatively frequent each making up about 20% of the third codon positions.

As shown in Table 13, the COI gene shows the highest similarity for amino acid sequence among all five protein genes. Indeed, this gene is thought as one of the most conserved mitochondrial protein genes. Comparison of this gene with its counterparts mouse, frog, *Drosophila yakuba*, sea urchin, and nematode reveals that the divergent regions are concentrated in several blocks besides those of single or few amino acid changes (Figure 20A). The longest divergent region was found at the 3' end of this gene, which was located from codon 456 to the end of the gene. Of a total of 50 amino acid codons, only 13 amino acids are identical to *Drosophila* and only three are conserved among all six taxa. Other divergent blocks were located within this gene, including blocks at codon positions 132, 170, 253, 327, and 402 downstream from the 5' end. In all cases, almost none of these amino acids of the chaetognath is same as those of the other five species. It is relevant to note that many conserved amino acids in the COI gene of the chaetognath are altered. For example, single amino acid changes were found in many sites, and blocks containing several amino acid changes were also observed in this gene (Figure 20A). At position 20 downstream from the 5' end, a leucine in the chaetognath, replaces a conserved phenylalanine found in corresponding sites of the COI gene of the other five species. Other sites conserved in the other five species include glycines at positions 25, 86, 260 and 278, proline at position 42, serine at position 140 and glutamate at position 263, have all

Table 15. Base composition at 3rd codon positions in *Sagitta* mitochondrial protein genes

	Base composition % of total			
Gene	Guanine	Adenine	Thymine	Cytosine
ND2	21.8	20.0	41.4	16.8
COI	19.1	23.4	39.6	17.9
COII	14.7	26.7	42.2	16.4
COIII	19.3	20.6	36.8	23.3
CytB	17.4	25.2	39.2	18.2

undergone replacements in the chaetognath. A block including three amino acids, serine, phenylalanine and serine, is located at position 443 downstream from the 5' end of the COI gene, where three different amino acids, alanine, tyrosine and threonine found in corresponding sites of four of the five species compared. Similarly, amino acid replacements have also been observed in the other four protein coding genes, COII, COIII, Cytochrome b, and ND2 (Figure 20). These observations suggest that many amino acids which are completely conserved in most animal mitochondria, have been subjected to replacements in the chaetognath.

In contrast to the conserved genes, such as COI, COII, COIII, and Cytochrome b, the ND2 gene shows the most divergent of amino acid sequence compared to those of other phyla. This results in the lowest similarity of amino acid sequence between the chaetognath and most other species (Table 13). Further comparison of the ND2 with its counterparts reveals that only few conserved amino acid blocks were observed throughout this gene (Figure 20B). For example, the first of these blocks, including three codons was found at codon position 49 downstream from the 5' end. The second of these blocks, containing three codons was located at codon position 62. Only two blocks with two amino acids were found at positions 112 and 162 downstream from the 5' end. A block located at codon position 155 has five amino acids conserved in five of six taxa, is the longest conserved region in the ND2 gene. As observed in the COI gene, many sites including single and multi-amino acid blocks which are conserved in the other five species have replacements in the chaetognath.

The COI gene is the only entire protein gene sequenced. It is also the most conserved gene of mitochondrial genome. In order to obtain the nature and frequency of specific base

substitutions, detailed analysis of the COI gene with its counterparts is presented in Table 16A. In this gene, the number of transitions ranges from 214 for *Drosophila yakuba* up to 272 for mouse. In all cases, transitions are less frequent than transversions. The numbers of A-G transitions are relatively uniform in all five comparisons, so much of the difference in total transitions results from the number of C-T substitutions. The number of C-T transitions between *Sagitta* and *Drosophila yakuba* is approximately same as that of A-G transitions and is the lowest among all comparisons. In all cases, more than 60-65% of total transversions are A-C and A-T transversions, and the number of A-T transversions is twofold higher than those of A-C transversions. The G-C and G-T transversions show the lowest values. The lowest number of G-C transversions is that between *Sagitta* and *Drosophila yakuba*, which is only 10% of total substitutions. The frequency of G-T transversions between the chaetognath and mouse is the lowest, only 11% of total base substitutions.

Corrected divergences (Jukes and Cantor 1969) for comparisons between the chaetognath and other taxa are shown in Table 16 and are lowest for the COI gene (45-57%) and highest for ND2 (Table 16B). Divergences between *Sagitta* and *Drosophila* are slightly lower than with other taxa, but all divergences are quite high.

Table 17 summarizes the base substitutions for all protein coding genes. The overall comparisons include almost 4,000 sites. Transversions comprise about 60% of all substitutions. Divergence estimates range from about 56% (*Sagitta-Drosophila*) to over 76% (*Sagitta*-nematode).

Table 16. Nature and frequency of specific base substitutions in five protein coding genes between *Sagitta*, a chaetognath, and *Drosophila*, mouse, frog, sea urchin and nematode

A. COI

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	102	120	108	110	124
C--T	112	152	163	153	116
Total	214	272	271	263	240
Transversions					
A--C	49	70	69	69	50
A--T	151	139	130	132	177
G--C	36	56	57	67	49
G--T	68	52	66	65	91
Total	304	317	322	333	367
Indels	4	4	4	4	4
Number of sites	1525	1528	1528	1528	1528
Total changes	522	593	597	600	611
% changes	34.2	38.8	39.1	39.3	40.0

Table 16. Nature and frequency of specific base substitutions in five protein coding genes between *Sagitta*, a chaetognath, and *Drosophila*, mouse, frog, sea urchin and nematode

B. ND2

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	68	73	64	53	53
C--T	66	101	86	75	61
Total	134	174	150	128	114
Transversions					
A--C	42	52	53	47	32
A--T	122	89	93	88	95
G--C	19	44	48	46	22
G--T	70	41	56	77	79
Total	253	226	250	258	228
Indels	7	7	7	8	13
Number of sites	664	664	664	665	565
Total changes	394	407	407	394	355
% changes	59.3	61.3	61.3	59.2	62.8

Table 16. Nature and frequency of specific base substitutions in five protein coding genes between *Sagitta*, a chaetognath, and *Drosophila*, mouse, frog, sea urchin and nematode

C. COIII

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	55	46	48	56	57
C--T	56	78	63	70	67
Total	111	124	111	126	124
Transversions					
A--C	40	40	46	50	39
A--T	88	71	78	63	88
G--C	19	38	36	35	32
G--T	41	30	30	34	67
Total	188	179	190	182	226
Indels	1	2	2	2	4
Number of sites	687	685	685	685	675
Total changes	300	305	303	310	354
% changes	43.7	44.5	44.2	45.3	52.4

Table 16. Nature and frequency of specific base substitutions in five protein coding genes between *Sagitta*, a chaetognath, and *Drosophila*, mouse, frog, sea urchin and nematode

D. COII

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	24	24	28	26	24
C--T	25	39	38	43	32
Total	49	63	66	69	56
Transversions					
A--C	11	17	20	19	20
A--T	33	38	31	30	53
G--C	7	13	12	14	13
G--T	19	14	14	18	38
Total	70	82	77	81	124
Indels	4	4	4	4	5
Number of sites	346	346	346	346	341
Total changes	123	149	147	154	185
% changes	35.5	43.1	42.5	44.5	54.3

Table 16. Nature and frequency of specific base substitutions in five protein coding genes between *Sagitta*, a chaetognath, and *Drosophila*, mouse, frog, sea urchin and nematode

E. Cyt B

	Species compared				
Type of changes	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	47	50	44	51	64
C--T	52	70	87	82	64
Total	99	120	131	133	128
Transversions					
A--C	25	41	38	39	38
A--T	70	63	71	62	95
G--C	16	25	25	36	21
G--T	32	27	40	40	60
Total	143	156	174	177	214
Indels	0	1	0	0	2
Number of sites	737	735	737	737	724
Total changes	242	277	305	310	344
% changes	32.8	37.7	41.4	42.1	47.5

Table 17. Summary of mutation types for protein coding genes, between *Sagitta* and members of five other taxa

Type of changes	Species compared				
	Drosophila	mouse	frog	urchin	nematode
Transitions					
A--G	296	313	292	296	322
C--T	311	440	437	423	340
Total	607	753	729	719	662
Transversions					
A--C	167	220	226	224	179
A--T	464	400	403	375	508
G--C	97	176	178	198	137
G--T	230	164	206	234	335
Total	958	960	1013	1031	1159
Indels	16	18	17	18	28
Number of sites	3959	3958	3960	3961	3833
Total changes	1581	1731	1759	1768	1849
% changes	39.9	43.7	44.4	44.6	48.2
Jukes-Cantor Divergence	56.5	65.0	66.7	67.2	76.2

Ribosomal RNA Gene

The complete sequence of 1,018 nucleotides for the large ribosomal RNA gene in the chaetognath mitochondria was determined (Figure 17A). This gene is much smaller than that of all vertebrate (Anderson et al. 1981; Anderson et al. 1982) and some invertebrate large ribosomal RNA genes reported to date and slightly greater than that of two nematodes (Okimoto et al. 1992). It is clear that the large ribosomal RNA gene is also smaller than that of the vestimentiferan although only a partial large ribosomal RNA sequence was obtained. The entire large RNA gene of the chaetognath is only 1,018 nucleotides long, but the partial sequence from the vestimentiferan is 1,117 nucleotides. Correct alignment is the most important step in nucleotide sequence data preparation for phylogenetic analysis regardless of method; it is equivalent to assigning character homologies. High levels of multiple substitutions between two taxa coupled with any degree of insertions or deletions will make DNA sequences difficult to align. Because of a number of deletions in the chaetognath large ribosomal RNA gene, it indeed is difficult to align among some species. Here, I tried to align only between two species, chaetognath and mouse. Alignment of the chaetognath large ribosomal RNA gene with its mouse counterpart is presented in Figure 24. The major difference in size between chaetognath and mouse mitochondrial ribosomal RNA gene results from the absence in the chaetognath rRNA gene of specific blocks. As shown in Figure 24, there are at least 22 deletion events, besides those of single or less than three nucleotides, in the chaetognath mitochondrial large ribosomal RNA gene. Although a small number of additions have been found between the chaetognath and mouse, the majority of changes are deletions. A total of 40 nucleotides of insertion are observed, more than half located near the 3' end of this gene. In contrast, a total of 604 nucleotides of

FIGURE 24.

Nucleotide sequence of *Sagitta* (Phylum Chaetognatha) mitochondrial 16S rRNA, aligned with its counterpart mouse. The top shows the complete sequence of large ribosomal RNA gene from *Sagitta*, the bottle is that of mouse. Gaps are indicated by dashes, residues homologous with those of Phylum Chaetognatha by dots. The positions for deletions ranging from single to three nucleotides are shown by asterisks, above the sequence. The positions for deletions more than three nucleotides are indicated by using DEL.+numbers. The positions for all insertions are shown by letter X.

```
DEL.1
----TTGAAAAAGATAAAGACCTGTTAAGTGATTGTAAATCTAGTTTAACGTCAAACATG 56
ACTAA.CCT.GCCC..GCCCTACACA..TAT.A.TAT.C.A.TA.A...ATCA......T

DEL.2                              X    *                    DEL.3
----------TTTTTTACACAGACATAAACTTC-AACCGGTGCATTCTAT---------- 95
TATCCTACTAAAAG.ATTGG...A.G...-...GT..ATC.AGGAG....AGAACTAGTA

DEL.3
------------------------------------------------------------
CCGCAAGGGAAAGATGAAAGACTAATTAAAAGTAAGAACAAGCAAAGATTAAACCTTGTA

                   XXX         DEL.4                DEL.5
CCTTTTGTATAATGAA---AGTAGGAT-------------AGAATT-------------- 125
.......C........CTA.C...A.AACTTCTAACTAAA......ACAGCTAGAAACCC

DEL.5                                                    DEL.6
-----------------------------------------------------TGTGG------ 130
CGAAACCAAACGAGCTACCTAAAAACAATTTTATGAATCAACTCGTCTA.....CAAAAT

                   DEL.7                                 **
----GATCTCTGTTTA-------AAGGATACATTAAATATTATAATGA--ACAGGTAGGA 177
AGTGAGAAGA.T....GGTAGAGGT.A.A.GCC...CG.GCT.GG...TAG.T...TACC

                        **                          DEL.8
CAAAAATTGTATATCGACCTC--AACAAACTAAATATACAAACAAGAAA----GAAA--- 228
......A..A..T.AAGTTCAATTTT.....TGC..A.A......C...ATCAA...GTA

DEL.9                   DEL.10                            DEL.
-----AGCTTAATCTTAAAAAC--------CTAGGTAAAACGATCTAGACCTTTAA---- 271
AGTTT..A...TAGCC....GAGGGACAGCTCTTC.GG....GAAA.A........TAGT

11                 ***        *  XXXX                      X
---GATATACTTATTTA---TTTAACC-TTCATATTAGGTTTATAACATTA-ACCATTTA 323
GAAT.AT..ACA.AAC.GCT.......A..----G....CC..A..GCAGCC....-A..


CAATATAGGATTTAATATAATTAATAGTTTCGATATAGCATCTACTTTGAAATAAAAATT 383
A.GA.AGC.T.CA.GCTC..CAT.A.A....A..TA.TTCCA..A...AC.CC..CTTCC

           *
TAAAAGTAAAAT-AAAATGGACTTCCGAAAGTTTCACAAAAACCTTTCCTTATAATATAC 442
....CT......TGGGT.AAT..ATAACTTTA.AG.TGC..CA..G.TAG...G.GTA..

                      *
AAGGTACACCCTGCTAAAGAAA-AGATGTAGCATAATTAATAGCCACTTTATTGGTGGAG 501
...AATTC.AA.T..CC...C.T.CGC...TA.C..C.CGG.TAAC.A..G..A..TA.T
```

```
                                                DEL.12
AGTCTGAAAGGGCACTCGAGAGGTTTCAGTATTACGTAT------AAAAAAAAATCTATA 555
CAGACT.T...CA.TAATCACAC.A.A.A..A.C.ACC.ATAACTTCTCTGTT.A.CCA.

                                      DEL.13
TATATTTATATCTAAAAAAAATATTGAAGG------------------------------ 585
C.CCGGA..GC.....GG...G..CC..AAAGATAAAAGGAACTCGGCAAACAAGAACCC

DEL.13              ***        DEL.14                 DEL.15
-----------CGAAAAT---ACCTCTA------CAATGAATAAAGGC------------ 613
CGCCTGTTTAC.A....CATC.......GCATTA...GT.T..G....ACTGCCTGCCCA

DEL.15                                    *       *
------------------------GATAAGACCC-TAAAAG-CTCTATCAATAGTTGT- 645
GTGACTAAAGTTTAACGGCCGCGGTATCCT....GTGC....GTAGC..A.TC.C....T

DEL.16       *       DEL.17
-------TGGGG-CAACGA------------------------------------------ 656
CCTTAAT.A...A.T.GC.TGAACGGCTAAACGAGGGTCCAACTGTCTCTTATCTTTAAT

DEL.17                                              DEL.18
-----------------------------GAAATATAACATT----------------- 669
CAGTGAAATTGACCTTTCAGTGAAGAGGCT.........T.A.AAGACGAGAAGACCCTA

DEL.18                                DEL.19
------------TCTTTCTTCTAGCTATTTA------------------------------ 688
TGGAGCTTAAAT.A.A.AACT..T.......ATTTATTAAACCTAATGGCCCAAAAACTA

DEL.19
------------------------------------------------------------
TAGTATAAGTTTGAAATTTCGGTTGGGGTGACCTCGGAGAATAAAAAATCCTCCGAATGA

DEL.19
------------------------------------------------------------
TTATAACCTAGACTTACAAGTCAAAGTAAAATCAACATATCTTATTGACCCAGATATATT

DEL.19         *                                  X        XX     **
----------TACT-AGTTACTTTAGGGATAACAGCGCAATATCCATTAACAAGA--TCA 735
TTGATCAACGG..CA......CC..................-C.T...T--....GT...

            X        **                     XX          DEL.20
TATTTATAATGGAGAGTGT--GACCTCGATGTTGAACTAAAGGAAAAC----TAGAGTAT 789
...CG.C...T-..G..T.AC.............G.TC--....C.T.CCAA.G.T...G

        *
ACACTAGT-GTCTATCGTCTGTTCGCCGATTTTACCTTTACGTGATTTGAGTTCAAACCG 848
.AG...T.AA.GGT....T.....AA.....AA.GTCC........C........G....
```

```
                                                                    171

                                  X                 XXXXX
GTGTAAACCAGGTTGGTTTCAACCTTCTTTAATATTATTATTTGGATAGTACGAAAGGAA 908
.A.C..T......C......T.T..A-....CG...TC.CCC-----............C


  DEL.21          **        XX     DEL.22               XXXXXXXXX
AT------TCGGTTAG--TTACTGATTTAAG----TTATTTTTAATATTTTCTTGTCGAG 956
.AGAGAAA.A.AGCCACC....AA.--....CGCTC.CAAC.....T.A.----------.

                                                          XXXXXXXXX
AAAAAGAACTTCTAATTCTGCGTATATAGTTATCAAAATATTTTTCTTGTAAAAAAAAGC 1016
..T..A.T..AAAT.AAA.ATA..CG..CACCCTCT..CC.AGAGAAG..T---------

XX
TA                                                                1018
--
```

deletion are required, and these domains are located in many regions throughout the entire gene. A domain of 159 nucleotides in length, located at nucleotide position 688 downstream from the 5' end is the longest of deletion between the chaetognath and mouse. As a result of deletion events, the large ribosomal RNA gene is much shorter than that of vertebrates and some invertebrates. In fact, more than one third of entire sequence coding for large ribosomal RNA gene has been lost, compared to the corresponding gene in vertebrates and some invertebrates. It is approximately 564, 622, 507, and 308 nucleotides shorter than those of corresponding large ribosomal RNA gene from mouse, frog, sea urchin, and *Drosophila yakuba*. Evidently the overall small size of the chaetognath large ribosomal gene can serve the same function as sequences in other animal species. This implies that there is a sequence limitation for the ribosomal RNA genes, which defines a universal core structure presumably essential for ribosomal functions.

As in all genes, the level of variability of particular rRNA gene segments is related to structural and functional constraints. For example, one of the most conserved segments in the small ribosomal RNA gene codes for the site of tRNA attachment (Simon et al. 1990). In spite of the small size, a number of deletions, and low G+C content of the chaetognath mitochondrial large ribosomal RNA gene, many conserved domains occur throughout the entire length of this gene between the chaetognath and mouse (Figure 24). These conserved regions may be important for assembly into the secondary structure required for this large ribosomal RNA.

Although ribosomal RNA genes in general have been characterized as highly conserved, data from members of the phyla Vestimentifera and Chaetognatha show that the entire ribosomal gene is not highly conserved, rather, it is a mosaic of alternating conserved and

variable segments.

Summary

Analysis of one third of the genome reveals that gene organization in the chaetognath mitochondrial DNA shows the considerable difference from that of other animal mtDNAs reported to date. This study has demonstrated that many rearrangements have occurred between the Phylum Chaetognatha and other phyla. The divergence from the vertebrate lineage has relocated the large ribosomal RNA gene adjacent to the structure gene for COI. The cytochrome b gene has been relocated to the region upstream the 5' end of the large ribosomal RNA gene.

Both insertion and deletion events have been observed in the chaetognath mitochondrial protein genes. In general, the protein genes are shorter than those of other animal species. The number of insertion or deletion change varies from gene to gene. It is found that the divergent regions were concentrated in several blocks, in addition to those involving single or a few amino acid codons in the COI gene. In contrast, only a few conserved amino acid blocks were observed in the ND2 gene. As observed in the vestimentiferan, a number of single and several amino acid mutations have also been found in the chaetognath mitochondrial protein genes. Many amino acids which are completely conserved in the corresponding protein genes of vertebrate and some invertebrate mtDNAs, have undergone replacements in the chaetognath during evolution. The nature and frequency of specific base substitutions for the COI gene from the chaetognath mtDNA was also determined. More than half of the detectable base substitutions are transversions, and more than 60-65% of total transversions are A-C and A-T transversions. The number of A-T

transversions is twofold higher than those of A-C transversions, while the G-C and G-T transversions show the lowest values.

A number of deletion events in large ribosomal RNA gene of the chaetognath mtDNA result in a reduction in total length of the large rRNA gene relative to those of vertebrates and some invertebrates. This difference reflects mainly the addition or deletion of entire domains, a common feature of rRNA evolution.

CONCLUSIONS

Based on investigation of sequences coding various functional genes from several animal species mitochondrial genomes, I have developed a suite of primers which have been successfully used for studying the mitochondrial genomes from two invertebrate Phyla: Vestimentifera and Chaetognatha. Using these primers, several fragments containing three functional classes of genes from both phyla, Vestimentifera and Chaetognatha have been amplified with PCR and sequenced.

Analysis of sequences representing about one third of each genome from both phyla mitochondrial DNAs reveals that the gene organization, lengths of genes, genetic code and the overall codon usage of both mitochondrial genomes show several differences from those of other animal species. Firstly, both genomes exhibit novel gene orders. The ATPase 6 gene is not located between the COII and COIII genes, but is replaced by insertion of two tRNA genes in the Phylum Vestimentifera. In the Phylum Chaetognatha, the gene order of the first fragment is cytochrome b, $tRNA^{Met}$, large ribosomal RNA, COI, and COII genes; the ND2 and COIII genes are located in the second segment and are contiguous. This gene organization shows the extensive differences from those of vertebrates and other invertebrates. Secondly, in both mtDNAs, AGA and AGG specify serine, while TGA specifies tryptophan. Thirdly, the codon ACA was found as the translation initiation codon of the cytochrome b gene from the chaetognath. The COI gene of the vestimentiferan mtDNA does not encode a complete stop codon but rather end with T or TA.

Many insertion or deletion events have been observed in these protein genes from both phyla. Thus these protein genes are shorter than those of other animal species. In all protein genes from both mtDNAs, a number of potential secondary structures have been

found, some of which may be functional in the mitochondrial system.

Five tRNA genes of the vestimentiferan and one tRNA gene of the chaetognath have been identified. Each of them can be folded into the four-arm cloverleaf structure. There are the highest average similarities for all five tRNA sequences between the vestimentiferan and sea urchin.

A number of deletion events in the ribosomal RNA genes from both mitochondrial genomes result in a reduction in length of these rRNAs compared to those of other animal species. Detailed sequence analysis revealed a number of significant examples complementarity between the ribosomal RNA and protein genes. This suggests that they might play an important role in the regulation of mitochondrial translation and transcription mechanisms.

Besides insertion and deletion events, a number of single and multi-amino acid mutations have been observed in these protein genes of both mtDNAs. Some amino acids which are completely conserved in the corresponding protein genes of vertebrate and most invertebrate mtDNAs, have undergone replacements in the Vestimentifera and Chaetognatha during evolution. Data from comparisons of different protein genes with their counterparts reveal that these genes exhibit strikingly different patterns or modes of evolution.

The evolution of the vestimentiferan mitochondrial tRNA genes involves mainly substitutions, although small insertion/deletion events also occur in the different functional regions. The best conserved region is anticodon loop. The degree of conservation of the 5' half was higher than that the 3' half. The DHU loop , TΨC loop and TΨC stem were the most variable regions. Data from the nature and frequency of specific base substitutions of these genes suggest that the rate of evolution varies among as well as within mitochondrial

tRNA genes.

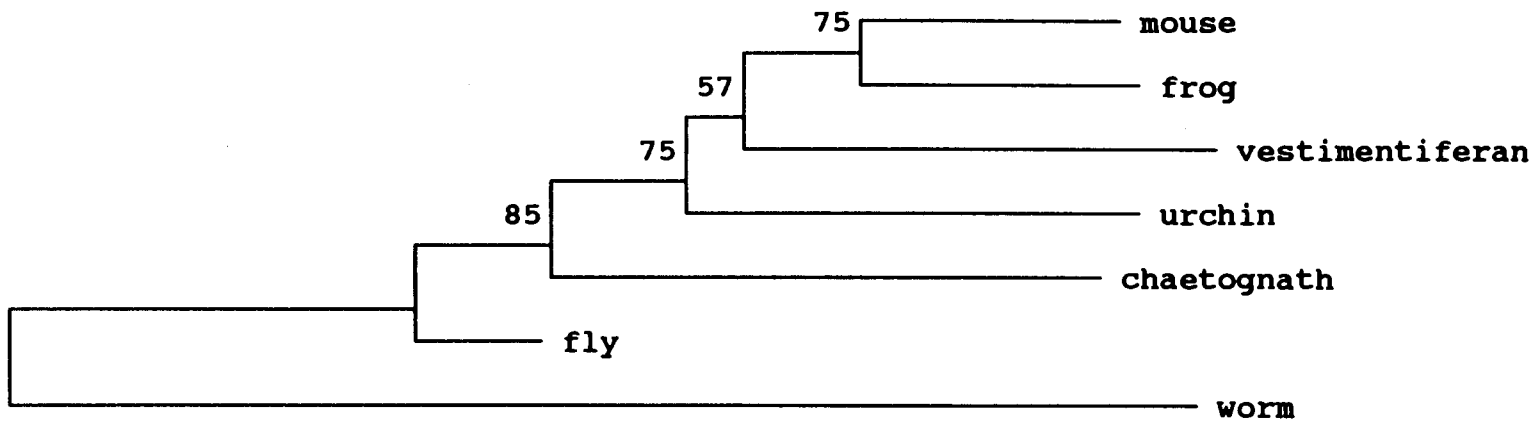
In this study, sequence data were obtained for the 5' ends of the COII and COIII genes for both the vestimentiferan and the chaetognath. Since the cytochrome oxidase subunits are among the most conserved of mitochondrial genes, and the 5' ends tend to yield relatively unambiguous alignments, an attempt was made to combine both phyla into a single phylogenetic tree. Sequences were aligned using the amino acid translation of the first 348 nucleotide sites for COII and the first 675 sites of the COIII gene, for *Ridgeia*, *Sagitta*, *Drosophila*, mouse, *Xenopus*, sea urchin and nematode. Pairwise corrected divergences (Jukes and Cantor 1969) were calculated for all seven taxa for both genes, using all nucleotide sites and using second codon positions. Neighbor-joining trees were then determined for each of the data sets.

The results are shown in Figures 25 and 26. Bootstrap values (out of 100 resamplings) are given at each node. The only phylogenetic relationships supported by all trees are the monophyly of the vertebrates (mouse and frog) and the placement of the pseudocoelomate nematode worm as an outgroup. The inconsistency of these results is not surprising, given the high degree of divergence at the nucleotide level. It is unlikely that nucleotide sequence data alone will provide the resolution necessary for determining relationships at the phylum level. It is more likely that a cladistic analysis of gene organization will provide these answers.

FIGURE 25

Neighbor-joining trees constructed from the nucleotide sequences of the COII gene. The nucleotide sequences of the COII gene include *Ridgeia* (Phylum Vestimentifera), *Sagitta* (Phylum Chaetognatha), mouse (Bibb et al., 1981), *Drosophila yakuba* (Clary and Wolstenholme 1985), *Xenopus* (Roe et al., 1985), sea urchin (Jacobs et al.,1988) and *Caenohabditis elegans* (Okimoto et al., 1992). The trees are based on (A) using all nucleotide sites and (B) using second codon position divergences only.

(A).


75
mouse
57
frog
75
vestimentiferan
85
urchin
chaetognath
fly
worm

(B).

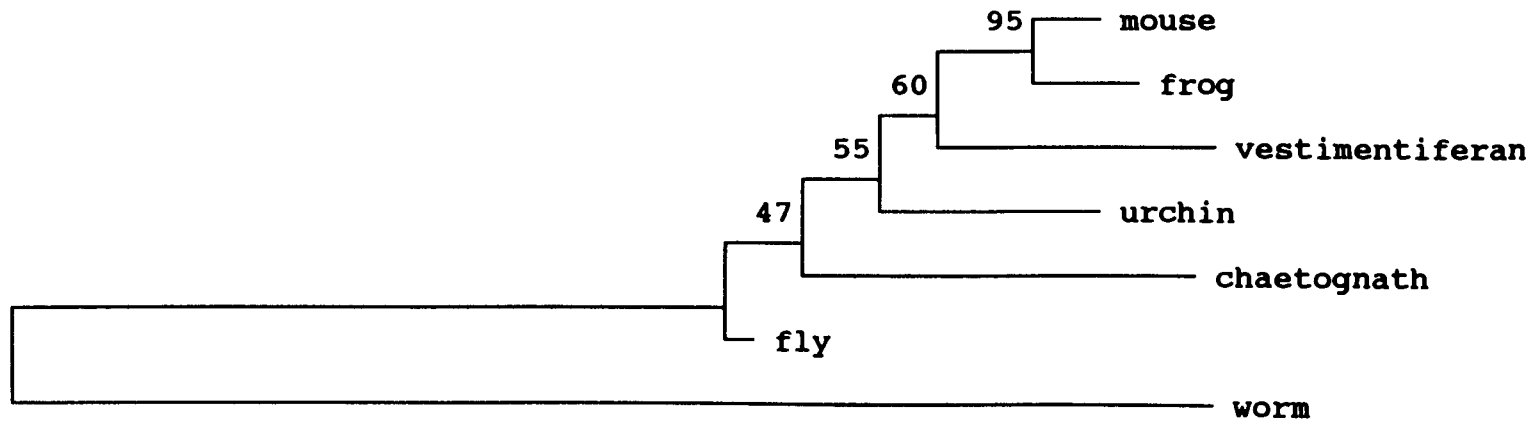
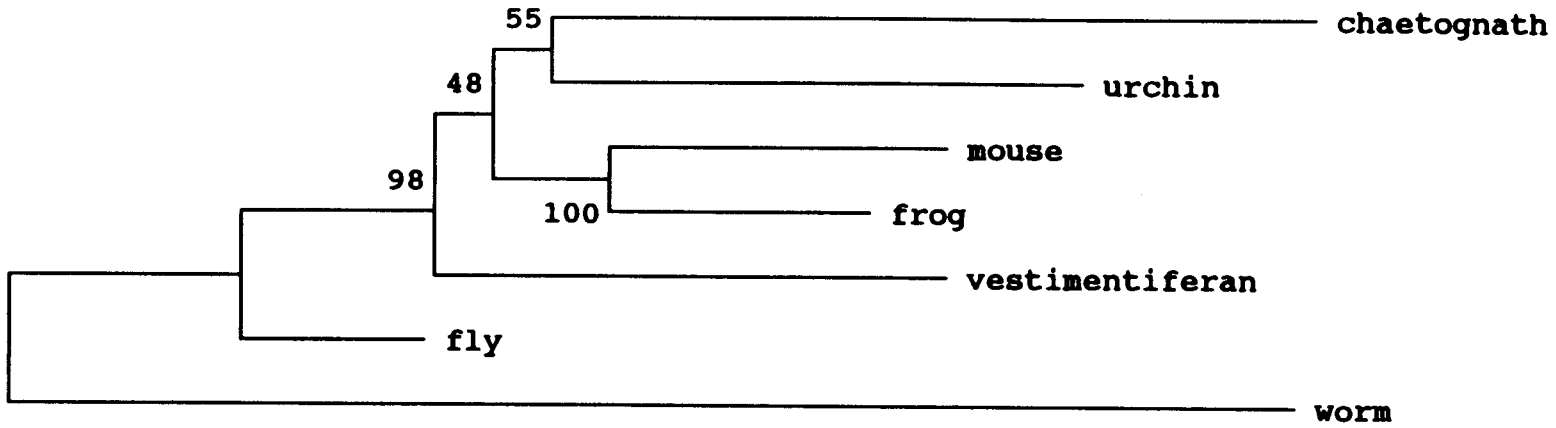

95
mouse
60
frog
55
vestimentiferan
47
urchin
chaetognath
fly
worm

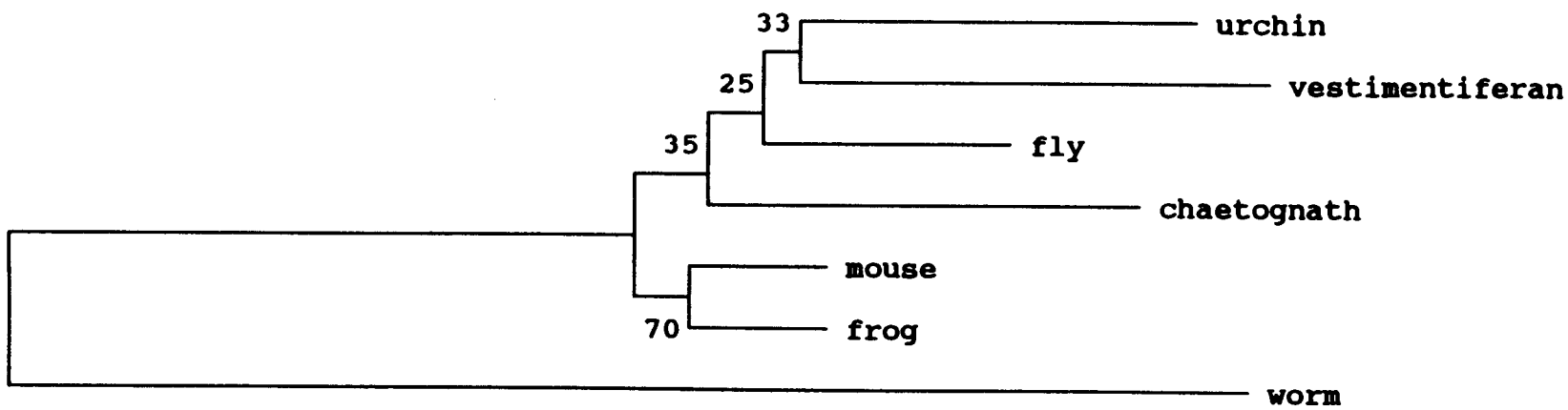
FIGURE 26

Neighbor-joining trees constructed from the nucleotide sequences of the COIII gene. The nucleotide sequences of the COIII gene include *Ridgeia* (Phylum Vestimentifera), *Sagitta* (Phylum Chaetognatha), mouse (Bibb et al., 1981), *Drosophila yakuba* (Clary and Wolstenholme 1985), *Xenopus* (Roe et al., 1985), sea urchin (Jacobs et al.,1988) and *Caenohabditis elegans* (Okimoto et al., 1992). The trees are based on (A) using all nucleotide sites and (B) using second codon position divergences only.

(A).


55
chaetognath
48
urchin
mouse
98
100
frog
vestimentiferan
fly
worm

(B).


33
urchin
25
vestimentiferan
35
fly
chaetognath
mouse
70
frog
worm

REFERENCES

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

Anderson S, de Bruijn MHL, Coulson AR, Eperon IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA. J Mol Biol 156:683-717

Aquadro C. F., Kaplan N. and Risko K. J. (1984) An analysis of the dynamics of mammalian mitochondrial DNA sequence evolution. Mol Biol Evol 1:239-252

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

Arnason U. and Johnsson E. (1992) The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitutina*. J Mol Evol 34:493-505

Atkinson T. and Smith M. (1985) Chapter 3. In Oligonucleotide Synthesis:a practical approach. Edt. by M. J. Gait, IRL press, New York. pp.35-81

Attardi G. (1985) Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. 93:93-145

Attardi G. and Schatz G. (1988) Biogenesis of mitochondria. Annu. Rev. Cell Biol. 4:289-333

Barrell B. G., Bankier A. T. and Drouin J. (1979) A different genetic code in human mitochondria. Nature 282:189-194

Barrell B. G., Anderson S., Bankier A. T., de Bruijn M. H. L., Chen E., Coulson A. R., Drouin J., Eperon I. C., Nierlich D. P., Roe B. A., Sanger F., Schreier P. H., Smith A.

J. H., Staden R. and Young I. G. (1980) Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc. Nalt. Acad. Sci. USA 77:3164-3166

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

Birnboim H. C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7:1513-1523

Bonitz S. G., Berlani R., Coruzzi G., Li M., Macino G., Nobrega F. G., Nobrega M. P., Thalenfeld B. E. and Tzagoloff A. (1980) Codon recognition rules in yeast mitochondria. Proc. Nalt. Acad. Sci. USA 77:3167-3170

Brown G. G. and Simpson M. V. (1982) Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc. Natl. Acad. Sci. USA 79:3246-3250

Brown W. M., George M. Jr. and Wilson A. C. (1979) Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76:1967-1971

Brown W. M. (1981) Mechanisms of evolution in animal mitochondrial DNA. Ann NY Acad Sci 361:119-134

Brown W. M., Prager E. M., Wang A. and Wilson A. C. (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evil 18:225-239

Brown W. M. (1983) In Evolution of Genes and Proteins (Koehn R. K. and Nei M., eds), pp. 62-68, Sinauer Associates, Sutherland, MA.

Brown W. M. (1985) The mitochondrial genome of animals. In: Molecular Evolutionary Genetics, ed. R. J. MacIntyre, pp.95-130. New York: Plenum

Cabot E. L. and Beckenbach A. T. (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. CABIOS 5:233-234

Cann R. L., Brown W. M. and Wilson A. C. (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106:479-499

Cann R. L., Stoneking M. and Wilson A. C. (1987) Mitochondrial DNA and human evolution. Nature 325:31-36

Cantatore P, Gadaleta MN, Roberti M, Saccone C, and Wilson A. C. (1987) Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature 329:853-855

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

Carr S. M., Brothers A. J. and Wilson A. C. (1987) Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. Evolution 41:176-188

Chomyn A. and Attardi G. (1987) Mitochondrial gene products. Curr. Top. Bioenerg. 15:295-329

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

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

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

Clayton D. A. (1982) Replication of mammalian mitochondrial DNA. Cell 28:693-705

Cutler E. B. (1975) Pogonophora and Protostomia - a procrustean bed? Z. Zool. Syst. Evol. Sonderheft 1:112-122

de Bruijn M. H. L. (1983) *Drosophila melanogaster* mitochondrial DNA: a novel organization and genetic code. Nature 304:234-241

Desjardins P. and Morais R. (1990) Sequence and gene organization of the chicken mitochondrial genome. J Mol Biol 212:599-634

Desjardins P. and Morais R. (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J Mol Biol 32:153-161

Felsenstein J. (1989) PHYLIP-Phylogeny inference package (version 3.2). Cladistics 5:164-166

Ferris S. D., Sage R. D., Prager E. M., Ritte U. and Wilson A. C. (1983) Mitochondrial DNA evolution in mice. Genetics 105:681-721

Fox T. D. (1979) Five TGA "stop" codons occur within the translated sequence of the yeast mitochondrial gene for cytochrome *c* oxidase subunit II. Proc. Nalt. Acad. Sci. USA 76:6534-6538

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

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

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

Gray M. W., Sankoff D. and Cedergreen R. I. (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

Guyer R. L. and Koshland Jr. DE (1990) The molecule of the year. Science 246:1543-1544

Hall B. D., Clarkson S. G. and Tocchini-Valenti G. (1982) Transcription initiation of eucaryotic transfer RNA genes. Cell 29:3-5

Heckman J. E., Sarnoff J., Alzner-De Weerd B., Yin S. and RajBhandary U. L. (1980) Novel features in the genetic code and codon reading pattern in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. Proc. Nalt. Acad. Sci. USA 77:3159-3163

Henicoff S. and Eghtedarzadeh M. K. (1987) Conserved arrangement of nested genes at the Drosophila *gart* locus. Genetics 117:711-725

Higuchi R., Von Beroldingen C. H., Sensabaugh G. F. and Erlich h. A. (1988) DNA typing from single hairs. Nature 332:543-546

Himeno H., Masaki H., Kawai T., Ohta T., Kumagi I., Miura I., and Watanabe K. (1987) Unusual genetic codes and a novel gene structure for $tRNA^{ser}$ AGY in starfish mitochondrial DNA. Gene 56:219-230

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

Hoffmann R. J., Boore J. L. and Brown W. M. (1992) A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics Vol 131 No2:397-412

Innis M. A., Gelfand D. H., Sninsky J. J. and White T. J. (1990) PCR Protocols. A guide to methods and applications. Academic Press, NY

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

Jacobs H., Asakawa S., Araki T., Miura K., Smith M. J. and Watanabe K. (1989) Conserved tRNA gene cluster in starfish mitochondrial DNA. Curr Genet 15:193-206

Jones M. L. (1985) On the vestimentifera, new phylum: six new species, and other taxa, from hydrothermal vents and elsewhere. Biol. Soc. Wash. Bull. 6:117-158

Jones M. L. (1985) Vestimentiferan pogonophorans: their biology and affinities. In: The Origins and RElationships of Lower Invertebrates, (eds. S. Conway Morris, J. D. Geoge, R. Gibson and H. M. Platt), pp. 327-342. Clarendon Press, Oxford.

Jukes T. H. and Cantor C. R. (1969) Evolution of protein molecules. In H. N. Munro, ed., *Mammalian Protein Metabolism*, pp. 21-132. New York: Academic Press.

Kocher T. D., Thomas W. K., Meyer A., Edwards S. V., Paabo S., Villablanca F. X. and Wilson A. C. (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86:6196-6200

Lewin B. (1990) Genes IV. New York: Oxford Univ. Press

Li H., Gyllensten U. B., Cui X., Saiki R. K., Erlich H. A. and Arnheim N. (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414-417

Maniatis T., Fritsch E. F. and Sambrook J. (1982) Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratories.

Mullis K. B., Faloona F. A., Scharf S., Saiki R. K., Horn G. and Erlich H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposium on Quantitative Biology. 51:263-273

Mullis K. B. and Faloona F. A. (1987) Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. Meth. Enzymol. 155:335

Mullis K. B. (1990) The unusual origin of the polymerase chain reaction. Sci. Am. 262:56

Nei M. (1987) Molecular Evolutionary Genetics. New York: Columbia Univ. Press

Ojala D., Montoya J., and Attardi G. (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature (London), 290:470-474

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

Okimoto R., Macfarlane J. L., Clary D. O., and Wolstenholme D. R. (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130:471-498

Olsen G. J. (1988) Phylogenetic analysis using ribosomal RNA. Methods Enzymology 164:793-812

Pääbo S., Gifford J. A. and Wilson A. C. (1988) Mitochondrial DNA sequences from a 7,000-year old brain. Nucl. Acids Res. 16:9775-9787

Pääbo S., Thomas W. K., Whitfield K. M., Kumazawa Y. and Wilson A. C. (1991) Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. 33:426-430

Ponce M. R. and Micol J. L. (1992) PCR amplification of long DNA fragments. Nucl. Acids Res. 20: 623

Powell J. R., Caccone A., Amato G. D. and Yoon C. (1986) Rate of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc Natl Acad Sci USA 83:9090-9093

Pumo D. E., Phillips C. J., Barcia M. and Millan C. (1992) Three patterns of mitochondrial DNA nucleotide divergence in the meadow vole, *Microtus pennsylvanicus*. J Mol Evol 34:163-174

Roe B. A., Ma D. P., Wilson R. K. and Wong F.-H. (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260:9759-9774

Saiki R. K., Scharf S. J., Faloona F., Mullis K. B., Horn G. T., Erlich H. A. and Arnheim N. (1985) Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350-1354

Saiki R. K., Gelfand D. H., Stoffel S., Scharf S. J., Higuchi R., Horn G. T., Mullis K. B., and Erlich H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487

Saitou N. and Nei M. (1987) The neighbor-joiring method: A new method for reconstructing phylogenetic trees. Mol. Boil. Evol. 4:406-425

Sankoff D., Cedergren R. and Abel Y. (1990) Genomic divergence through gene rearrangement. Methods Enzymol. 183:426-438

Siewing R. (1975) Thoughts about the phylogenetic-systematic position of Pogonophora. Z. Zool. Syst. Evol. Sonderheft 1:127-138

Simon C., Paabo S., Kocher T. and Wilson A. C. (1990) Evolution of the mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction. In: Clegg M. O'Brien S. (eds), Molecular Evolution. UCLA Symposis on Molecular and Cellular Biology, New Series, Vol. 122, Alan R. Liss, Inc., NY, p235

Smith M. J., Banfield D. K., Doteval K., Gorski S., and Kowbel D. J. (1989) Gene arrangement in sea star mitochondrial DNA demonstrates a major event during echinoderm evolution. Gene 76:181-185

Smith M. J., Banfield D. K., Doteval K., Gorski S., and Kowbel D. J. (1990) Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. J. Mol. Evol. 31:195-204

Solignac M., Monnerot M. and Mounolou J-C. (1986) Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J Mol Evol 23:53-60

Sprinzl M., Hartman T., Meissner F., Moll J. and Vorderwulbecke T. (1987) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 15:r53-r188

Tamura K. and Aotsuka T. (1988) Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. Biochem. Genet. 26:815-819

Thomas W. K. and Beckenbach A. T. (1989) Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. J. Mol. Evol. 29:233-245

Thomas W. K., Maa J. and Wilson A. C. (1989) Shifting constraints on tRNA genes during mitochondrial DNA evolution in animals. New Biologist 1:93-100

Thomas W. K. and Wilson A. C. (1991) Mode and tempo of molecular evolution in the nematode *Caenorhabditis*: cytochrome oxidase II and calmodulin sequences. Genetics 128:269-279

Wilson A. C., Cann R. L., Carr S. M., George M., Gyllensten U. B., Helm-Bychowski K. M., Higuchi R. G., Palumbi S. R., Prager E. M., Sage R. D. and Stoneking M. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26:375-400

Wolstenholme D. R. and Clary D. O. (1985) Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109:725-744

Wolstenholme D. R., Clary D.O., Macfarlane J. L., Wahleithner J. A. and Wilcox L. (1985) In Achieverments and Perspectives of Mitochondrial Research. vol. II (Quagliariello E., Slater E. C., Palmieri F., Saccone C. & Kroon A. M. eds) pp.61-69. Elsevier, Amsterdam.

Wolstenholme D. R. Macfarlane J. L. Okimoto R. Clary D.O. and Wahleithner J. A. (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA 84:1324-1328

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