Molecular Marine Biology and Biotechnology (1994) 3(5), 294-299

DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates

O. Folmer, M. Black, W. Hoeh,* R. Lutz, and R. Vrijenhoek+

Center for Theoretical and Applied Genetics, and Institute of Marine and Coastal Science, Rutgers University, New Brunswick, New Jersey 08903-231

Abstract

We describe "universal" DNA primers for polymerase chain reaction (PCR) amplification of a 710-bp fragment of the mitochondrial cytochrome c oxidase subunit I gene *(COI)* from *11* invertebrate phyla: Echinodermata, Mollusca, Annelida, Pogonophora, Arthropoda, Nemertinea, Echiura, Sipuncula, Platyhelminthes, Tardigrada, and Coelenterata, as well as the putative phylum Vestimentifera. Preliminary comparisons revealed that these *COI* primers generate informative sequences for phylogenetic analyses at the species and higher taxonomic levels.

Introduction

The purpose of this short communication is to describe "universal" DNA primers for the polymerase chain reaction (PCR) amplification of a 710-bp fragment of the mitochondrial cytochrome c oxidase subunit I gene *(COI).* This study was motivated by the recent discoveries of more than 230 new invertebrate species, comprising new genera, families, classes, orders, and potentially a new phylum, from deep-sea hydrothermal vent and cold-water sulfide or methane seep communities (Tunnicliffe, 1991). Our goal was to develop molecular techniques for phylogenetic studies of these diverse organisms. We focused on the mitochondrial cytochrome c oxidase subunit I *(COI)* gene because it appears to be among the most conservative protein-coding genes in the mitochondrial genome of animals (Brown, 1985), which was preferable for the evolutionary time depths likely to be found in our studies. We quickly became aware of the broad utility of these *COI* primers for broader systematic studies of metazoan invertebrates, including acoelomates, pseudocoelomates, and coelomate protostomes and deuterostomes.

Results

To design candidate primers, we compared published DNA sequences from the following species: blue mussel, *Mytilus edulis;* fruitfly, *Drosophila yakuba;* honeybee, *Apis mellifera;* mosquito, *Anopheles gambiae;* brine shrimp, *Artemia franciscana;* nematodes, *Ascaris suum* and *Caenorhabditis elegans;* sea urchin, *Strongylocentrotus purpuratus;* carp, *Cyprinus carpio;* frog, *Xenopus laevis;* chicken, *Gallus gallus;* mouse, *Mus musculus; cow, Bos taurus;* fin whale, *Balaenoptera physalus;* and human, *Homo sapiens* (Figure *1).* Several highly conserved regions of these *COI* genes were used as the targets for primer designs.

Altogether, three coding-strand and six anti-coding-strand primers were tested (Table *1)* for amplification efficiency. The following primer pair consistently amplified a 710-bp fragment of *COI* across the broadest array of invertebrates:

LCO1490: 5'-ggtcaacaaatcataaagatattgg-3'

HC02198: 5'-taaacttcagggtgaccaaaaaatca-3'

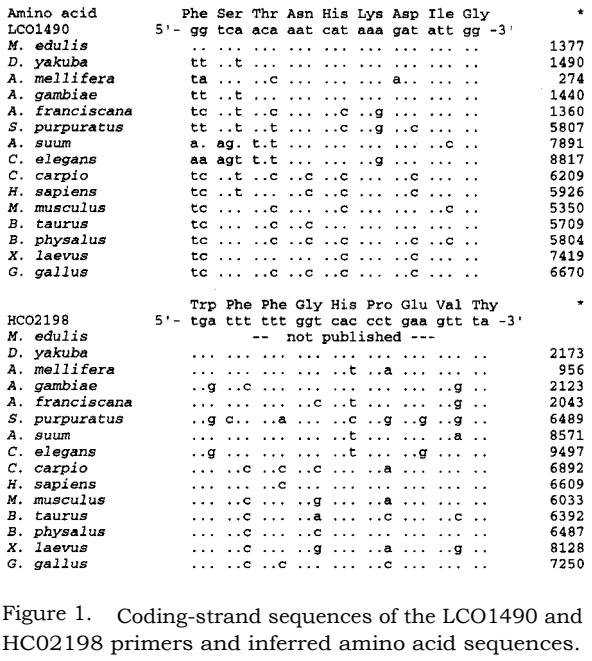
In the code names above, L and H refer to light and heavy DNA strands, CO refers to cytochrome oxidase, and the numbers (1490 and 2198) refer to the position of the *D. yakuba 5'* nucleotide.

We also present the primers as coding-strand sequences, along with their inferred amino acids (Figure *1).* The usefulness of these primers results from the high degree of sequence conservation in their respective 3' ends across the 15 taxa. The 3' end of each primer is on a second-position nucleotide. All other pairwise primer combinations amplified fewer taxa or gave additional nonspecific products under less stringent amplification conditions.

The LCO1490 and HC02198 amplified DNA from more than 80 invertebrate species from 11

*Present address: Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada.

+Correspondence should be sent to this author.



```
Amino acid          Phe Ser Thr Asn His Lys Asp Ile Gly          *
LCO1490        5'- gg tca aca aat cat aaa gat att gg -3'
M. edulis           .. ... ... ... ... ... ... ... ..        1377
D. yakuba           tt ..t ... ... ... ... ... ... ..        1490
A. mellifera        ta ... ..c ... ... ... a.. ... ..         274
A. gambiae          tt ..t ... ... ... ... ... ... ..        1440
A. franciscana      tc ..t ..c ... ..c ..g ... ... ..        1360
S. purpuratus       tt ..t ..t ... ..c ..g ..c ... ..        5807
A. suum             a. ag. t.t ... ... ... ... ..c ..        7891
C. elegans          aa agt t.t ... ... ..g ... ... ..        8817
C. carpio           tc ..t ..c ..c ..c ... ..c ... ..        6209
H. sapiens          tc ..t ... ..c ..c ... ..c ... ..        5926
M. musculus         tc ... ..c ... ..c ... ... ..c ..        5350
B. taurus           tc ... ..c ..c ... ... ... ... ..        5709
B. physalus         tc ... ..c ..c ..c ... ..c ..c ..        5804
X. laevus           tc ... ... ... ..c ... ..c ... ..        7419
G. gallus           tc ... ..c ..c ..c ... ..c ... ..        6670

                    Trp Phe Phe Gly His Pro Glu Val Thy          *
HCO2198        5'- tga ttt ttt ggt cac cct gaa gtt ta -3'
M. edulis                -- not published ---
D. yakuba           ... ... ... ... ... ... ... ... ..        2173
A. mellifera        ... ... ... ... ..t ..a ... ... ..         956
A. gambiae          ..g ..c ... ... ... ... ... ..g ..        2123
A. franciscana      ... ... ... ..c ..t ... ... ..g ..        2043
S. purpuratus       ..g c.. ..a ... ..c ..g ..g ..g ..        6489
A. suum             ... ... ... ... ..t ... ... ..a ..        8571
C. elegans          ..g ... ... ... ..t ... ..g ... ..        9497
C. carpio           ... ..c ..c ..c ... ..a ... ... ..        6892
H. sapiens          ... ... ..c ... ... ... ... ... ..        6609
M. musculus         ... ..c ... ..g ... ..a ... ... ..        6033
B. taurus           ... ..c ... ..a ... ..c ... ..c ..        6392
B. physalus         ... ..c ... ..c ... ... ... ... ..        6487
X. laevus           ... ..c ... ..g ... ..a ... ..g ..        8128
G. gallus           ... ..c ..c ... ... ..c ... ... ..        7250
```

Figure 1. Coding-strand sequences of the LCO1490 and HC02198 primers and inferred amino acid sequences. Dots represent identical nucleotides at a given position compared with *Drosophila yakuba.* *Position as listed in GenBank. Accession numbers and primary references for GenBank sequences are as follows: *Mytilus edulis,* M83761/M83762 (Hoffmann et al., 1992); *Drosophila yakuba,* X03240 (Clary and Wolstenholme, 1985); *Apis mellifera,* M23409 (Crozier et al., 1989); *Anopheles gambiae,* L20934 (Beard et al., 1993); *Artemia franciscana (J.R.* Valverde, direct submission to GenBank access number X69067); *Strongylocentrotus* purpuratus, X12631 (Jacobs et al., 1988); *Ascaris* suum, X54252, and *Caenorhabditis elegans,* X54253 (Okimoto et al., 1990); *Cyprinus carpio,* X61010 (Chang and Huang, 1991); *Homo sapiens,* M12548 (Anderson et al., 1981); *Mus musculus,* V00711 (Bibb et al., 1981); *Bos taurus,* V00654 (Anderson et al., 1982); *Balaenoptera physalus,* X61145 (Arnason et al., 1991); *Xenopus laevis,* X02890 (Roe et al., 1985); and *Gallus gallus,* X52392 (Desjardins and Morais, 1990).

phyla (Table 2). The PCR products of species from five phyla (Mollusca, Annelida, Arthropoda, Vestimentifera, and Coelenterata) are illustrated in Figure 2. Except for *Hydra,* all products resulted from a single PCR amplification. The *Hydra* sample was reamplified to provide sufficient product for direct sequencing. For several species, initial amplification produced multiple PCR products. In these cases, target DNA for sequencing was obtained by raising the annealing temperature, or gel-isolating the initial 710-bp fragment and reamplifying it.

To verify that the amplified fragment is indeed *COI,* we obtained a minimum of 200 by of sequence from all species listed in Table 2 (except those marked with an asterisk). Typically, cycle-sequencing with these primers produced a readable sequence of at least 651 bp, equivalent to 219 inferred amino acid residues. To demonstrate that the products are *COI,* we provide four new sequences (in reading frame) from work in progress on deep-sea invertebrates (Figure 3). Comparisons of these sequences with *COI* from *D. yakuba* reveal that most variation occurs at the third-position nucleotides. Ongoing analyses of this *COI* fragment from a diverse array of bivalve mollusks and vestimentiferan tube worms suggest that phylogenetic resolution at the phylum and class level can be obtained from inferred amino sequences. Intermediate-level resolution (family to genus) is retained in first- and second-position nucleotides. Third-position substitutions are saturated at these higher levels, but retain informative polymorphisms within at least one bivalve species, *Bathymodiolus thermophilus.*

Discussion

The universal DNA primers, LCO1490 and HCO2198, amplified a 710-bp region of the mitochondrial cytochrome oxidase subunit I gene from a broad range of metazoan invertebrates. We are presently using these primers to examine phylogenetic relations among the following taxa: (1) tube worms (Vestimentifera) and other protostome worms (Pogonophora and Annelida); (2) deep-sea marine bivalve mollusks (Mytilidae and Vesicomyidae); (3) freshwater bivalve mollusks (Unionidae, Dreissenidae, and Corbiculidae); (4) vent-associated

Table 1. Other COI primers tested in this study, presented relative to the coding strand of *Drosophila yakuba.*

Primer	Sequence	Position
LCO1606	tta att gga gat gat caa at	
D. yakuba	... ...	1606
LCO1495	aca aat cat aaa gat att gg	
D. yakuba	... ...	1495
LCO1828	gga gct ggt aca ggt tga ac	
D. yakuba	... ...	1828
HCO1862	tga act gtt tac cct cct tt	
D. yakuba	... ...t	1862
HCO2042	tga tca gta ctt att aca gc	
D. yakuba	... ...	2042
HCO2799	cca cga cgg tac ata gat tat gct g	
D. yakuba	..t ... ..t ..t tc. ... ..c c.. .	2799
HCO2911	gaa gct tta gtg tgt caa cga ggg g	
D. yakuba	... ag. ... ... .c. ... ... caa .	2911
HCO2192	tga ttt ttt ggt cac cct ga	
D. yakuba	... ...	2192
HCO2768	caa cat ttt tta gga tta gc	
D. yakuba	... ...	2768

Table 2. Species representing eleven different phyla for which the LCO1490 and HC02198 primers amplified and sequenced the 710-bp mitochondrial COI fragment.

Phylum	Class	Order	Genus/species
Echinodermata	Crinoidea		Unidentified deep-sea species*
Mollusca	Polyplacophora	Ischnochitonida	*Katharina* sp.
	Aplacophora		*Chaetoderma* sp.
	Gastropoda	Archaeogastropoda	*Lepetodrilus elevatus*
			Astraea sp.
			Collisella sp.
		Stylommatophora	*Biomphalaria* sp.*
	Bivalvia	Nuculoida	*Yoldia scissurata*
			Nuculoma tenuis
		Solemyoida	*Solemya velum*
		Mytiloida	*Bathymodiolus thermophilus*
			7 New vent and seep spp.
			Mytilus edulis
			Modiolus modiolus
			Idas washingtonia
		Eulamellibrachia	*Calyptogena magnifica*
			C. pacifica
			C. ponderosa
			Vesicomya cordata
			16+ New species
		Ostreoidia	*Hinnites multirugosus*
		Unionoida	*Utterbakia* (6+ spp.)
		Myoida	*Mya arenaria*
		Veneroida	*Mercenaria mercenaria*
			Corbicula flumenea
			Dressera polymorpha
			Ensis directus
		Pholadomyoida	*Pandora gouldiana*
	Scaphopoda		*Dentalium* sp.
	Cephalopoda	Teuthoidea	*Loligo paeleii*
			Octopus sp.
Annelida	Polychaetes	Phyllodocida	*Branchipolynoe symmytilida*
		Terebellida	*Amphisamytha galapagensis*
			Paralvinella palmiformis
			Alvinella pompejana
	Oligochaete	Haplotaxida	*Tubifex tubifex*
	Hirudinea	Gnathobdellida	*Hirudo medicinalus*
Pogonophora	Perviata		*Galatheolinum brachiosum*
			Unidentified Loihi Smnt. sp.
Vestimentifera†	Axonobranchia	Riftiida	*Riftia pachyptila*
	Basibranchia		
		Tevniida	*Ridgeia piscesae*
			Tevnia jerichonana
			Oasisia alvinae
		Lamellibranchiida	*Escarpia* sp.
			Lamellibrachia barhami
Arthropoda	Malacostraca	Decapoda	*Rimicaris exoculata*
			Chorocaris chacei
			Alvinocaris markensis
			Bythograea thermydron
			Munidopsis (*lentigo?*)
	Pycnogonida		Unidentified Californian species
			Ammothea sp.
Nemertinea	Anopla	Heteronemertini	*Cerebratulus* sp.
Echiura		Xenopneusta	*Urechis* sp.

* Amplified, but not sequenced to date
† Jones (1985)

Table 2. *Continued*

Phylum	Class	Order	Genus/species
Sipuncula	Phascolosomida	Phascolosomiformes	*Phascolosoma* sp.
Platyhelminthes	Turbellaria	Polycladida	Unidentified marine species
	Trematoda		*Fasciola hepatica*
Tardigrada	Eutardigrada		*Adorybrotus coronifer**
Coelenterata	Hydrozoa	Anthomedusae	*Hydra littoralis*

arthropods (Caridae); and (5) parasitic platyhelminths (Trematoda). We also are investigating the utility of this *COI* fragment for larval identifications in several of these groups. Independent laboratories have verified the utility of the LCO1490 and HCO2198 primers for amplification and sequencing of *COI* from (1) oysters, genera *Crassostrea (Y-P.* Hu, Louisiana State University, and M. Hare, University of Georgia) and *Ostrea* (Diarmaid O'Foigel, University of South Carolina); (2) scallops, genus *Placopecten* (P. Gaffney, University of Delaware); (3) hard clams, genus *Mercenaria (D.* O'Foigel); (4) archaeogastropod limpets (A. MacArthur, University of Victoria); (5) arachnids (A. Tan, University of Hawaii); and (6) marine hydrozoans (S. Karl, University of South Florida).

Experimental Procedures

Whole cell DNA was extracted from either fresh tissue or tissue frozen at - 80°C immediately after collection of a specimen. We used a conventional hexadecyl-trimethyl-ammonium bromide (CTAB) protocol, modified from Doyle and Dickson (1987). Typically, 1 mm^3 of tissue was extracted and the DNA resuspended in 75 to 150 μl (dependent upon the size of the pelleted DNA) of sterile distilled water. In our experience, DNA extracted by this protocol and stored at - 20°C remains intact for at least three years.

Figure 2. Agarose gel of PCR products from seven different species of invertebrates. All PCR products except lane 7 are directly amplified from total DNA extraction. Lane L, Phi-X/HaeIII ladder. Lane 1, blue mussel, *Mytilus edulis.* Lane 2, squid, *Loligo pealeii.* Lane 3, polychaete *Paralvinella palmiformis.* Lane 4, oligochaete *Tubifex tubifex.* Lane 5, shrimp, *Rimicaris exoculata.* Lane 6, tube worm, *Riftia pachyptila.* Lane 7, reamplification of hydra, *Hydra littoralis.* Lane 8, negative control PCR reaction with all components except template DNA.

Polymerase chain reaction

We typically used 1 μl of the DNA extract as template for a 50-μ1 PCR reaction, using 4 units of Taq polymerase (Promega, Madison, WI) per reaction. Each 50-μ1 reaction consisted of 5 p.l of lox buffer (provided by the manufacturer), 5 μl of $MgCl_2$ (0.025 mol/liter, both solutions supplied with the polymerase), 2.5 μl of each of the two primer stock solutions (10 μmol/liter), 5 p.l C, T, A, G nucleotide mix (Boehringer Mannheim, Indianapolis, IN, 2 μmol/liter for each nucleotide), and 29 μl sterile distilled water. Reactions were amplified through 35 cycles at the following parameters: one minute at 95°C, one minute at 40°C, and one and a half minutes at 72°C, followed by a final extension step at 72°C for seven minutes. Amplifications were confirmed by standard submarine gel electrophoresis, using 2% w/v low-melting agarose/TBE gels (NuSieve, FMC BioProducts), stained with ethidium bromide.

Sequencing

Most templates could be sequenced from a single round of amplification. Occasionally, templates provided too little product from a single amplification. In such cases, the first amplification product was gel-isolated and used as template for a reamplification with a higher annealing temperature (50°C, all other parameters being held the same). In all instances, the PCR product for sequencing was obtained by running the entire reaction volume on a 2% low-melting agarose gel, using wide-tooth combs. The reaction product was excised from the gel and subsequently purified utilizing Wizard-PCR kits (Promega).

We used -y-^{33}P (NEN Dupont) end-labeled versions of the LCO1490 and HCO2198 primers for cycle-sequencing (Perkin-Elmer Cetus, Amplitaq

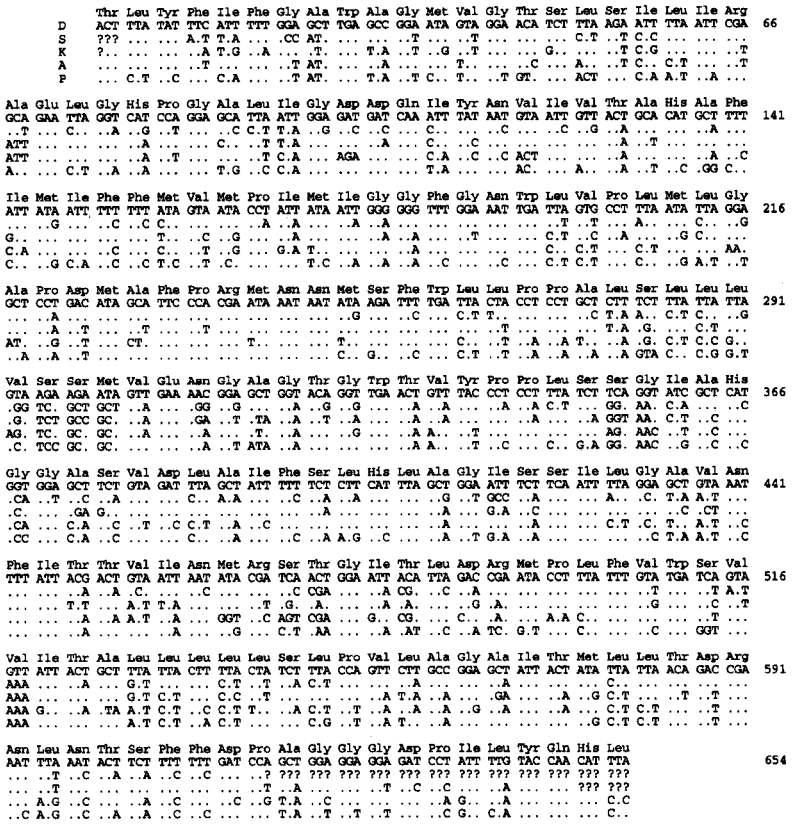


Figure 3. Four new cytochrome oxidase subunit I nucleotide sequences from marine invertebrates shown in reference to *Drosophila yakuba. D, D. yakuba;* S, *Solemya velum* (Mollusca: Bivalvia); *K, Katharina* sp. (Mollusca: Polyplacophora); *A, Amphisamytha galapagensis* (Annelida: Polychaeta: Ampharetidae), and P, *Paralvinella palmiformis* (Annelida: Polychaeta: Alvinellidae). Nucleotide #1 corresponds to position *#1516* in the published *D. yakuba* sequence.

Cycle-sequencing Kit, protocol according to the manufacturer) of the double-stranded PCR products. Two electrophoretic analyses were required to sequence the complete fragment in each direction. First, we used a 6% denaturing (50% w/v urea) polyacrylamide gel (19:1 acrylamide to bis-acrylamide ratio) in a 40-cm-tall, wedge (0.4-1.2-mm) gel configuration to obtain approximately 250 to 300 by of readable sequence. Second, we used a 5% denaturing polyacrylamide gel in an 88-cm-tall, straight (0.4-mm) configuration, to obtain an additional 350 to 425 by of sequence.

Acknowledgments

Our thanks to A. Trivedi and C. Di Meo for assistance in the laboratory. Dr. S. Karl's advice was greatly appreciated, particularly during the early phase of this work. This is contribution No. 94-26 of the Institute of Marine and Coastal Sciences, Rutgers University, and New Jersey Agricultural Experiment Station Publication No. 2-67175-8-94, supported by state funds and National Science Foundation grants OCE89-17311 and OCE93-02205 to R.C.V. and R.A.L.

References

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., and Young, I.G. *(1981).* Sequence and organization of the human mitochondrial genome. *Nature 290:457-465.*

Anderson, S., de Bruijn, M.H., Coulson, A.R., Eperon, I.C., Sanger, F., and Young, I.G. *(1982).* Complete sequence of bovine mitochondrial DNA: conserved features of the mitochondrial genome. *J Mol Biol 156:683-717.*

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

Beard, C.B., Hamm, D.M., and Collins, F.H. *(1993).* The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol Biol 2:103-109.*

Bibb, M.J., van Etten, R.A., Wright, C.T., Walberg, M.W., and Clayton, D.A. *(1981).* Sequence and gene organization of mouse mitochondrial DNA. *Cell 26:167-180.*

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

Chang, Y.S., and Huang, F.L. *(1991).* GenBank Accession number *X61010.*

Clary, D.O., and Wolstenholme, D.R. *(1985).* The mitochondria) DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization and genetic code. *j Mol Evol 22:252-271.*

Crozier, R.H., Crozier, Y.C., and Mackinlay, A.G. *(1989).* The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Mol Biol Evol 6:399-411.*

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

Doyle, J.J., and Dickson, E. *(1987).* Preservation of plant samples for DNA restriction endonuclease analysis. Taxon *36:715-722.*

Hoffmann, R.J., Boore, J.L., and Brown, *W. (1992).* A novel mitochondrial genome organization for the blue mussel *Mytilus edulis. Genetics 313:397-412.*

Jacobs, H.T., Elliott, D.J., Veerabhadracharya B.M., and Farquharson, A. *(1988).* Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. *JMol Biol 202:185-217.*

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

Okimoto, R., Macfarlane, J.L., and Wolstenholme, D.R. *(1990).* Evidence for 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.*

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

Tunnicliffe, *V. (1991).* The biology of hydrothermal vents: ecology and evolution. *Oceanogr Mar Biol Annu Rev 29:319-407.*