THE MYSTERY OF THE CHAETOGNATH: A MOLECULAR

PHYLOGENETIC APPROACH USING PELAGIC CHAETOGNATH SPECIES

ON PELICAN ISLAND, GALVESTON, TEXAS

A Thesis

by

LEAH NICOLE TOWERS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2010

Major Subject: Zoology

The Mystery of the Chaetognatha: A Molecular Phylogenetic Approach Using Pelagic

Chaetognath Species on Pelican Island, Galveston, Texas

THE MYSTERY OF THE CHAETOGNATHA: A MOLECULAR

PHYLOGENTIC APPROACH USING PELAGIC CHAETOGNATH SPEICES

ON PELICAN ISLAND, GALVESTON, TEXAS

A Thesis

by

LEAH NICOLE TOWERS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Robin Brinkmeyer
	Michael Benedik
Committee Member,	Antonietta Quigg
Head of Department,	U.J. McMahan

December 2010

Major Subject: Zoology

ABSTRACT

The Mystery of the Chaetognatha: A Molecular Phylogenetic Approach Using Pelagic

Chaetognath Species on Pelican Island, Galveston, Texas. (December 2010)

Leah Nicole Towers, B.S., Texas A&M University at Galveston

Co-Chairs of Advisory Committee: Dr. Michael Benedik
Dr. Robin Brinkmeyer

The phylum Chaetognatha is a mysterious group of organisms that has eluded scientists for more than a century because of their unique morphology and developmental characteristics, i.e. protostome (mouth develops from blastopore; e.g. mollusks, annelids, arthropods) versus deuterostome (anus develops from blastopore; e.g. echinoderms and chordates) offer few clues to their evolutionary origins. Some early morphological studies argued that chaetognaths were derived mollusks or nematodes according to gross ultrastructural data, while other studies focused on the coelomic cavity. 33

Although 18S rRNA is widely used in molecular phylogeny studies, it has limits such as long- branch chain attractions and a slow rate of evolutionary change. Long-branch chain attractions are a phenomenon in phylogenetic analyses when rapidly evolving lineages are inferred to be closely related, regardless of their true evolutionary relationships. Hence other genes are used in this study to complement the 18S rRNA such as the cytochrome oxidase genes. The cytochrome oxidase genes are highly

conserved throughout all eukaryotic organisms and they are less ambiguous to align as compared to the ribosomal genes, making them better phylogenetic markers as compared to the 18S rRNA gene.

This study focuses on using a molecular approach (ARDRA, PCR, phylogenetic tree reconstruction) to determine the phylogeny of pelagic chaetognaths found on Pelican Island, Galveston, Texas. 18S rRNA, Cytochrome Oxidase I and Cytochrome Oxidase II genes were used to help decipher the phylogeny of this group.

All analyzed genes in this study (18S rRNA, COI, and COII) grouped the Pelican Island chaetognaths with the protostomes. The maximum parsimony bootstrap tree for the 18S rRNA gene, grouped the samples closest to the arthropods (protostome). For the COI and COII genes, the minimum evolution bootstrap tree grouped the 8 collected samples more closely to two other protostome phyla: the mollusks and annelids (COI) while bootstrapping with the COII grouped the samples with the nematodes (with >66% bootstrap). My findings are significant because they reveal phylogenetic results of a protostome lineage for the Chaetognatha using 3 genes, one of which (COII) has not been greatly studied for the Chaetognatha.

DEDICATION

To my parents: My deepest thanks for making me the person I am today. Without your love, guidance and support through all the years I would not have accomplished what I have so far in life. Thanks again. I love both of you very much.

ACKNOWLEDGEMENTS

NOMENCLATURE

ARDRA   Amplified Ribosomal DNA Analysis

COI   Cytochrome Oxidase I Gene

COII   Cytochrome Oxidase II Gene

MEGA   Molecular Evolutionary Genetics Analysis

PCR   Polymerase Chain Reaction

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Introduction

Chaetognatha (meaning hair-jaws; arrow worms) is a phylum of marine vermiform predatory organisms. They are either planktonic or benthic, are approximately 0.5 to 12 cm in length, and are characterized in part by the presence of chitinous spines around the mouth and peculiar ciliary organs located dorsally (Fig. I.1). They are found in almost every marine ecosystem including the open ocean, tide pools, polar waters, marine caves, coastal lagoons, and the deep sea (Bone et al. 1991). The oldest fossil record for Chaetognaths dates back to the Cambrian Period 542 million years ago (Vannier et al. 2007). A single class (Sagittoidea) and three orders (Monophragmophora, Biphragmorphora, and Aphragmophora) are included within this phylum. One hundred-twenty-five species are placed within 10 families and 24 genera (Fig. I. 2).

Chaetognaths have perplexed biologists since their discovery in 1768 (Halanych 1996) because their unique morphology and developmental characteristics i.e. protostome (mouth develops from blastopore; e.g. mollusks, annelids, arthropods) versus deuterostome (anus develops from blastopore; e.g. echinoderms and chordates) offer few clues to their evolutionary origins.

This thesis follows the style of Evolution.

Molecular phylogenies reconstructed using ribosomal RNA (i.e. 18S and 28S) and whole genome using the expressed sequence tag (EST) approach places Chaetognatha close to the base of the protostome tree (Marlétaz et al. 2006, 2008; Telford and Holland 1993). This may explain their deuterostome embryonic characters. If chaetognaths branched off from the protostomes before they evolved their distinctive protostome embryonic characters, they may have retained deuterostome characters inherited from early bilaterian ancestors. Thus, chaetognaths may be a useful model for the ancestral bilaterian (Papillon et al. 2004; Marlétaz et al. 2008). In the past century, the affinity of the chaetognaths to one ancestral group versus another has resulted primarily from controversy over four characters; the coelomic condition, the absence of circular muscle, a tripartite *Bauplan* (i.e. body plan or anatomy), and the fate of the blastopore during embryological development.
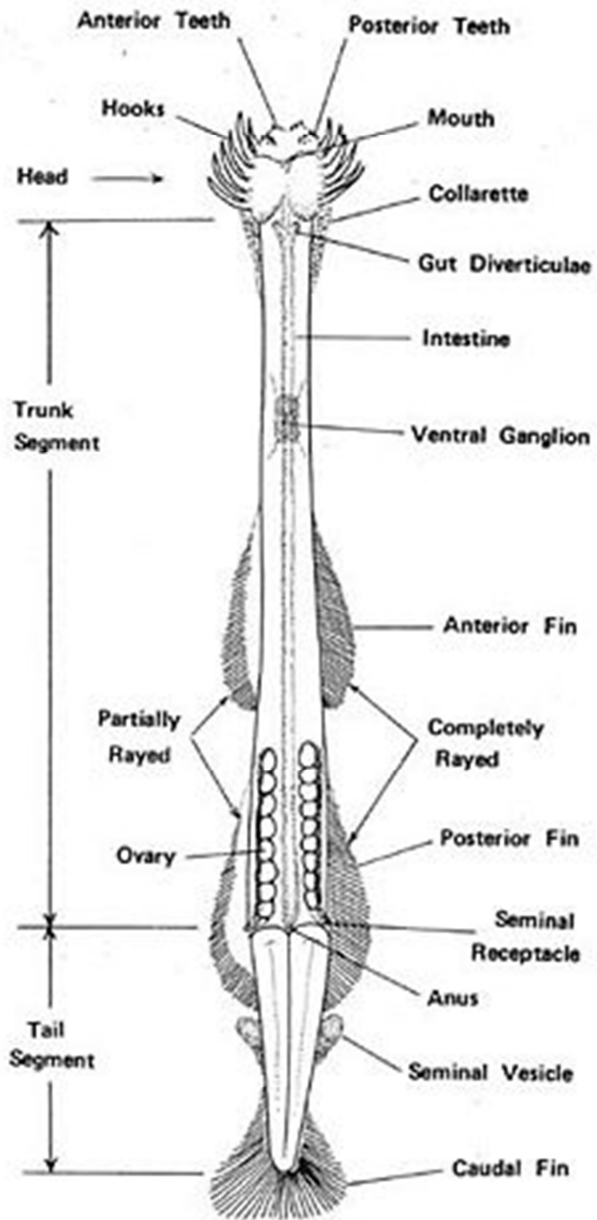
Fig. I. 1. Schematic of Chaetognath
Bauplan. (McLelland 1989)
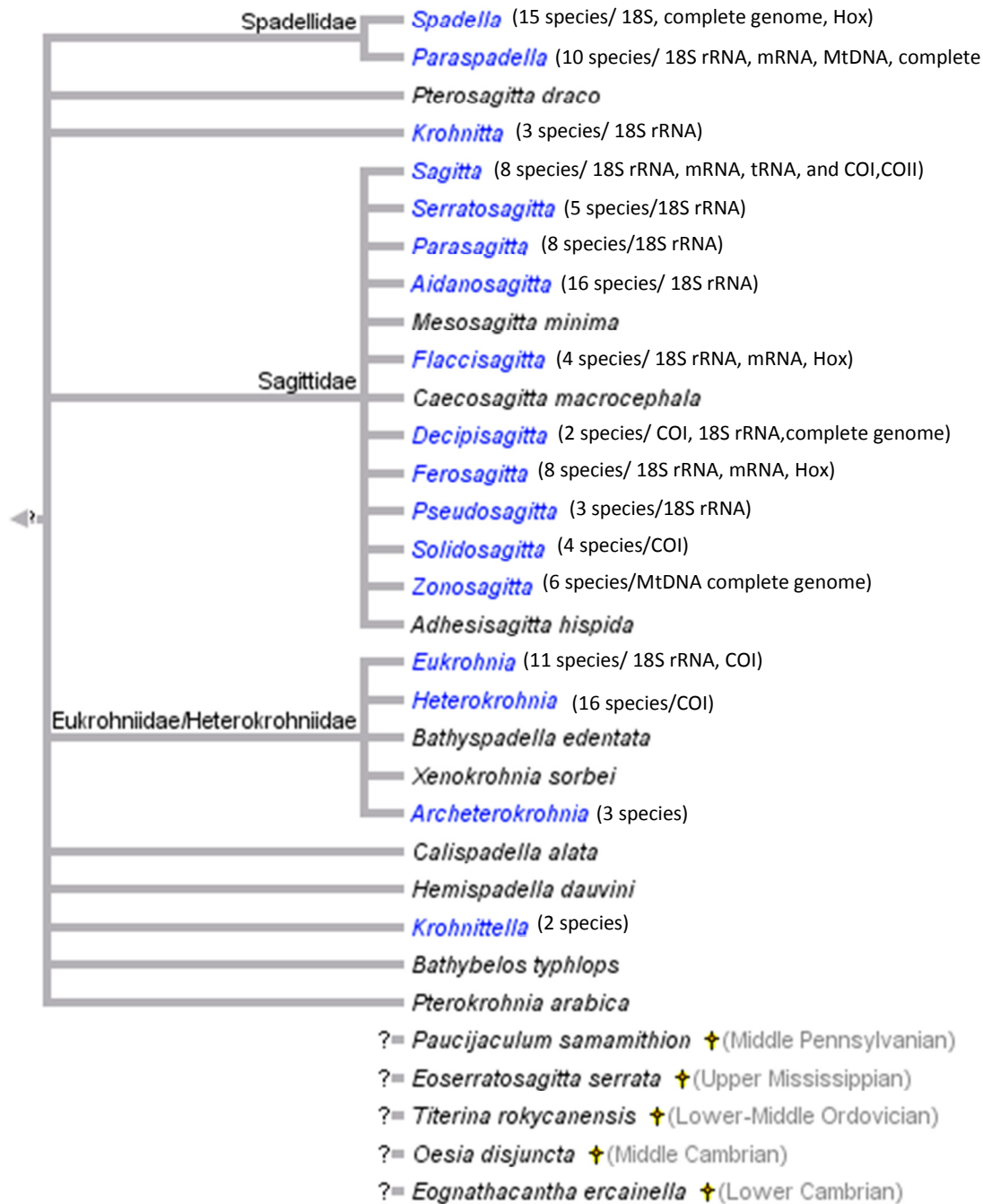
Fig. I.2. Chaetognath 'tree of life' including families (Spadellidae, Sagittidae, and Eukrohnidae/Heterokrohnidae) with respective genera. Number of species per genus and availability of gene sequences in GenBank are in parentheses). Dendrogram is from the Tree of Life Web Project (2006).

1.2 The Morphology Debate

In 1907, Gunther proposed a protostome affinity by suggesting chaetognaths are derived mollusks comparing several morphological characteristics both gross and ultrastructural (Gunther 1907). In the decades to follow, Hyman (1959) and Wilmer (1990) concluded that chaetognaths had a basal deuterostome lineage based upon ultrastructural characteristics. This newer hypothesis emphasized the importance of the tripartite coelomic arrangement. However, because of the lack of synampomorphies (i.e. a character trait that is shared with two or more taxa) with other phyla, a myriad of hypotheses have since been proposed concerning chaetognath origins. In 1968, Ghiradelli hypothesized that chaetognaths are pseudocoelomate and closely related to the nematodes (Ghiradelli 1968; Fig. I.3). This hypothesis was originally supported by the apparent lack of circular muscle in the body wall in both taxa. In 1982, Welsch and Störch described the body cavity of the chaetognath *Sagitta elegans* as coelomate, based on the presence of a thin epithelium lining the body cavity (Welsch and Störch 1982). Shortly after that, Nielsen (1985) argued that the chaetognaths were more closely related to the acoelomate acanthocephalans. The debate continued over whether chaetognaths were coelomates or pseudocoelomates and the fact that chaetognaths did not satisfy Hyman's classical definition of a coelom as both "bounded on all sides by tissue of entomesodermal origin and lines by peritoneum" accounted for much of the continued skepticism over the interpretation of chaetognaths as coelomates (Shinn and Roberts 1994). Shinn and Roberts (1994) argued that Hyman's definition was irrelevant because the coelomic lining of chaetognaths is ultrastructurally very similar to the coelomic

lining in other coelomates (e.g. in many small polychaetes, pterobranchs, and enteropneusts and in the tentacular region of lophophorates). Their ultrastructural and developmental observations of hatchling chaetognaths determined the presence of a true coelem (Shinn and Roberts 1994).

Gross morphological and knowledge only of *Spadella spp.*, a benthic species of Chaetognatha, lead Hyman (1951b) to conclude that *Spadella* was closely related to, or the ancestor of nematodes (acoelomate) since marine nematodes are benthic infauna, and their presence in the water column is limited. Thus, the most parsimonious assumption for Hyman was that the chaetognath-nematode ancestor was a benthic organism. Moreover, the common perception of nematodes, during that time period, as a cigar-shaped organism with a smooth cuticle similar to *Spadella* seemed to support his conclusion. However, these similarities were based upon the extensive study of one genus of nematodes. In fact, many marine groups of nematodes are considerably different. The Epsilonematidae and the Draconematidae both have several free –living species that are covered with a combination of spines, bristles, cuticular folds, and/or annuli (Bone et al. 1983). Many years later, Halanych (1996) hypothesized that the pelagic chaetognaths arose from benthic species. Based upon Hyman's assumption that the common ancestor of chaetognaths and nematodes was benthic, several morphological modifications, such as swimming or buoyancy control would be strongly favored and allow for a successful invasion of the pelagic environment. Chaeatognaths have well-developed fins, and some nematodes have caudal alae and bursae (Halanych 1996). Halanych also hypothesized that if the chaetognath-nematode ancestor was a

benthic organism, selective pressures may have acted to increase the surface area of the

structures similar to alae and to modify the hardness and size of the anterior spines,

resulting in a chaetognath lineage (Halanych 1996).
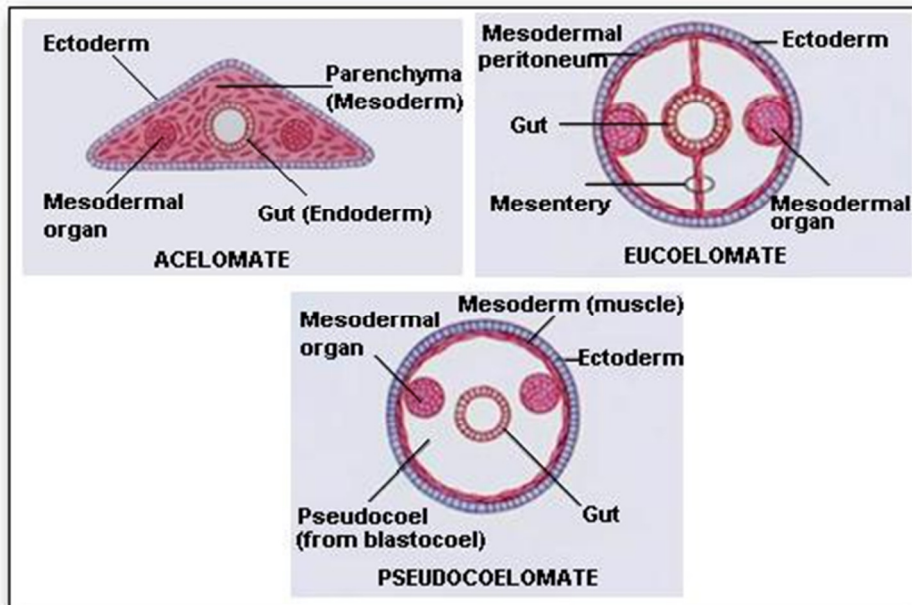
Fig. I. 3. Schematic of various coelomic conditions.
(http://www.tutorvista.com/content/biology/biology-iii/animal-
kingdom/animal-organisation.php)

The first cladisitc analysis, based upon gross and ultrastructural morphology, of the

diverse metazoan phyla placed chaetognaths within the clade of aschelminths (Meglitsch

and Schram 1991) Nielsen (1985), using morphological characteristics, also placed the

chaetognaths in a clade of aschelminths that included Gastrotricha, Nematoda, Nematomorpha, Priapulida, Kinorhyncha, Loricifera, Rotifera, and Acanthocephala; in his cladistic analysis, however, Chaetognatha constituted the sister group to Syndermata but did not group with the remaining aschelminths. Zrzavy et al (1998) used morphological and 18s rRNA genes to place the Chaetognatha as the sister group of a clade containing all other protostome phyla except lophophorates i.e. mollusks, annelids, and brachiopods (Fig. I.4.A).Other analyses (as reviewed by Eernisse 1998) placed the chaetognaths either as a sister group to Nematomorpha, within the Ecdysozoa, with the Nematoda at the base of the tree, or in a clade containing Nematomorpha+Nematoda+Chaetognatha within Ecdysozoa (Fig. I.4.B; Eernisse1998). Littlewood et al. (1998) placed Chaetognaths as a sister group to Gnathosomulida with both as a sister group to Nematoda, within Ecdysozoa.
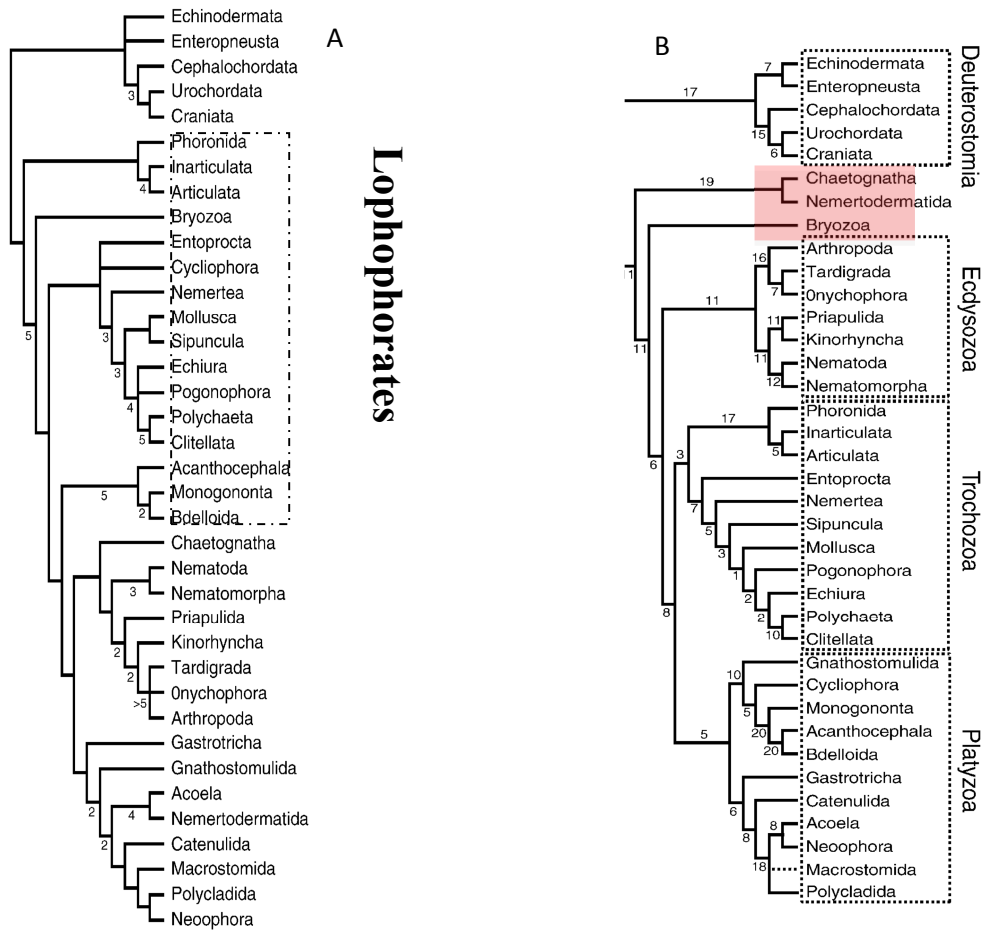
Fig.I. 4. A) Strict consensus of 16 trees of 455 steps (consistency index = 0.466; retention index = 0.701)based on the morphological data of Zrzav´y et al. (1998) with Bremer support values >1 are indicatedB) Tree derived using 18s and morphological data from Eernisse (1998) with chaetognaths grouping with nematodes.

1.3 Embryology

Hyman (1959) used embryological characters, including blastopore fate, as the

primary reason for the designation of chaetognaths as deuterostomes (Fig. I.5). In

chaetognaths, the blastopore is located posteriorly before it is lost, and the anus forms de

novo (Hyman 1959). Chaetognaths share common characteristics with deuterostomes during their ontogeny: radial cleavage, a blastopore at the rear end of the body, and a post-anal tail, however, they do not pass through the dipleura stage that is seen in every deuterostome phylum (Hyman 1959). Chaetognaths are also direct developers that hatch from eggs as small adults and do not exhibit larval stages as is the norm for most aquatic invertebrates (Ball and Miller 2006). Willmer (1990) suggested that the development of the chaetognath nervous system is more like that of protostomes based upon patterns of evolutionary characteristics among invertebrates. Harzch & Müller (2007) reexamined the development of the central nerve center of *Sagitta* and also concluded that its placement is with the protostomes.

Role of Genetics in Determining the Chaetognath Lineage

1.4 18S rRNA

In the absence of unambiguous morphological information; DNA sequence data can be a very useful tool for determining the phylogenetic placement of a group of organisms. DNA encoding for 18S ribosomal RNA (rRNA) gene has been widely used to reconstruct ancient phylogenetic events (hundreds of millions of years ago) because of its slow rate of change and conserved evolution (Field et al. 1988). Phylogenetic analyses of 18S rRNA suggest that the fate of the blastopore has been more variable across animal phyla and that embryology is not reliable for cladistic placement. Telford and Holland's study of chaetognath 18S rRNA (1993) could not place them within the deuterostome group and provided the first molecular evidence that contradicted

Hyman's (1959) conclusions. They proposed that the chaetognaths likely evolved
independently during an early stage of metazoan evolution and that the similarities in
terms of embryology to deuterostomes may be the result of convergence (Telford and
Holland 1993).
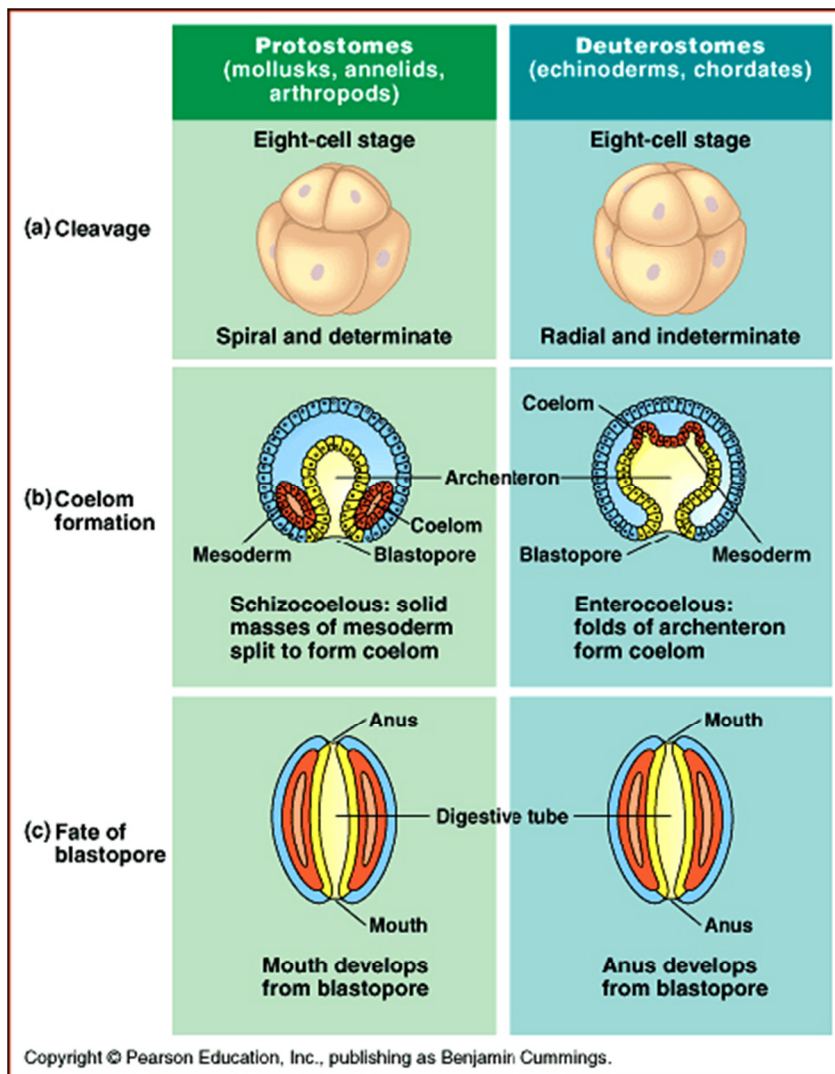
Fig. I.5. Schematic of Protostome vs. Deuterostome
Development (Arendt et.al. 2001)

Interestingly, however, their 18S rRNA sequence was from a pelagic chaetognath, *Sagitta elegans,* whereas Hyman (1959) had studied a benthic species. An extensive survey of 18S rRNA in vertebrates and invertebrates by Wada and Satoh (1994) also found no evidence to place the chaetognaths within the deuterostome group.

The use of 18S rRNA for phylogenetic analysis has been widespread. Studies of 18S r RNA sequences have been used to infer phylogenetic history across a very broad spectrum, from studies among the basal lineages of life to relationships among closely related species and populations. The reasons for the systematic versatility of18S rRNA include the numerous rates of evolution among different regions of 18S rRNA (both among and within genes), the presence of many copies of most 18S rRNA sequences per genome, and the pattern of concerted evolution that occurs among repeated copies. These features facilitate the analysis of 18S rRNA by direct RNA sequencing, DNA sequencing (either by cloning or amplification), and restriction enzyme methodologies (Hillis &Dixon 1991).

Chaetognaths have extremely divergent 18S rRNA sequences in comparison with other metazoans, and in all the analyses published thus far, have tended to group with other divergent sequences such as those of nematodes or gnathostomulids. Although 18S rRNA has been the gene of choice for phylogenetic analysis because of the large number of sequences available and because its properties are well known, there are several problems associated with using this gene. The 18S rRNA gene is difficult to align among diverse metazoan taxa, particularly when sequences from rapidly evolving taxa are included. For chaetognaths, the limited number of species analyzed to date appear to

have rapidly evolving 18S rRNA genes, making it difficult to accurately place them within Metazoan phylogeny due to the 'long branch attraction' phenomenon that, no matter the method for tree reconstruction, groups all species having long branches together regardless of their true relationship (Telford and Copley 2005). One solution is to identify and sequence genes that evolve at a rate comparable to that in other metazoans (Garey et al. 1998).

1.5 Cytochrome Oxidase Gene

The cytochrome oxidase mitochondrial gene encodes for metallo-hemo-proteins (i.e. enzymes) that catalyze the last step of cellular respiration in eukaryotes and prokaryotes. The three main subunits of this gene, COI, COII, and COIII, which represent the catalytic core of the enzyme are conserved structurally and functionally in all organisms making them good targets to identify the origin of the Chaetognatha (Bonnefoy et al. 1994).

Cytochrome c oxidase subunit 1 (COI) is the terminal catalyst in the mitochondrial respiratory chain and is involved in electron transport and proton translocation across the membrane. The CO1 gene is often used as the mitochondrial marker for evolutionary study because it is the largest of the three subunits and the protein sequence contains highly conserved functional domains and variable regions (Morlais & Severson 2002). Villa et al. (2006) also found that the cytochrome oxidase II genes, encoding for metabolic and structural proteins, are also excellent targets for phylogenetic analyses. These genes are conserved and the alignment of their sequences is less ambiguous compared to rRNA (Villa et al. 2006).

COII genes have been used for *Sagitta setosa* to determine phylogeographic distribution (Peijnenburg et al. 2004), and only until very recently, COI sequences for 14 species of Chaetognaths were published for the Barcode of Life Project (Jennings et al. 2010). However, the database of COI and COII genes for Chaetognatha as well as other organisms is not yet large or consistent enough to resolve the placement of the chaetognaths. Interestingly, Barthelemy et al. (2006) examined the complete mitochondrial genome (mtDNA) of two chaetognaths, *Spadella cephaloptera* and *Paraspadella gotoi* and determined them to be highly unusual in that their mtDNA is thus far the smallest for all metazoans due to missing genes and appears to have an otherwise gene arrangement that suggests a phylogenetic affinity with annelids and mollusks.

1.6 Homeobox 'Hox' Genes

Hox genes are a group of related genes that specify the anterior- posterior axis and segment identity of metazoan organisms during early embryonic development. These genes are critical for the proper number and placement of embryonic segment structures. Hox genes are defined by a DNA sequence known as the homeobox, which is a sequence of 180 nucleotides that encode for a protein domain known as the homeodomain. Recently, Hox genes have provided an insight into some problematic phyla that have been puzzling zoologists and embryologists for several decades (Papillon et al. 2003). In 2003, Papillon's group discovered a Hox mosaic median/posterior gene in *S. cephaloptera* which was the first set of Homeobox genes reported for any

chaetognath species. These genes were of interest in respect to metazoan phylogeny and the reconstruction of the evolutionary history of the Hox cluster. They also examined this newly discovered gene further and found that it displayed specific median and posterior signatures of the *S. cephaloptera.* If this gene is considered a conversion product, the phylogenetic position of chaetognaths would remain enigmatic. However, another view is that the chaetognath set of Hox genes has retained features of an ancestral state of the Hox cluster before the complete divergence of median and posterior genes. Thus, this mosaic gene provides some evidence that Chaetognatha could be an early off-shoot of the triploblastic lineage that predates the deuterostome/protostome split (Papillon et al. 2003). Another study by Matus et al. (2007) isolated eight Hox genes, one Parahox gene, and *Mox*, a related homeodomain gene, from the pelagic chaetognath, *Flaccisagitta enflata*. Although chaetognath **c**entral class Hox genes lack the Lox5 or "spiralian" parapeptide, a diagnostic amino-acid motif that has been utilized previously to assign lophotrochozoan affinity, they do possess a central class Hox gene that has a partial "Ubd-A peptide" found in both ecdysozoan and lophotrochozoan *Ubx/Abd-A/Lox2/Lox4* genes. Additionally, they reported the presence of two distinct chaetognath posterior Hox genes that possess both ecdysozoan and lophotrochozoan signature amino-acid motifs.

1.7 Whole Genomic Analysis

Whole genomic sequencing analysis has become popular recently to help scientists determine the placement of several organisms in the phylogenetic tree of life. Using the whole genome to infer phylogeny is beneficial for organisms that display either peculiar morphological characteristics or have key phylogenetic positions, like the Chaetognatha. Marlétaz and colleagues (2006) used the expressed sequence tag (EST) approach to create a transcriptome database for *Spadella cephaloptera*. Later, they (Marlétaz et al. 2008) used this database to perform comparisons between organisms. This approach using EST initially suggested that the gene repertory shared by all metazoans is larger than expected (Marlétaz et al. 2008). The identification of a core set of metazoan conserved genes from a large number of organisms provided marker genes for phylogenomic analyses and signature genes as rare genomic changes, which could lead to a reevaluation of animal phylogeny (Marlétaz et al. 2008). The transcript comparisons at various taxonomic scales emphasized the conservation of a core gene set and phylogenomic analysis confirmed the basal position of chaetognaths among protostomes (Marlétaz et al. 2008). To avoid problems associated with phylogenetic reconstruction stemming from long- branch attractions (which infers rapidly evolving lineages to be closely related regardless of their true evolutionary relationships), they focused on short branch sequences to infer phylogenies across various phyla. In doing this, they found the topology obtained after analysis of the dataset supports the branching of chaetognaths with protostomes. Furthermore, striking genetic heterogeneity was detected within the sample population at the nuclear and mitochondrial levels that cannot be explained by

cryptic speciation. These findings reveal both shared ancestral and unique derived

characteristics of the chaetognath genome. These data strongly suggest that the

chaetognath genome is likely the product of an original evolutionary history. The

illustration of these found features among the chaetognaths make these organisms a

pivotal model for comparative genomics that could eventually provide insight into the

evolution of animal genomes (Marlétaz et al. 2008). They also discovered the

Guanidinoacetate N-methyltransferase gene in the chaetognath EST database collection.

This gene is present in cnidarians and deuterostomes, but lost in protostomes, suggesting

that chaetognaths are a stem of the protostome group (Marlétaz et al. 2006; Fig.I.6).
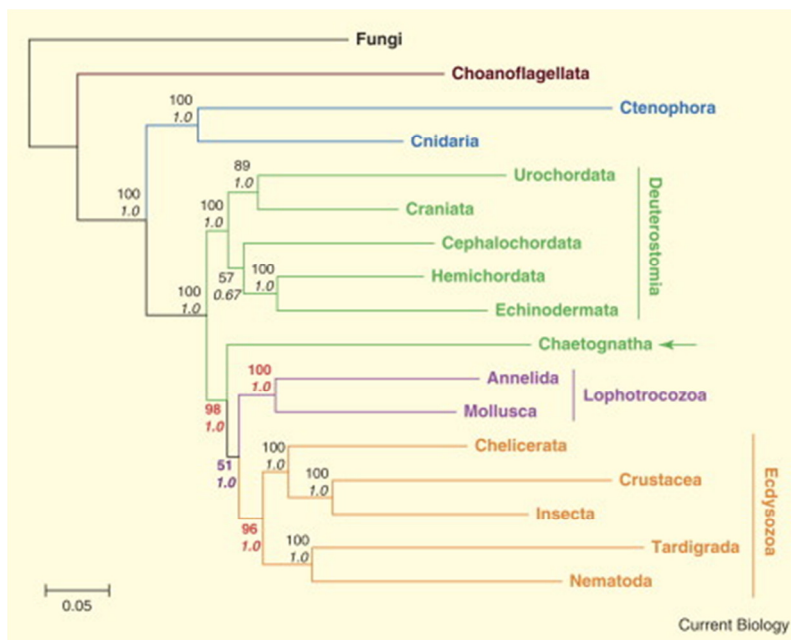
Fig. I. 6. Chaetognaths as basal protostomes. A rooted
maximum likelihood tree based on the analysis of a concatenated
79 proteins and 11,667 positions in a ribosomal protein data set.
The green line from deuterostomes to chaetognaths indicates the
ancestral state of embryological deuterostomy for bilaterians.
(Marlétaz et al. 2006).

Even though several different types of studies have been used over the years to resolve the chaetognath debate, their placement still remains a mystery. Depending on what type of study you perform (molecular, morphological or embryological) the ambiguities in this organism seem to still challenge scientists. Figure I.7 illustrates the history of chaetognath placement according to analysis of different genes i.e. 18S rRNA, COI, whole genome, and mtDNA as reviewed by Harzch and Müller (2007). A major issue is the inconsistency of the class of genes that have been sequenced in each of the genera (Fig. I.2) making direct phylogenetic analyses inadequate.
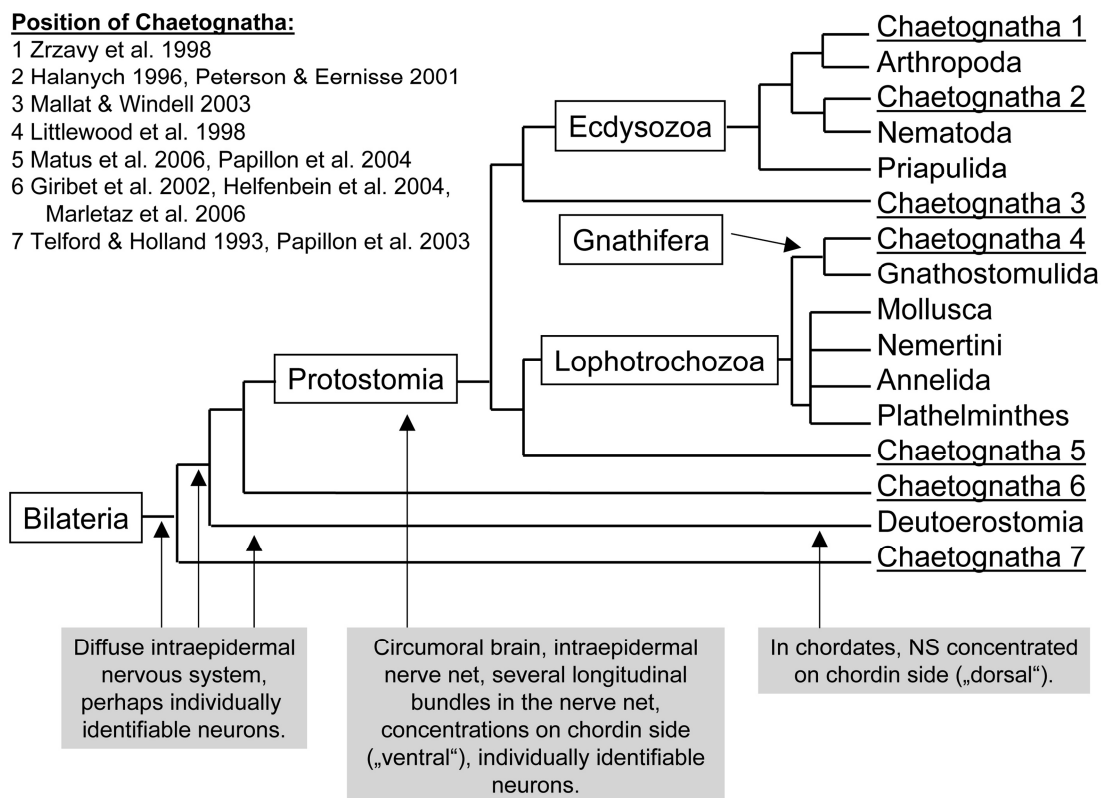
Fig. I. 7. History of placement of chaetognaths as determined by analysis of different genes, i.e. 18S rRNA, COI, whole genome, or mtDNA (Harzch and Müller, 2007).

1.8 Research Objectives

For my Master's thesis research, I characterized 18S rRNA, COI, and COII genes of purportedly new species of pelagic chaetognaths to aid in determining their placement in the phylogenetic tree of life.

Hypothesis 1: **The analysis of pelagic chaetognath 18S rRNA sequences will not be the best phylogenetic marker due to their high rate of evolutionary change and long-branch chain attraction.**

Objective 1. 18s rRNA Genes

Most 18S rRNA analyses of chaetognaths are limited to only *Spadella spp.*, a benthic chaetognath, and *Sagitta spp.*, a pelagic chaetognath. According to previous studies (Halanych 1996), the 18S rRNA genes for chaetognaths are highly diverse indicating rapid evolution. By including additional sequences from purportedly different species of chaetognaths, more similarities between this group may be identified and will resolve their origins. Moreover, benthic and pelagic chaetognaths may represent very different branches on the tree. Meaning that the usage of 18S rRNA may not be the best marker to infer phylogenetic placement of these organisms. In part, due to the fact that they are very large (over 1500 bp in length), and they have a tendency to have long- branch chain attractions which can hinder the phylogenetic inference (Halanych 1996).

Hypothesis 2: **Phylogenetic analysis of pelagic chaetognaths using the highly conserved COI and COII genes will confirm most closely related phyla as compared to 18S rRNA because the sequence alignment is less ambiguous.**

Objective 2. Cytochrome Oxidase Genes

Although 18s rRNA has been widely used in phylogenetic studies, the evolution of one gene may not represent the evolution of the entire genome. Therefore it is necessary to separately sample as many additional independent genes as possible and compare the phylogenies derived from these genes to see whether they support or contradict each other. Moreover, genes coding for metabolic and structural proteins such as cytochrome oxidase I and II are receiving increasing attention. These genes are conserved and the alignment of their sequences is less ambiguous compared to 18S rRNA (Villa et al. 2006). By using the smaller cytochrome oxidase genes, I hope to gain a better understanding of the phylogenetic placement of this mysterious organism.

CHAPTER II

MATERIALS AND METHODS

2.1 Sampling

Sampling was conducted at Texas A&M University at Galveston (Fig.II.1) small boat

basin. Samples were collected on May 2008 and, May 2009. The May 2009 sampling

was taken after Hurricane Ike hit Galveston Island in September 2008. Samples were

collected both times using a zooplankton net with a mesh size of 333μM. A 7% solution

of cold magnesium chloride was added to the water sample to inhibit the movement of

chaetognaths while individually selecting them under a dissecting microscope. Each

chaetognath was rinsed with filtered seawater before placing them into 0.2mL PCR tubes

for DNA extraction. After DNA extraction, samples were kept in -20° C freezer until

further analysis was performed.

2.2 DNA Extraction

The REDExtract-N-Amp Tissue PCR kit from Sigma (catalog # XNAT) was used for

DNA extraction and PCR amplification. The DNA extraction from animal tissue

protocol was used with a few modifications. The protocol uses 100μL of Extraction

Solution and 25μL of Tissue Preparation Solution, only 1/5 of each volume was used for

each extraction. After incubating the tissue in the mixture above for 10 minutes at room

temperature, the samples were then incubated at 95°C for 3 minutes. To finalize the

extraction, 20µL of Neutralization Solution B was added to the sample. After this step

the DNA is ready for PCR amplification.
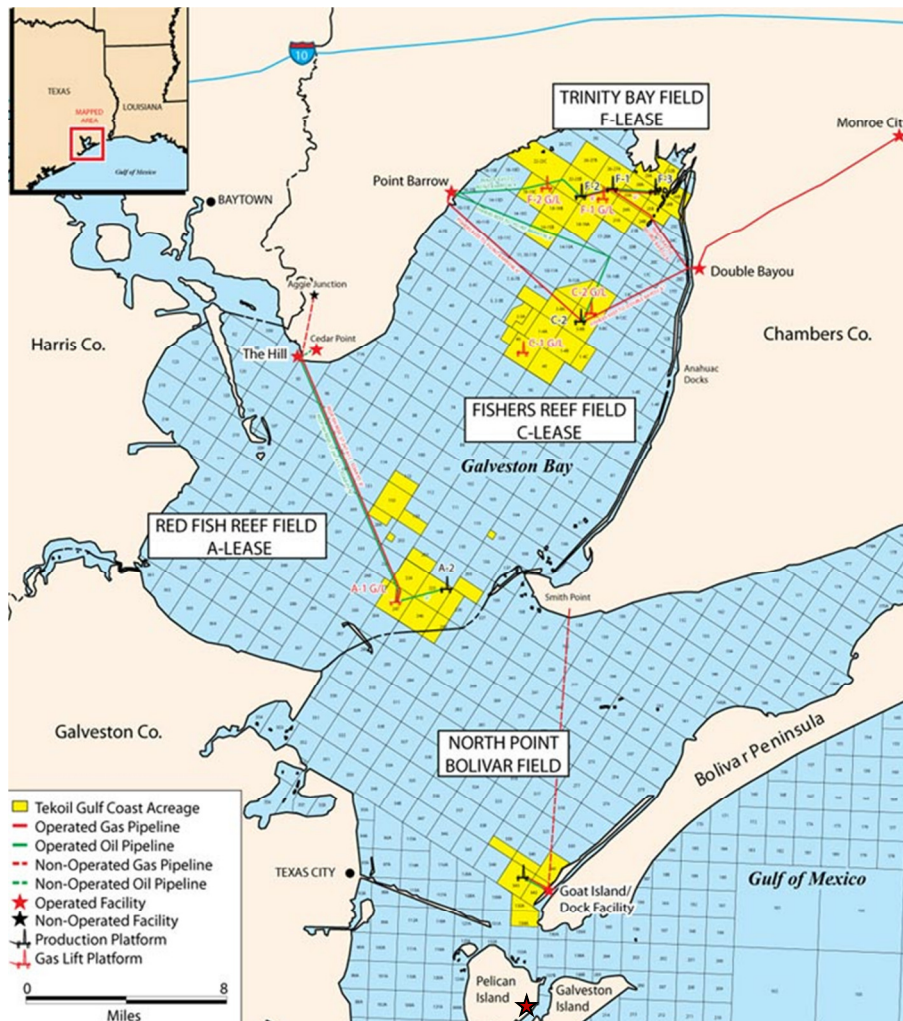
Fig. II.1. Map of Galveston Bay with Pelican Island (29° 17'8" N, 94° 49' 38" W).

 The dark red Star indicates sampling area.

(www.tekoilandgas.com/maps/operations_area_map)

2.3 PCR Modifications

After several attempts to achieve amplification of my three primers (COI, COII, and 18S), I was able to resolve my PCR program with some minor modifications. For both the cytochrome oxidase 1 and cytochrome oxidase II primers, I had an initial denaturing step of 95° C for 2 minutes. Then another denaturation step of 95° C for 30 seconds, followed by the annealing step of 50° C for 1 minute. After this, an elongation step of 72°C for 45 seconds, these three steps being repeated for 35 cycles. And lastly, an extension step of 72°C for 7 minutes. The 18S rRNA was the same except the annealing temp was changed from 50°C to 60° C.

Table II. 1. Primers for amplification of 18s rRNA and Cytochrome oxidase I and Cytochrome oxidase II Genes

Name	Target gene	5'---->3'	Reference
18S F	18s rRNA	CTGGTTGATCCTGCCAG	Wada & Satoh 1993
18S R	18s rRNA	CACCTACGGAACCTTG	Wada & Satoh 1993
LCO1 490	Cytochrome oxidase I	GGTCAACAAATCATAAAG ATATTGG	Cooper et al. 2007
HCO2 198	Cytochrome oxidase I	TAAACTTCAGGGTGACCAA AAAATCA	Cooper et al. 2007
COII F	Cytochrome oxidase II	GGAGCATCTCCTTTAATAG AACA	Peijnenburg et al. 2004
CoII R	Cytochrome oxidase II	CCACAATTTCTGAACATTG ACCA	Peijnenburg et al. 2004

2.4 ARDRA

Amplified Ribosomal DNA Restriction Analysis (ARDRA) was used to analyze each

DNA sample. The amplicons were digested using the restriction enzymes HaeIII and

RsaI (Promega) and run on an 8% polyacrylamide gel. HaeIII is a high-frequency cutting

restriction enzyme that has a high average number of restriction sites and comes from

the organism, *Haemophilius aegyptius*. RsaI was used as a secondary restriction

endonuclease to increase fingerprint resolution. RsaI comes from *Rhodobacter*

sphaeroides. The ARDRA digestion mixture used was 15µl of PCR product (should be

600-800ng estimated visually), .75µl HaeII (=7.5 U), and .75µl RsaI (=7.5U) and then

add 1.5µl 10 Buffer C (provided with restriction enzymes), and finally add PCR water to

have a final volume of 25µl. The gels were stained with ethidium bromide for 20

minutes and pictures were taken using the gel documenter. The stained gels were using

the BioNumerics software (Applied Maths: Sint-Martens-Latem, Belgium).The

recognition sequences for both of these enzymes are: HaeIII (5'GG^CC3') and RsaI

(5'GT^AC3').

2.5 Sequencing

The PCR products from cytochrome oxidase 1(COI) and cytochrome oxidase 2

(COII) and 18S rRNA genes were sequenced using the ABI 3130 sequencer 4 capillary

36cm system. The default AB SEQ POP 7 36BDV3 protocol setting was also used to run

the samples. The PCR products were cleaned using the ABI sequencing protocol. This

sequencing protocol has three steps (see below). All genes were analyzed using the

MEGA 4 software analysis tool, using the Clustal IW with the following settings: Gap

opening penalty of 5 and Gap extension penalty of 4 and a 5% delay of divergent cutoff.

These setting were used for both the pairwise and multiple alignment parameters. To my

knowledge, no known database is available for the COI and COII genes for

chaetognaths. An extensive search of the COI and COII genes was conducted and

imported into a database. The goal is to utilize these databases as a tool to elucidate the

chaetognath's ancestry.

The sequencing protocol has three parts. Part I: The Exo-Sap It protocol requires 5µL

of PCR product with 2µL of Exo-Sap It solution (USB cat # 78250). The Exo-Sap It

solution is 0.7µL of 10 X SAP (Shrimp Alkaline Phosphatase) buffer, 0.5µL of Exo 1,

.05µL of Shrimp Alkaline Phosphatase and 0.3µL of dd water (PCR grade). Mix this

solution thoroughly and run the Exo-Sap It program on the thermocycler. The incubation

times for this program are 37°C for 15 minutes followed by an 80°C incubation for 15

minutes. Hold the samples at 10°C until the solutions are ready for Part II. PartII is

known as the Cycle Sequence, this step amplifies the cleaned product from Part I. Take

2µL of the cleaned product (from Part I.) and add 8µL of the cycle sequence mix

(Applied BioSystems cat #4337455). This 8 µL sequence mix contains 1µL of dd water,

1µL of the Big Dye v 3.1, 2µL of the 5x Big Dye Sequence Buffer and 4µL of the

specific primer you would like to sequence at a dilution of 1:3. Then mix thoroughly and

run the Cycle Sequence PCR program. This PCR program has an initial denaturation

step of 95°C for 10 seconds then an annealing step of 50°C for 5 seconds, followed by

and elongation step of 60° C for 1 minute and 15 seconds. Repeat these steps for 14

cycles. The next step has the same initial denaturing and annealing temperatures and times, but add 15 seconds to the elongation step still at 60° C. Repeat this for 4 cycles. Lastly, use the same initial denaturation and annealing temperatures and times, but add 30 seconds making the elongation of 60°C for 2 minutes. Repeat this for 4 cycles. Hold the PCR program at 4°C. The third part uses the BigDye Xterminator Purification kit (Applied BioSystems cat# 4376486). This last process uses a BDX cleaning method. Add 11µL of the SAM solution and 2.5µL of the BDX cleaning bead solution (found in the BigDye Xterminator Purification kit) directly to the cleaned samples, then vortex on a shaker at 1900 rpm for 30 minutes. Centrifuge the samples at 1,000 X g for 2 minutes. This centrifugation allows the BDX beads to fall to the bottom. Only load the supernatant into the 96- well plate for sequencing, taking caution to not pipette any beads. The beads will hinder your results if added to the samples to be sequenced.

2.6 Phylogenetic Analyses

Phylogenies, or evolutionary trees, are the basic structures necessary to think clearly about differences between species, and to analyze those differences statistically. They have been around for over 140 years, but statistical, computational, and algorithmic work on them is barely 40 years old. Phylogenies are inferred with various kinds of data. Some of the most widely used ones are discretely coded characters, molecular sequences, gene frequencies, and quantitative traits (Felsenstein 2004). A large (and ever increasing) number of methods have been described to convert information in sequences

into an evolutionary tree, which raises the inevitable question of how to choose the right tree with all the possibilities (Page and Holmes 1998).

Distances versus Discrete Characters

Distance methods first convert aligned sequences into a pairwise distance matrix, and then input that matrix into a tree building method. The discrete methods consider each nucleotide site directly. The trees obtained by parsimony (a discrete method) and minimum evolution (a distance method) are identical in topology and branch lengths. The parsimony analysis identifies seven substitutions and places them on the five branches of the tree. The distance tree apportions the observed distances between the sequences over the branches of the tree. While both methods arrive at the same estimates of the lengths in one branch, the parsimony method requires one change at each of the seven sites for a total of seven changes. The parsimony method also gives us which site contributes to the length of each branch. Also, discrete methods allow us to infer the attributes of extinct ancestors, which give us insight into molecular evolution (Page and Holmes 1998).

Sequences

1	2	3	4	5	6	7
T	T	A	T	T	A	A
A	A	T	T	T	A	A
A	A	A	A	A	T	T
A	A	A	A	A	A	A

Distances

```
2 |  3
3 |  5   4
4 |  5   4   2
  +——1———2———3
```

Parsimony

A ╱ 1
  ╱2   4   5   6 ╱ C
B ╱ 3   ╱ 7   D

Distance

A

2   C

2   1

B   1   1   D
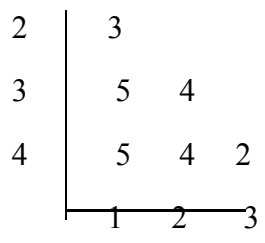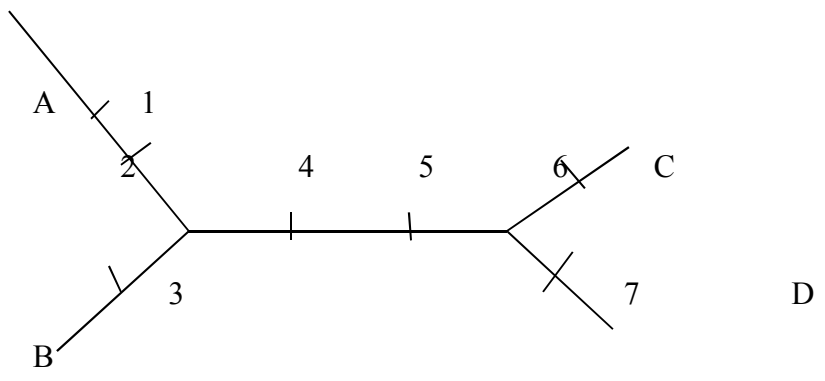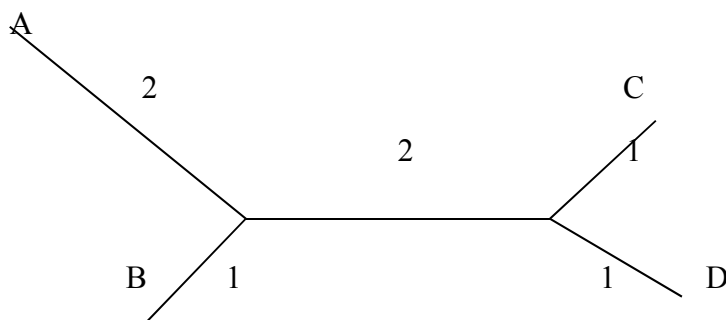
Clustering Methods versus Search Methods

Cluster methods follow a set of steps (an algorithm) and arrive at a tree. Cluster

methods are easy to implement and always produce a single tree, but have limitations.

This method does not allow for the evaluation of competing hypotheses. Optimal criteria

is another tree building method that assigns each tree a "score" or rank which is a

function of the relationship between tree and data. Examples of this method are

maximum parsimony and maximum likelihood. Optimality methods allow us to evaluate

the quality of any tree, hence we can compare how well competing hypotheses of

evolutionary relationships fit the data (Page and Holmes 1998).

Distance Methods

Distance methods are based on the idea that we know the actual evolutionary distance

between all members of a set of sequences. However, distances are rarely, if ever, exact

tree metrics. For this reason, we must use goodness of fit measures when working with

distance methods. One class of the goodness of fit method seeks the metric tree that best accounts for the observed distances. The second class of the goodness of fit method seeks the tree whose sum of branch lengths is the minimum (known as minimum evolution).

Minimum Evolution

The minimum evolution tree method is similar to parsimony, but the length (L) is computed from pairwise distances between the sequences rather than from the fit of the individual nucleotide sites to a tree. To use this method we need to be able to compute the length of any tree, which are not always biologically valid (Page and Holmes 1998). In the minimum evolution method the tree is fit to the data, and the branch lengths are determined, using the unweighted least squares method. The least squares trees are determined for different topologies, and the choice is made among them by choosing the one of shortest total length. Thus this method makes partial use of the least square criterion. It uses two criteria at the same time, one for choosing branch lengths, another for choosing the tree topology. The minimum evolution method was first used my Kidd and Sgaramella-Zonta (1971), who used the sum of the absolute values of branch lengths. Its present day use comes from its independent invention by Rzhetsky and Nei (1994). They used the sum of the branch lengths (Felsenstein 2004).

Minimum Evolution requires an amount of computation similar to least squares, since it uses least squares to evaluate branch lengths for each tree topology. Bryant and Waddell (1998) used methods for speeding up the least squares calculations that also speed up the minimum evolution methods as well. In 1974, Beyer et al. used an early

variant of minimum evolution that did not use least squares to infer the branch lengths. Instead, required that the path lengths between all pairs of species remain longer than, or equal to, the observed distances (Felsenstein 2004).

Unweighted Pair Group Method with Arithmetic Means (UPGMA)

UPGMA is an example of clustering method. This method first finds the pair of taxa with the smallest distance between them and defines the branching between them as half of that distance-in effect placing a node at the midpoint of the branch. It then combines the two taxa into a "cluster" and rewrites the matrix with the distance from the cluster to each remaining taxa. Since the "cluster" serves as a substitute for two taxa, the number of entries in the matrix is now reduced by one. This process is repeated on the new matrix and reiterated until the matrix consists of a single entry. That set of matrices is then used to build up the tree by starting at the root and moving out to the first two nodes represented by the last two clusters (Hall 2008).

Neighbor Joining

Neighbor joining (NJ) is a widely used method for tree building which combines computational speed with uniqueness of result- most implementations give a single tree. Neighbor joining is a clustering method rather than and optimality method, and hence suffers from the limitation that it does not optimize a criterion of fit between tree and data. However, it is a good method for estimating the minimum evolution tree (Page and Holmes 1998). Neighbor joining, like the least squares methods, is guaranteed to recover the true tree if the distance matrix happens to be an exact reflection of a tree. Neighbor joining is also useful to rapidly search for a good tree that can then be improved by other

criteria. Pearson, Robins, and Zhang (1999) use it while retaining nearly-tied trees, choosing among them by minimum evolution or least squares criteria. Ota and Li (2000) use neighbor joining and bootstrapping to find an initial tree and identify which regions of it are candidates for rearrangement. Then they use maximum likelihood for further search. This results in a substantial improvement in speed over pure likelihood methods (Felsenstein 2004).

Two modifications of neighbor joining have been developed to allow for differential weighting in the algorithm to take into account differences in statistical noise. Gascuel (1997) has modified the neighbor joining algorithm to allow for the variances and covariances of the distances, in a simple model of sequence evolution. Gascuel's method, called BIONJ, thus comes closer to what generalized least squares would give, though it is still an approximation. Bruno, Soccci, and Halpern (2000) developed a weighted neighbor joining method which uses weights in the formula. This method also uses the exact formula for the variance of a Jukes-Cantor distance, which is approximate for other models of DNA change, but more correctly with the very high variances of distances when tips are far apart on the tree (Felsenstein 2004).

Objections to Distance Methods

When considering distance methods you must distinguish between methods of constructing the trees and methods for obtaining the distances. If the estimates of evolutionary distance are poor then the performance of a distance method maybe adversely affected, which may not be a true reflection of the merits of the tree building method itself. If the original data are in the form of distances, such as those obtained

from DNA hybridization studies, then the only option is to use a distance method. However, if the sequences are available they can be analyzed directly or convert into distances. Once converted into distances, the evolution of the individual sites, or categories of sites on a tree, cannot be traced. This leads only to an overall estimate of the relationship between the tree and the data (Page and Holmes 1998).

Discrete Methods

In contrast to distance methods, discrete methods operate directly on the sequences, or on functions derived from the sequences, rather than on pairwise distances. By doing this, they avoid the loss of information that occurs when sequences are converted into distances. The two major discrete methods are maximum parsimony (MP) and maximum likelihood (ML). Maximum parsimony chooses the tree, or trees, that require the fewest evolutionary changes. Maximum likelihood chooses the tree, or trees, that of all trees is the one that is most likely to have produced the observed data (Page and Holmes 1998).

Maximum Parsimony

The data for maximum parsimony comprise individual nucleotide sites. For each site the goal is to reconstruct the evolution of that site on a tree subject to the constraint of invoking the fewest possible evolutionary changes (Page and Holmes 1998). Most biologists are familiar with the usual notion of parsimony in science, which essentially maintains that simpler hypotheses are preferable to more complicated ones. Methods for estimating trees under the criterion of parsimony equate "simplicity" with the explanation of attributes shared among taxa as due to their inheritance from a common ancestor. In general, parsimony methods for inferring phylogenies operate by selecting

trees that minimize the total tree length: the number of evolutionary steps (transformations from one character state to another) required to explain a given set of data. Parsimony methods must distinguish between the optimality criterion (minimal tree length under a specified set of restrictions on permissible character-state changes) and the actual algorithm used to search for optimal trees. Early descriptions of parsimony methods were presented in a way that tended to obscure the boundaries between criteria and algorithms (Hillis et al. 1996). Parsimony is advantageous because it is relatively straightforward to understand and it makes few assumptions about the evolutionary process. The principal objection to parsimony is that under some models of evolution it is not consistent. This inconsistency is termed "long branches attract". This scenario happens when there are two unrelated sequences that are each separated from their ancestor by a long edge. The problem of long branches attracting is most likely to occur when rates of evolution show considerable variation among sequences, or where the sequences being analyzed are quite divergent (Page and Holmes 1998).

Maximum Likelihood

Maximum likelihood methods of phylogenetic inference evaluate a hypothesis about evolutionary history in terms of the probability that a proposed model of evolutionary process and hypothesized history would give rise to the observed data. It is thought that a history with a higher probability of giving rise to the current state of affairs is a preferable hypothesis to one with a lower probability of reaching the observed state. Maximum likelihood estimation was first used in phylogenetic inference by Cavalli-Sforza and Edwards (1967). However, because they did not use sequence data, this work

remained relatively obscure. In 1992, Thorne and Kishino applied the maximum likelihood method to amino acid sequence data. And later in 1993, Felsenstein brought maximum likelihood framework to nucleotide-bases phylogenetic inference (Hillis et al. 1996).

In addition to its consistency properties, maximum likelihood is useful because it often yields estimates that have lower variance than other methods (i.e., it is frequently the estimation method least affected by sampling error). It also tends to be robust to many violations of the assumptions used in its models. Part of its power in this respect is that many models of sequence evolution that assume identical distributions across sites can safely assume that the actual substitution processes taking place at different sites have much in common, even if they are not exactly identical. A few parameters describe the major components that determine evolution. The overall results of both improved compensation for superimposed changes and of sampling variance is that even with very short sequences, maximum likelihood tree inference tends to outperform alternative methods (like parsimony or additive distances) when evaluated under many models of sequence evolution (Hillis et al. 1996).

Maximum likelihood requires three elements, a model of sequence evolution, a tree and the observed data. The tree specifies both the topology (branching order) and the branch lengths. The maximum likelihood method must solve two problems: (1) for a given tree topology, what set of branch lengths make the observed data most likely; and (2) which tree of all the possible trees has the greatest likelihood (Page and Holmes 1998).

Phylogenetic analysis seeks to infer the history that is most consistent with a set of observed data. To apply a maximum likelihood approach, a concrete model of the evolutionary process that accounts for the conversion of one sequence into another must be specified. This method also evaluates the probability that the chosen evolutionary model will have generated the observed sequences; phylogenies are then inferred by finding those trees that yield the highest likelihoods (Hillis et al. 1996).

Objections to Likelihood

Likelihood requires an explicit model of evolution, which may be seen as both a strength and a weakness. It is a strength because it makes us aware of the assumptions being made- in other methods these assumptions are often only implicit and hence may be overlooked. The information we require to infer the tree presupposes that we have the tree in the first place. One approach is to choose the combination of model and parameters that maximizes the likelihood. This requires that we search for the best model as well as the best tree, which greatly increases the computational difficulty. Computing the likelihood itself is computationally time consuming. It has been shown recently that more than one maximal likelihood value may exist for a given tree (Page and Holmes 1998).

Bootstrapping

The simplest test of phylogenetic accuracy is the bootstrap; it is rare now to see a tree without it. Bootstrapping essentially tests whether your whole dataset is supporting your tree: or if the tree is just a marginal winner among many nearly equal alternatives. This is done by taking random subsamples of the dataset, building trees from each of these

and calculating the frequency with which the various parts of your tree are reproduced in each of the random subsamples. If group X is found in every subsample tree, then its bootstrap support is100%, if it is found in only two-thirds of the subsample trees, its bootstrap support is 67%. Each of the subsamples is the same size as the original, which is accomplished by allowing repeat sampling of sites; that is, random sampling with replacement. It is a simple test, but bootstrap analyses of known phylogenies (viral populations evolved in the laboratory) show that it is a generally dependable measure of phylogenetic accuracy, and that the values of 60% or higher are likely to indicate reliable groupings (Baldauf 2003). In a sense the bootstrap mimics the first method of estimating sampling error, but instead of sampling from the population it re-samples from our sample. Each re-sampling is a pseudoreplicate. From each pseudoreplicate we derive an estimate of the parameter that is being measured, such as the mean height of a population. Bootstrapping can be applied to phylogenies by generating pseudoreplicates from the sequence data (Baldauf 2003).

The interpretation of bootstrap values has been both murky and controversial. Felsenstein (1985) proposed that bootstrap values of 95% or greater be considered statistically significant and indicate "support" for a clade; alternative nodes can be rejected if they occur in less than 5% of the bootstrap estimates. However, bootstrap confidence levels apply to single nodes—they are not joint confidence statements. Thus, although two clades may each be supported at 95% and are thus not contradictory, the confidence interval that includes both clades may be only 90%, and the joint confidence drops as additional nodes are considered. Joint confidence will thus be necessarily low

for a large tree, even if all nodes are strongly supported. A majority-rule consensus tree summarizing all of the bootstrap replicates provides a set of non contradictory nodes, each with a rejection probability below 50%, and can be interpreted as an "overall bootstrap estimate of the phylogeny" (Soltis and Soltis 2003).

One important caveat concerning the bootstrap is that this technique makes the assumption that nucleotide sites are independent and identically distributed. This means that each site is independent of every other site, and that there is a single distribution of rate of evolutionary change across all the sites. A further consideration is that the results of bootstrapping are often summarized using a majority-rule consensus tree showing the frequency of each split that occurs in at least half of bootstrap trees. If one or more sequences have uncertain relationships they may appear in very different positions in the bootstrap tree (they float over the tree), resulting in a general lowering of bootstrap values for those parts of the tree over which the sequences float. Hence, parts of the tree which are actually quite robust may have spuriously low bootstrap values. This problem can be addressed using other consensus methods (Page and Holmes 1998).

CHAPTER III

RESULTS

3.1 Preliminary Results

I initially chose 4 chaetognath samples to send off to Yale University for sequencing

after amplification with COI and 18S rRNA primers. (Refer to Appendix A. Table A.1.).

I wanted to confirm the samples as chaetognaths and also BLAST the sequences into

GenBank to confirm most closely related species for each sample (Refer to Appendix A.

Tables A.2-5). After confirming most closely related species, trees were generated for

each initial sample using the Tree view feature of the BLAST Web service (Refer to

Appendix A. Figs. A.1-4). This tree was calculated from a global multiple sequence

alignment and is therefore more accurate than the distance tree that can be generated

from the BLAST results (*www.ncbi.nlm.nih.gov).* The trees generated from each initial

sample showed that the samples grouped with various protostome phyla (Refer to

Appendix A. Figs. A.1-4). According to the preliminary results I concluded that more

samples needed to be analyzed using more genes and also more statistical methods for

phylogenetic tree reconstruction.

3.2 ARDRA Analyses

After analyzing the initial results, 32 chaetognath samples were collected during a

second field trip to better identify the placement of these organisms. Samples were

amplified using the three genes, 18S rRNA, COI, and COII (Refer to Appendix B. Figs.

B.1-3). The ARDRA analyses showed very little differentiation in the chaetognath

samples (Refer to Appendix B. Figs. B. 4-9). The goal was to obtain the most variable

samples to assure accurate phylogenetic placement, so the ARDRA gels were then put

into the BioNumerics software analysis tool to determine the various grouping patterns

(Refer to Appendix B. Figs. B. 10-18). The BioNumerics software tool was able to

identify several different banding patterns, of which 8 of the most variable were chosen

to use in further phylogenetic tree reconstructions. The samples that represented the most

variability were: 4, 6, 14, 15, 16, 21, 30, and 32. These samples were chosen according

to the groupings of the BioNumerics software. These samples appeared to group

differently (According to Appendix B. Figs. B 10-18), suggesting, they could possibly

be different species according to their BioNumerics fingerprint. By choosing 8 variable

samples, I hoped to gather the best phylogenetic placement of the chaetognaths.

3.3. Confirmation of Pelican Island Samples

Using the BLAST search tool (*www.ncbi.nlm.nih.gov)* the 18S rRNA, COI, and COII

sequences for chaetognath samples were compared with those in GenBank to determine

closest relative (Table III.1). For each gene class, the closest relative was *Sagitta* spp;

18S rRNA (86-100%, COI(83-100%), COII(83-100%). The current number of recorded

chaetognath species in the Gulf of Mexico is twenty-four (Tunnel et al. 2009). The most

common species found in the Gulf of Mexico were, *Flaccisagitta enflata, F. hexaptera,*

F.lyra, Serratosagitta sp., Pterosagitta draco, Mesosagitta decipiens, M.minima,

Krohnitta pacifica, K. subtilis, Sagitta sp., and *Ferosagitta hispida*. These data were

used to determine which sequences to download for tree reconstructions.

Table III.1 Closest Relatives for All 32 Pelican Island Chaetognath
Samples for All Primers Used.

Sample ID	Accession Number	Description	% Similarity
18S-4	DQ463220.1	*Sagitta setosa* microsatellite 18S rRNA gene	100
18S-6	D14363.1	*Sagitta crassa naikaiensis* 18s rRNA gene	94
18S-14	EU074279.1	*Sagitta setosa* 18s rDNA	100
18S-15	AY942657.1	*Sagitta elegans* haplotype ss 99ns2	100
18S-16	DQ4632181.1	*Sagitta setosa* microsatellite set 1	100
18S-21	Z19551.1	*Sagitta elagans* gene encoding 18s rDNA gene	88
18S-30	Z19551.1	*Sagitta elegans* gene encoding 18s rDNA gene	86
18S-32	AY922316.1	*Sagitta sp.* GG-2005 18s rDNA gene	95
COI-4	EU407234.1	*Sagitta bedoti* COI gene	94
COI-6	FJ648784.1	*Sagitta bedoti* NIOBZ2 COI gene	94
COI-14	DQ86280	*Sagitta bedoti* isolate 15	83

Table III.1 continued

	9.1	COI gene	
COI-15	DQ862797.1	*Sagitta bedoti* isolate 3 COI gene	100
COI-16	DQ862801.1	*Sagitta bedoti* isolate 7 COI gene	100
COI-21	DQ862801.1	*Sagitta bedoti* isolate 7 COI gene	100
COI-30	FJ648784.1	*Sagitta bedoti* NIOBZ2 COI gene	94
COI-32	EU407234.1	*Sagitta bedoti* COI gene	94
COII-4	AY585623.1	*Sagitta setosa* haplotype ss C1 COII gene	92
COII-6	AY585599.1	*Sagitta setosa* haplotype N34 COII gene	87
COII 14	DQ486112.1	*Sagitta setosa* haplotype ss T6 60/H23 COIIgene	85
COII-15	DQ486112.1	*Sagitta setosa* haplotype ss T6 60/H23 COII gene	85
COII-16	AY585583.1	*Sagitta setosa* haplotype ss T4/8 COII gene	86
COII-21	DQ486112.1	*Sagitta setosa* haplotype ss T6 60/H23 COII gene	83
COII 30	DQ486119.1	*Sagitta setosa* haploytpe ss S3 13/H31 COII gene	100
COII 32	DQ463221.1	*Sagitta setosa* microsatellite	100

3.4 Placement of Pelican Island Chaetognaths in 18S rRNA Trees

The red ellipses in each figure below exhibit the placement of my 8 chaetognath samples. Figure III.1 illustrates an 18S sequence tree reconstructed using the Neighbor Joining method in MEGA 4.0 using the pairwise deletion option which includes all gaps and missing data before the calculation begins. The chaetognath sample sequences(denoted as 18S-#) were dispersed among 5 groups; The Nematodes (18S- 16), and the Arthropods (18S -4), the Annelids (18S- 6,14,15,21,30,32),the Chaetognaths (18S -4,6, 21,30,32), and the Mollusks(18S-15). In Figure III.2, the Minimum Evolution method for tree reconstruction (also using the pairwise deletion option in MEGA 4.0) placed the 18S sequences among 5 groups, just like the Neighbor joining method. The Nematodes (18S- 16), and the Arthropods (18S -4), the Annelids (18S- 6,14,15,21,30,32),the Chaetognaths (18S- 4,6, 21,30,32), and the Mollusks(18S-15). In Figure III.3, the Maximum Parsimony method for tree reconstruction placed the 18S sequences among 4 groups; the Annelids (18S- 6,14,21.30,32), the Arthropods (18S-4,15,16), the Nematodes (18S-4,15,16), and the Chaetognaths (18S- 6,14,21,30,32). For this tree reconstruction, the maximum parsimony method was used with the use all sites option (instead of complete deletion), which accounts for all gaps and missing data. Also, all nucleotide changes were weighted equally (standard parsimony) so that all nucleotide differences were included in the calculation. In Figure III.4, the unweighted pair group method with arithmetic mean (UPGMA) placed the 18S sequences in 4 groups; the Cnidarians (18S -6,14,21,30,32), the Arthropods (All 8 samples), the Nematodes (18S-6,14,21,30,32), and the Mollusks (18S- 16). This UPGMA tree

reconstruction utilized the pairwise deletion option in the MEGA software, which maintains alignment gaps and missing-information before the calculation begins. In Figures III.5-8, the same methods were used above, but a bootstrap was also calculated on each tree. The bootstrap is the best test for phylogenetic accuracy. The bootstrap tests whether your whole dataset is supporting your tree (Baldauf 2003). The neighbor joining bootstrap method was calculated using 100 replicates and a pairwise deletion (like that of the non bootstrap neighbor joining), the minimum evolution bootstrap tree was calculated using 100 replicates and a pairwise deletion (like that of the non bootstrap minimum evolution tree), the maximum parsimony bootstrap tree was calculated using 200 replicates and the use all sites options (like that of the non bootstrap maximum parsimony tree: which has an equal weighting of nucleotide changes), and lastly, the UPGMA bootstrap tree was calculated using 100 replicates and the pairwise deletion option (like that of the non bootstrap UPGMA tree). The results are the same (as compared to the groupings for the non bootstrapped trees) for the neighbor joining bootstrap method, the minimum evolution bootstrap method, and the UPGMA bootstrap method. The maximum parsimony bootstrap tree (Fig.III.7) was the only bootstrap tree that differed. The maximum parsimony bootstrap tree placed the chaetognaths in 3 groups, unlike the original 4 groups from the non bootstrapped tree. 18S -4, 15, 16 grouped with the Arthropods and Annelids, while 18S- 6,14,21,30,32 grouped with two chaetognath genera, *Krohnitta sp,* and *Sagitta sp.* All 18S sequence samples also either grouped with the Pelican Island sequence samples or the known chaetognath sequences downloaded from GenBank.

Fig. III.1. Evolutionary relationships of 129 taxa 18s rRNA genes sequences. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 96.68169421 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Fig. III.2. Evolutionary relationships of 129 taxa 18s rRNA gene sequences. The evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length = 96.32552746 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All position containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .
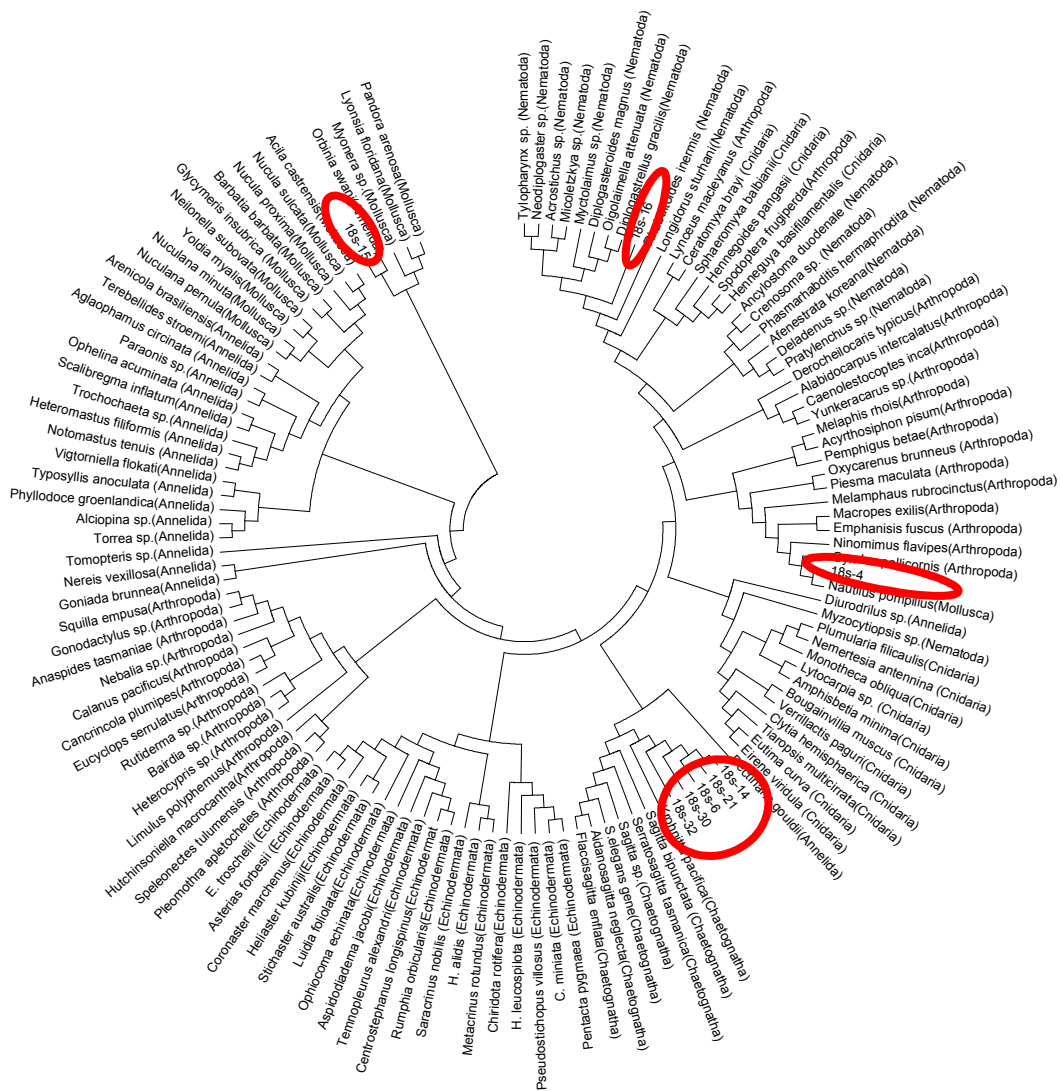
Fig. III.3. Evolutionary relationships of 129 taxa 18s r RNA gene sequences. The evolutionary history was inferred using the Maximum Parsimony method**.** Tree #1 out of 62 most parsimonious trees (length = 838) is shown. The consistency index (which calculates the minimum number of steps in the cladogram to the actual number of steps present) is 0.078759 , the retention index (measures the amount of synapomorphy on the tree) is 0.620079 , and the composite index is 0.048837 for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 2 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 22 positions in the final dataset, out of which 22 were parsimony informative. Phylogenetic analyses were conducted in MEGA4 .
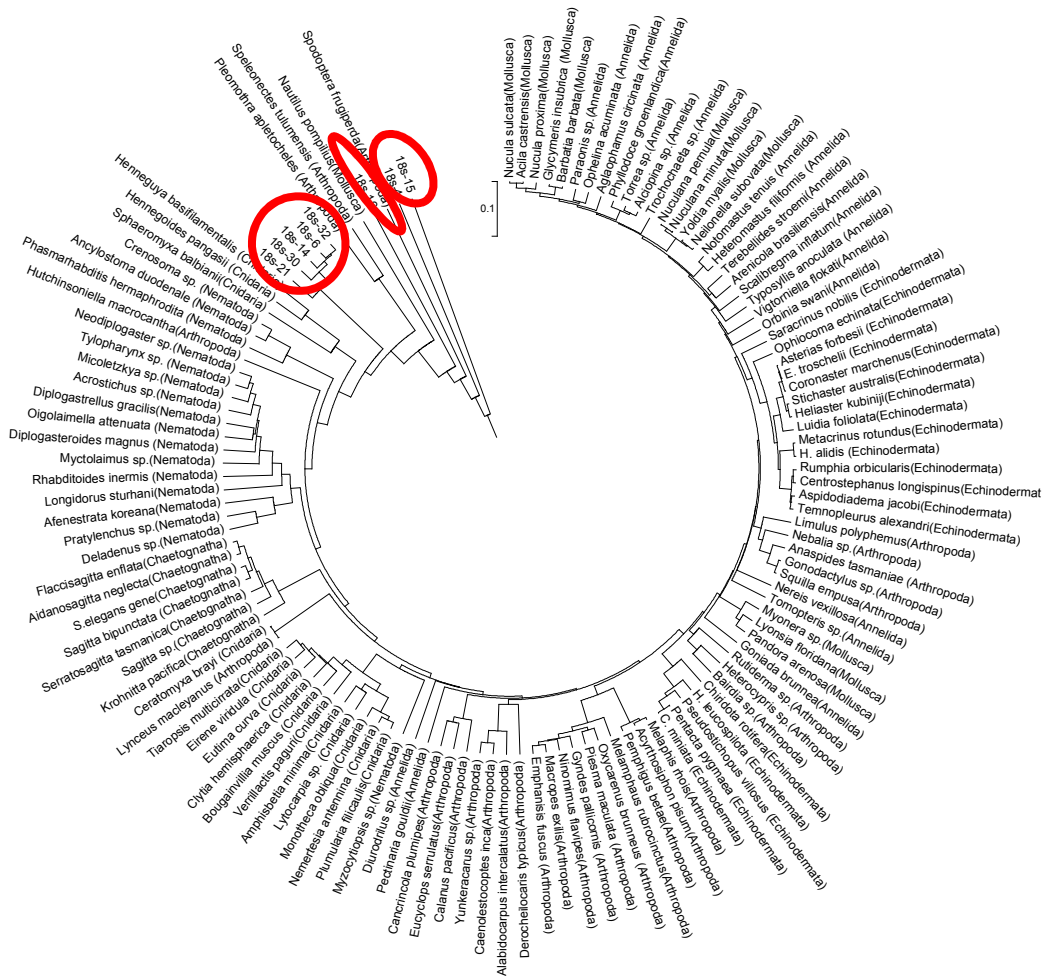
Fig. III.4 Evolutionary relationships of 129 taxa 18s rRNA gene sequences. The evolutionary history was inferred using the UPGMA method. The optimal tree with the sum of branch length = 96.36478115 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Fig. III.5. Evolutionary relationships of 129 taxa 18s rRNA gene sequence using Bootstrap method. The evolutionary history was inferred using the Neighbor-Joining method . The optimal tree with the sum of branch length = 96.68169421 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA.

Fig. III.6. Evolutionary relationships of 129 taxa 18s rRNA sequences using Bootstrap method. The evolutionary history was inferred using the Minimum Evolution method . The optimal tree with the sum of branch length = 96.32552746 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.
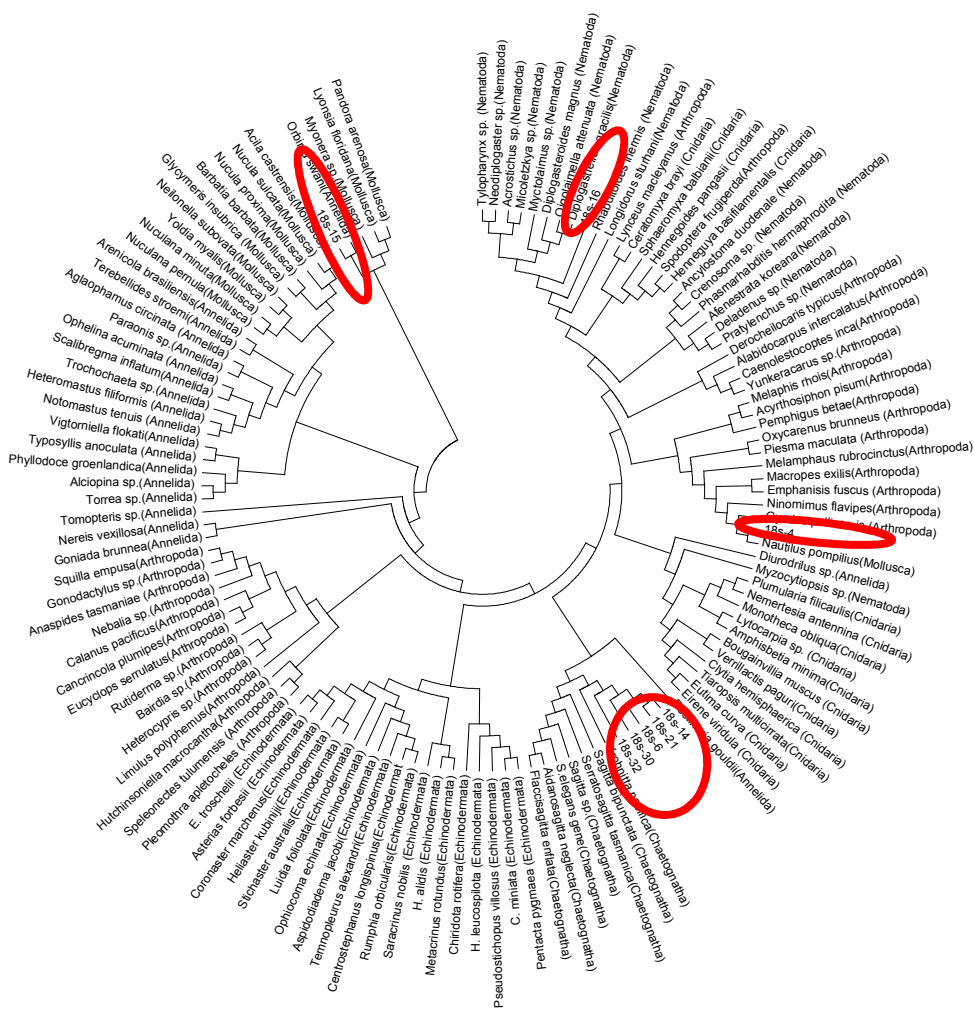
Fig. III.7. Evolutionary relationships of 129 taxa of 18s rRNA gene sequences using Bootstrap method. The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 80 most parsimonious trees (length = 839) is shown. The consistency index (calculates the minimum number of steps in a clade to the actual number of steps in a clad)is 0.078665, the retention index which measures the amount of synapomorphy on the tree) is 0.619587, and the composite index is 0.048740 for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (200 replicates) is shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 22 positions in the final dataset, out of which 22 were parsimony informative. Phylogenetic analyses were conducted in MEGA 4.
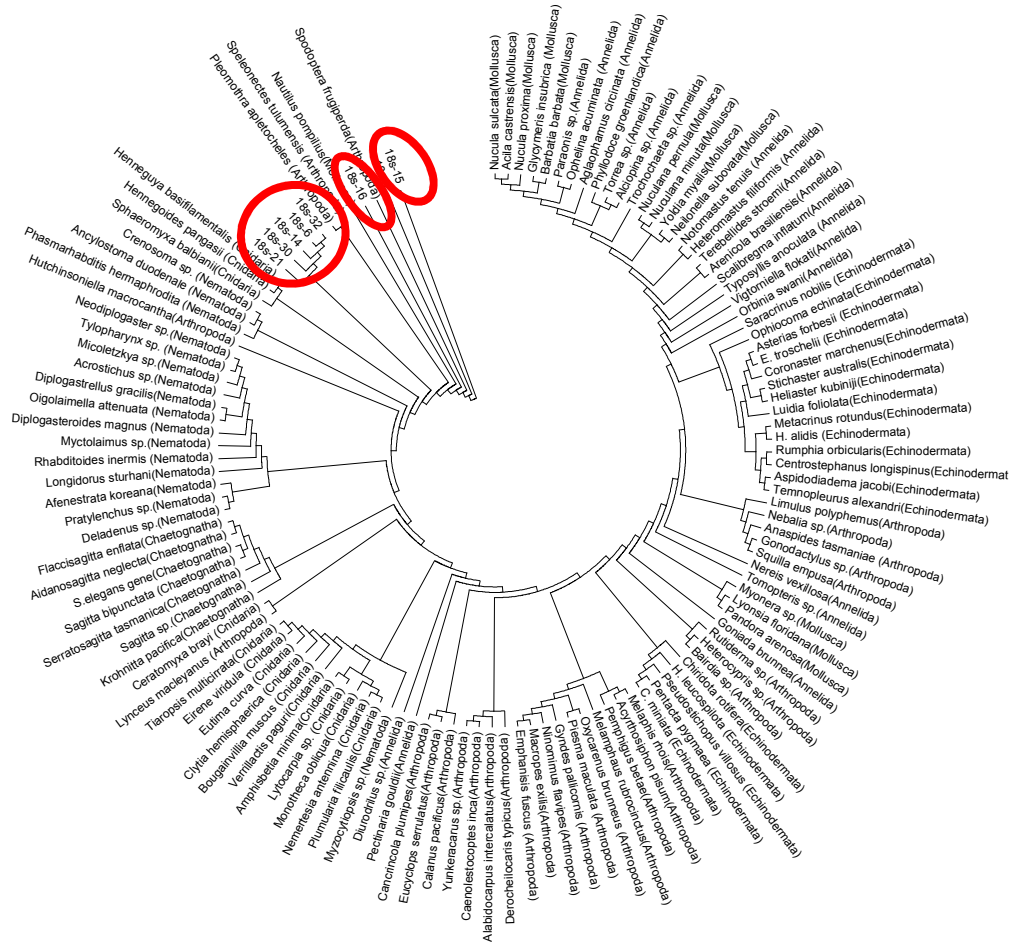
Fig. III.8. Evolutionary relationships of 129 taxa of 18s rRNA genes sequences using Bootstrap method. The evolutionary history was inferred using the UPGMA method. The optimal tree with the sum of branch length = 96.36478115 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 2487 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

3.5 Placement of the Pelican Island Chaetognaths in the COI Trees

The red ellipses in each figure below exhibit the placement of my 8 chaetognath samples based on COI genes. Figure III.9. illustrates a cytochrome oxidase 1 sequence tree reconstructed using the Neighbor Joining method , calculated in MEGA 4.0 similar to that of the 18S, using the pairwise deletion option. The chaetognath sample sequences (denoted as COI-#) were dispersed among 2 groups; the Annelids (COI- 4,6,30,32), and the Cnidarians (COI- 4,6, 14,15,16,21,30,32). In Figure III.10, the Minimum Evolution method for tree reconstruction (also using the pairwise deletion option in MEGA 4.0) placed the COI sequences among 3 groups; the Annelids and Cnidarians (COI- 4, 6, 30, and 32) and the Mollusks (COI -14,15,16,21). In Figure III.11, the Maximum Parsimony method for tree reconstruction (using the use all sites option in MEGA 4.0) placed the COI sequences among 5 groups; the Annelids and Cnidarians (COI- 4,6,30,32), the Arthropods (COI- 21,16), the Nematodes (COI -14,15,16), and the Mollusks (COI-21). In Figure III.12, the unweighted pair group method with arithmetic mean (UPGMA) places the COI sequences in 3 groups; the Annelids and Cnidarians (COI -4,6,30,32) and the Mollusks (COI- 14,15,16,21). For this UPGMA tree reconstruction the same pairwise deletion option was applied similar to the 18S rRNA tree reconstruction.

In Figures III.13-16, the same methods were used above, but a bootstrap was also calculated on each tree. Similar to the 18S bootstrap tree reconstructions, the neighbor joining, minimum evolution and UPGMA method calculated the trees using 100 replicates and the pairwise deletion method. The maximum parsimony bootstrap tree reconstructions calculated the tree using 200 replicates and the use all sites option, which accounts for missing data or gaps in the alignment. The results are the same for every tree except the Maximum Parsimony bootstrap reconstructed tree, which grouped the COI sequences in only two groups instead of the original five. According to Figure III.15, the maximum parsimony bootstrap analysis placed the COI samples with Annelids (COI- 4,6,30,32) and Cnidarians (All eight samples). The COI gene sequence samples also all grouped with either other Pelican Island chaetognath samples or known chaetognath sequences downloaded from GenBank.

Fig. III.9. Evolutionary relationships of 139 taxa using COI Gene sequences. The evolutionary history was inferred using the Neighbor-Joining method . The optimal tree with the sum of branch length = 131.78341619 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1534 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

Fig. III.10. Evolutionary relationships of 139 taxa using the COI gene sequences. The evolutionary history was inferred using the Minimum Evolution method . The optimal tree with the sum of branch length = 123.91399611 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1534 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.
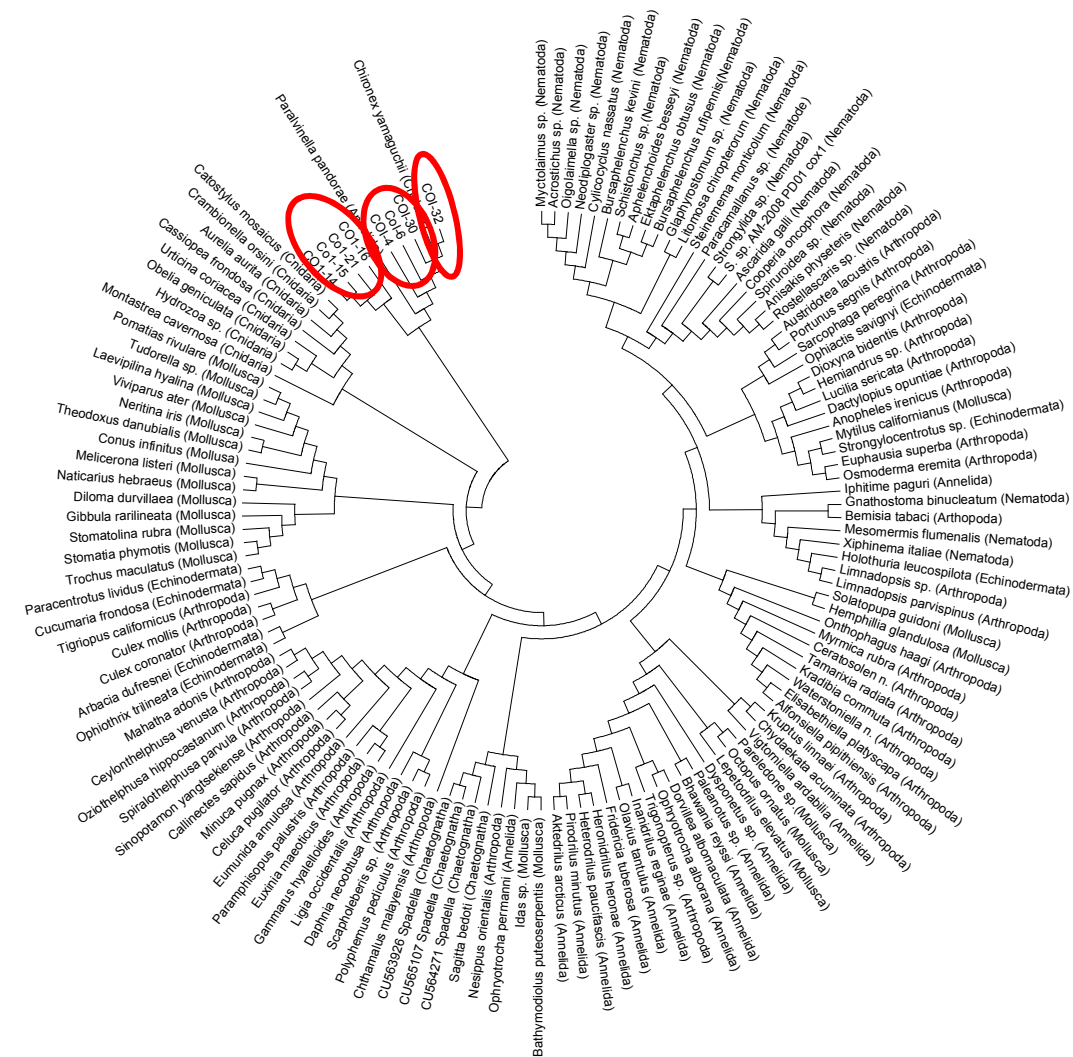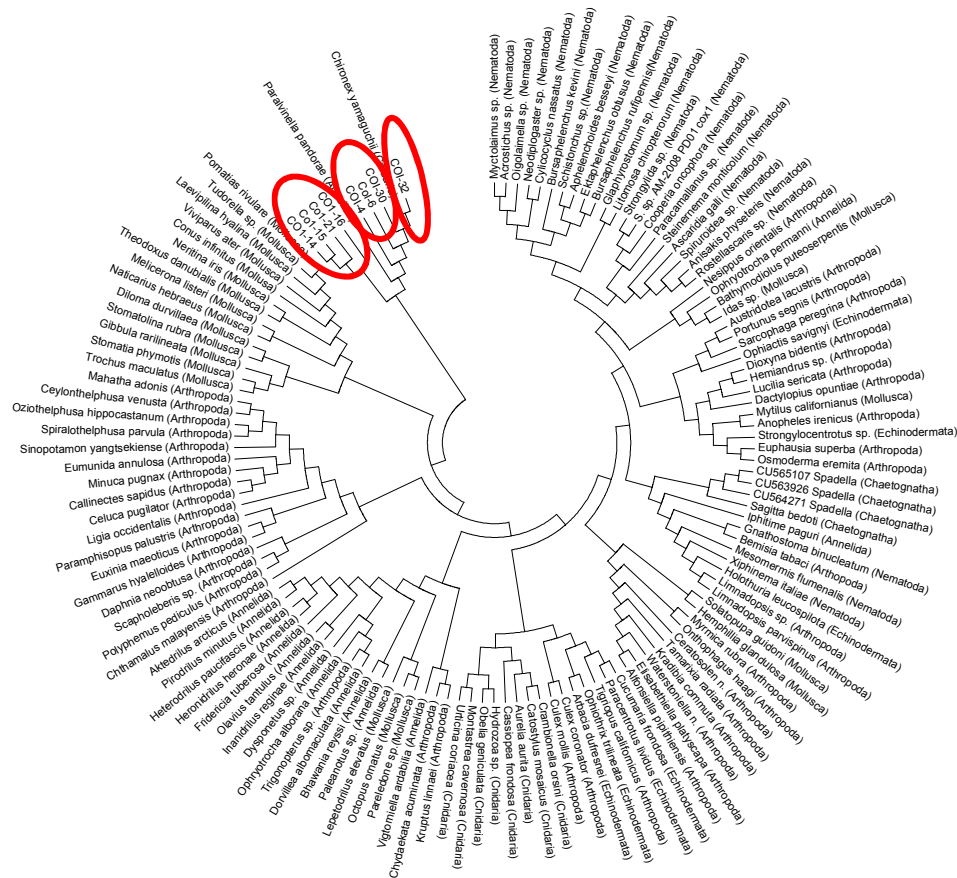
Fig.III.11. Evolutionary relationships of 139 taxa using the COI gene sequences. The evolutionary history was inferred using the Maximum Parsimony method . Tree #1 out of 14 most parsimonious trees (length = 4143) is shown. The consistency index (which calculates the minimum number of steps in the clade to the actual number of steps in the clade) is 0.065894 , the retention index(which measures the amount of synapomorphy on the tree) is 0.482689 , and the composite index is 0.031806 for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 2 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 91 positions in the final dataset, out of which 91 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.
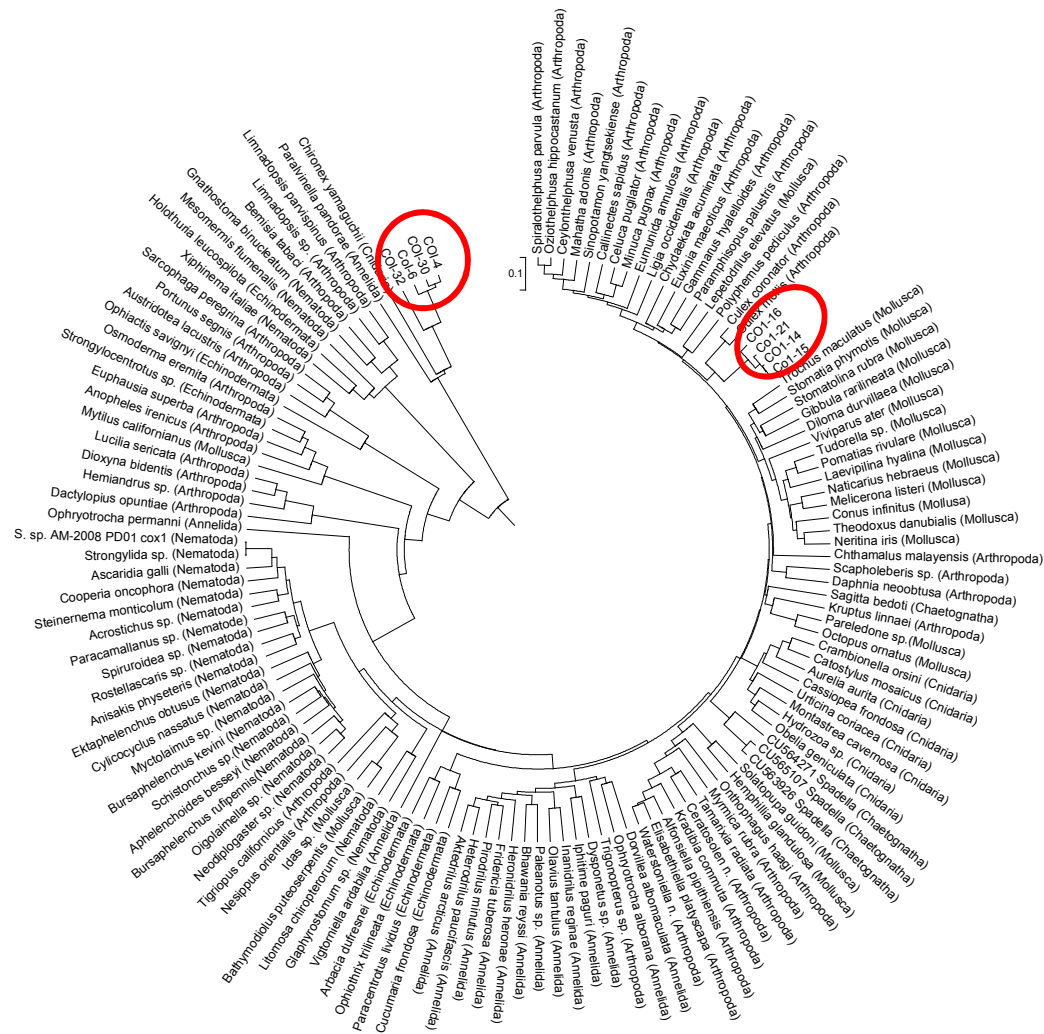
Fig.III.12. Evolutionary relationships of 139 taxa using the COI gene sequences. The evolutionary history was inferred using the UPGMA method . The optimal tree with the sum of branch length = 139.37333686 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 91 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

Fig. III.13. Evolutionary relationships of 139 taxa using COI gene sequences and Bootstrap method. The evolutionary history was inferred using the Neighbor-Joining method . The optimal tree with the sum of branch length = 127.41238157 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches . The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset. There were a total of 91 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Fig. III.14. Evolutionary relationships of 139 taxa using COI gene sequences and Bootstrap method. The evolutionary history was inferred using the Minimum Evolution method . The optimal tree with the sum of branch length = 123.91399611 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches . The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 3. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1534 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Fig. III.15. Evolutionary relationships of 139 taxa using COI gene sequences and Bootstrap method. The evolutionary history was inferred using the Maximum Parsimony method . Tree #1 out of 14 most parsimonious trees (length = 4143) is shown. The consistency index is 0.065894 , the retention index (which measures the amount of synapomorphy on the tree) is 0.482689 , and the composite index (which calculates the minimum number of steps on the cladogram to the actual number of steps on the cladogram) is 0.031806 for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 2 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 91 positions in the final dataset, out of which 91 were parsimony informative. Phylogenetic analyses were conducted in MEGA4 .
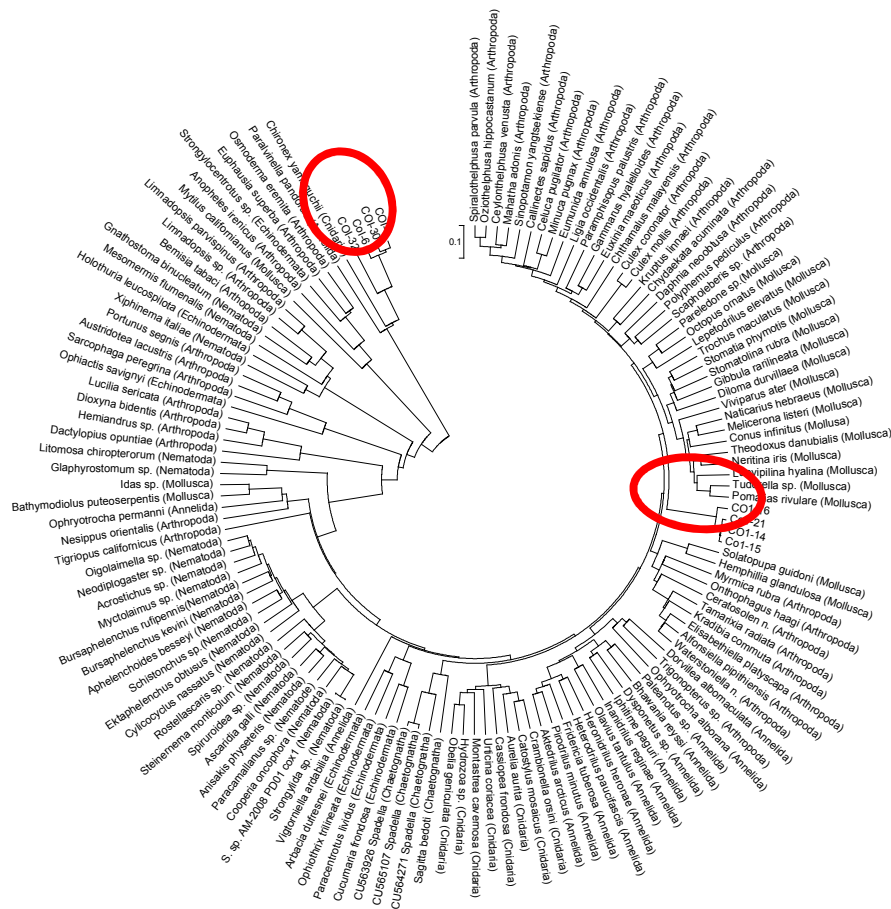
Fig. III.16. Evolutionary relationships of 139 taxa using COI gene sequences and Bootstrap method. The evolutionary history was inferred using the UPGMA method . The optimal tree with the sum of branch length = 134.28344700 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method] and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1534 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

3.6 Placement of the Pelican Island Chaetognaths in COII Trees

The red ellipses in each figure below exhibit the placement of my 8 chaetognath samples based on COII genes. Figure III.17 illustrates a cytochrome oxidase II sequence tree reconstructed using the Neighbor Joining method in MEGA 4.0 calculated using the pairwise deletion option (similar to 18S and COI genes). The chaetognath sample sequences(denoted as COII-#) were dispersed among 3 groups; the Nematodes (COII-4,6,30,32), and the Arthropods (COII- 4,6, 14,15,16,21,30,32). In Figure III.18, the Minimum Evolution method for tree reconstruction (using the pairwise deletion option similar to that of the 18S and COI genes) placed the COII sequences among 1 group: the Nematodes. In Figure III.19, the Maximum Parsimony method for tree reconstruction places the COII sequences among 3 groups; the Annelids (COII- 4,6,15,30,32), the Arthropods (COII – 14,16, 21), and the Mollusks (COII-21). For this tree reconstruction, the maximum parsimony method was used with the use all sites option (instead of complete deletion), which accounts for all gaps and missing data. Also, all nucleotide changes were weighted equally (standard parsimony) so that all nucleotide differences were included in the calculation (calculated similarly to that of the 18S and COI genes). In Figure III.20, the unweighted pair group method with arithmetic mean (UPGMA) place the COII sequences in 2 groups; the Annelids (COII -4,6,30,32) and the Arthropods (COI- 14,15,16,21). This UPGMA tree reconstruction method used the pairwise deletion option similar to both the 18S rRNA and COI genes. In Figures III.21-24, the same methods were used above, but a bootstrap was also calculated on each tree. For the calculated bootstrap neighbor joining, minimum evolution, and UPGMA trees,

100 replicates and the pairwise deletion option was used. Again, this maximum parsimony bootstrap tree was reconstructed using 200 replicates and equal weighting of the nucleotides. The results are the same for the neighbor joining bootstrap method. The Minimum Evolution COII bootstrap tree (Fig.III.22) placed the sequences among 3 groups, unlike the original single group from the non bootstrap tree. The Minimum evolution bootstrap tree also placed COII- 14,15,16,21 with Arthropods and Nematodes. This tree reconstruction also configured COII- 4, 6, 30, 32 with the original grouping within Nematodes.   Figure III.23, the Maximum parsimony bootstrap tree reconstruction, differs from the original non bootstrap tree. This tree reconstruction placed the sequence data into 3 groups; the Arthropods (COII 4, 6) the Nematodes (COII-30, 32), and the Chaetognath *Sagitta sp.* specifically for the COII -14, 15,16, 21. The UPGMA bootstrap tree reconstruction of the COII sequences (Fig.III.24) also differs from the original non bootstrap tree. This bootstrap tree placed the sequence data into 2 groups (like the original), but the dispersal of groupings is quite different. COII 4,6,30, and 32 are grouped with Annelids and Arthropods, COII 14, 16 only grouped with *Sagitta sp.*, and COII 15 grouped with Annelids and *Sagitta sp.*, while COII 21 grouped with *Sagitta sp.* and Arthropods. All COII sequence samples also either grouped with the Pelican Island sequence samples or the known chaetognath sequences downloaded from GenBank.

Fig. III.17. Evolutionary relationships of 151 taxa using COII gene sequences. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 198.33207828 are shown.. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

Fig. III.18. Evolutionary relationships of 151 taxa using COII gene sequences. The evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length = 146.49506232 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 3. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .
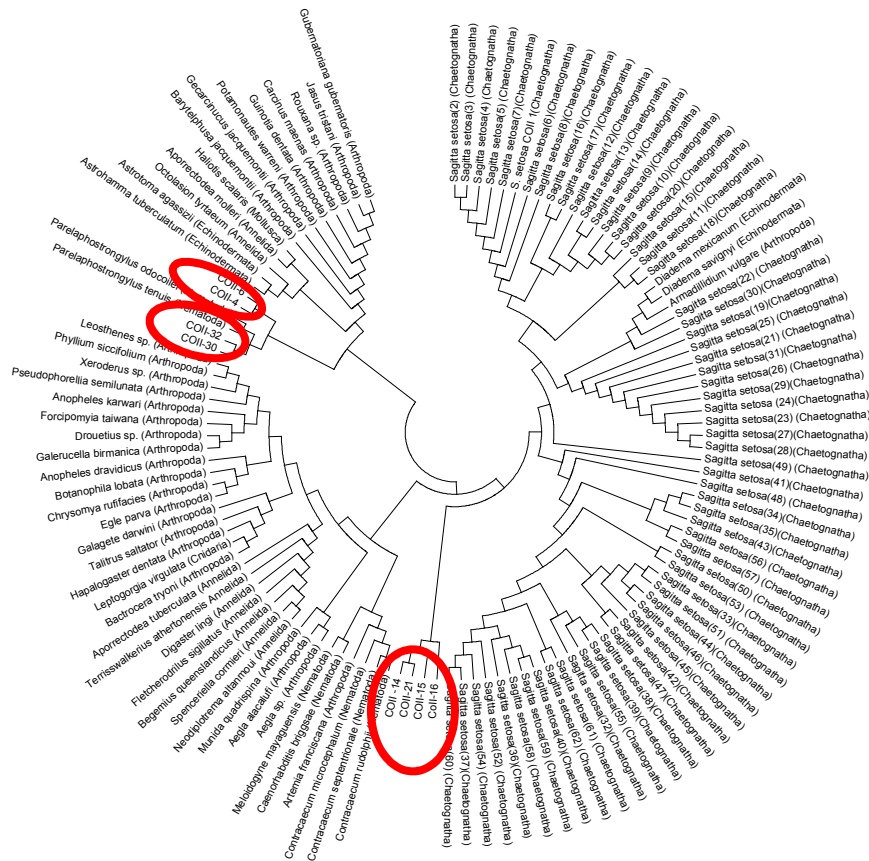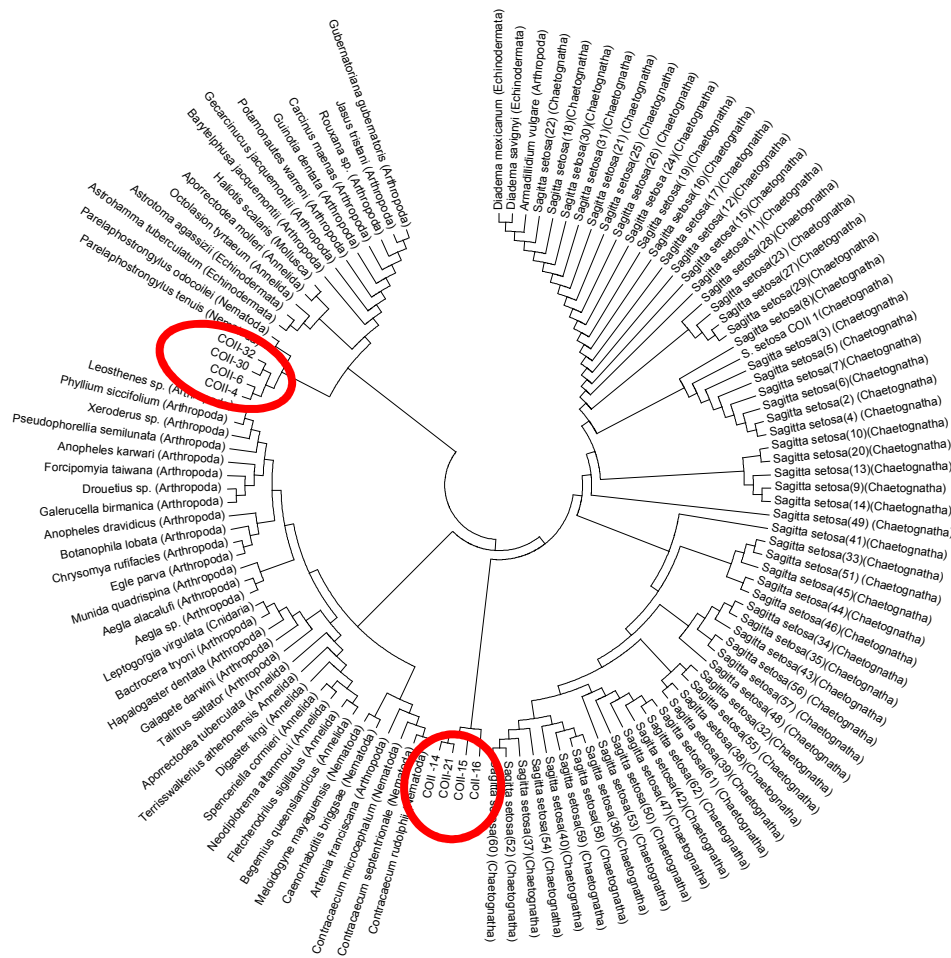
Fig. III.19. Evolutionary relationships of 151 taxa using COII gene sequences. The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 151 most parsimonious trees (length = 480) is shown. The consistency index (which calculates the minimum steps of the cladogram to the actual steps present on the cladogram) is 0.125000 , the retention index (which measures the amount of synapomorphy on the tree) is 0.667458 , and the composite index is 0.083432 for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 2 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 20 positions in the final dataset, out of which 20 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.

Fig. III.20. Evolutionary relationships of 151 taxa using COII gene sequences. The evolutionary history was inferred using the UPGMA method . The optimal tree with the sum of branch length = 192.77225383 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.

Fig. III.21. Evolutionary relationships of 151 taxa using COII gene sequences and Bootstrap method. The evolutionary history was inferred using the Neighbor-Joining method . The optimal tree with the sum of branch length = 198.33207828 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches . The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA 4.

Fig. III.22. Evolutionary relationships of 151 taxa using COII gene sequences and Bootstrap method. The evolutionary history was inferred using the Minimum Evolution method . The optimal tree with the sum of branch length = 146.49506232 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches . The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

Fig. III.23. Evolutionary relationships of 151 taxa using COII gene sequences and Bootstrap method. The evolutionary history was inferred using the Maximum Parsimony method . Tree #1 out of 84 most parsimonious trees (length = 4154) is shown. The consistency index (which calculates the minimum number of steps of the cladogram to the actual number of steps on the cladogram) is 0.105441, the retention index(which measures the amount of synapomorphy on the tree) is 0.598747, and the composite index is 0.063132 for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). The codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset . There were a total of 146 positions in the final dataset, out of which 146 were parsimony informative. Phylogenetic analyses were conducted in MEGA4 .

Fig. III.24. Evolutionary relationships of 151 taxa using COII gene sequences and Bootstrap method. The evolutionary history was inferred using the UPGMA method. The optimal tree with the sum of branch length = 192.77225383 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 1065 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 .

CHAPTER IV

DISCUSSIONS AND CONCLUSIONS

In the past 100 plus years, many attempts have been made to ally the chaetognaths to a bewildering variety of taxa. Proposed relatives have included nematodes, mollusks, various arthropods, acanthocephalans, rotifers, and chordates (Telford and Holland 1993). In 1968, Ghirardelli concluded that the Chaetognatha are distant relatives of the three major deuterostome phyla (Hemichordata, Echinodermata, and Chordata), which themselves are convincingly linked by an array of morphological, physiological, embryological, and molecular characters (Ghirardelli 1968). More recently, the metazoan phylogeny surveys using molecular data, sometimes combined with morphological characters, suggested various affinities, mostly among protostome**s:** analysis of intermediate filament sequences placed the Chaetognatha within lophotrochozoans (Erber et al.1998) and 18S rRNA analyses placed them within ecdysozoans (Halanych 1996; Littlewood et al. 1998; Zrzàvy et al. 1998), or within the basal ecdysozoans (Peterson and Eernisse 2001), between deuterostomes and protostome**s** (Giribet et al. 2000), or as an early offshoot of the bilaterian lineage (Telford and Holland 1993; Wada and Satoh 1994). However, the sequences used in these studies are extremely divergent, and the phylogenetic position of Chaetognath**a** remains dubious because of long-branch attraction artifacts. Moreover, each of these studies examined different genera of Chaetognatha so that no direct comparison is possible. More recently, analysis of *Hox* genes placed the Chaetognatha in a basal

position among the Bilateria (Papillon et al. 2003). This study also analyzed and compared other nuclear markers (myosin, elongation factor, 18S rRNA) isolated from several genera however placement of the Chaetognatha could not be resolved.

In the current study, DNA sequences from three different genes, COI, COII, and 18S rRNA, from eight individual specimens of Chaetognatha were analyzed and phylogenetic trees were reconstructed with four different algorithms (i.e. Neighbor Joining, Minimum Evolution, UPGMA, and Maximum Parsimony) that analyzed discreet (i.e. each nucleotide site) versus distance (i.e. branch length) characters. Comparison of tree reconstructions by these algorithms evaluated the placement in the tree, closest neighbor, and number of branches separating each of the eight Pelican Island chaetognaths; I determined that at least this genus (i.e. *Sagitta*) is most closely related to the protostome groups. Interestingly, all tree reconstructions produced from the various algorithms in the MEGA 4.0 software, dispersed the Arthropoda throughout each tree. The Phylum Arthropoda is very large and includes the insects as well as crustaceans, which could account for the wide dispersal among the various other protostome phyla (Nematodes, Annelids, Cnidaria). Also due to the diversity of this enormous phylum, the long-branch chain attractions could also explain the varying dispersal among the produced tree reconstructions, or that these phyla share a common ancestor.

4.1 18S rRNA Tree Reconstructions

Placements of the Pelican Island chaetognath 18S rRNA sequences by the Neighbor

Joining, and Minimum Evolution algorithms were almost identical and for each

sequence the nearest neighbor was the same. This result is not surprising since both of

these algorithms reconstructed the tree based on distance measures and the Neighbor

Joining method is actually a simplified version of the Minimum Evolution method.

Unlike these algorithms that placed the Pelican Island chaetognath sequences into four

different clusters, the UPGMA tree reconstruction placed them into three clusters that

were more closely placed together and with fewer bifurcations but longer branch length

separations. The Maximum Parsimony algorithm tree reconstruction appears to be the

best inference of the 'true' placement for my Pelican Island chaetognath sequences. This

tree placed the Pelican Island chaetognaths in two clusters with shorter branch length

separations than the Neighbor Joining, Maximum Evolution, and UPGMA trees. The

18S rRNA gene is approximately 1800 bp and more impacted by degradation than

shorter bp genes making alignment difficult. These inconsistencies, compensated by

gaps in the alignments, would directly affect the calculation of distance trees such as

Neighbor Joining and Maximum Evolution. Maximum Parsimony compares only

nucleotide positions that are filled. Additionally, due to the inconsistencies of sequence

quality, the option to eliminate gaps and missing nucleotides (Complete Deletion option)

was applied. Ultimately, only 22 nucleotide positions out of approximately 1800 were

parsimony informative. However, despite the paucity of data, Maximum Parsimony

produced the shortest branch lengths and tighter clustering of my sequences.

Maximum Parsimony placed the eight Pelican Island chaetognath samples closest to the protostome phylum of the arthropods (Appendix C. Table and Fig.C.7.). My findings support the findings of Zrzavy et.al (1998) and Peterson and Eernisse (2001), and Helmkampf et al. (2008). These three groups analyzed 18S rRNA sequences and found that the chaetognaths did not group with deuterostomes but allied with the Ecdysozoa (molting animals), which includes the arthropods. Telford and Holland (1993) using 18S rRNA data found little support to link the chaetognath to the deuterostomes, but proposed a protostome lineage at the base of the coelomates. Another study analyzing the 18S rRNA found that the chaetognaths grouped with another protostome phylum the nematodes (Halanych 1996).

There was no difference between the non-bootstrapped and bootstrapped tree reconstructions. However, the bootstrap is an essential method in phylogenetic studies to overcome assumptions about the sampled population. Since the samples in this study were randomly selected, assumptions about the normality of the population cannot be assumed, therefore, the bootstrap method was applied to test whether the whole data set is supporting the tree (Efron and Tibshirani 1993). The bootstrap essentially tests for reliability of the tree estimate (Hall 2008). Although all trees constructed grouped the eight Pelican Island chaetognath samples with protostomes, the maximum parsimony bootstrap tree clustered them closest together, allowing for a better understanding of their phylogenetic placement. The Maximum Parsimony bootstrap tree grouped the Pelican Island chaetognath samples with arthropods (30%, protostomes) and chaetognath sequences from other studies (60%, Refer to Fig. C.7.). Although the UPGMA bootstrap

tree method also grouped the Pelican Island chaetognath samples with arthropods (31%, protostomes) and other chaetognath samples (27%, Refer to Fig.C.8.), this method assumes a constant rate of evolution (molecular clock hypothesis), and is not a well-regarded method for inferring phylogenetic trees unless this assumption has been tested and justified for the data set being used (Legendre 1998). The 18S rRNA sequences analyzed in this study would be difficult to test for a constant rate of evolution because the sequences are highly divergent. The other bootstrapped trees (minimum evolution and neighbor joining) grouped my chaetognath sequences with annelids (30%, protostomes) and other chaetognath sequences (50%, Refer to Figs. C.5., C.6.), but the clustering of the Pelican Island chaetognath sequences was not as significant (Figs. III. 21, 22).

After analyzing the other bootstrap tree reconstructions for the 18S rRNA gene (which grouped with Arthropods (UPGMA, like that of the Maximum Parsimony bootstrap), and Annelids (Neighbor Joining and Minimum Evolution), the Pelican Island chaetognath sequences did not cluster together as tightly, except for the 18S rRNA bootstrap UPGMA tree. When comparing other trees besides the Maximum Parsimony tree, the sum of the branch lengths is used to determine the best phylogeny. So, when comparing the sum of the branch lengths for the other bootstrap trees, the UPGMA bootstrap tree has a branch length sum of 96.3 (Fig. III.8), compared to the sum of 96.6 (for 18S rRNA bootstrap Neighbor Joining; Fig.III.5) and although the 18S rRNA Minimum Evolution bootstrap (Fig.III.6) sum of branch lengths is the same as the UPGMA, this tree could not be the 'true' tree based on the eight Pelican Island

chaetognath samples clustering into four groups, instead of two like that of the bootstrap

Maximum Parsimony tree and three groups in the UPGMA tree. The Maximum

Parsimony bootstrap tree was chosen as the best tree representative for this gene because

parsimony does not require us to make any explicit assumptions about the process (like

the affect of long branch chain attractions), and thus can be applied when we have data

that cannot easily be modeled due to extremely divergent sequences like those found in

18S rRNA (Thollesson 2001).

I hypothesized that using the 18S rRNA would not be the best gene to help elucidate

the phylogeny of the chaetognath. In previous studies by Halanych (1996), Zrazvy

(1998) and Giribet et al. (2000), the 18s rRNA sequences used in these studies are

extremely divergent, and Halanych and Zrazvy admitted that the phylogenetic position

of the Chaetognatha remained dubious because of long-branch attraction artifacts. These

authors found that the Chaetognatha grouped with various protostome phyla. Halanych's

group found the chaetognaths to group with the nematodes, whereas, Zrzavy found the

chaetognaths to group with all protostome phyla except for the Lophophorates (i.e.

mollusks, annelids and brachiopods). Giribet et al. (2000) found the chaetognaths to

group with the Nemertodematida, a basal protostome bilaterian. The position of the

Chaetongnatha remained dubious because all of these studies found that the

chaetognaths group with different phyla. Even though the quality of the 18S rRNA

sequences produced from the eight collected samples was poor, the dubious nature of the

chaetognath's alignment is consistent with my 18S rRNA findings. Since the 18S rRNA

sequences are extremely divergent and also have a slow rate of evolutionary change

(Halanych 1996), the phylogenetic data that was produced from the Pelican Island

chaetognath samples and the various other 18S rRNA samples downloaded from

GenBank, yielded various phylogenetic relationships. Further, the ribosomal genes

copies in chaetognaths appear to be split into two highly divergent 'classes' whose

paralog vs. pseudogene status remains unclear. This possibly non-homogeneous

duplication of nuclear ribosomal genes complicates their use in genetic analyses

(Jennings 2010). Most of the 18S rRNA trees reconstructed placed the eight Pelican

Island chaetognath samples with either the protostome phylum Annelida (bootstrap

Neighbor Joining and bootstrap Minimum Evolution 50%), or the protostome phylum

Arthropoda (bootstrap Maximum Parsimony and bootstrap UPGMA 30%), again

selecting the Maximum Parsimony bootstrap tree because of the tight clustering of the

eight Pelican Island chaetognath sequences. This means that the 18S rRNA data is not

the best gene marker to identify the phylogenetic relationships of the Chaetognatha.

Although 18S rRNA is one of the most widely used genes in eukaryotic phylogenetic

studies, this gene proved to only be informative at the level of recognized families

(Bleidorn et al. 2005; Burnette et al. 2005), especially when exploring such enigmatic

groups like the chaetognaths. Furthermore, using the highly ambiguous alignments and

long- branch chain attractions of the 18S rRNA in this study may be meaningless, when

used alone, in a phylogenetic context.

4.2 Cytochrome Oxidase I and II Tree Reconstructions

The cytochrome oxidase gene data results prove to be a convincing argument for my hypotheses. COI is the terminal catalyst in the mitochondrial respiratory chain and is involved in electron transport and proton translocation across the membrane. The CO1 gene is often used as the mitochondrial marker for evolutionary study because it is the largest of the three subunits and the protein sequence contains highly conserved functional domains and variable regions (Morlais & Severson 2002). Villa et al. (2006) also found that the cytochrome oxidase II genes, encoding for metabolic and structural proteins, are also excellent targets for phylogenetic analyses. These genes are conserved and the alignment of their sequences is less ambiguous compared to rRNA (Villa et al. 2006). The COI (708bp) and COII (551 bp) genes allow for a less ambiguous alignment and therefore a better analysis of the phylogenetic trees obtained for each gene from the MEGA 4 software analysis tool.

After analyzing each COI phylogenetic tree reconstruction method, I concluded that the best phylogenetic tree to infer the placement of the eight Pelican Island chaetognath samples is the minimum evolution bootstrap tree (Fig.III.14). This tree clustered Pelican Island chaetognath COI sequences the closest together, representing a better inference of phylogeny. The Minimum Evolution bootstrap tree grouped the Pelican Island chaetognaths closely to two protostome phyla the mollusks and annelids (Appendix C. Table and Fig. C.14). This finding supports those of Marlétaz et al. 2006 (Fig.I.6) who sequenced 11,526 expressed sequence tags (EST) from a cDNA library of the benthic chaetognath *Spadella sp.* They found after phylogenetic tree reconstruction using the

rooted maximum likelihood method that the chaetognaths grouped closest to the

Lophotrochozoa (Annelids and Mollusks), which is represented in Fig.I.6. The

Minimum Evolution bootstrap tree reconstruction method was chosen as the best

representation for the COI genes, because the bootstrap is an essential method in

phylogenetic studies to overcome assumptions about the sampled population. Although

all trees constructed grouped the eight Pelican Island chaetognaths with other

protostomes, the Minimum Evolution bootstrap tree clustered the Pelican Island

chaetognath samples the closest together. The Minimum evolution bootstrap tree

grouped the Pelican Island chaetognath COI sequences with Mollusks (35%) and other

chaetognath (34%, Refer to Fig.C.14.) Although the UPGMA tree grouped the eight

Pelican Island chaetognaths with Mollusks (38%) and other chaetognaths (33%, Refer to

Fig.C.12.), this method assumes a constant rate of evolution (molecular clock

hypothesis), and is not a well-regarded method for inferring phylogenetic trees unless

this assumption has been tested and justified for the data set being used (Legendre 1998).

The majority of the other trees grouped the Pelican Island chaetognath samples closely

to the protostome phyla Cnidaria (Figs. C.9., C.13., C.15., C.16.), and although these

tree reconstructions did not cluster them as tightly as the Minimum Evolution bootstrap

method there may be some reasoning behind these findings. According to Papillon's

group in 2004, the chaetognaths are the first metazoans that could possess an mtDNA

genome with a complete absence of tRNA genes. Total lack of tRNA genes has

previously been reported only for some protozoans (*Plasmodium*), and some tRNA genes

are lacking in green alga (*Pedinomonas*) and angiosperm plants (*Arabidopsis*). The

metazoans closest to this situation are the cnidarians, for instance, *Metridium senile* (Beagley et al. 1998) and *Acropora tenuis* (Van Oppen et al. 2002), where only tryptophan and methionine tRNAs are observed and the other necessary tRNAs are imported nuclear products (Boore 1999). As in cnidarians, paucity of mtRNA genes in *S. cephaloptera* is probably a derived condition rather than a conserved primitive state for multicellular animals. This could explain why most of the COI tree reconstructions grouped the eight Pelican Island chaetognath samples with cnidarians.

Similar to COI, I concluded the best phylogenetic tree to infer the placement of the eight Pelican Island chaetognath COII sequences is the Minimum Evolution bootstrap tree (Fig.III.22). This tree reconstruction grouped the Pelican Island chaetognath samples closest to the protostome phyla Nematoda (Appendix C. Table and Fig.C.22). This supports the findings of Halanych (1996). Although Halanych used 18S rRNA downloaded GenBank sequences for several invertebrate phyla and compared those to known 18s rRNA chaetognath sequences in GenBank, his study paired the chaetognaths more closely to the nematodes than any other invertebrate phyla.

After comparing each tree reconstruction method for the COII gene, I realized that there was not one tree representative that clustered the eight Pelican Island chaetognath samples most closely together, so the bootstrap values were used to determine the best tree reconstruction method. Analysis of the bootstrap values of the most closely related phyla for all four bootstrap tree methods found that the Minimum Evolution tree reconstruction had the highest value (66%) compared to Neighbor Joining (20%), Maximum Parsimony (36%) and UPGMA (58%). This finding infers more confidence

for the relatedness of the Pelican Island chaetognaths to the nematodes. Although the bootstrap value produced for the Minimum Evolution tree was 66%, the UPGMA bootstrap value produced was 58%. The UPGMA bootstrap tree grouped the closest with the protostome phyla Arthropoda, which supports the maximum parsimony bootstrap tree 18S rRNA findings of this study.

The Maximum Parsimony bootstrap tree was found to be the best tree representative for the 18S rRNA gene, but the Minimum Evolution bootstrap tree appears to be the best for the cytochrome oxidase genes. The Minimum Evolution algorithm uses distance measures that correct for multiple hits at the same sites, and a topology showing the smallest value of the sum of all branches is chosen as the best estimate of the correct tree, and is more appropriate for shorter genes (COI, COII) than much larger genes like 18S rRNA. When comparing all bootstrap tree COI gene sum of branch lengths, the lowest branch length sum was that of the minimum evolution bootstrap tree (Fig.III.14; branch length: 123.9), compared to 127.4 for the neighbor joining bootstrap tree (Fig.III.13) and 134.2 for the UPGMA bootstrap tree (Fig.III.16), revealing that the Minimum Evolution bootstrap tree was the closest 'true' phylogenetic placement of the Chaetognatha within the Mollusks and Annelids.

The other bootstrapped trees reconstructed (Neighbor Joining, Maximum Parsimony, and UPGMA) grouped the chaetognaths more closely with the Cnidarians, but the sum of the branch lengths were larger for the Neighbor Joining and UPGMA meaning that the phylogenetic relationship to Cnidarians is not as strong as to the Mollusks and Annelids.

The bootstrap Minimum Evolution tree was also found as the best representative for the COII gene. Again, the sums of the branch lengths were analyzed to determine the closest 'true' phylogeny. The Minimum Evolution bootstrap tree sum of branch length was 146.5 (Fig.III.22), compared to 198.3 (Fig.III.21, neighbor joining bootstrap tree), and 192.8 (Fig.III.24, UPGMA bootstrap tree). Revealing that the Minimum Evolution bootstrap tree determined the closest 'true' phylogenetic placement of the Chaetognatha with the Nematodes, since the sum of the branch lengths was the lowest. The other bootstrap trees calculated for the COII gene placed the chaetognaths more closely to the Arthropods, which supports the findings of the 18S rRNA, but again after analyzing the sum of the branch lengths the Minimum Evolution bootstrap tree that placed the chaetognaths with the Nematodes appears to be the best tree representative for this gene.

Another comparison that can be made is the retention index that measures the amount of synapomorphy (a trait that is shared by two or more taxa and their most recent common ancestor, whose ancestor in turn does not possess the trait) on the tree. The higher the retention index the more closely related are each taxa represented. The retention index of the maximum parsimony bootstrap COII tree was 59.8% (Fig.III.23), as compared to 62 % for the 18S rRNA gene, and 48.2% for the COI gene). This finding revealed again that when comparing the three genes the maximum parsimony bootstrap tree is the closest 'true' tree for 18S rRNA.

Even though the Minimum Evolution bootstrap tree for both the COI and COII genes revealed the best phylogenetic tree reconstruction, the genes grouped with different phyla (COI-Annelids and Mollusks, and COII-Nematodes). This might be explained by

the highly unusual mtDNA found is the various chaetognath species. Faure and

Casanova (2006) found three striking mtDNA features when analyzing the

Chaetognatha. One is that they have the smallest mtDNAs yet known in any metazoan.

Also, among the usual 13 protein-encoding genes, atp6 and atp8 are lacking. Lastly, 21

of the 22 mitochondrial tRNA genes normally present in the bilaterians are missing in

the mtDNA of *P. gotoi* and no tRNA has been found by Papillon et al. (2004) in that of

S. cephaloptera. This variation in these two genera of Chaetognatha could explain the

different groupings of the COI and COII gene. Also all of the Pelican Island chaetognath

samples were pelagic (unlike the *P.gotoi* and *S. cephaloptera* above). When

downloading the available sequences in GenBank, the pelagic chaetognath sequences for

S.setosa were the only ones available for the COII gene, whereas the COI gene database

in GenBank had many genera including various benthic chaetognaths (like

S.cephaloptera) of which to compare the unusual mtDNA better, leading to the various

grouping of the COI and COII genes.

   The usage of the cytochrome oxidase genes in this study revealed positive

contributions throughout each tree reconstruction method. Each method either revealed

tighter groupings of the 8 Pelican Island chaetognath samples (COI) or stronger

bootstrap support values (COII) compared to those of the 18S rRNA phylogenetic trees.

The more substantial positive contributions of the cytochrome oxidase genes are

concentrated on more recent divergences (Halanych and Janosik 2006), leading to a

better understanding of the phylogenetic placement of the chaetognaths. The 18S rRNA

gene uses a highly conserved and deeply rooted divergence, which can sometimes lead

to the samples falsely grouping with other extremely divergent phyla (Telford and

Copely 2005).

The cytochrome oxidase genes were chosen for this study because they have been

found to encode for structural and metabolic proteins, which are excellent targets for

phylogenetic studies; mainly because they are highly conserved and the alignment of

their sequences is less ambiguous compared to rRNA (Villa et al. 2006). The GenBank

sequence database available for the COI gene is much larger than the COII sequence

database, which was a problem when downloading comparable sequences for alignment.

The COI gene has been studied and analyzed more frequently, so for a better

understanding of the chaetognath phylogenetic placement, the COII gene was also

analyzed in this study. By using similar genes (COI & COII), I attempted to produce

similar results (phylogenetic trees), but after analyzing each tree produced from the COI

and COII genes the results varied (COI grouping mostly with protostome phyla Cnidaria,

Mollusk, and Annelida, and COII grouping mostly with protostome phyla Arthropoda

and Nematoda: Refer to Appendix C Figs. C.9-C.24). Downloaded sequences from 139

taxa were analyzed using the various MEGA 4.0 tree reconstruction methods for the COI

gene, including both pelagic (*S.setosa, S.bedoti*), and benthic (*Spadella sp.)* chaetognath

species (Figs. III.9-16), but even though 151 COII taxa were analyzed (Figs.III, 17-24),

62 of those are from the same pelagic species of chaetognath (*S.Setosa*); this was due to

the minimal amount of available COII sequences from the GenBank online database. All

62 COII *S.setosa* sequences were used in this study to demonstrate the various gene

variants among the Chaetognatha. The lack of similar amounts of COII sequences could

explain the variable tree reconstruction results that were seen in this study among the cytochrome oxidase genes.

4.3 Conclusions

The tree reconstructions that were produced in this study were highly variable. Two different methods (Maximum Parsimony and Minimum Evolution) were chosen as the best algorithms for each of the three genes analyzed, and each tree grouped with various protostome phyla (arthropods, nematodes, annelids, and mollusks). This may be explained by the differences in the genes. The DNA encoding for 18S rRNA is not expressed as a protein since 18S rRNA is a ribosomal subunit. Whereas the COI and COII genes analyzed in this study encode for cytochrome oxidases i.e. proteins. Evolutionary processes on the history of individual genes such as convergence, i.e the same sequence change appearing independently in different lineages either because of similar selective pressures or by chance; changes in evolutionary rates, i.e. certain organisms evolve faster than others; horizontal gene transfer, i.e. sequences being transferred from one species to another by mechanisms other than vertical, linear descent; and timing, i.e. two lineages radiate from a third in relatively close succession before enough differences may have accumulated between them to be able to discern the order of emergence (Lynch 1999). Additional causes for differences in the placement of the Pelican Island chaetognaths may also be attributed to methodological issues such as different parameters being applied to the algorithms used to weigh sequence similarities.

Thus, even in the best scenarios, absolutely congruent phylogenies from the analysis of individual genes are *not* expected.

Although molecular phylogenetic tree inconsistencies are hardly a fundamental theoretical concern for evolutionary biology, if persistent they could still cause practical problems in assessing certain evolutionary relationships. However, a number of new approaches have recently emerged that address these difficulties. These methods include the combination of large sets of sequence information from genomic databases, as well as the use of genetic features, such as large-scale structural changes or the mapping of mobile genetic elements, that are less prone to convergence and selection-related artifacts (Lokas and Carroll 2006). The comparison of several different tree reconstructions (using different algorithms, i.e. Neighbor Joining, Minimum Evolution, Maximum Parsimony or UPGMA) is necessary because, depending on the genes used (whether large or small or divergent or similar), the evolutionary relationships in each tree reconstruction attempt to reveal the 'true' phylogeny.

The findings of this study support the morphological data findings of several groups. In 1907, Gunther proposed a protostome affinity for the chaetognaths using gross morphological studies to link them to ancestral mollusks, which supports my COI minimum evolution bootstrap tree results (grouping closest with mollusks). Zrzavy et al. (1998) placed the chaetognaths in a clade grouped with all protostome phyla, except the Lophophorates. Two other studies, Eernisse (1998) and Littlewood et al. (1998) placed the chaetognaths in a clade within the Ecdysozoa with a nematode base, which supports

my COII Minimum Evolution bootstrap tree results (grouping closely with nematodes) and my 18S rRNA Maximum Parsimony tree results (grouping closely with arthropods).

The early embryological findings of Hyman (1959) show that the chaetognaths share common characteristics with deuterostomes during their ontogeny: radial cleavage, a blastopore at the rear end of the body, and a post-anal tail, however, they do not pass through the dipleura stage that is seen in every deuterostome phylum. More recent studies (Wilmer 1990, Harzch & Müller 2007), examined the development of the central nervous system and concluded the chaetognaths were more closely related to the protostomes, which supports my findings. All of the 8 samples grouped with protostomes in every tree reconstruction method.

The role of genetics in determining the chaetognath lineage is very important. All three of the genes used in this study grouped the eight Pelican Island chaetognaths with the protostomes. This supports the findings of the previous studies of Wada and Satoh (1994) and Telford and Holland (1993), which by using the 18S rRNA gene they were able to link the chaetognaths to protostomes. Another very important aspect of using molecular methods for phylogeny is the usage of mtDNA (like the COI and COII used in this study). In 2006, Barthelemy et al. used mtDNA to link the chaetognaths to annelids and mollusks, which support the COI minimum evolution bootstrap tree produced in this study. The Hox gene study (Papillon 2003) also revealed a basal protostome lineage for chaetognaths due to the retained ancestral features of this gene. Another study from Matus et al. (2007) found a Hox gene that possesses both ecdysozoan and lophotrochozoan signature amino-acid motifs, which support all of my results. The COI

Minimum Evolution bootstrap, tree which groups chaetognaths with mollusks and annelids (lophotrochozoans) and also the COII Minimum evolution bootstrap tree which groups chaetoghmaths with nematodes (ecdysozoans) and finally the 18S rRNA Maximum Parsimony tree which groups chaetognaths with arthropods (ecdysozoans). Whole genomic analysis also found the chaetognaths to group with protostomes. Marlétaz et al. (2006) used whole genome sequence analysis to group the chaetognaths more closely to the annelids and mollusks, again supporting the COI minimum evolution bootstrap tree in this study.

The combination of 18S rRNA, COI and COII gene sequence data was essential for the resolution and recovery of stable clades at various depths of the trees. Furthermore, combining three genes produced a phylogenetic 'signal' present in data sets of different genes, which can be hidden in a single gene analysis to emerge (Zanol et al. 2009). This study revealed the various clades that could be formed depending on the genes and tree reconstruction method performed.

4.4. Future Directions

Although this study helped to shed some light on the enigmatic Chaetognatha, additional work is necessary. This study used 3 different genes to help elucidate the chaetognath's phylogenetic placement, which most of the other studies focused only on one gene. In the future, when using molecular techniques to resolve chaetognath phylogeny, more than one sequence alignment tool should be used to test for better accuracy of ambiguous alignments and divergent sequences. The chaetognath sequences

downloaded from GenBank were useful, however they were limited and inconsistent in that more than one gene for a single genus were not always available making comparisons of tree reconstruction algorithms more difficult. Also when dealing with large genes, like that of the 18S rRNA it is important to be mindful of the ambiguous alignments that can occur because of the extremely divergent sequences and the rapidly evolving nature of this genes which causes 'long branch chain attractions' that can ultimately interfere with the phylogenetic tree reconstruction and produce false or misleading clades. Strategies for dealing with ambiguous alignments and divergent sequences would be to only download sequences with similar base pairs for the specific gene and also sequences with fewer gaps, which can cause the alignment software to match up sequences that are typically not paired together in a phylogenetic tree.

This study focused on only 8 samples (from one genus of Chaetognatha) collected at one time at the same location. In the future more samples from various chaetognath families and genera as well as multiple geographic locations should be analyzed to create a better understanding of the placement of this group.

LITERATURE CITED

Arendt, D.,U. Technau and J.Wittbrodt. 2001. Evolution of the bilaterian foregut. Nature.409: 81–85.

Ball, E.E and D.J. Miller. 2006. Phylogeny: the continuing classificatory conundrum of chaetognaths. Curr.Biol.16: R593-596.

Barthelemy R., F. Petenian, J. Vannier,J.P. Casanova and E. Faure.2006. Evolutionary History of the Chaetognaths Inferred from Actin and 18S-28S rRNA Paralogous Genes.Int.J.Zool.Res.2:284- 300.

Beagley, C. T., R. Okimoto and D. R. Wolstenholme. 1998. The mitochondrial genome of thesea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics. 148:1091–1108.

Bone, Q., H.Kapp and A.C. Pierrot-Bults. 1991. Introduction and relationships of the group. Pp. 1-17 *in* Bone, Q., H.Kapp, A.C. Pierrot-Bults, eds. The biology of the Chaetognaths. Oxford Science Publications, New York.

Baldauf, S. L. June 2003. Phylogeny for the faint of heart: a tutorial. TIG. 19: 345-351.

Beyer, W.A., M.L. Stein, T.F. Smith and S.M. Ulam. 1974. A molecular sequence metric and evolutionary trees. Math. Biosci. 19:9-25.

Bleidorn, C., L. Vogt and T. Bartolomaeus. 2005. Molecular phylogeny of lugworms (Annelida, Arenicolidae) inferred from three genes. Mol. Phylogenet. Evol. 34: 673–679.

Bone, Q., K.P. Ryan and A.L. Pulsford. 1983. The structure and composition of the teeth and grasping spines of chaetognaths. J. Mar. Biol. Assoc. U.K. 63:929-939.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767–1780.

Bryant, D. and P. Waddell. 1998. Rapid evaluation of least squares and minimum evolution criteria on phylogenetic trees. Mol. Biol. Evol. 15:1346-1359.

Bruno W.J., N. D. Socci and A. L. Halpern.2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol. Biol. Evol. 17: 189-197.

Burnette, A.B., T.H. Struck and K.M.Halanych. 2005. Holopelagic *Poeobius meseres* ("Poeobiidae", Annelida) is derived from benthic flabelligerid worms. Biol. Bull. 208:213–220.

Cavalli-Sforza, L.L. and A.W.F. Edwards.1967. Phylogenetic analysis: models and estimation procedures. Am.J.Hum.Genet. 19:233-257.

Cooper, S. J.B., J. H. Bradbury, K. M. Saint, R. Leys, A. D. Austin and W. F. Humphreyes. 2007. Subterranean archipelago in the Australian arid zone: mitochondrial DNA: Phylogeography of amphipod from central Western Australia Mol. Ecol.16:1533-1544.

Eernisse, D.J. 1998. Arthropod and annelid relationships re-examined. Pp. 43-56 *in* . R.A. Fortney and R.H. Thomas, eds. Arthropod relationships. Chapman and Hall, London.

Efron, B. and R.J.Tibshirani. 1993. An introduction to the bootstrap. Chapman Hall, New York.

Erber, A., D. Riemer, M. Bovenschulte and K. Weber. 1998. Molecular phylogeny of metazoan intermediate filament proteins. J. Mol. Evol. 47:751–762.

Faure, E. and J.P. Casanova. 2006. Comparison of chaetognath mitochondrial genomes and phylogenetical implications. Mitochondrion 6:258-262.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Felsenstein, J. 1993. Phylogeny Inference Package (PHYLIP). Version 3.5. University of Washington, Seattle.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates Inc., Sunderland, MA.

Field,K.G.,G.J. Olsen, D.J. Lane, S.J. Giovannoni, M.T. Ghiselin, E.C. Raff, N.R. Pace and R.A. Raff. 1988.Molecular phylogeny of the animal kingdom. Science 239:748-753.

Garey, J. and A. Schmidt-Rhaesa. 1998. The essential role of "minor" phyla in molecular studies of animal evolution. Am.Zool. 38: 907-917.

Gascuel. O.1997. BIONJ: an improved version of the NJ algorithm bases on a simple model of sequence data. Mol. Biol.Evol. 14:685-695.

Giribet, G., D. L. Distel, M. Polz, W. Sterrer and W. C. Wheeler. 2000.Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelmintes, and Chaetognatha: a combined ppproach of 18SrDNA sequence morphology. Syst.Biol.49: 539-562.

Ghirardelli, E. 1968. Some aspects of the biology of the chaetognaths. Adv. Mar.Biol. 6:271-375.

Gunther, R.T. 1907. The Chaetognatha, or primitive Mollusca. Q.J. Microsc.Sci.57:357-394.

Halanych, K.M. 1996.Testing hypotheses of chaetognath origins: Long branches revealed by 18S Ribosomal DNA. 1996. Sys.Biol.45:223-246.

Halanych, K.M. and A.M. Janosik.2006. A review of molecular markers used for annelid phylogenetics. Integr. Comp. Biol. 46:533–543.

Hall, B. G. 2008. Phylogenetic trees made easy. A how-to manual. 3rd ed. Sinnauer Associates Inc, Sunderland, M A.

Harzch, S. and C.H.G. Müller.2007 A new look at the ventral nerve center of Sagitta: implications for the phylogenetic position of Chaetognatha (arrow worms) and the evolution of the bilaterian nervous system. Front. Zool. 4:1-15.

Helmkampf, M., I. Bruchhaus and B. Hausdorf. 2008. Multigene analysis of lophophorate and chaetognath phylogenetic relationships. Mol. Phylo. Evo. 46:206-214.

Hillis, D., C. Moritz and B. Mable. 1996. Molecular sytematics. 2nd ed. Sinauer Associates, Inc., Sunderland, MA.

Hyman, L.H. 1951. The invertebrates: Acanthocephala, Aschelminthes, and Entoprocta. The Pseudocoelomates Bilateria, Volume III. McGraw-Hill, New York.

Hyman, L.H.1959. The invertebrates: Smaller coelomic groups, Volume V. McGraw-Hill, New York.

Jennings R.M., A. Bucklin and A. Pierrot-Bults. 2010. Barcoding of arrow worms

    (Phylum Chaetognatha) from three oceans: genetic diversity and evolution within

    an enigmatic phylum. PLoS ONE 5:e9949. doi:10.1371/journal.pone.0009949

Johnson, R.A. and D.W. Wichern. 1988. Applied multivariate statistical analysis.

    Prentice-Hall, Upper Saddle River, NJ.

Peijnenburg C.A., T.A. Katja, .J. Johannes, A.J. Breeuwer, C. Annelies, C. Pierrot-Bults

    and B.J. Menken. 2004. Phylogeography of the plankton Chaetognatha *Sagitta*

    setosa reveals isolation in European seas. Evolution 58:1472-1487.

Kendall, M. G. and Gibbons, J. D. 1990. Rank correlation methods. (5th ed.) Edward

    Arnold, London.

Kidd, K. K. and L.A. Sgaramella-Zonta. 1971. Phylogenetic analysis: concepts and

    methods. Am. J. Hum. Genet. 23:235–252.

Legendre, L. 1998. Numerical ecology. (2nd ed.) Elsevier, Boston.

Littlewood, D.T., M.J. Telford, K.A. Clough and K. Rohde.1998. Gnathostomulida-an

    enigmatic metazoan from both morphological and molecular perspectives. Mol.

    Phylogenet. Evol. 9:72-79.

Lokas A. and S.B. Carroll.2006. Bushes in the tree of life. PLoS Biol .4:e352.

Lynch, M. 1999. The age and relationship of the major animal phyla. Evolution 53:

    319-325.

Marletaz, F.E. E. Martin, Y.Perez, D.Papillon, X. Caubit, C.J. Lowe, B. Freeman, L.Fasano, C. Dossat,P. Wincker, J.Weissenbach and Y. Le Parco. 2006. Chaetognaths Phylogenomics: a protostome with deuterostome - like development. Curr. Biol. 16: 575-576.

Marletaz, F.E. E. Martin, Y.Perez, D.Papillon, X. Caubit, C.J. Lowe, B. Freeman, L.Fasano, C. Dossat, P. Wincker, J.Weissenbach and Y. Le Parco. 2008. Chaetognath transcriptome reveals ancestral and unique features among bilaterians.Genome Biol. 9:R94.

Matus, D.Q., K.M. Halanych, M.Q. Martindale. 2007. The *Hox* gene complement of a pelagic chaetognath, *Flaccisagitta enflata*. Int. and Comp. Biol. 47:854-864.

Meglitsch, F.R. and P.A. Schram. 1991. Invertebrate zoology. Oxford University Press, Oxford, UK.

McLelland, J.A. 1989. An illustrated key to the Chaetognatha of the northern Gulf of Mexico with notes on their distribution. Gulf. Res. Repts. 8:145-172.

Morlais, I. and D.W. Severson. 2002. Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome C Oxidase subunit I (COI) from *Aedes aegypti*. DNA Seq.13:123-127.

National Center for Biotechnology Information. www.ncbi.nlm.nih.gov. (Accessed on 8/12/10).

Nielsen, C.1985. Animal phylogeny in the light of trochea theory. Biol.J.Linn.Soc.25:243-299.

Ota, S. and W.H. Li. 2000. NJML: A hybrid algorithm for the neighbor-joining and
maximum likelihood methods. Mol. Biol. & Evol. 17: 1401-1409.

Page, D.M. and E. C. Holmes.1998. Molecular evolution. A phylogenetic approach.
Blackwell Science Ltd, Oxford, UK.

Papillon, D., Y. Perez, L. Fasano, Y. Le Parco and X. Caubit.2003. *Hox* gene survey in
the chaetognath *Spadella cephaloptera*: evolutionary implications. Dev. Genes Evol.
213: 142- 148.

Papillon, D.,Y.Perez, X. Caubit and Y. Le Parco. 2004. Identification of Chaetognaths as
protostomes is supported by the analysis of their mitochondrial genome. Mol. Biol.
Of Evo. 21:2122-2129.

Pearson W.R ,G. Robins and T. Zhang. 1999. Generalized neighbor-joining: more
reliable phylogenetic tree reconstruction. Mol. Biol. Evol. 16: 806-816.

Peterson, K.J. and D.J. Eernisse. 2001. Animal phylogeny and the ancestry of
bilaterians: inferences from morphology and 18S rDNA gene sequences.
Evol.Dev. 3: 170-205.

Rzhetsky A. and M.Nei. 1994. METREE: a program package for inferring and testing
minimum-evolution trees. Computer Applications in the Biosciences. 10:4: 409-412.

Shinn, G.L. and M.E. Roberts. 1994. Ultrastructure of hatchling chaetognaths
(*Ferosagitta hispida*): Epithelial arrangement of mesoderm and its phylogenetic
implications. J. Morphol. 219:143-163.

Soltis, P. S. and D. E. Soltis. 2003. Applying the bootstrap in phylogeny construction.
Stat. Sci. 18:256-267.

Telford, M. J. and P.W.H. Holland. 1993. The phylogenetic affinities of the chaetognaths: A molecular analysis. Mol. Biol. Evol. 10:660-676.

Telford, M.J. and R.R. Copley. 2005. Animal phylogeny: fatal attraction. Current Biol.15:296-299.

Thollesson, M. 2001. Bioinformatics. (1st ed.) Blackwell Science Ltd, Oxford, UK.

Thorne. J .L and H.Kishino. 1992. Freeing phylogenies from artifacts of alignment. Mol. Biol. Evol. 9: 1148-1162.

Tree of Life Web Project .2006. Chaetognatha. Arrow Worms. Version 06 September 2006 (temporary).http://tolweb.org/Chaetognatha/2478/2006.09.06 in The Tree of Life Web Project, http://tolweb.org/.

Tunnell, J.W. , D.L. Felder and S.A. Earle. 2009. Gulf of Mexico, origin, waters, and biota: biodiversity. (1st ed.) Texas A&M University Press, College Station.

Vannier J, M. Steiner, E. Renvoise, S.X. Hu and J.P. Casanova. 2007. Early Cambrian origin of modern food webs: evidence from predator arrow worms. Proc. R. Soc. Lond. (Biol).274:627–633.

Van Oppen, M. J., J. Catmull, B. J. McDonald, N. R. Hislop, P. J. Hagerman and D. J. Miller. 2002. The mitochondrial genome of *Acropora tenuis* (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. J. Mol. Evol. 55:1–13.

Wada, H.and N.Satoh. 1994. Details of the evolutionary history from invertebrates to vertebrates , as deduced from the sequences of 18S rDNA. Proc.Natl.Acad.Sci. USA. 91:1801-1804.

Welsch, U.and V. Storch. 1982. The structure of the coelomic epithelium of *Sagitta*

    elegans (Chaetognatha). Zoomorphology 100: 217-222.

Willmer,P.1990. Invertebrate relationships: patterns in animal evolution. Cambridge

    Univ. Press, New York.

Villa, N. O., K.T. Asano and H. Suga. 2006. Phylogenetic relationships of *Pythium* and

    Phytophthora species based on ITS rDNA, cytochrome oxidase II and β –tubulin

    gene sequences. Mycologia 98: 410-422.

Zanol, J., K.M. Halanych, T.H. Struck and K.Fauchald. 2009. Phylogeny of the bristle

    worm family Eunicidae (Eunicida, Annelida) and the phylogenetic utility of

    noncongruent 16S, COI and 18S in combined analyses. Mol. Phyl. Evol. Vol. 55:660-

    676.

Zrzavy, J., S. Mihulka, P. Kepka, A. Bezdek and D. Tietz. 1998. Phylogeny of the

    Metazoan based on morphological and 18S ribosomal DNA

    evidence.Cladistics.14:249-285.

APPENDIX A

PRELIMINARY OBSERVATIONS

Initially, four chaetognath samples were sent to Yale University for analysis (Table A.1). The LCO1490 (Cytochrome oxidase I) and the 18S primers were initially used to discover the best primer set to use for the chaetognath observations. The sequencing data was analyzed using the GenBank database BLAST tool on the NCBI website. After obtaining the results from the BLAST (*www.ncbi.nlm.nih.gov)*, tables were constructed showing the closest identified organism to the analyzed sequences. Also, phylogenetic trees were obtained for each sample using the distance tree tool from GenBank. After observation of the initial data shown below, I concluded that further samples needed to be collected to infer more about the phylogeny of the chaetognaths. These four initial samples helped to show the reasoning behind why these creatures still elude scientists phylogenetically. By studying the trees from the initial data, you can determine that each one gives us a varied answer as to what origin the chaetognaths fall into. Figures A.1-4 show varied groupings for each of the initial chaetognath samples. In Figure A.1, the initial sequenced sample 1 grouped with arrow worms (Chaetognatha) and also closely with Arthropoda. In Figure A.2, the initial sequenced sample 3 grouped the closest with Arthropoda and Mollusca. The groupings of Figure A.3 showed very different results for the initial sequence sample 5. This initial sample grouped closely with spiders (Arthropoda), and other Arthropod groups. But after analyzing the tree further, I found the strange pairing of Phylum Mollusca with Phylum Chordata and the arrow worm

(Chaetognatha) not grouping with my initial sample 5 sequence. The initial sample 14

(Figure A.4) resulted in an even more interesting tree. This tree was constructed using

the 18S primer set instead of the cytochrome oxidase 1 primer set. This tree

reconstruction paired my initial sample 14 with Arthropoda and Annelida, but by

examining the tree further, you can see that Phylum Arthropoda was separated by the

Phylum Annelida into two groups. My preliminary findings using the initial four

samples made me realize the phylogenetic conundrum that the Phylum Chaetongnatha

displays. It was these initial findings that led me to delve deeper into the mystery of this

group. By analyzing more chaetognath samples and exploring different primer sets, I

hope to help to better understand the placement of this mysterious group of organisms.

Table A.1. Initial Primers Used in Preliminary Experiment

Sample Name	Primer used	Sample ID
1	LCO 1490	18
3	LCO1490	30
5	LCO1490	37
14	18S	55

Table A.2. Sample 1 Results (Closest Relatives by BLAST Tool in GenBank)

| Accession Number	Scientific Name	Common Name	Max ID
EU407234.1	*Sagitta bedoti* cytochrome oxidase subunit I	chaetognath	81%
FJ648784.1	*Sagitta bedoti* voucher NIOBZ 2 COI	chaetognath	81%
FJ648783.1	*Sagitta bedoti* voucher NIOBZ 1 COI	chaetognath	81%
DQ862801.	*Sagitta bedoti* isolate 7 COI	chaetognath	81%
DQ862797.1	*Sagitta bedoti* isolate 3 COI	chaetognath	81%
DQ862795.1	*Sagitta bedoti* isolate 1 COI	chaetognath	81%
DQ862809.1	*Sagitta bedoti* isolate 15 COI	chaetognath	80%
DQ862798.1	*Sagitta bedoti* isolate 4 COI	chaetognath	80%
DQ862806.1	*Sagitta bedoti* isolate 12 COI	chaetognath	80%
DQ862805.1	*Sagitta bedoti* isolate 11 COI	chaetognath	80%
DQ862804.1	*Sagitta bedoti* isolate 10 COI	chaetognath	80%
DQ862802.1	*Sagitta bedoti* isolate 8 COI	chaetognath	80%
DQ862800.1	*Sagitta bedoti* isolate 6 COI*	chaetognath	80%
DQ862807.1	*Sagitta bedoti* isolate 13 COI	chaetognath	80%
DQ862808.1	*Sagitta bedoti* isolate 14 COI	chaetognath	80%
DQ862799.1	*Sagitta bedoti* isolate 5 COI	chaetognath	80%
DQ862796.1	*Sagitta bedoti* isolate 2 COI	chaetognath	80%
DQ862803.1	*Sagitta bedoti isolate 9 COI*	chaetognath	79%
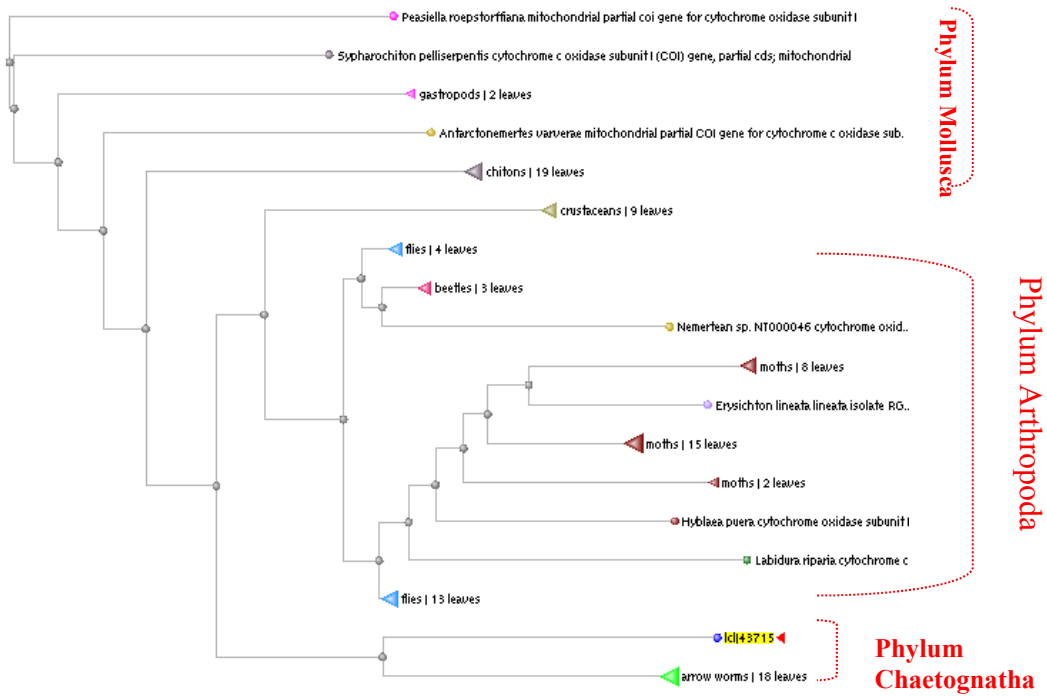
Fig. A.1. Phylogenetic Tree of Sample 1 Results from GenBank Tree Construction Tool. The sequenced sample highlighted in yellow is most closely related to arrow worms (Chaetognatha).

Table A.3. Sequence 3 Results (Closest Relatives by BLAST Tool in GenBank)

| Accession Number	Scientific Name	Common Name	Max ID
DQ862801.1	*Sagitta bedoti* isolate 7 COI	Chaetognath	78%
DQ862795.1	*Sagitta bedoti* isolate 1 COI	Chaetognath	78%
EU407234.1	*Sagitta bedoti* cytochrome oxidase subunit 1	Chaetognath	78%
DQ862797.1	*Sagitta bedoti* isolate 3 COI	Chaetognath	78%
FJ648784.1	*Sagitta bedoti* voucher NIOBZ 2 COI	Chaetognath	78%
FJ648783.1	*Sagitta bedoti* voucher NIOBZ 1 COI	Chaetognath	78%
DQ862807.1	*Sagitta bedoti* isolate 13 COI	Chaetognath	77%
DQ862809.1	*Sagitta bedoti* isolate 15 COI	Chaetognath	77%
DQ862798.1	*Sagitta bedoti* isolate 4 COI	Chaetognath	77%
DQ862802.1	*Sagitta bedoti* isolate 8 COI	Chaetognath	77%
DQ862800.1	*Sagitta bedoti* isolate 6 COI	Chaetognath	77%
DQ862799.1	*Sagitta bedoti* isolate 5 COI	Chaetognath	77%
DQ862806.1	*Sagitta bedoti* isolate 12 COI	Chaetognath	77%
DQ862805.1	*Sagitta bedoti* isolate 11 COI	Chaetognath	77%
DQ862796.1	*Sagitta bedoti* isolate 2 COI	Chaetognath	77%
DQ862808.1	*Sagitta bedoti* isolate 14 COI	Chaetognath	77%
DQ862804.1	*Sagitta bedoti* isolate 10 COI	Chaetognath	77%
DQ862803.1	*Sagitta bedoti* isolate 9 COI	Chaetognath	76%
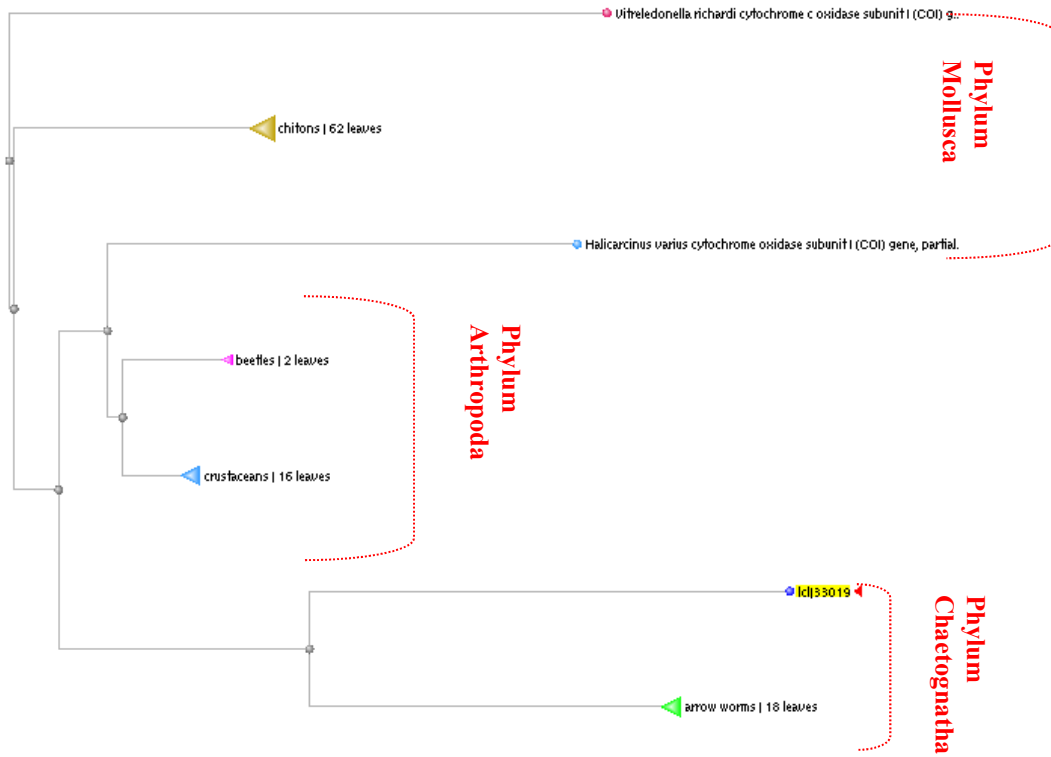
Fig. A.2. Phylogenetic Tree of Sequence 3 Results from GenBank Tree Construction Tool. The highlighted yellow region is most closely related to arrow worms (Chaetognatha) and the next closest relative is Arthropoda.

Table A.4. Sequence 5 Results (Closest Relatives by BLAST Tool in GenBank)

Accession Number	Scientific Name	Common Name	Max ID
EU495066.1	*Dendropoma petraeum voucher* MNCN Dp90Malt COI	Gastropod	81%
EU407234.1	*Sagitta bedoti* COI gene	Chaetognath	84%
DQ862801.1	*Sagitta bedoti* isolate 7 COI	Chaetognath	84%
DQ862797.1	*Sagitta bedoti* isolate 3 COI	Chaetognath	84%
DQ862795.1	*Sagitta bedoti* isolate 1 COI	Chaetognath	84%
FJ648784.1	*Sagitta bedoti voucher* NIOBZ 2 COI	Chaetognath	83%
FJ648783.1	*Sagitta bedoti voucher* NIOBZ 1 COI	Chaetognath	83%
EU367572.1	*Phytomyza sp.*"cimicifuga" COI	British fly	87%
DQ207224.1	*Palmadusta artuffeli* haplotype 1089 COI	Gastropod	86%
DQ862809.1	*Sagitta bedoti* isolate 15 COI	Chaetognath	84%
DQ862806.1	*Sagitta bedoti* isolate 12 COI	Chaetognath	84%
DQ862805.1	*Sagitta bedoti* isolate 11 COI	Chaetognath	84%
DQ862802.1	*Sagitta bedoti* isolate 8 COI	Chaetognath	84%
DQ862800.1	*Sagitta bedoti* isoalte 6 COI	Chaetognath	84%
DQ862798.1	*Sagitta bedoti* isolate 4 COI	Chaetognath	84%
DQ862796.1	*Sagitta bedoti* isolate 2 COI	Chaetognath	84%
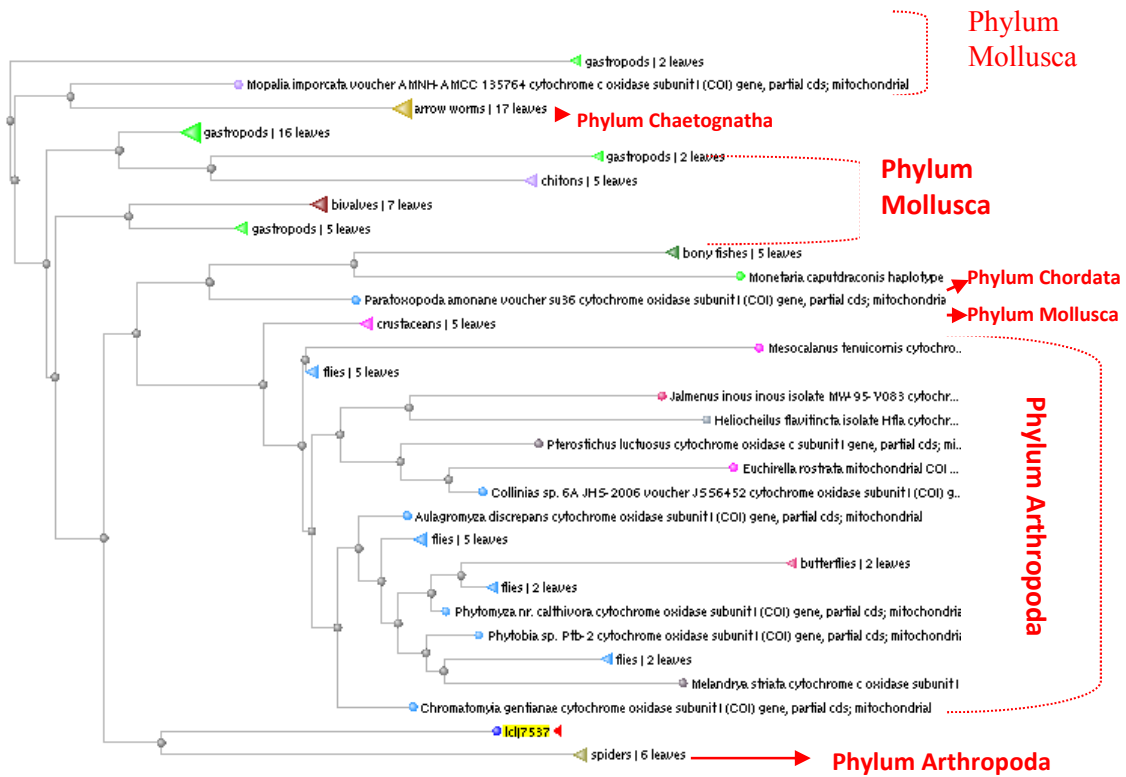
Fig. A.3. Phylogenetic Tree of Sequence 5 Results from GenBank Tree Construction Tool. The yellow highlighted sequenced sample is most closely related to spiders (Arthropoda).

Table A.5. Sequence 14 Results (Closest Relatives by BLAST Tool in GenBank

Accession Number	Scientific Name
Z70526.1	*H.tridens tientsinensis* 18S rRNA gene
Z70525.1	*H.tridens shenei* 18S rRNA gene
EU380303.1	*Ameira scotti* 18S ribosomal RNA gene, partial sequence
DQ538502.1	*Pandarus smithi* 18S ribosomal RNA gene, partial sequence
AY627028.1	*Clavella addunca* 18S small subunit ribosomal RNA gene, partial sequence
AF005442.1	*Belisarius xambeui* 18S ribosomal RNA gene, complete sequence
FJ222167.1	Uncultured marine eukaryote clone 18S ribosomal RNA gene, partial sequence
FJ222137.1	Uncultured marine eukaryote clone 18S ribosomal R gene, partial sequence
FJ221714.1	Uncultured marine eukaryote clone 18S ribosomal RNA gene, partial sequence
EF622736.1	*Nemurella pictetii* 18S ribosomal RNA gene, partial sequence
EF526976.1	Uncultured marine eukaryote clone SA1_1B02 18S ribosomal RNA gene, partial sequence
AF012482.1	*Mecyclothorax vulcans* 18S ribosomal RNA gene, complete sequence
AY527056.1	*Potamodrilus fluviatilis* 18S ribosomal RNA gene, partial sequence
AY527055.1	*Mooreonuphis stigmatis* 18S ribosomal RNA gene, partial sequence
AY527049.1	*Rheomorpha neiswestonovae* 18S ribosomal RNA gene, partial sequence
AF202985.1	*Bactrurus pseudomucronatus* 18S ribosomal RNA gene, complete sequence

Uncultured marine eukaryote clone BTPL20040810.0146 18S ribosomal RNA gene,.
Pyramimonas gelidicola strain AnM0046 small subunit ribosomal RNA gene, partial ..
Uncultured marine eukaryote clone SA1_1B02 18S ribosomal RNA gene, partial seq.
Babesia canis 18S ribosomal RNA
Aora gracilis 18S ribosomal RNA gene, complete sequence

Tectocepheus sarekensis isola
mites & ticks | 9 leaves
Coelotes terrestris 18S rR...
Atopochthonius artiodactylus
sea spiders | 6 leaves

crustaceans | 4 leaves

H.tridens shenei 18S rRNA gene
P.pisum gene for 18S ribosomal RNA
Clavella adduncа 18S small subunit ribosomal RNA gene, partial sequence
Ameira scotti 18S ribosomal RNA gene, partial sequence
Pandarus smithi 18S ribosomal RNA gene, partial sequence

stoneflies | 4 leaves
Oxya chinensis 18S ribosomal RNA gene, complete sequence
H.tridens tientsinensis 18S rRNA gene
stoneflies | 2 leaves

Belisarius xambeui 18S ribosomal RNA gene, complete sequence
Nemurella pictetii 18S ribosomal RNA gene, partial sequence
Bactrurus pseudomucronatus 18S ribosomal RNA gene, complete sequence
Bactrurus brachycaudus isolate 92- 18s 18S ribosomal RNA gene, complete sequence
Uncultured marine eukaryote clone BTPL20040810.0096 18S ribosomal RNA gene, partial sequence
Uncultured marine eukaryote clone BTQB20040719.0044 18S ribosomal RNA gene, partial sequence
Mecyclothorax vulcans 18S ribosomal RNA gene, complete sequence
Potamodrilus fluviatilis 18S ribosomal RNA gene, partial sequence
Mooreonuphis stigmatis 18S ribosomal RNA gene, partial sequence
Rheomorpha neiswestonovae 18S ribosomal RNA gene, partial sequence
Uncultured eukaryote clone TW5130 18S ribosomal RNA gene, partial sequence
Bactrurus mucronatus 18S ribosomal RNA gene, complete sequence
Uncultured eukaryote clone TW5110 18S ribosomal RNA gene, partial sequence
Uncultured eukaryote clone TW5109 18S ribosomal RNA gene, partial sequence
Dorvillea cf. similis 18S ribosomal RNA gene, partial sequence
Hyalinoecia tubicola 18S ribosomal RNA gene, partial sequence
Nematonereis unicornis 18S ribosomal RNA gene, partial sequence
Eunice sp. SHP- 2002 18S ribosomal RNA gene, partial sequence
Eurycletodes laticauda 18S ribosomal RNA gene, partial sequence
Itunella muelleri 18S ribosomal RNA gene, partial sequence
Argestigens sp. Greenland- RJH- 2007 18S ribosomal RNA gene, partial sequence
Sarsameira sp. Greenland- RJH- 2007 18S ribosomal RNA gene, partial sequence
Paramenophia sp. New Caledonia- RJH- 2007 18S ribosomal RNA gene, partial sequence
Sewellia tropica 18S ribosomal RNA gene, partial sequence
Diarthrodes sp. 2 New Caledonia- RJH- 2007 18S ribosomal RNA gene, partial sequence
Diarthrodes sp. 1 New Caledonia- RJH- 2007 18S ribosomal RNA gene, partial sequence
Dactylopusia sp. New Caledonia- RJH- 2007 18S ribosomal RNA gene, partial sequence
Paracyclopina nana 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, 2...
Lepeophtheirus salmonis external transcribed spacer, partial sequence; and 18S ribosomal RNA gene, internal transcribed spacer 1, 5....
Ectocyclops polyspinosus 18S rRNA gene
Macrocyclops albidus 18S rRNA gene
Aphotopontius mammillatus 18S ribosomal RNA gene, partial sequence
Aeolosoma sp. 18S rRNA gene
Hatschekia sp. RJH- 2006 18S ribosomal RNA gene, partial sequence
Macrocyclops albidus 18S ribosomal RNA gene, partial sequence
Lepeophtheirus hippoglossi 18S ribosomal RNA gene, partial sequence
Dissonus manteri 18S ribosomal RNA gene, partial sequence
Kroyeria sp. RJH- 2006 18S ribosomal RNA gene, partial sequence
Apocyclops royi 18S small subunit ribosomal RNA gene, partial sequence

Xestospongia muta 18S ribosomal RNA gene, complete sequence ➔ **Phylum Porifera**
Uncultured marine eukaryote clone NA1_1D3 18S ribosomal RNA ..
Macrobothridium sp. 18S small subunit ribosomal RNA gene, comp.

Pycnogonum litorale 18S ribosomal RNA gene, partial sequence
Pontoeciella abyssicola 18S small subunit ribosomal RNA gene, partial sequence
Hatschekia pagrosomi 18S small subunit ribosomal RNA gene, partial sequence
Ecbathyrion prolixicauda 18S small subunit ribosomal RNA gene, partial sequence
Rhogobius contractus 18S small subunit ribosomal RNA gene, partial sequence
Caligus elongatus 18S small subunit ribosomal RNA gene, partial sequence
Gloiopotes watsoni 18S small subunit ribosomal RNA gene, partial sequence
Asterocheridae sp. 1 New Caledonia- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Pachos sp. Japan- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Xarifia sp. New Caledonia- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Critomolgus sp. 2 New Caledonia- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Critomolgus sp. 1 New Caledonia- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Chondracanthus lophii 18S ribosomal RNA gene, complete sequence
Lernentoma asellina 18S small subunit ribosomal RNA gene, partial sequence
Acanthocyclops viridis 18S small subunit ribosomal RNA gene, partial sequence
Cyclops sp. UK- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
Euryte sp. New Caledonia- RJH- 2004 18S small subunit ribosomal RNA gene, partial sequence
lcl|44401
Uncultured eukaryote clone TW5120 18S ribosomal RNA gene, partial sequence

Phylum Arthropoda

Phylum Annelida

Phylum Arthropoda

➔ **Phylum Platyhelminthes**

Phylum Arthropoda

Fig. A.4. Phylogenetic Tree of Sequence 14 Results GenBank Tree Construction Tool. The highlighted yellow region is my sequenced sample. This tree reveals the divergence of the various 18S rRNA sequences by examining the Phylum Arthropoda being split into various clusters. The arthropods have been known to show extremely divergent sequences because of the extreme diversity among the phylum.

APPENDIX B

ARDRA ANALYSES

In May 2009, 32 individual chaetognaths were selected for DNA extraction. All 32

samples went through PCR amplification for each of the three primers and the samples

were then run on a 1% agarose gel at 100 volts for 1 hour (Figs. B.1-3). Then the

amplified products went through ARDRA analyses to further indentify the similarities

and/or differences of each sample (Figs. B 4-9).To further investigate, the ARDRA gels

from each primer where analyzed using the BioNumerics software to evaluate any

similarities or differences in each sample (Figs. B. 10-18). I wanted to see if there was a

difference between the three statistical analyses, so for each primer set, I chose the

Pearson Correlation, Ranked Pearson Correlation, and the Jaccard method. All

dendrograms were constructed using the UPGMA method. The Pearson Correlation is

the most common measure of correlation and uses two variables to reflect the degree of

which the variables are related (Page and Holmes 1998). The Ranked Pearson

Correlation also provides a relation of variables, but is often used when the Pearson

Correlation gives misleading results (Kendall and Gibbons 1990). The Jaccard index,

also known as the Jaccard similarity coefficient, is used for comparing the similarity and

diversity of the sample set (Johnson and Wichern 1988). These gels were then grouped

according to the dendrograms from the BioNumerics software. This grouping also

helped to determine which samples would be picked for sequencing. After running the

ARDRA gels, I chose 8 samples that appeared to be different in the gel banding patterns.

The arrows on each dendrogram represent samples that were unique and could not be

grouped. The 8 samples I chose that appeared to be different according to their gel

banding patterns were samples 4, 6, 14, 15,16, 21, 30, 32.
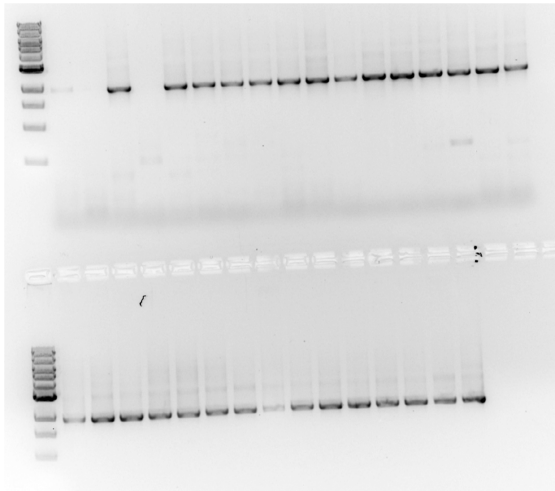
Fig. B. 1. Amplified products of 32 chaetognath samples for 18s primer (~1900bp)
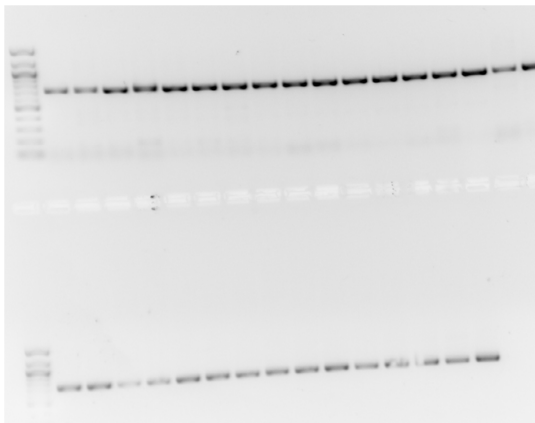
Fig. B.2. Amplified products of 32 chaetognath samples for COI primer (708 bp)
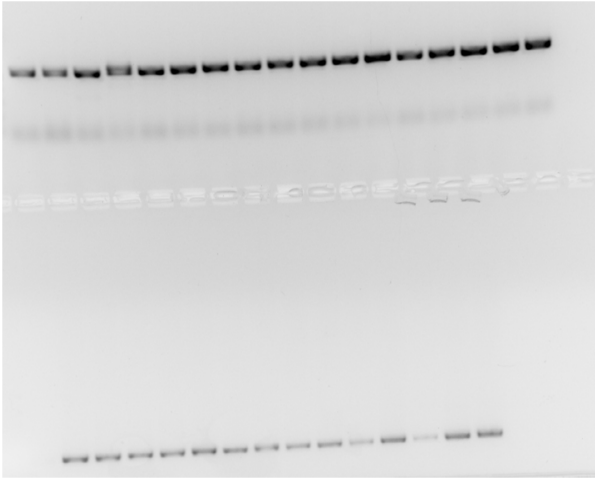
Fig.B.3. Amplified products of 32 chaetognath samples for COII primer (551bp)
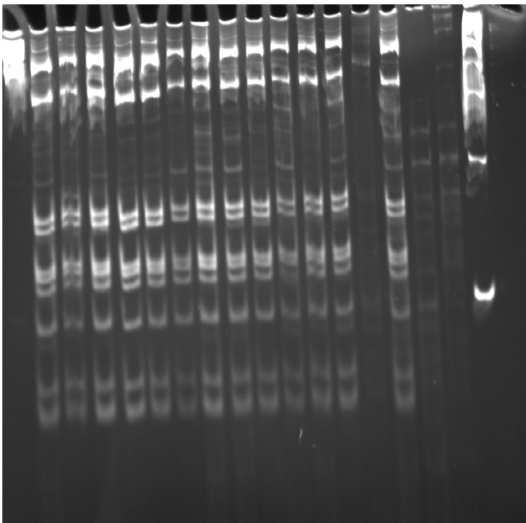
Fig.B.4. 18s ARDRA gel samples 1-16
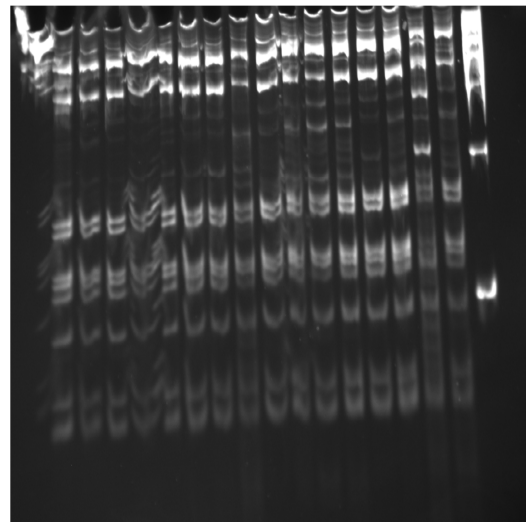
Fig.B.5. 18s ARDRA gel samples 17-32
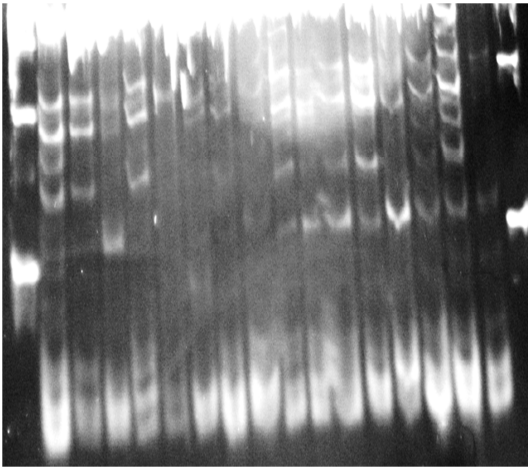
Fig.B.6. COI ARDRA gels samples 1-16
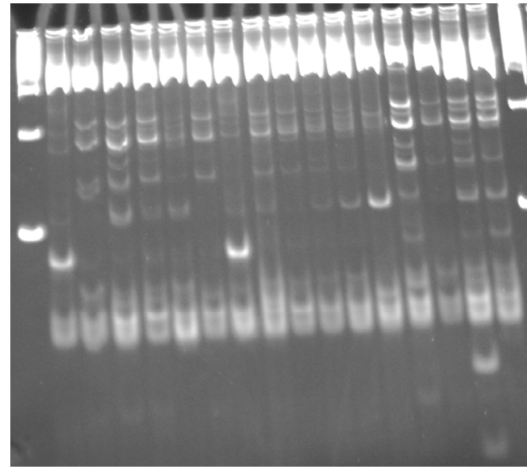
Fig.B.7. COI ARDRA gels samples 17-32
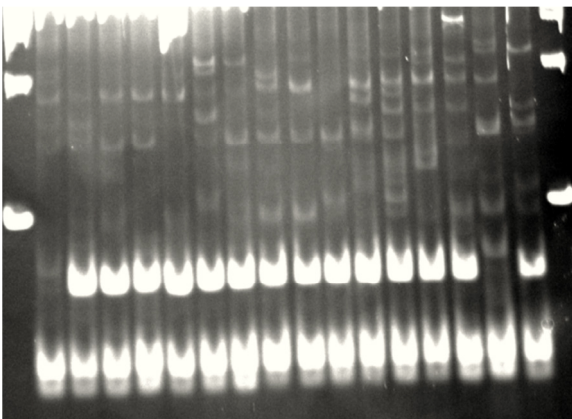
Fig.B.8. COII ARDRA gels samples 1-16
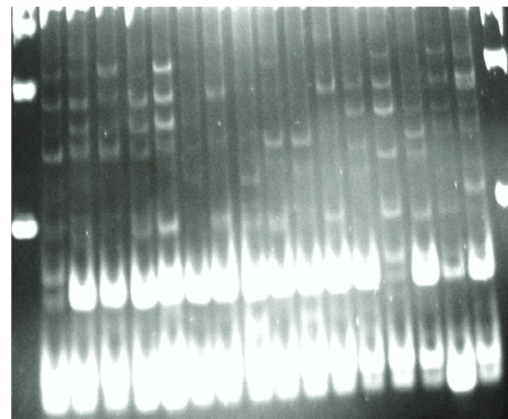
Fig.B.9. COII ARDRA gels samples 17-32
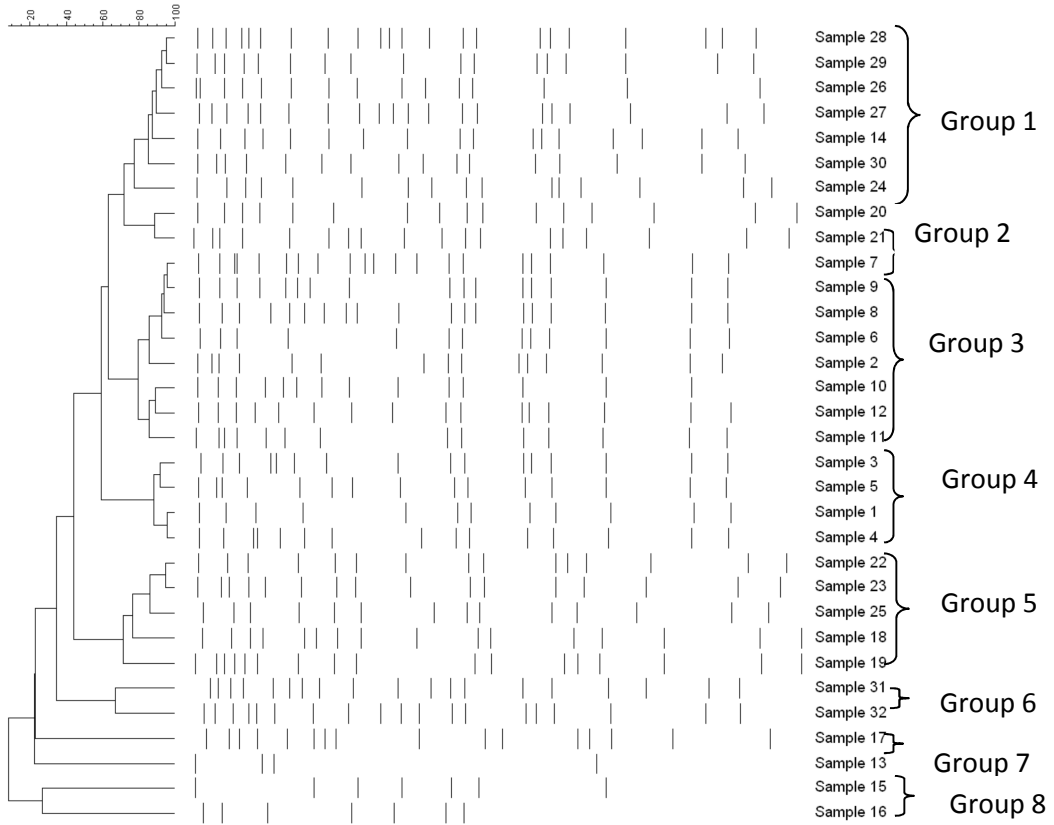
Fig.B.10. 18s samples/ Pearson Correlation/UPGMA
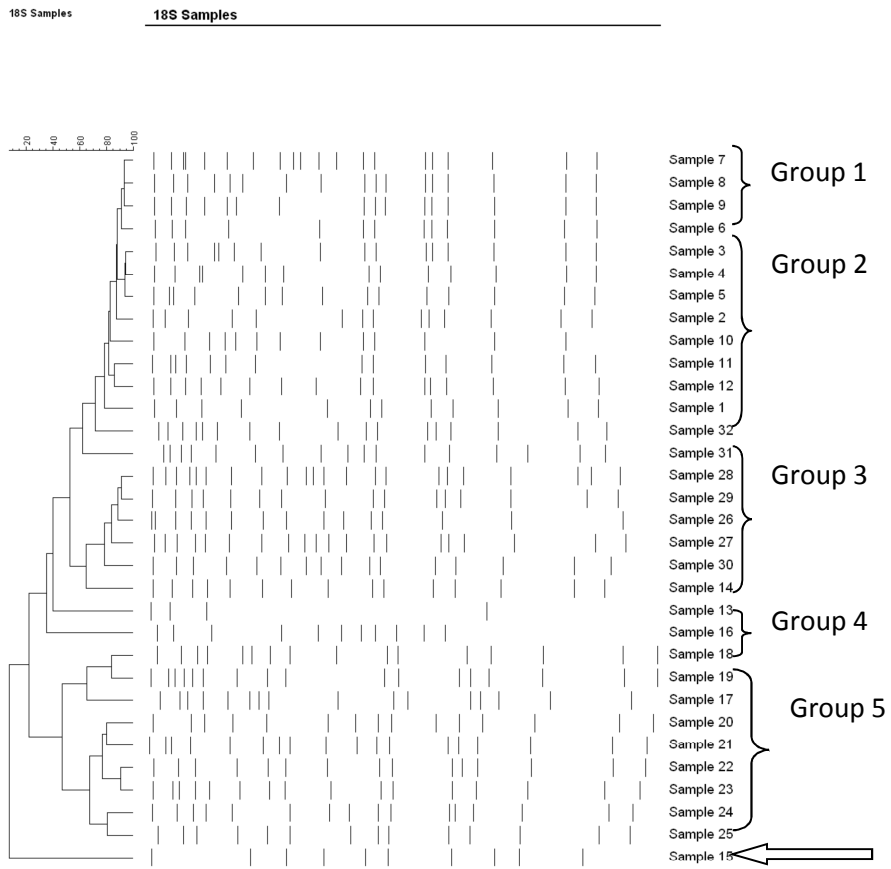
Fig.B.11. 18s samples/Ranked Pearson Correlation/UPGMA

18s jacc upgma (32 entries)

18S Samples
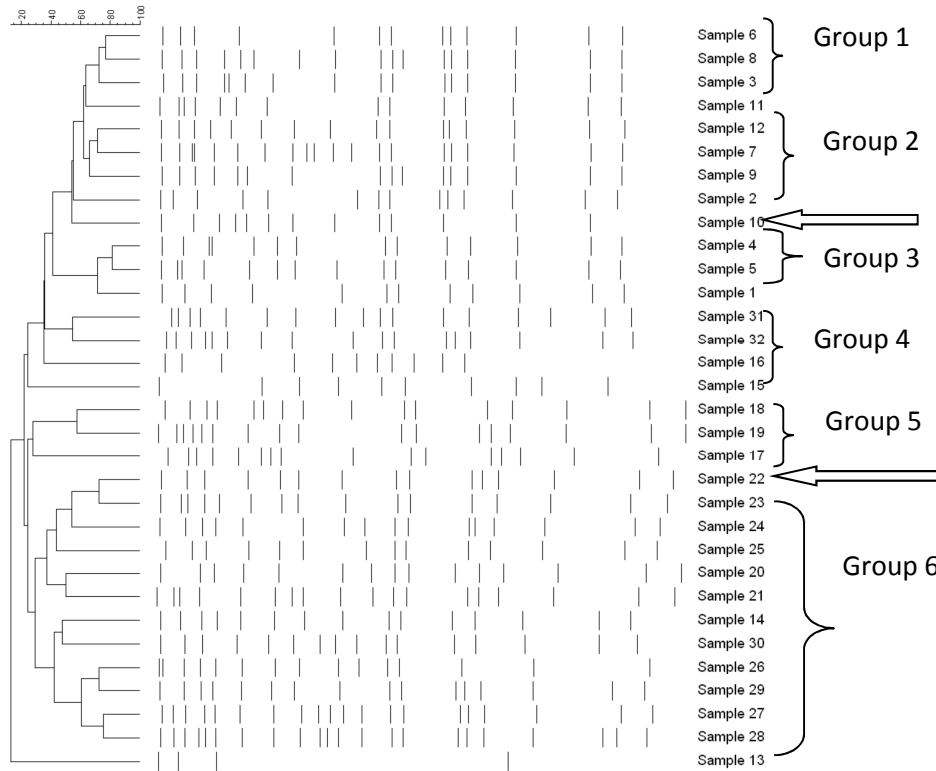
Fig.B.12. 18s Jaccard/UPGMA

CO1 Samples
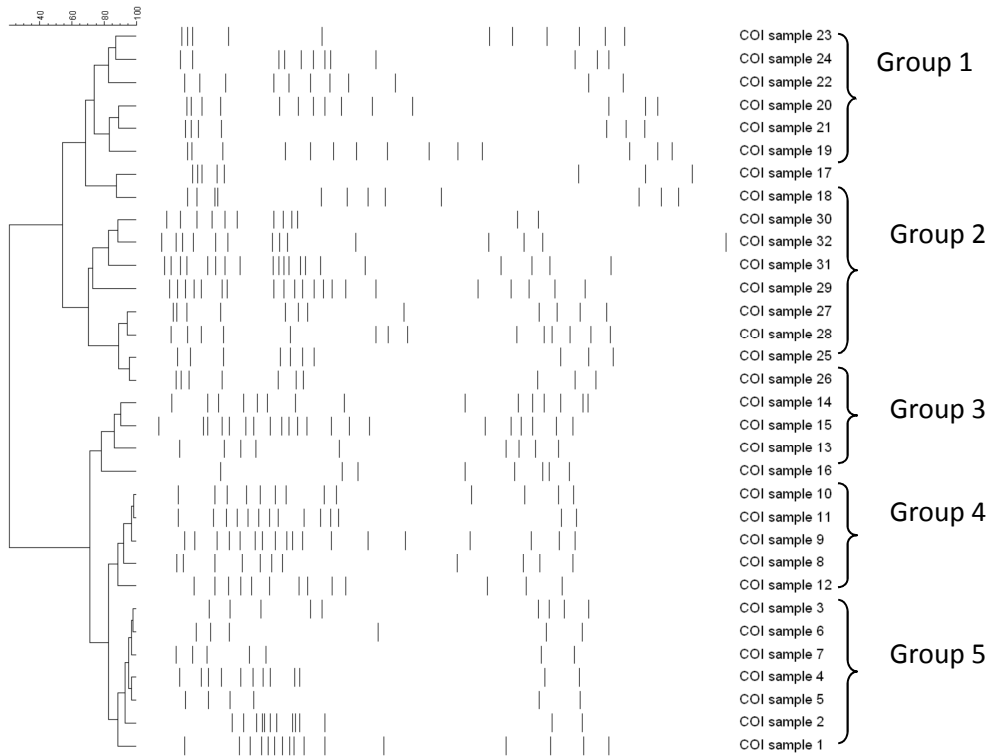
Fig.B.13. COI samples/ Pearson correlation/UPGMA
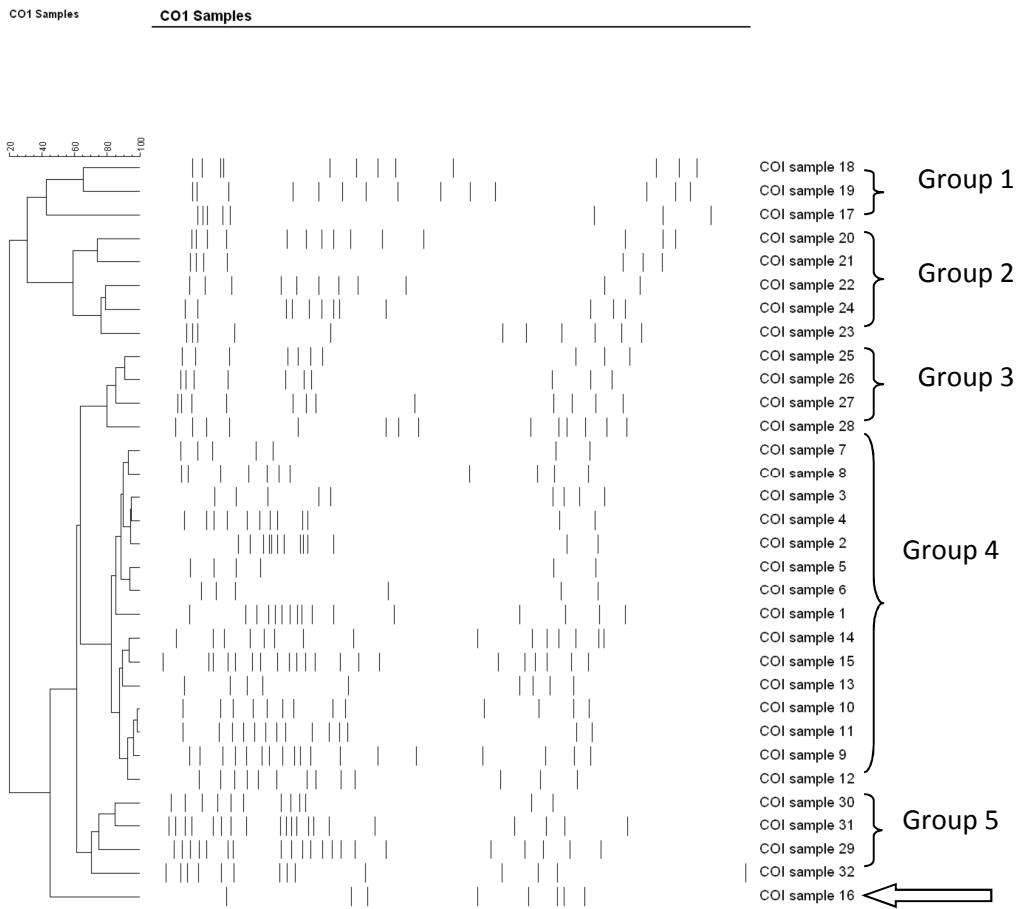
Fig.B.14. COI samples/ Ranked Pearson Correlation/UPGMA
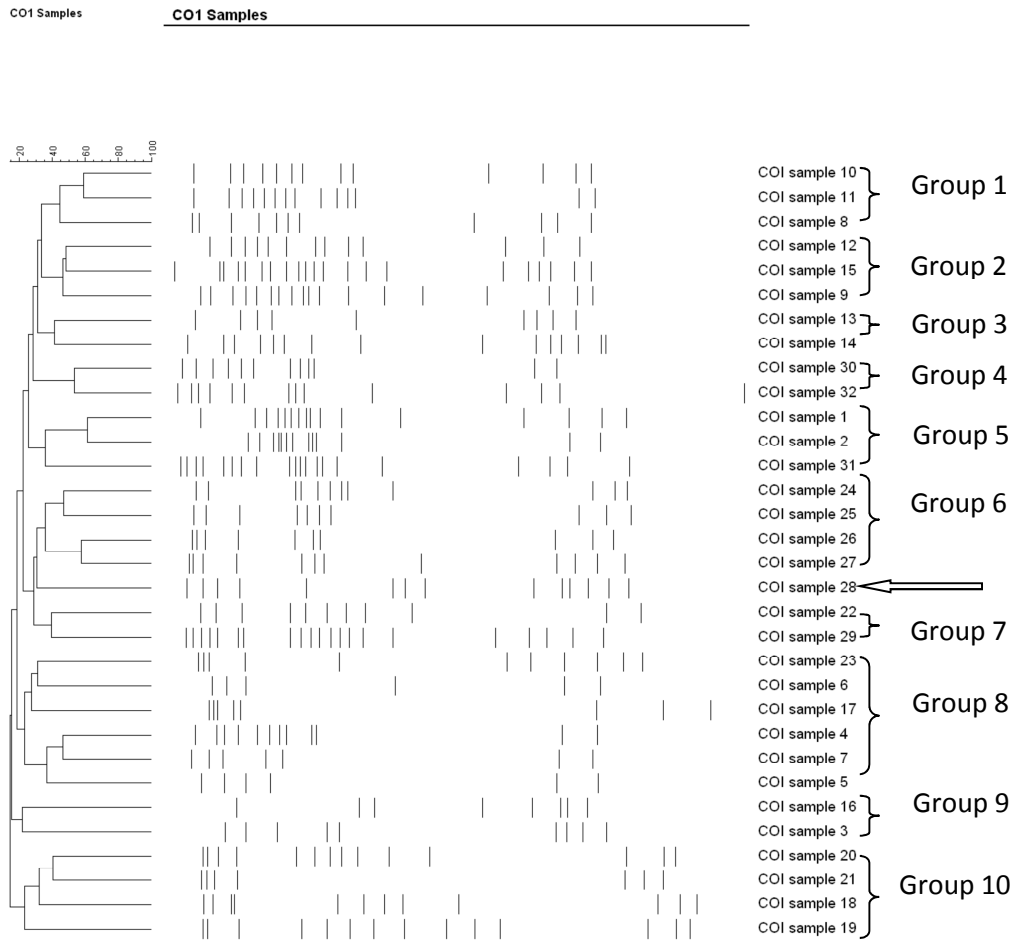
Fig.B.15. COI samples/Jaccard/UPGMA
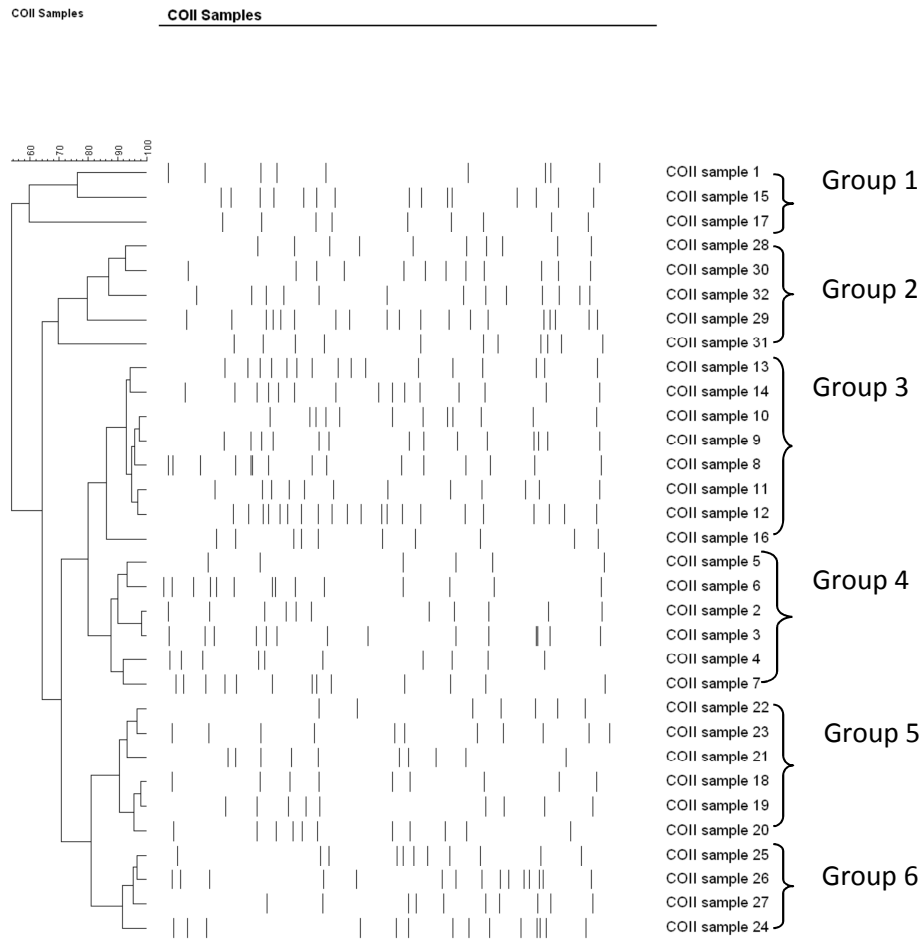
Fig.B.16. COII samples/Pearson Correlation/UPGMA
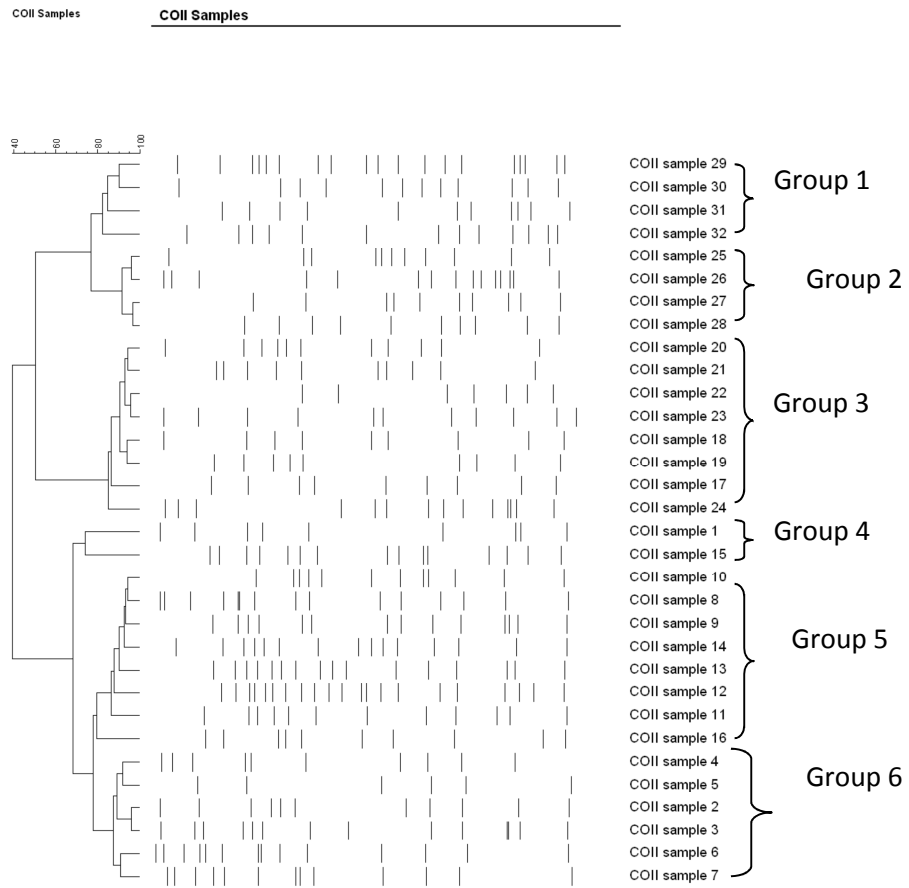
Fig.B.17. COII samples/Ranked Pearson Correlation/UPGMA
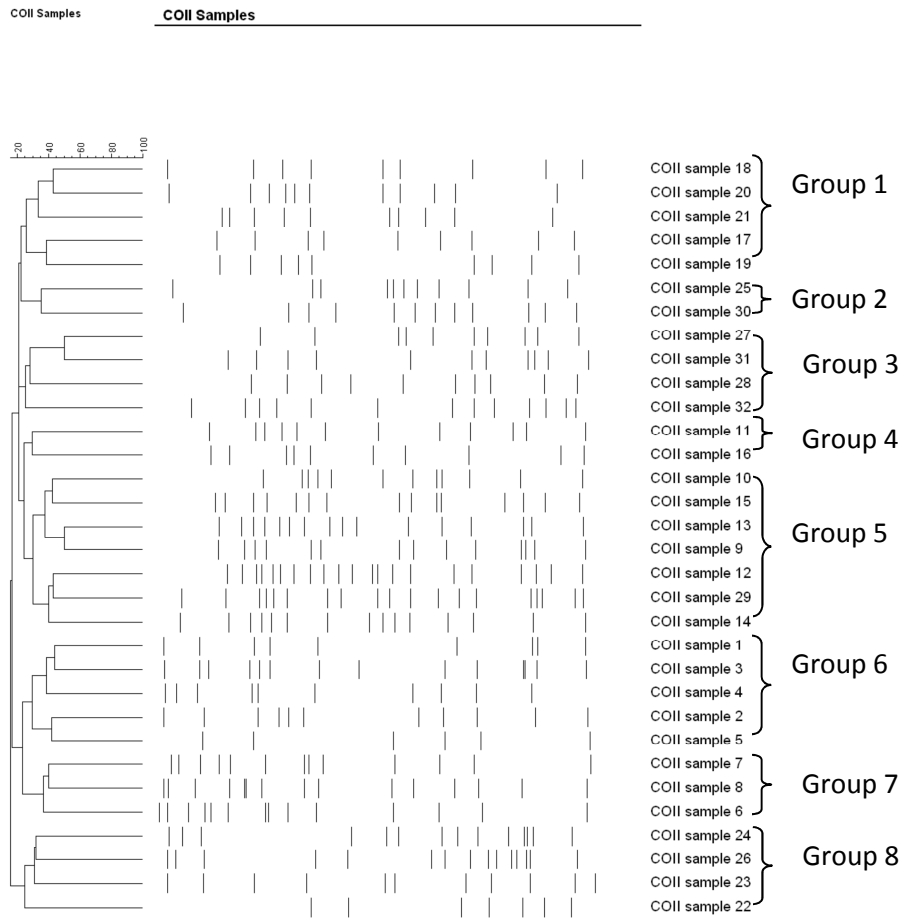
Fig.B.18. COII Samples/ Jaccard/ UPGMA

APPENDIX C

TABLES AND GRAPHIC REPRESENTATIONS

OF EACH PHYLOGENETIC TREE

The following tables and pie charts represent each tree reconstruction results using all three genes. For each table a lineage column was added to better understand the groupings that were found. If the 8 samples were grouped with chaetognaths, a question mark was used, because the lineage of the chaetognaths remained unknown until further analyzing of each tree reconstruction method was performed. The pie charts reveal the percentages of each phylum that grouped with my 8 collected samples.

18S rRNA Data

Table C.1. 18S samples using the Neighbor joining method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with *Nautilus s*p., *Ninominus s*p.	Mollusca/Arthropoda	Protostome/Protostome
18S-6	Groups with 14, 21,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
18S-14	Groups with 6, 21,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
18S-15	Groups with *Obininia sp., Myonera sp.*	Annelida/ Mollusca	Protostome/Protostome
18S-16	Groups with *Oigolamella sp., Diplogaster sp., Myctolaimus sp.*	Nematoda/Nematoda/Nematoda	Protostome
18S-21	Groups with 6, 14,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
18S-30	Groups with 6, 14, 21, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
18S-32	Groups with6, 14, 21,30, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome

18S Neighbor Joining

Fig.C.1. This method pairs the 8 samples more closely to Chaetognaths (46%) and Annelids (27%).

Table C.2. 18S samples using Minimum Evolution method showing the various phyla groupings.

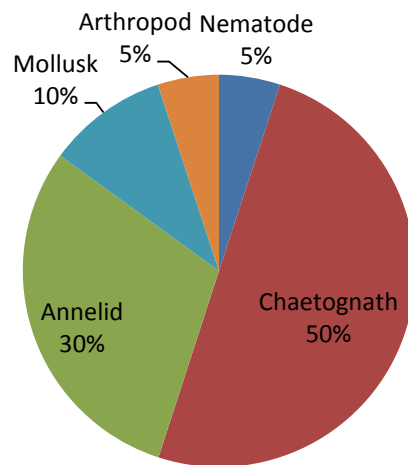| Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with *Nautilus sp., Gyndes sp.*	Mollusca/Arthropoda	Protostome/Protostome
18S-6	Groups with 14, 21,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/Annelida	?/?/Protostome
18S-14	Groups with 6, 21,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/Annelida	?/?/Protostome
18S-15	Groups with *Obininia sp., Myonera sp.*	Annelida/ Mollusca	Protostome/Protostome
18S-16	Groups with *Oigolamella sp., Diplogaster sp., Myctolaimus sp.*	Nematoda/Nematoda/Nematoda	Protostome
18S-21	Groups with 6, 14,30, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/Annelida	?/?/Protostome
18S-30	Groups with 6, 14, 21, 32, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/Annelida	?/?/Protostome
18S-32	Groups with6, 14, 21,30, *Krohnitta sp., Pectinaria sp.*	Chaetognatha/Chaetognatha/Annelida	?/?/Protostome

18S Minimum Evolution

Fig. C.2. This method pairs the 8 samples more closely to Chaetognaths (46%) and Annelids (27%).

Table C.3. 18S samples using Maximum Parsimony method showing the various phyla groupings.

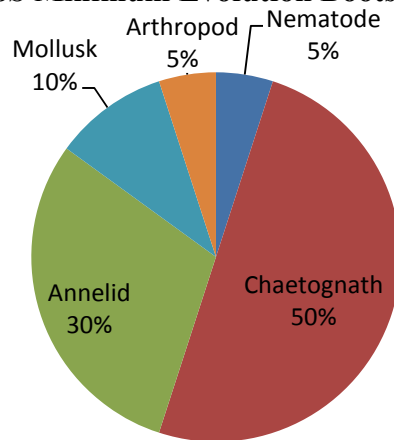| Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with 15,16, *Caenolestocoptes sp., Longidorus sp*.	Arthropoda/Nematoda	Protostome/Protostome
18S-6	Groups with 14,21,30,32, *Pectinaria sp., Krohnitta sp.*	Chaetognatha/Annelida/ Chaetognatha	?/Protostome/?
18S-14	Groups with 6,21,30,32, *Pectinaria sp., Krohnitta sp.*	Chaetognatha/Annelida/ Chaetognatha	?/Protostome/?
18S-15	Groups with 4,16, *Caenolestocoptes sp., Longidorus sp.*	Arthropoda/Nematoda	Protostome/Protostome
18S-16	Groups with 4,15, *Caenolestocoptes sp., Longidorus sp.*	Arthropoda/Nematoda	Protostome/Protostome
18S-21	Groups with 6,14,30,32, *Pectinaria sp., Krohnitta sp.*	Chaetognatha/Annelida/ Chaetognatha	?/Protostome/?
18S-30	Groups with 6,14,21,32, *Pectinaria sp., Krohnitta sp.*	Chaetognatha/Annelida/ Chaetognatha	?/Protostome/?
18S-32	Groups with 6,14,21,30, *Pectinaria sp., Krohnitta sp.*	Chaetognatha/Annelida/ Chaetognatha	?/Protostome/?

18S Maximum Parsimony
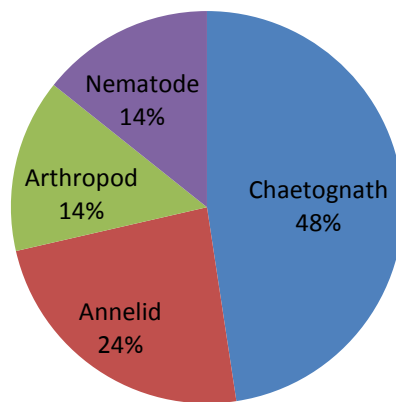
Fig.C.3. This method pairs the 8 samples more closely to Chaetognaths (48%) and Annelids (24%).

Table C.4.18S samples using the UPGMA method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with 15, *Spodotera sp.*	Chaetognatha/Arthropoda	?/Protostome
18S-6	Groups with 14,21,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-14	Groups with 6,21,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-15	Groups with 4, *Spodotera sp.*	Chaetognatha/Arthropoda	?/Protostome
18S-16	Groups with *Nautilus sp., Speleonectes sp.*	Mollusca/Arthropoda	Protostome/Protostome
18S-21	Groups with 6,14,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-30	Groups with 6,14,21,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-32	Groups with 6,14,21,30, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome

.

Fig.C.4.This method pairs the 8 samples more closely to Arthropods (31%) and Chaetognaths (27%).

Table C.5.18S samples using Neighbor Joining Bootstrap method showing various phyla groupings.

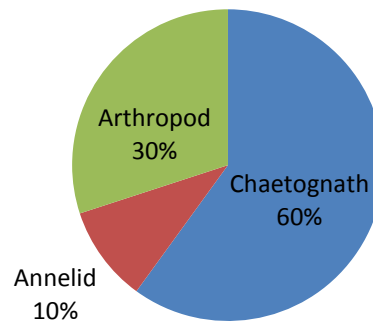| Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with *Nautilus sp., Ninomimus sp.*	Mollusca/Arthropoda	Protostome/Protostome
18S-6	Groups with 14,21,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-14	Groups with 6,21,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-15	Groups with *Orbinia sp., Myonera sp.*	Annelida/ Mollusca	Protostome/Protostome
18S-16	Groups with *Myctolaimus sp., Rhabditoides sp.*	Nematoda/Nematoda	Protostome/Protostome
18S-21	Groups with 6,14,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-30	Groups with 6,14,21,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-32	Groups with 6,14,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome

18S Neighbor Joining Bootstrap

Fig.C.5.This method pairs the 8 samples more closely to Chaetognaths (50%) and Annelids (30%).

Table C.6.18S samples using Minimum Evolution Bootstrap method showing various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with *Nautilus sp., Gyndes sp.*	Mollusca/Arthropoda	Protostome/Protostome
18S-6	Groups with 14,21,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-14	Groups with 6,21,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-15	Groups with *Orbinia sp., Myonera sp.*	Annelida/ Mollusca	Protostome/Protostome
18S-16	Groups with *Diplogaster sp.*	Nematoda	Protostome
18S-21	Groups with 6,14,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-30	Groups with 6,14,21,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome
18S-32	Groups with 6,14,30,32, *Krohnitta sp., Pectinaria sp.*	Chaetognath/Chaetognath/ Annelida	?/?/Protostome

18S Minimum Evolution Bootstrap

Fig C.6.This method pairs the 8 samples more closely to Chaetognaths (50%) and Annelids (30%).

Table C.7.18S samples using Maximum Parsimony Bootstrap method showing various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with 15,16, *Hutchinsoniella sp., Pleomothra sp., Speleonectes sp., Pectinaria sp.*	Chaetognatha/ Arthropoda/Arthropoda/ Arthropoda/Annelida	?/ Protostome/Protostome/ Protostome/Protostome
18S-6	Groups with 14,21, 30,32, *Krohnitta sp., Sagitta sp.*	Chaetognatha/Chaetognatha/ Chaetognatha	?/?/?
18S-14	Groups with 6,21, 30,32, *Krohnitta sp., Sagitta sp.*	Chaetognatha/Chaetognatha/ Chaetognatha	?/?/?
18S-15	Groups with 4,16, *Hutchinsoniella sp., Pleomothra sp., Speleonectes sp., Pectinaria sp.*	Chaetognatha/ Arthropoda/Arthropoda/ Arthropoda/Annelida	?/ Protostome/Protostome/ Protostome/Protostome
18S-16	Groups with 4,15, *Hutchinsoniella sp., Pleomothra sp., Speleonectes sp., Pectinaria sp.*	Chaetognatha/ Arthropoda/Arthropoda/ Arthropoda/Annelida	?/ Protostome/Protostome/ Protostome/Protostome
18S-21	Groups with 6,14, 30,32, *Krohnitta sp., Sagitta sp.*	Chaetognatha/Chaetognatha/ Chaetognatha	?/?/?
18S-30	Groups with 6,14, 21,32, *Krohnitta sp., Sagitta sp.*	Chaetognatha/Chaetognatha/ Chaetognatha	?/?/?
18S-32	Groups with 6,14, 21,30, *Krohnitta sp., Sagitta sp.*	Chaetognatha/Chaetognatha/ Chaetognatha	?/?/?

18S Maximum Parsimony Bootstrap

Fig.C.7.This method pairs the 8 samples more closely to Chaetognaths (60%) and Arthropods (30%).

Table C.8.18S samples using UPGMA Bootstrap method showing various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
18S-4	Groups with 15, *Spodotera sp.*	Chaetognatha/Arthropoda	?/Protostome
18S-6	Groups with 14,21,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-14	Groups with 6,21,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-15	Groups with 4, *Spodotera sp.*	Chaetognatha/Arthropoda	?/Protostome
18S-16	Groups with *Nautilus sp., Speleonectes sp.*	Mollusca/Arthropoda	Protostome/Protostome
18S-21	Groups with 6,14,30,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-30	Groups with 6,14,21,32, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome
18S-32	Groups with 6,14,21,30, *Henneguya sp., Phasmarhabditis sp., Pleomothra sp.*	Chaetognatha/Cnidaria/ Nematoda/Arthropoda	?/Protostome/Protostome/ Protostome/Protostome

18S UPGMA Bootstrap
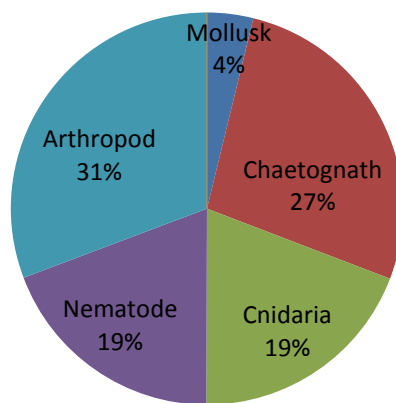
Fig.C.8. This method pairs the 8 samples more closely to Arthropods (31%) and Chaetognaths (27%).

Table C.9. COI Neighbor Joining method showing the various phyla groupings.

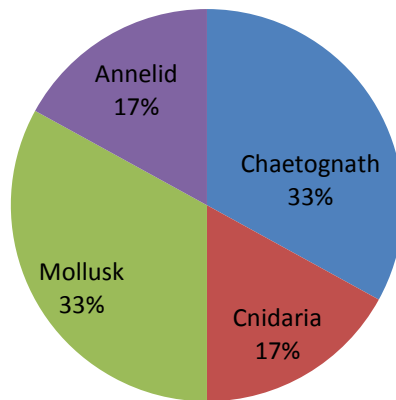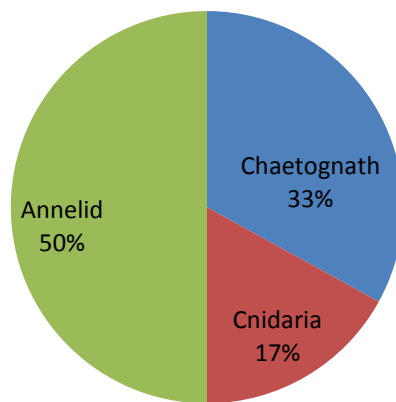| Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6,30,32, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/ Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp. and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/ Protostome/Protostome
COI-14	Groups with15, 16, 21, *Obelia sp., and Monastrea sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/ Protostome/Protostome
COI-15	Groups with 14, 16, 21, *Obelia sp., and Monastrea sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-16	Groups with 14,15,21, *Obelia sp., and Monastrea sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-21	Groups with 14,15,16, *Obelia sp. and Monastrea sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-30	Groups with 4, 6, 32, *Paralvinella sp. , and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Neighbor Joining
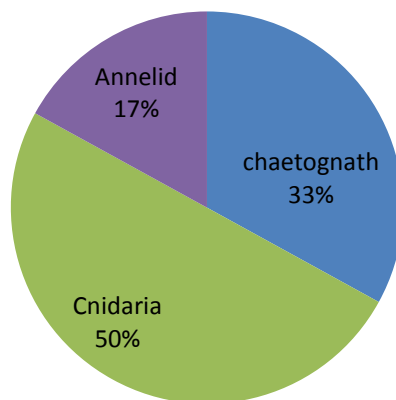
Fig.C.9. This method pairs the 8 samples more closely to Cnidaria (50%) and Chaetognaths (33%).

Table C.10. COI Minimum evolution method showing various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30,32, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with 15,16, 21, *Pomatias sp., and Tudorella sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-15	Groups with 14,16, 21, *Pomatias sp., and Tudorella sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-16	Groups with 14, 15, 21, *Pomatias sp., and Tudorella sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-21	Groups with 14,15,16, *Pomatias sp., and Tudorella sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Paralvinella sp., and Chironex sp. s*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Minimum Evolution

Fig.C.10. This method pairs the 8 samples more closely to Mollusks (33%) and Chaetognaths (33%).

Table C.11. COI Maximum Parsimony method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30, 32, *Parlavinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp., and Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with *Gnathostoma sp., and Neodiplogaster sp.*	Nematoda/Nematoda	Protostome/Protostome
COI-15	Groups with *Aurelia sp., Bursaphelenchus sp.*	Cnidaria/Nematoda	Protostome/Protostome
COI-16	Groups with *Spiralothelphusa sp., Anisakis sp.*	Arthropoda/Nematoda	Protostome/Protostome
COI-21	Groups with *Dactylopius sp., Mytilus sp.*	Arthropoda/Mollusca	Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/ Protostome/Protostome
COI-32	Groups with 4, 6, 30, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Maximum Parsimony

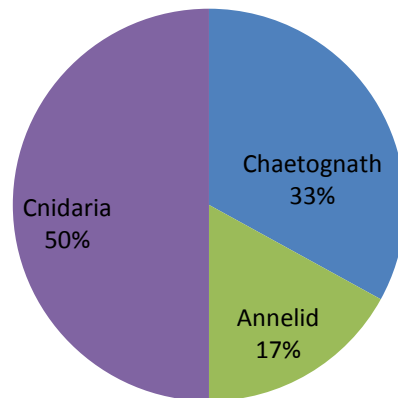

Fig C.11.This method pairs the 8 samples more closely to Annelids (50%) and Chaetognaths (33%).

Table C.12. COI UPGMA method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30, 32, *Chironex sp., Paralvinella sp.*	Chaetognatha/Cnidaria/ Annelida	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Chironex sp., Paralvinella sp.*	Chaetognatha/Cnidaria/ Annelida	?/Protostome/Protostome
COI-14	Groups with 15, 16, 21, *Tudorella sp., Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-15	Groups with 14,16, 21, *Tudorella sp., Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-16	Groups with 14,15, 21, *Tudorealla sp., Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-21	Groups with 14, 15, 16, *Tudorella sp, Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Chironex sp., Paralvinella sp.*	Chaetognatha/Cnidaria/ Annelida	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Chironex sp., Paralvinella sp.*	Chaetognatha/Cnidaria/ Mollusca	?/Protostome/Protostome

COI UPGMA
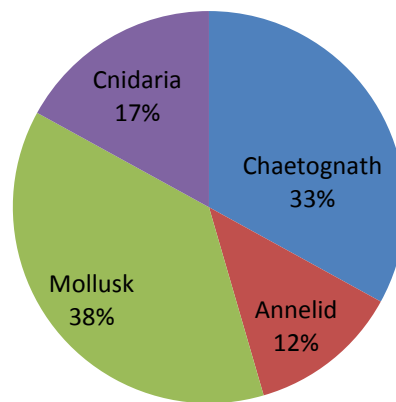
Fig.C.12. This method pairs the 8 samples more closely to Mollusks (38%) and Chaetognaths (33%).

Table C.13.COI Neighbor Joining Bootstrap method showing the various phyla groupings.

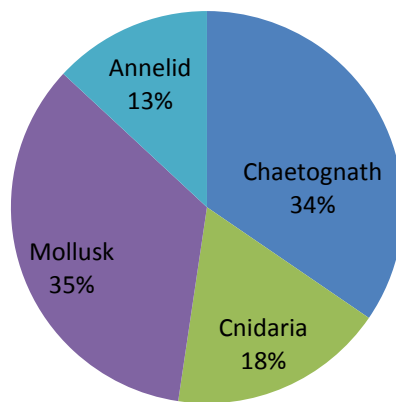| Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with 15,16, 21, *Monastrea sp., Obelia sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-15	Groups with 14, 16,21 , *Monastrea sp., Obelia sp*.	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-16	Groups with 14, 15, 21, *Monastrea sp., Obelia sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-21	Groups with 14,15, 16, *Monastrea sp., Obelia sp*.	Chaetognath/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Neighbor Joining Bootstrap

Fig.C.13.This method pairs the 8 samples more closely to Cnidaria (50%) and Chaetognaths (33%).

Table C.14. COI Minimum Evolution Bootstrap method showing the various phyla groupings.

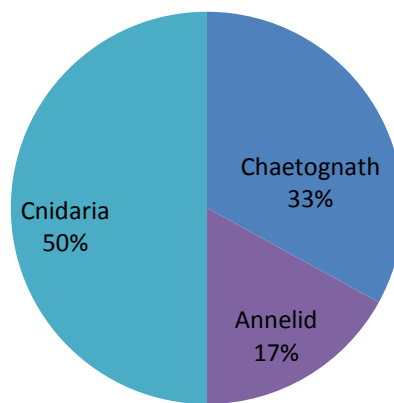| Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30, 32,*Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with 15,16 ,21, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-15	Groups with 14, 16, 21, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?Protostome/Protostome
COI-16	Groups with 14,15, 21, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-21	Groups with 14,15, 16, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Minimum Evolution Bootstrap

Fig.C.14.This method pairs the 8 samples more closely to Mollusks (35%) and Chaetognaths (34%).

Table C.15. COI Maximum Parsimony Bootstrap method showing the various phyla groupings.

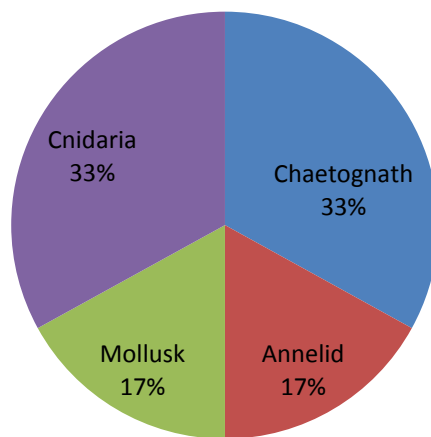| Sample ID	Closes grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30,32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with 15,16, 21, *Monastrea sp., Urticina sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-15	Groups with 14,16, 21, *Monastrea sp., Urticina sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-16	Groups with 14,15, 21, *Monastrea sp., Urticina sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-21	Groups with 14,15, 21, *Monastrea sp., Urticina sp.*	Chaetognatha/Cnidaria/ Cnidaria	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4, 6, 30, *Paralvineall sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

COI Maximum Parsimony Bootstrap

Fig.C.15. This method pairs the 8 samples more closely to Cnidaria (50%) and Chaetognaths (33%).

Table C.16.COI UPGMA Bootstrap method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COI-4	Groups with 6, 30, 32, *Paralvinella sp., Chironex sp.*	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-6	Groups with 4, 30,32, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-14	Groups with 15,16, 21, *Tudorella sp., Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-15	Groups with 14,15, 21, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-16	Groups with 14,15, 21, *Tudorella sp., Pomatias sp*.	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-21	Groups with 14,15,16, *Tudorella sp., Pomatias sp.*	Chaetognatha/Mollusca/ Mollusca	?/Protostome/Protostome
COI-30	Groups with 4,6, 32, *Paralvinella sp., Chironex sp*.	Cheatognatha/Annelida/ Cnidaria	?/Protostome/Protostome
COI-32	Groups with 4,6, 30, *Paralvinella sp., Chironex sp*.	Chaetognatha/Annelida/ Cnidaria	?/Protostome/Protostome

Fig.C.16. This method pairs the 8 samples more closely to Cnidaria (33%) and Chaetognaths (33%).

Table C.17.COII samples using Neighbor Joining method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-6	Groups with 4,30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-14	Groups with 15,16, 21, *Aegla sp., Munida sp., and Sagitta sp.*	Chaetognatha/Arthropoda/ Arthropoda/Chaetognatha	?/Protostome/Protostome /?
COII-15	Groups with 14,16,21, *Aegla sp., Munida sp., and Sagitta sp.*	Chaetognatha/Arthropoda/ Arthropoda/Chaetognatha	?/Protostome/Protostome /?
COII-16	Groups with 14,15, 21, *Aegla sp., Munida sp., and Sagitta sp.*	Chaetognatha/Arthropoda/ Arthropoda/Chaetognatha	?/Protostome/Protostome /?
COII-21	Groups with 14,15,16, *Aegla sp., Munida sp., an Sagitta sp.*	Chaetognatha/Arthropoda/ Arthropoda/Chaetognatha	?/Protostome/Protostome /?
COII-30	Groups with 4,6, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-32	Groups with 4,6, 30, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome

COII Neighbor Joining
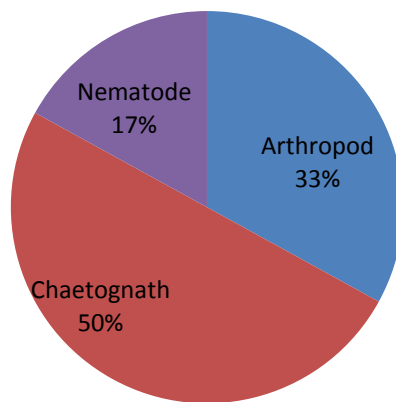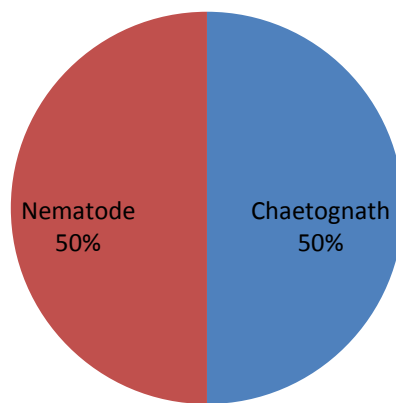

Fig.C.17.This method pairs the 8 samples more closely to Chaetognaths (50%) and Arthropods (33%).

Table C.18. COII samples using Minimum Evolution method showing the various phyla groupings.

Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-6	Groups with 4, 30, 32, *Parelphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-14	Groups with 15,16, 21, *Contracaecum sp.*	Chaetognatha/Nematoda	?/Protostome
COII-15	Groups with 14,16, 21, *Contracaecum sp.*	Chaetognatha/Nematoda	?/Protostome
COII-16	Groups with 14,15,21, *Contracaecum sp.*	Chaetognatha/Nematoda	?/Protostome
COII-21	Groups with 14,15, 16, *Contracaecum sp.*	Chaetognatha/Nematoda	?/Protostome
COII-30	Groups with 4,6, 32, *Parelpshostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-32	Groups with 4,6, 30, *Parelphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome

COII Minimum Evolution

Fig.C.18.This method pairs the 8 samples more closely to Chaetognaths (50%) and Nematodes (50%).

Table C.19.COII samples using Maximum Parsimony method showing the various phyla groupings.

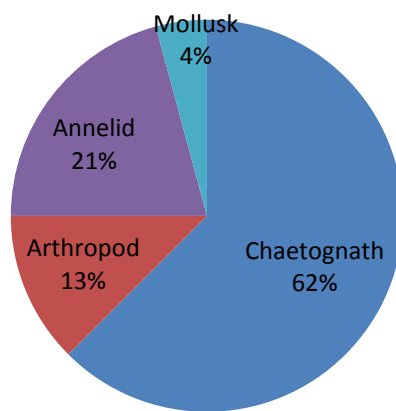| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6,15,30,32, *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
COII-6	Groups with 4, 15, 30,32, *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
COII-14	Groups with 16, *Sagitta sp., Carcinus sp.*	Chaetognatha/Chaetognatha/ Arthropoda	?/?/Protostome
COII-15	Groups with 4,6, 30, 32, *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
COII-16	Groups with 14, *Sagitta sp., Carcinus sp.*	Chaetognatha/Chaetognatha/ Arthropoda	?/?/Protostome
COII-21	Groups with *Sagitta sp, Guinotia sp., Haliotis sp.*	Chaetognatha/Arthropoda/ Mollusca	?/Protostome/ Protostome
COII-30	Groups with 4,6, 15, 32, *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome
COII-32	Groups with 4,6,15, 30, *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Chaetognatha/ Annelida	?/?/Protostome

COII Maximum Parsimony

Fig.C.19.This method pairs the 8 samples more closely to Chaetognaths (62%) and Annelids (21%).

Table C.20. COII samples using UPGMA method showing the various phyla groupings.

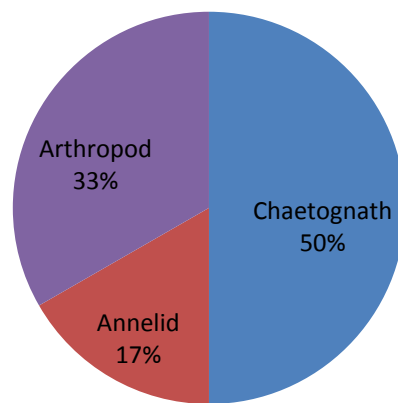| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Annelida	?/Protostome
COII-6	Groups with 4, 30, 32, *Parelaphostongylus sp.*	Chaetognatha/Annelida	?/Protostome
COII-14	Groups with 15,16,21, *Sagitta sp., Haplogaster sp., Galagete sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/ Protostome
COII-15	Groups with 14,16, 21, *Sagitta sp., Haplogaster sp., Galagete sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/ Protostome
COII-16	Groups with 14,15, 21, *Sagitta sp., Haplogaster sp., Galagete sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/ Protostome
COII-21	Groups with 14,15,16, *Sagitta sp., Haplogaster sp., Galagete sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/ Protostome
COII-30	Groups with 4,6, 32, *Parelaphostrongylus sp.*	Chaetognatha/Annelida	?/Protostome
COII-32	Groups with 4,6, 30, *Parelaphostrongylus sp.*	Chaetognatha/Annelida	?/Protostome

COII UPGMA

Fig.C.20.This method pairs the 8 samples more closely to Chaetognaths (50%) and Arthropods (33%).

Table C.21. COII samples using Neighbor Joining Bootstrap method showing the various phyla groupings.

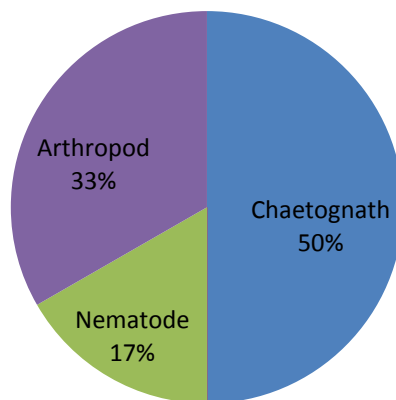| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-6	Groups with 4, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-14	Groups with 15,16,21, *Sagitta sp., Aegla sp., Munida sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/Protostome
COII-15	Groups with 14,16,21, *Sagitta sp., Aegla sp., Munida sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/Protostome
COII-16	Groups with 14,15,21, *Sagitta sp., Aegla sp., Munida sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/Protostome
COII-21	Groups with 14,15,16, *Sagitta sp., Aegla sp., Munida sp.*	Chaetognatha/Chaetognatha/ Arthropoda/Arthropoda	?/?/Protostome/Protostome
COII-30	Groups with 4,6, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-32	Groups with 4,6, 30, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome

COII Neighbor Joining Bootstrap

Fig.C.21.This method pairs the 8 samples more closely to Chaetognaths (50%) and Arthropods (33%).

Table C.22. COII samples using Minimum Evolution Bootstrap method showing the various phyla groupings.

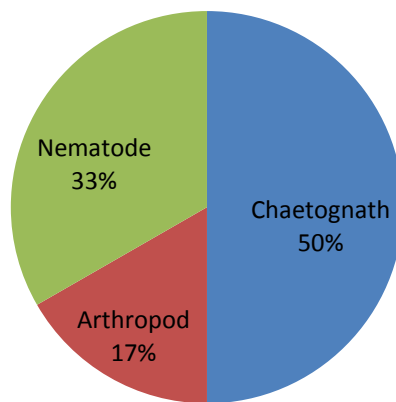| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	/Protostome
COII-6	Groups with 4,30,32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	/Protostome
COII-14	Groups with 15, 16, 21, *Sagitta sp,Contracaecum sp., Artemia sp.*	Chaetognatha/Chaetognatha/ Nematoda/Arthropoda	?/?/Protostome/Protostome
COII-15	Groups with 14,16, 21,*Sagitta sp., Contracaecum sp., Artemia sp.*	Chaetognatha/Chaetognatha/ Nematoda/Arthropoda	?/?/Protostome/Protostome
COII-16	Groups with 14,15, 21,*Sagitta sp., Contracaecum sp., Artemia sp.*	Chaetognatha/Chaetognatha/ Nematoda/Arthropoda	?/?/Protostome/Protostome
COII-21	Groups with 14,15, 16, *Sagitta sp., Contracaecum sp., Artemia sp.*	Chaetognatha/Chaetognatha/ Nematoda/Arthropoda	?/?/Protostome/Protostome
COII-30	Groups with 4,6, 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	/Protostome
COII-32	Groups with 4,6, 30, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	/Protostome

COII Minimum Evolution Bootstrap

Fig C.22.This method pairs the 8 samples more closely to Chaetognaths (50%) and Nematodes (33%).

Table C.23.COII samples using Maximum Parsimony Bootstrap method showing the various phyla groupings.

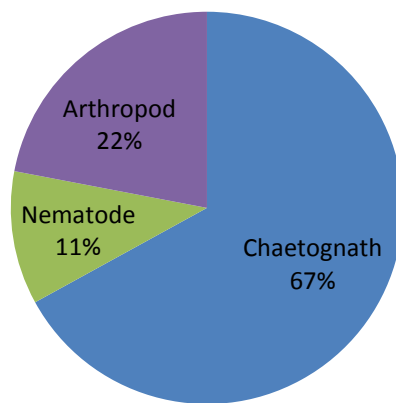| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, *Aegla sp.,Munida sp.*	Chaetognatha/Arthropoda/ Arthropoda	?/Protostome/Protostome
COII-6	Groups with 4, *Aegla sp.,Munida sp.*	Chaetognatha/Arthropoda/ Arthropoda	?/Protostome/Protostome
COII-14	Groups with 15,16,21, and *Sagitta sp.*	Chaetognatha/Chaetognatha	?/?
COII-15	Groups with 14,16,21, and *Sagitta sp.*	Chaetognatha/Chaetognatha	?/?
COII-16	Groups with 14,15,21, and *Sagitta sp.*	Chaetognatha/Chaetognatha	?/?
COII-21	Groups with 14,15,16, and *Sagitta sp.*	Chaetognatha/Chaetognatha	?/?
COII-30	Groups with 32, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome
COII-32	Groups with 30, *Parelaphostrongylus sp.*	Chaetognatha/Nematoda	?/Protostome

COII Maximum Parsimony Bootstrap

Fig.C.23.This method pairs the 8 samples more closely to Chaetognaths (67%) and Arthropods (22%).

Table C.24. COII samples using UPGMA Bootstrap method showing the various phyla groupings.

| Sample ID	Closest grouping	Associated phylum	Lineage
COII-4	Groups with 6, 30, 32, *Sagitta sp., Fletcherodrilus sp., Anopheles sp.*	Chaetognatha/Chaetognatha/ Annelida/Arthropoda	?/?/Protostome/Protostome
COII-6	Groups with 4,30, 32, *Sagitta sp., Fletcherodrilus sp., Anopheles sp.*	Chaetognatha/Chaetognatha/ Annelida/Arthropoda	?/?/Protostome/Protostome
COII-14	Groups with *Sagitta sp.*	Chaetognatha	?
COII-15	Groups with *Sagitta sp., Aporrectodea sp.*	Chaetognatha/Annelida	?/Protostome
COII-16	Groups with *Sagitta sp.*	Chaetognatha	?
COII-21	Groups with *Sagitta sp., Carcinus sp.*	Chaetognatha/Arthropoda	?/Protostome
COII-30	Groups with 4,6, 32, *Sagitta sp., Fletcherodrilus sp., Anopheles sp.*	Chaetognatha/Chaetognatha/ Annelida/Arthropoda	?/?/Protostome/Protostome
COII-32	Groups with 4,6, 30, *Sagitta sp.,Fletcherodrilus sp., Anopheles sp.*	Chaetognatha/Chaetognatha/ Annelida/Arthropoda	?/?/Protostome/Protostome

COII UPGMA Bootstrap
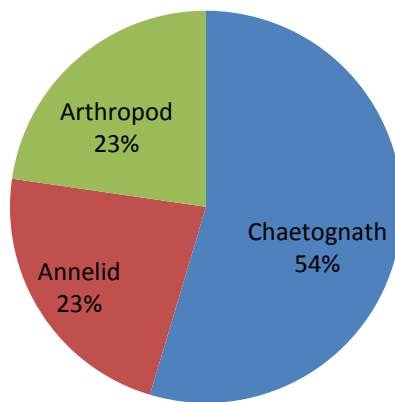
Fig.C.24.This method pairs the 8 samples more closely to Chaetognaths (54%) and Arthropods (23%) and Annelids (23%).

VITA

Name:	Leah Nicole Towers
Address:	Department of Marine Sciences Texas A&M University at Galveston 5007 Avenue U, Galveston, TX 77551
Email Address:	leah.towers@comcast.net
Education:	B.S., Marine Biology, Texas A&M University at Galveston, 2001 M.S., Biology, Texas A&M University, 2010