J Mol Evol (1997) 44:135–144

JOURNAL OF MOLECULAR EVOLUTION

Evolution of 28S Ribosomal DNA in Chaetognaths: Duplicate Genes and Molecular Phylogeny

Maximilian J. Telford,* Peter W.H. Holland**

Molecular Zoology Laboratory, Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, United Kingdom

Received: 19 March 1996 / Accepted: 5 August 1996

Abstract. The chaetognaths are an extraordinarily homogeneous phylum of animals at the morphological level, with a bauplan that can be traced back to the Cambrian. Despite the attention of zoologists for over two centuries, there is little agreement on classification within the phylum. We have used a molecular biological approach to investigate the phylogeny of extant chaetognaths. A rapidly evolving expansion segment toward the 5′ end of 28S ribosomal DNA (rDNA) was amplified using the polymerase chain reaction (PCR), cloned, and sequenced from 26 chaetognath samples representing 18 species. An unusual finding was the presence of two distinct classes of 28S rDNA gene in chaetognaths; our analyses suggest these arose by a gene (or gene cluster) duplication in a common ancestor of extant chaetognaths. The two classes of chaetognath 28S rDNA have been subject to different rates of molecular evolution; we present evidence that both are expressed and functional. In phylogenetic reconstructions, the two classes of 28S rDNA yield trees that root each other; these clearly demonstrate that the Aphragmophora and Phragmophora are natural groups. Within the Aphragmophora, we find good support for the groupings denoted *Solidosagitta, Parasagitta,* and *Pseudosagitta.* The relationships between several well-supported groups within the Aphragmophora are uncertain; we suggest this reflects rapid, recent radiation during chaetognath evolution.

Key words: Chaetognaths — Duplicate genes — Molecular phylogeny

Introduction

Chaetognaths comprise a small phylum of vermiform marine invertebrates found as common predators within the plankton. They play an ecologically important role in the transfer of energy between trophic levels. Chaetognaths have been known to zoologists at least since the 18th century (Slabber 1778); recent molecular analyses suggest they evolved from a lineage that diverged very early in metazoan radiation (Telford and Holland 1993; Wada and Satoh 1994; Halanych 1996). A remarkable feature of the phylum is the almost invariant body plan to which all extant species adhere; indeed, this organization can be traced back 530 million years, to chaetognath fossils from Burgess Shale deposits (D.H. Collins, personal communication). Homogeneity of body form, however, leads to great difficulties in erecting a stable and evolutionarily meaningful classification. Chaetognaths display few variable traits, and much of what variation exists is quantitative rather than qualitative (variation in relative proportions rather than novel structures). Conse-

Abbreviations: FM, Fitch-Margoliash method; KITSCH, FM with contemporary tips method; ML, maximum likelihood method; MLK, ML with contemporary tips; MP, maximum parsimony method; NJ, neighbor-joining method; UPGMA, unweighted pair-group method with arithmetic mean; PCR, polymerase chain reaction; rDNA, ribosomal DNA

**Present address:* Molecular Biology Unit, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom

***Present address:* School of Animal and Microbial Sciences, The University of Reading, Whiteknights, Reading, RG6 6AJ, United Kingdom

Correspondence to: M.J. Telford; e-mail: mjt@nhm.ac.uk

quently, there is considerable disagreement between taxonomic schemes suggested by different workers.

At present, the consensus is to recognize just ten genera within the phylum (Bone et al. 1991); although Casanova (1986) further erects the genus *Archeterokrohnia* from within *Heterokrohnia,* while Kassatkina (1982) proposes a further five genera which are so far unverified by other workers (Bone et al. 1991). A second genus commonly perceived as containing subgroups is *Sagitta*; it contains approximately half the currently described species of chaetognaths. Some workers raise the genus *Sagitta* to the level of family (Sagittidae), raising groups within it to the level of genus (Bieri 1991; Tokioka 1965a,b). Even these workers do not agree on the number and composition of genera within the Sagittidae. In summary, it is difficult to disagree with Bieri's comment that ''the state of chaetognath systematics is close to chaotic'' (Bieri 1991).

Since morphological homogeneity is at the root of the problem, there is a need for investigation of chaetognath phylogeny based on nonmorphological characters. Nucleic acid sequence data have significant potential in such situations. The genes encoding 18S and 28S ribosomal RNA have been widely and successfully used in molecular phylogenetic studies at all taxonomic levels, since they are ubiquitous, homologous in all organisms, and have discrete sites that accumulate mutations at a range of evolutionary rates (Hillis and Dixon 1991). However, if phylogenetically informative data are to be obtained from DNA sequence comparisons, careful choice of gene, or region of gene, is vital. Regions that accumulate mutations very slowly (in relation to the time scale of evolution being investigated) will display few differences between taxa; sites that evolve too quickly will lose phylogenetic information through multiple substitutions and tend to be hard to align.

Unfortunately, we do not know the precise age of the evolutionary radiations of the chaetognaths; Bieri (1991) mentions three fossils related to modern species in the Mazon Creek fauna (Upper Pennsylvanian) and suggests that the radiation giving rise to the extant species might have been as long ago as the Precambrian. Comparison of genes is the most direct approach to determining the rough date of the radiation that produced the extant chaetognath groups. As a pilot study, therefore, we compared 550 nucleotides at the 5′ end of the 18S ribosomal DNA (rDNA) from three chaetognath species: *Sagitta elegans* (Telford and Holland 1993), *Sagitta setosa*, and *Spadella cephaloptera* (P.W.H.H. and N.A. Williams, unpublished data). Since the genera *Spadella* and *Sagitta* are considered distantly related chaetognaths, this comparison can be used to estimate the level of DNA sequence conservation across the phylum. Phylogenetic reconstructions grouped the three species extremely closely, relative to their divergence from other metazoa, and considerably closer than are *Xenopus* and human sequences (data not shown). This suggests that accurate resolution of chaetognath phylogeny requires a relatively rapidly diverging region of DNA. This result contradicts Bieri's estimate based on fossils.

Accordingly, one of the variable expansion segments (the D2 domain) of the 28S rDNA gene was selected for PCR amplification, cloning, and sequence determination from a diversity of chaetognath species. We show that chaetognaths possess two distinct classes of 28S rDNA; we compare their patterns of molecular evolution and describe insights into chaetognath phylogeny.

Materials and Methods

Samples and DNA Extraction. Eighteen chaetognath species were included in the present study, representing a broad spread across the currently recognized diversity of the phylum. For five species, specimens were collected and analyzed from more than one geographical site; in view of the uncertain taxonomic status of some species, such samples were treated separately. This gave a total of 26 independent samples for analysis. All specimens were collected and identified by experienced chaetognath specialists (members of ''The Chaetognath Group''); Table 1 summarizes the collection data. Animals were preserved in 5% SDS, 250 mM EDTA, 50 mM Tris.Cl pH 8, and transported at room temperature prior to DNA extraction as described by Holland (1993).

DNA Amplification, Cloning, and Sequencing. Two PCR primers, MT3 and MT4, were designed to amplify a variable region of the 28S rDNA gene, spanning the D2 expansion segment. The OLIGO 2.0 program was used to ensure that the primers could not form significant secondary structure or stable primer dimers and that they had similar melting temperatures. MT3 and MT4 primers, designed for the present study, have also been used for taxonomic and identification purposes in vestimentiferan tube worms (Williams et al. 1993).

MT3 (5′ primer) 5′ AAAGGATCCGATAGYSRACAAGTACCG 3′
MT4 (3′ primer) 5′ CCCAAGCTTGGTCCGTGTTTCAAGAC 3′
(Y = C or T; R = A or G; S = C or G)

The PCR cycling parameters were: 95°C, 3 min, then 35 cycles of 95°C for 1 min, 50°C for 1 min, 72°C for 2 min. Reactions were performed in a volume of 20 μl, essentially as described by Holland (1993). Products were separated by agarose gel electrophoresis; the amplified band of approximately 500 bp was purified, ligated into *Sma*I-cut pUC18 vector, and transformed into competent *E. coli*–strain DH5α.

Table 1 shows the species collected, their origin, the collector/identifier, and their Genbank accession number. Two or three clones were sequenced from most samples to take account of polymorphisms or errors introduced by Taq DNA polymerase; however, if initial clones differed at only 2% of sites or fewer, these were treated as ambiguities and coded appropriately in phylogenetic analyses. For three species, sequencing of initial clones revealed the presence of two distinct ''classes'' of 28S rDNA gene. Alignment of these sequences to those obtained from all other chaetognath species in the study revealed that every clone fell into one or other class. Of the 55 recombinant clones sequenced in total, 41 fell into one sequence class (denoted class I) and 14 into the other (class II).

Sequence Alignment and Phylogenetic Analyses. Initial sequence alignment was performed using CLUSTAL V (Higgins et al. 1992),

Table 1. Details of sequences obtained

Species/sequence	Collector	Region	Genbank accession
Class II			
Spadella cephaloptera/D	Dixon	Plymouth	Z77129
Eukrohnia fowleri/TH	Thuesen	E. Pacific	Z77123
Eukrohnia fowleri/Kb	Kapp	E. Atlantic	Z77125
Eukrohnia fowleri/Ka	Kapp	E. Atlantic	Z77124
Eukrohnia fowleri/T	Terazaki	W. Pacific	Z77126
Sagitta bipunctata/C	Casanova	Atlantic	Z77127
Sagitta macrocephala/T	Terazaki	W. Pacific	Z77128
Sagitta robusta/T	Terazaki	W. Pacific	Z77130
Sagitta serratodentata/C	Casanova	Atlantic	Z77104
Class I			
Eukrohnia fowleri/K	Kapp	E. Atlantic	Z77103
Eukrohnia hamata/K	Kapp	E. Atlantic	Z77105
Eukrohnia hamata/KU	Kurbjeweit	Antarctica	Z77106
Sagitta crassa/N	Nagasawa	Tokyo Bay	Z77107
Sagitta elegans/B	Bone	Plymouth	Z77108
Sagitta enflata/C	Casanova	Atlantic	Z77109
Sagitta enflata/T	Terazaki	W. Pacific	Z77110
Sagitta ferox/T	Terazaki	W. Pacific	Z77111
Sagitta gazellae/KU	Kurbjeweit	Antarctic	Z77112
Sagitta lyra/TH	Thuesen	E. Pacific	Z77114
Sagitta marri/KU	Kurbjeweit	Antarctic	Z77117
Sagitta maxima/KU	Kurbjeweit	Antarctic	Z77118
Sagitta serratodentata/C	Casanova	Atlantic	Z77119
Sagitta setosa/C	Casanova	Atlantic	Z77120
Sagitta setosa/O	Øresland	W. of Sweden	Z77121
Sagitta zetesios/TH	Thuesen	E. Pacific	Z77122
Sagitta hexaptera/T	Terazaki	W. Pacific	Z77113
Sagitta macrocephala/TH	Thuesen	E. Pacific	Z77115
Sagitta macrocephala/T	Terazaki	W. Pacific	Z77116

followed by manual adjustment taking into account secondary structure predictions, determined using the MFOLD program within the GCG package. Phylogenetic analyses were performed on five separate DNA sequence alignments, each subsequent alignment including progressively more closely related taxa and hence a greater number of confidently aligned sites. This strategy enables more extensive data sets to be used to root phylogenetic trees, without discarding sites that are phylogenetically informative at higher taxonomic levels (for example, Williams et al. 1993). Alignments and positions used in each analysis are available on request. Maximum parsimony (MP) analyses were performed using PAUP 3.1.1 (Swofford 1993). Exhaustive searches were used only on the smallest data set; otherwise the random addition option was used, with ten replicates. The robustness of each node was assessed by 500 bootstrap replications. The midpoint rooting option was used as only one of several guides to determine root positions of trees. Four distance matrix methods were also used: Fitch-Margoliash (FM), KITSCH (FM with a molecular clock assumption), neighbor-joining (NJ), and UPGMA. Each was implemented using PHYLIP 3.4 (Felsenstein 1989) from a distance matrix calculated using the ML distance correction. The global option was also used when using FM and KITSCH.

Results and Discussion

Paralogous 28S rRNA Genes in Chaetognaths

The PCR (Saiki et al. 1988) was used to amplify approximately 500 base pairs (bp) of 28S rDNA from 26 chaetognath samples representing 18 species. The region targeted for amplification was the D2 expansion segment; this is thought to encode a surface loop on the 28S rRNA molecule immediately 3′ to a conserved region that interacts with 5.8S rRNA. The D2 expansion segment has been found to show relatively high substitution rates in other taxa—for example, the vestimentiferan tube worms (Williams et al. 1993).

Alignment of the DNA sequences revealed an unexpected result. The chaetognath 28S rDNA sequences can be allocated to two very distinct classes. Comparing confidently aligned and hence homologous positions, there are on average eight differences between species within class I, 24 differences between species within class II, and 34 differences between species of different class. There is no correlation between chaetognath genus and our isolation of either class I or class II sequence; class I clones were isolated from 15 divergent species, and class II from five divergent species. Furthermore, in three species both classes were readily found (*Eukrohnia fowlerii, Sagitta macrocephala, Sagitta serratodentata*). This distribution, coupled with molecular phylogenetic analyses (see later), strongly suggests that both classes of 28S rDNA gene are present in the genomes of all extant chaetognaths.

By analogy to other metazoa, each class probably rep-

resents a tandem ribosomal gene cluster, homogenized through molecular drive (Dover 1982;1986). However, the rate of sequence homogenization through gene conversion between the class I and class II gene clusters might be close to zero, since they appear to have evolved separately during the evolutionary radiation of the chaetognaths.

There are few precedents for the presence of distinct rDNA sequence classes within a species though the twin 5S rDNA gene clusters found in *Xenopus* are well known (Fedoroff 1979). The *Xenopus* clusters are evolving separately and are expressed differentially, one solely in the oocyte and the other in somatic tissues in later embryonic and adult stages. Carranza et al. (1996) report two types of 18S rDNA in a platyhelminth.

Two Functional Gene Clusters or Ribosomal Pseudogenes?

A duplication of the 28S rDNA gene (or gene cluster) in an ancestor of extant chaetognaths could have given rise either to two functional 28S rDNA genes/clusters or to one functional gene/cluster and one pseudogene/pseudogene cluster. Sequence information can be used to resolve these alternatives, since pseudogenes will not be subject to the same intense selection pressures as functional genes. At protein-coding loci, pseudogenes may be recognized by the presence of stop codons, frame shifts, lack of third-position substitution bias, or changes in otherwise-invariant amino acid residues (e.g., Arctander 1995). These indicators cannot be used in non–protein-coding genes (such as 28S rDNA); however, selection does operate to maintain secondary structure of functional RNA molecules. Within stem-loop structures of ribosomal RNA molecules, a change in one base on a stem is often compensated by a change in its base-pair partner on the other side of the stem (although such compensation is not perfect; Hillis and Dixon 1991). A pseudo-rRNA gene would be expected to have elevated levels of uncompensated base changes within stem regions.

There is no published report of a conserved secondary structure for the D2 expansion segment of 28S rRNA, presumably due to its high rate of evolution and large divergence between phyla. We therefore performed secondary structure predictions for the alignable regions of class I and class II chaetognath D2 expansion segments. Three stem structures were identified in common between the two classes. The pattern of nucleotide substitution during chaetognath evolution was then examined across all stem regions.

In rRNA stem structures, base pairings occur between guanine and cytosine and between adenine and uracil (the RNA equivalent of thymine). In rRNA, however, pairings between guanine and uracil are also stable. If selection is maintaining secondary structure then the ratio of structure *maintaining* changes (such as G:C → G:U) to structure *destroying* changes (G:C → G:A or G:C → G:G) should be higher than that expected (based on analysis of the background rate of nucleotide substitutions). We performed this comparison as well as a comparison between the changes G:U → G:C with G:U → G:G or G:U → G:A. (A:U pairings were all but unchanging in this data set and so changes between A:U and G:U and vice versa were not considered).

Table 2. Chi-square tests[a]

Class	Change	Observed #	Expected #	χ^2	*P*
I	GC to GT vs NC	6:0	4.5:1.5	2	n.s.
	GT to GC vs NC	10:0	6.8:3.2	4.7	<0.05
	Combined	16:0	11.3:4.7	6.65	<0.01
II	GC to GT vs NC	9:0	6.7:2.3	3.1	<0.1
	GT to GC vs NC	6:0	4.1:1.9	2.8	<0.1
	Combined	15:0	10.8:4.2	5.8	<0.01

[a] Chi-square tests showing that class I and II genes have significantly higher ratio of structure-conserving (GC to GT or GT to GC) to non-conserving (NC) changes than would be expected in a gene that is not selected to maintain secondary structure. See text for detailed explanation

The ''expected'' ratios of structure-conserving:structure-destroying were calculated as follows (using the example of change from an initial G:C pair). The relative probabilities of change from C to T, C to G, and C to A are equivalent to those for G:C to G:T, G:C to G:G, and G:C to G:A; these may be estimated directly from the sequence data by summing the actual number of such changes on the most parsimonious tree. This provides a very conservative estimate as it is based on *all* positions, including secondary structure regions (identified or not). These latter are expected to have a bias toward structure-conserving base changes. The observed ratios of structure-conserving to structure-destroying changes *in the secondary structure regions* were then compared to the expected ratios, employing a chi-squared test to test for a significant bias toward structure-conserving changes. The results (Table 2) show that both genes (or gene clusters) face selection pressure to maintain stem:loop structures. We conclude that they are both expressed and functional.

Hypotheses to account for the existence of the 28S rDNA expansion segments suggest that there is little or no selective pressure acting upon them. Clarke et al. (1984) suggest that the expansion segments are the remnants of mobile elements that have been inserted into the transcribed regions of the ribosomal genes or that they are remnants of linkers that connected different functional domains that have subsequently been eliminated from all but the nuclear-encoded eukaryotic ribosomes (Clarke 1987). Our analyses, on the contrary, support the hypothesis that these regions are under selection and that the structure of expansion segments is important to the function of 28S rRNA. The same conclusion was reached

by Larson and Wilson (1989) on the basis of rates of sequence change in ''variable'' regions of salamander 28S rRNA.

It has been noted previously (Vawter and Brown 1993) that transition rates are elevated in secondary structure regions of ribosomal genes due to the fact that single substitutions that conserve secondary structure in a base pair are all transitions (see above). We point out that this will only be true of genes under selection; hence a transition bias (compared to noncoding, nonstem structure or nonribosomal regions) could potentially be used to identify expressed ribosomal genes without first identifying secondary structures. Such biases in transition rates in stem regions might also usefully be taken into account when reconstructing phylogenies from rRNA genes.

Class I and Class II 28S rDNA Genes Evolve at Different Rates

The above analyses suggest that both class I and class II 28S rDNA genes (or gene clusters) are expressed and functional. We might expect, therefore, that the genes comprise a mosaic of conserved, essential sites and variable, neutrally evolving sites. The rate of nucleotide change in DNA sequences is usually considered to be a function of DNA replication error rate and generation time (or germ cell generations per unit time). Since these factors are equal for a given genome, it is surprising that the class I and class II 28S rDNA genes of chaetognaths have experienced very different rates of molecular evolution. For example, the number of nucleotide differences between *Eukrohnia fowlerii* and *Sagitta macrocephala* over well-aligned, homologous regions is 11 differences for the class I 28S rDNA D2 segment but 23 differences for the class II D2 segment.

How can this curious observation be explained? The most likely explanation is that the two genes/clusters are experiencing different selection pressures; in one (class I) there is stronger stabilizing selection, which slows sequence change. Another possibility is that there is some effect of cluster size on the rate of substitution. In this context, it is intriguing that preliminary evidence (based on the relative numbers of clones obtained from each class) suggests that the faster-evolving class II sequences derive from a smaller rDNA cluster than the class I sequences.

Class I and Class II 28S rDNA Clones Derive from Chaetognaths

An alternative explanation for the very different rates of substitution in the two classes of 28S rDNA is that both are not, as assumed above, derived from the chaetognath genome. If one class was derived from the genome of an endoparasite of chaetognaths, different evolutionary rates could be explained by selection differences. (Great differences between animal taxa in rates of evolution of rRNA genes are well documented.) We consider this unlikely, since it demands the presence of a geographically widespread taxon of parasites, in which related (but never identical) parasites are present in related chaetognath species. The same argument is even stronger for food items because—to explain the congruence of the two phylogenies—we would have to believe that the chaetognaths have a single source of food with which they coevolve.

Convincing evidence against the parasite hypothesis also comes from sequence comparison with other eukaryotes. Diagnostic residues at the 5′ and 3′ regions of the amplified region confirm that both classes of 28S rDNA cloned are derived from metazoans. Sequence alignment reveals that the two classes are far closer to each other than to other metazoan 28S rDNA sequences previously reported. For example, an alignment between the two classes contains 209 confidently aligned sites, but introduction of the six closest sequences found in GenBank reduces this to only 106 confidently aligned sites. Phylogenetic analyses using the latter alignment clearly place the class I and class II sequences as close sister groups, divergent from known metazoans (data not shown). An early divergence from other metazoans was also found by analysis of chaetognath 18S rDNA (Telford and Holland 1993; Wada and Satoh 1994). These results, therefore, give strong support to the view that class I and class II 28S rDNA clones both derive from chaetognaths.

Further evidence in support of this conclusion includes an unusual feature of 28S rDNA from the chaetognath *Eukrohnia* (see below).

Eukrohnia: *Convergence or Conversion in 28S rDNA?*

The class I and class II 28S rDNA genes originated by a duplication event that predated the radiation of extant chaetognaths. During subsequent diversification of chaetognaths, the class I and class II sequences have evolved at different rates and have not been homogenized to uniformity within each species. This implies that intragenomic exchange of information does not occur between the classes in most chaetognath species.

Examination of 28S rDNA sequences from *Eukrohnia fowlerii* and *E. hamata* revealed a possible example of intragenomic exchange between class I and class II genes. The genus *Eukrohnia* yielded clones from both classes, each from several geographical sites. Alignment to all other chaetognath 28S rDNA sequences revealed a similar-sized insertion is present in an identical position in both class I and class II genes only in the *Eukrohnia* sequences (Fig. 1). There are three possible explanations for this observation. First, the ''insertion'' could be a primitive feature in 28S rDNA genes of chaetognaths,

```
Class II
Sp.cephaloptera.D   GGATTCAGCTCAGGGG----------------------------------------------------------------------------GCCA-CTCTTKKKTGCACTTCC
E.fowleri.TH        GGATTCAGCTCAGGGGATCTGTTCTGTTCTGGCCGGGCGGAGTCCAGGG-CGG-CTGCGCCAACTCTGGATCTCTG--CCCGGTCT---CTCGGACGGTTTCTTGGGTGCACTTCC
E.fowleri.Kb        GGATTCAGCTCAGGGGATCTGTTCTGTTCTGGCCGGGTGGAGTCCGGGGTCGGTCTCGGCCGACTCTGGATCTCTG--CCCGGTCT---CTCGGACGGTTTCTTGGGTACACTTCC
E.fowleri.Ka        GGATTCAGCTCAGGGGATCTGTTCTGTTCTGGCCGGGCAGAGTCCGGGGTTGGTCTCGGCCGACTCTGGATCTCTG--CCCGGTCT---CTCGGACGGTTTCTTGGGTGCACTTCC
E.fowleri.T         GGATTCAGCTCAGGGGATCTGTTCTGTTCTCCNCGGGCAGAGTCCGGGGTCGGTYTCGGCCGACTCTGGATCTCTG--CCCGGTCT---CTCGGACGGTTTCTTGGGTGCACTTYC
S.bipunctata.C      GGATTCAACCCGAGGG----------------------------------------------------------------------------TAGC-CCTTTGGGTGCACTTCC
S.macrocephala.T    GGATTCAGCNCGAGGG----------------------------------------------------------------------------TCGC-CTCTCGGGTGCACTTCC
S.robusta.T         GGATTCAGCTCGGGGG----------------------------------------------------------------------------TAAT-TCTCCGGGTGCACTTCC
S.serratodentata.C  GGATTCAGCCCGAGGG----------------------------------------------------------------------------TCGC-CCTTCGGGTCGACTTCC

Class I
E.fowleri.K         GGATTCAGCTCGCGGGG-CCGGGCTCGCCCTCG--GTCTGGGGTCCTGGCGTCTCG-----------GCGCCAGGTCCCCGGCCTGGGGCGGGTCCTCCTCGCGGGTTGCACTTCC
E.hamata.K          GGATTCAGCTCGCAGGG-TCGGGTGCGTCCGGG--GTTCTGGGGCCTGACTTGTCA-----------GGTCCCTGGCCTCGGTCT---CGCGCCTGCCTTCCGGG-TGCACTTCC
E.hamata.KU         GGATTCAGCTCGCAGGG-TCGGGTGCGTCCGGG--GTTCTGGNGCCTGACTTGTCA-----------GGTCCCTGGCCTCGGTCT---CGCGCCTGCCTTCCGGN-TGCACTTCC
S.crassa.N          GGATTCAGCTCCGGAGG----------------------------------------------------------------------------CAA---CTCCGGGGTGCACTTCC
S.elegans.B         GGATTCAGCTCCGGGGG----------------------------------------------------------------------------TAA---CCTCGGGGTGCACTTCC
S.enflata.C         GGATTCAGCTCCGGGGG----------------------------------------------------------------------------CAA---CCTCGGGNNGCACTTCC
S.enflata.T         GGATTCAGCTCCGGGGG----------------------------------------------------------------------------CAA---CCTCGGGGTGCACTTCC
S.ferox.T           GGATTCAGCTCCGGAGG----------------------------------------------------------------------------CAA---CTTCGGGGTGCACTTCC
S.gazellae.KU       GGATTCAGCTCCGGAGG----------------------------------------------------------------------------CRA---CTTCGGGGTGCACTTCC
S.lyra.TH           GGATTCAGCTCCGGAGG----------------------------------------------------------------------------CAA---CTTCGGGGTGCACTTCC
S.marri.KU          GGATTCAGCCCCGGGGG----------------------------------------------------------------------------CAA---CCTCGGGGTGCACTTCC
S.maxima.KU         GGATTCAGCTCCGGAGG----------------------------------------------------------------------------CAA---CTTCGGGGTGCACTTCC
S.serratodentata.C  GGATTCAGCTCGGNGAG----------------------------------------------------------------------------AAA---TCTCCGGGTGCACTTCC
S.setosa.C          GGATTCAGCTCCGAGGG----------------------------------------------------------------------------CAA---CTTCGGGGTGCACTTCC
S.setosa.O          GGATTCAGCTCCGAGGG----------------------------------------------------------*---------CAA---CTTCGGGGTGCACTTCC
S.zetesios.TH       GGATTCAGCCCCGGGGG----------------------------------------------------------------------------CAA---CCTCGGGGTGCACTTCC
S.hexaptera.T       GGATTCAGCTCCGGGGG----------------------------------------------------------------------------CAA---CCTCGGGGTGCACTTCC
S.macrocephala.TH   GGATTCAGCTCCGGGGG----------------------------------------------------------------------------CGANNCCCTCGGGGTGCACTTCC
S.macrocephala.T    GGATTCAGCTCCGGGGG----------------------------------------------------------------------------CGAAAGCCTCGGGGTGCACTTCC
```

Fig. 1. Region of alignment covering the insertion found in both classes of *Eukrohnia* sequence. The insertion in each class is similar sized although there is no convincing evidence for homology from sequence similarity despite attempts to align the inserts from the two classes.

predating the class I/class II duplication. This seems unlikely since it implies this region was precisely deleted on at least three (and probably more) independent occasions in chaetognath evolution (see later for phylogenetic trees). Second, the insertion could have arisen independently in the class I and class II 28S rDNA genes of *Eukrohnia*. This could be rationalized if the molecules were facing similar selection pressures, perhaps to interact with a divergent protein or RNA molecule specific to *Eukrohnia* cells. The third and most likely explanation is that the insertion occurred in only one of the gene classes and was then transferred to the other by intragenomic transmission (a gene conversion event). This would imply that either class I or class II 28S rDNA of *Eukrohnia* is chimeric.

High sequence similarity in this region between the two classes would support the latter hypothesis, but such similarity is not evident in the species studied here. Interestingly, whichever of the hypotheses is correct, the observation lends further strong support to the contention that both class I and class II genes derive from chaetognaths.

Phylogenetic Conclusions I: Phragmophora and Aphragmophora

The highest taxonomic division within the class Sagittoidae within the Chaetognatha is generally proposed to be into the orders Phragmophora and Aphragmophora (Tokioka 1965a,b, 1974). The former (including the genera *Spadella* and *Eukrohnia*) are chaetognaths possessing an internal sheet of transverse musculature or phragma not found in the Aphragmophora (the genus *Sagitta* in this study). To investigate if this high-level split is justified, it is necessary to determine a rooted phylogenetic tree for chaetognaths. However, the D2 domain of chaetognath 28S rDNA is extremely divergent from that of other metazoa. Since very distant outgroups can cause incorrect rooting of trees, we decided not to include non-chaetognath sequences in phylogenetic analyses. Two alternative approaches were therefore used to obtain rooted trees. As a first approach to estimate root position, we imposed a molecular clock assumption onto distance matrix methods for phylogeny reconstruction (KITSCH or UPGMA) or used the midpoint rooting option following MP analysis with PAUP (conceptually similar to assuming a clock). Each method was applied in turn to two separate DNA sequence alignments: one comprising the 19 class I 28S rDNA sequences and one comprising the nine class II sequences. The two classes were treated separately in these analyses to maximize the number of confidently aligned sites. For the larger class I alignment, the two distance methods placed a root between Phragmophora and Aphragmophora. (MP analyses on all class I sequences gave 36 equally most parsimonious trees and the root position varied.) The root was also placed between the Phragmophora and Aphragmophora when UPGMA and midpoint-rooted MP were applied to the smaller class II data set, although in this case KITSCH placed the root within Aphragmophora (separating *Sagitta robusta* from other chaetognaths).

A second approach to determine root position did not rely on a molecular clock assumption. The two classes of chaetognath 28S rDNA are far closer to each other than to any other metazoan sequences and arose by duplication prior to the radiation of extant chaetognaths. Each class, therefore, can be used as an outgroup to determine the position of the root within the other class. This procedure is formally identical to that used recently to determine a root for eukaryote, eubacterial, and archaebacterial divergence, based on duplications of the elongation factors TU and G or of amino-acyl tRNA synthetase genes (Creti et al. 1994; Brown and Doolittle 1995). For this approach, we used an alignment comprising all 28 chaetognath 28S rDNA sequences determined; this included 208 confidently aligned sites, of which 46 were informative. The most parsimonious tree determined from this alignment is shown in Fig. 2. This tree places the root between the Aphragmophora and Phragmophora for both classes of 28S rDNA gene. Applying distance matrix methods (FITCH, NJ) to the same data set produced an identical root position for the class I genes although the root was within the Aphragmophora for the smaller set of class II genes (separating *Sagitta robusta* from other chaetognaths). We believe this latter result to be an artefact caused by an unusually long branch leading to *S. robusta*, which causes it to be placed artificially close to the root of the tree.

In summary, the most consistent root position result produced by both strategies was between the Aphragmophora and the Phragmophora. This supports the division of chaetognaths into these two orders (Tokioka 1965a,b).

Phylogenetic Conclusions II: Radiation of the Sagittidae

The Aphragmophora comprise the larger of the two chaetognath orders, and most taxonomic debate has focused on relationships within this group. The genus *Sagitta* contains the majority of species and some authors raise *Sagitta* to family level (Sagittidae), within which multiple genera are recognized (Bieri 1991; Tokioka 1965a,b). Bieri's classification implies a close relationship between particular species; for example, between *S. elegans* and *S. setosa* (into *Parasagitta*) or *S. maxima, S. gazellae*, and *S. lyra* (into *Pseudosagitta*). Here, for the sake of clarity, we persist with the traditional nomenclature and indicate Bieri's generic names by quotation marks.

We investigated the phylogeny within the genus *Sagitta* using the 28S rDNA sequence data obtained in this study. These analyses were performed using the class I

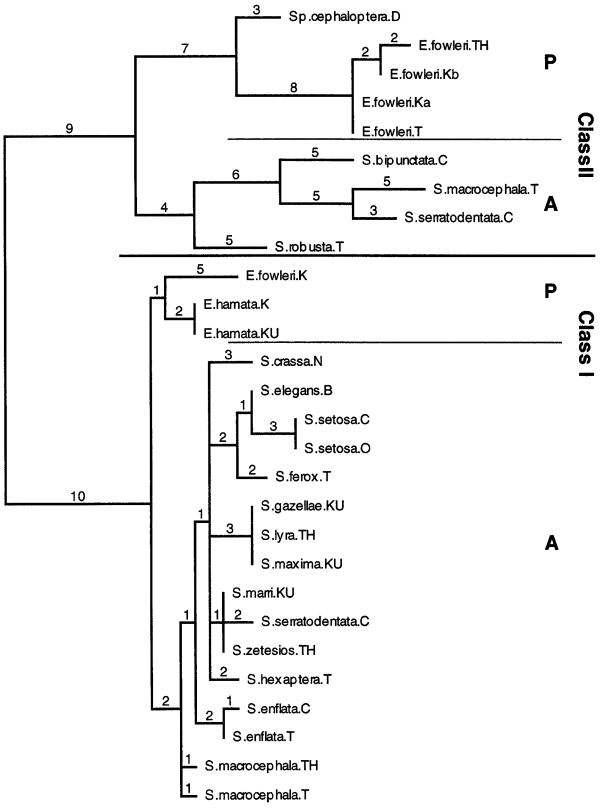

Fig. 2. A strict consensus of the 86 equally most parsimonious trees (109 steps C.I. = 0.642) for both classes of chaetognath D2 region; 208 bases were used in the alignment of which 46 were phylogenetically informative for parsimony. The trees were found using PAUP 3.1.1 with a heuristic search using ten replicates of the random addition option and TBR branch swapping. The deep division between class I and class II genes is shown. A = Aphragmophora and P = Phragmophora. The class II genes are evolving far more quickly than the class I genes. The class I sequences from *Sagitta* are seen to be very close, suggesting a rapid radiation. *Numbers on branches* indicate number of inferred changes. This tree confirms that Aphragmophora and Phragmophora are natural groups, but it should not be consulted for relationships within the class I *Sagitta* (see text). *Capital letters* after species name indicate collectors (see Table 1). *Ka* and *Kb* are two separate class II sequences derived from *E. fowleri* samples collected by H. Kapp.

sequences only, since these were obtained from a wider diversity of relevant species than were class II. Restricting analysis to one class of 28S rDNA also increased the number of confidently aligned sites for phylogenetic analysis (for this reason, Fig. 2 should not be referred to for phylogeny within the genus *Sagitta*). We first attempted to find the position of the root within *Sagitta*. The previous analyses revealed that Aphragmophora and Phragmophora are natural monophyletic groups; hence, the *Eukrohnia* class I sequences are ideal outgroups for determining root position within a *Sagitta* class I phylogeny. Analysis of a class I sequence alignment using six different phylogenetic reconstruction methods did not reveal a consistent or convincing root position within the genus *Sagitta*. MP and NJ placed this root between *S. serratodentata* and the rest; KITSCH and UPGMA divided *S. maxima, S. gazellae, S. lyra,* and the two *S. macrocephala* from the rest; FM and MP separated *S. hexaptera, S. crasa, S. ferox,* and the two *S. enflata* from the rest. We cannot present a strong argument in favor of one root position over the others; we suggest the uncertainty reflects a rapid evolutionary radiation after the origin of the genus *Sagitta*.

Although root position at the base of *Sagitta* cannot be found, insight into relationships between particular *Sagitta* species is still possible. We determined a sequence alignment for class I 28S rDNA clones from *Sagitta* species only (to maximize aligned sites) and produced unrooted phylogenetic trees using three reconstruction methods that do not assume a molecular clock (MP, NJ, FITCH). Figure 3 shows a strict consensus between the three trees. Several groupings of *Sagitta* species are consistently found by all reconstruction methods; most of these groupings were shown to be robust, as judged by bootstrap resampling.

A grouping of *S. maxima, S. gazellae,* and *S. lyra* is extremely well supported; strong support is also obtained for grouping *S. elegans* with *S. setosa.* These groups are

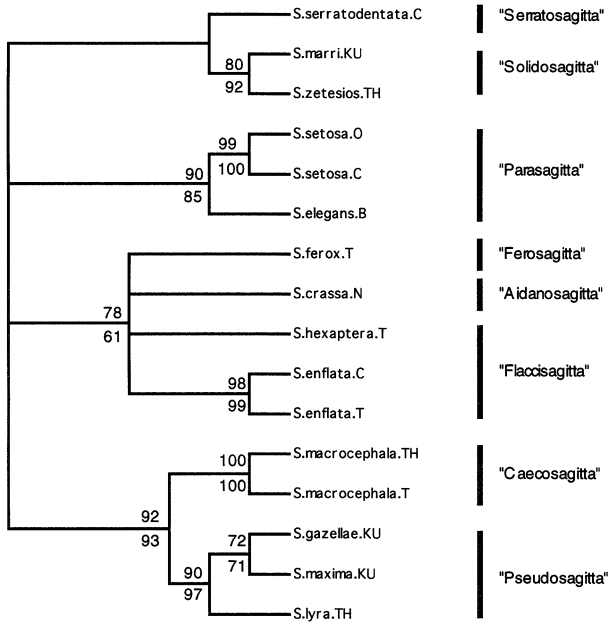

Fig. 3. Unrooted, strict consensus of NJ, FM, and MP analyses for the class I gene of the genus *Sagitta*. There were 369 aligned positions of which 46 were informative for parsimony. The distance matrix for FM and NJ was constructed using an ML distance correction (PHYLIP version 3.5) and the Jumble and Global Rearrangements options were used for FM. The MP analysis used a heuristic search with 10 random addition repetitions and gave two equally parsimonious trees (83 steps C.I. = 0.675) (PAUP version 3.1.1). Several groups within this clade are well supported although the relationships between these groups are uncertain as indicated by the polytomy. Bootstrap values are given for parsimony (*above line*) and NJ (*below line*). The *names on the right* are generic names for the indicted species based on Bieri's analyses of morphology. *Capital letters* after species name indicate collectors (see Table 1).

equivalent to the genera ''*Pseudosagitta*'' and ''*Parasagitta*,'' respectively, as proposed by Bieri (1991). Moderate support is obtained for a relationship between the ''*Flaccisagitta*,'' ''*Aidanosagitta*,'' and ''*Ferosagitta*'' of Beiri, and possibly the ''*Caecosagitta*'' with ''*Pseudosagitta*.'' The relationships between these well-supported groupings are uncertain in the consensus unrooted tree, again suggesting a rapid evolutionary radiation within the genus *Sagitta*.

Three Stages in Chaetognath Evolution

Using DNA sequence data to infer dates in the evolutionary history of organisms is fraught with difficulty, and to a large extent dependent on relative constancy of substitution rates during evolution. There has certainly been a significant change in rate of substitution in one class (if not both) of chaetognath 28S rDNA genes since the duplication event; for this reason we believe it pointless to speculate about dates for the duplication and for subsequent speciation events. We can, however, show that the evolutionary history of the chaetognaths involved three distinct stages. First, the lineage leading to chaetognaths separated from other phyla early in metazoan radiation, probably in the Precambrian. This is supported by the extreme divergence of both classes of chaetognath 28S rDNA from that of other metazoa, and by previous analyses of 18S rDNA (Telford and Holland 1993; Wada and Satoh 1994). In this lineage a 28S rDNA gene or possibly whole ribosomal gene cluster duplication occurred in the ancestor of all extant chaetognaths. Second, the extant lineages of chaetognath diverged from this common ancestor relatively recently. This divergence involved separation of the lineages leading to the extant Phragmophora and Aphragmophora. Evidence that this occurred much more recently than the origin of chaetognaths includes the similarity in 28S and 18S rDNA sequences from the genera *Sagitta*, *Spadella*, and *Eukrohnia*. Finally, subsequent to the separation of the Phragmophora and Aphragmophora, the genus *Sagitta* underwent a rapid evolutionary radiation, founding several subgroups which have subsequently speciated over a longer time period.

Acknowledgments. This work was initiated as part of an NERC CASE Studentship to M.J.T., held between the University of Oxford and the Marine Biological Association, Plymouth, U.K. We thank Dr. Quentin Bone, F.R.S., for his advice and encouragement, Nic Williams for help in the pilot study of 18S rDNA variation, and A.C. Pierrot-Bults and R. Thomas for helpful comments on the manuscript. We are grateful to many members of the International Chaetognath Study Group who kindly supplied specimens for DNA extraction; particular thanks are due to Drs. Jean-Paul Casanova, David Dixon, Taichiro Goto, Helga Kapp, Frank Kurbjeweit, Professor S. Nagasawa, Vidar Øresland, Professor M. Terazaki, and Eric Thuesen. The International Chaetognath Study Group can be contacted through H. Kapp at Zoologisches Institut u. Zool. Museum, Universität Hamburg, Martin-Luther-King-Platz 3, 2000 Hamburg 13, Germany.

References

Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proc R Soc Lond [Biol] 262:13–19

Bieri R (1991) Systematics of the Chaetognatha. In: Bone Q, Kapp H, Pierrot-Bults AC, (eds) The biology of chaetognaths. Oxford University Press, Oxford

Bone Q, Kapp H, Pierrot-Bults AC (1991) The biology of chaetognaths. Oxford University Press, Oxford

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Carranza S, Giribet G, Ribera C, Baguña J, Riutort M (1996) Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (*schmidtea*) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida). Mol Biol Evol 13:824–832

Casanova J-P (1986) Découverte en Mediteranée d′un chaetognathe nouveau du genre archaique profond *Archeterokrohnia:* description et signification biogéographique. Rapport et Procès-verbaux des Réunions de la commission internationale pour l′Exploration scientifique de la Mer mediteranée. 30:196

Clarke CG (1987) On the evolution of ribosomal RNA. J Mol Evol 25:343–350

Clarke CG, Tague BW, Ware VC, Gerbi SA (1984) *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. N.A.R. 12:6197–620

Creti R, Ceccarelli E, Bocchetta M, Sanangelantoni AM, Tiboni O, Palm P, Cammarano P (1994) Evolution of translational elongation factor (EF) sequences: reliability of global phylogenies inferred from EF-1α (Tu) and EF-2(G) proteins. Proc Natl Acad Sci USA 91:3255–3259

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Dover GA (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165

Fedoroff NV (1979) On spacers. Cell 16:697–710

Felsenstein J (1989) PHYLIP—phylogeny inference package. Cladistics 5:164–166

Halanych KM (1996) Testing hypotheses of chaetognath origins: long branches revealed by 18S ribosomal DNA. System Biol 45:223–246

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal V—improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Q Rev Biol 66 (4):411–453

Holland PWH (1993) Cloning genes using the polymerase chain reaction. In: Stern CD, Holland PWH (eds) Developmental biology; a practical approach. O.U.P., Oxford

Kassatkina AP (1982) Chaetognaths in the seas of the USSR and adjacent waters. Nauka Press, Leningrad

Larson A, Wilson AC (1989) Patterns of ribosomal RNA evolution in salamanders. Mol Biol Evol 6:131–154

Saiki RK, Grelfland DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491

Slabber M (1778) Natuurkundige Verlustigingen behelzende microskopise Waarneemingen van in-en uitlandse Water-en Land-dieren. In: Bosch J (ed) Haarlem

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign

Telford MJ, Holland PWH (1993) The phylogenetic affinities of the chaetognaths: a molecular analysis. Mol Biol Evol 10(3):660–676

Tokioka T (1965a) The taxonomical outline of the chaetognatha. Publ Seto Mar Biol Lab XII (5):335–357

Tokioka T (1965b) Supplementary notes on the systematics of chaetognatha. Publ Seto Mar Biol Lab XIII (3):231–242

Tokioka T (1974) On the specific validity of species pairs or trios of plankton animals distributed respectively in different but adjoining water masses as seen in chaetognaths. Publ Seto Mar Biol Lab 21:269–279

Vawter L, Brown WM (1993) Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134:597–608

Wada H, Satoh N (1994) Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. PNAS USA 91:1801–1804

Williams NA, Dixon DR, Southward EC, Holland PWH (1993) Molecular evolution and diversification of the vestimentiferan tube worms. J Mar Biol Assoc UK 73:437–452