



International Journal of
**Zoological
Research**

ISSN 1811-9778



Academic
Journals Inc.

www.academicjournals.com

Evolutionary History of the Chaetognaths Inferred from Actin and 18S-28S rRNA Paralogous Genes

¹R. Barthélémy, ¹F. Péténian, ²J. Vannier, ¹J.P. Casanova and ¹E. Faure
¹E.R. Biodiversity and Environment, Case 5, Pl. V. Hugo, Université de Provence,
13331, Marseilles Cedex 3, France
²Université Claude Bernard LYON 1,
UMR 5125 PEPS Paléoenvironnements et Paléobiosphère,
Campus Scientifique de la Doua, Bâtiment GEODE, 2 rue Raphaël Dubois,
69622 Villeurbanne, France

Abstract: The chaetognaths constitute a small and enigmatic phylum of marine invertebrates whose phylogenetic affinities remain uncertain. Our phylogenetical investigations inferred from partial paralogous 18S-28S rRNA genes suggest that the event resulting in the presence of two classes of rRNA genes would have occurred at approximately 300-400 million years and prior to the radiation of extant chaetognath, whereas the taxon, according to both molecular and paleontological data, would be dated from at least the Early Cambrian. These divergent rRNA genes could be the result of a whole ribosomal cluster duplication or of an allopolyploid event during a crisis period, since the fossil are lacking posteriorly to the post-Carboniferous period (*c.a.*, 300 million years). In addition, actin phylogeny evidenced that the cytoplasmic chaetognath actin clustered with the cytoplasmic insect actins, while the muscular chaetognath actins are placed basal to all muscular vertebrate actins. The present study suggests that the gene conversion mechanisms could be inefficient in this taxon; this could explain the conservation of extremely divergent paralogous sequences in the chaetognath genomes which could be correlated to the difficulties to identify a sister group between chaetognaths and other taxa among metazoans.

Key words: Chaetognaths, 18S rRNA, 28S rRNA, D2 region, actin, paralogous genes, molecular clock

Introduction

Chaetognaths are a small marine phylum composed of approximately 120 species. Most of them are planktonic and a few are benthic (Tokioka, 1965). The bauplan is built around a hydroskeleton, the body is divided into three parts: a small head armed with two sets of grasping spines (hooks) around the mouth; a trunk containing the gut and ovaries; a tail segment with testes (Ghirardelli, 1968; Casanova, 1999). In the last few decades, the position of the chaetognaths within the animal kingdom have been strongly debated because of embryological and morphological features shared with the two main branches of Bilateria, the deuterostomes and the protostomes. Indeed, for example, chaetognaths display some embryological characters considered atypical of deuterostomes (mouth not arising from the blastopore and mesoderm formed by enterocoely), whereas their morphology recalls the organization of protostomes (ventral nerve cord and chitinous structures) (Ghirardelli, 1968). In addition, most of the morphological and embryological informations are ambiguous or conflicting. This is partially due to the lack of recognisable synapomorphies with other phyla. Nucleic acid sequence

Corresponding Author: E. Faure, E.R. Biodiversity and Environment, Case 5, Pl. V. Hugo, Université de Provence, 13331, Marseilles cedex 3, France Tel: 33 491 10 61 77 Fax: 00 33 491 10 62 65

data have significant potential in such situations, however, classical phylogenetic molecular markers such as small subunit ribosomal RNA nuclear 18S (SSU rRNA) sequences or intermediate filaments did not help convincingly to define the chaetognath affinities (Erber *et al.*, 1998; Halanych, 1996; Mallatt and Winchell, 2002; Telford and Holland, 1993; Wada and Satoh, 1994). Finally, the analyses of the mitochondrial genomes of *Spadella cephaloptera* (Papillon *et al.*, 2004) and *Paraspadella gotoi* (Helfenbein *et al.*, 2004) supported close relationships with the protostomes, while a *Hox* gene survey suggested a basal position among the Bilateria (Papillon *et al.*, 2003). Moreover, a recent synopsis on the Chaetognatha (Ghirardelli and Gamulin, 2004) stressed again on the obscurity of their origin. In the future, the research projects undertaken independently by both Genoscope (France, http://www.cns.fr/externe/English/Projets/Projet_HT/HT.html) and a consortium of American laboratories (http://www.auburn.edu/academic/science_math/biology/faculty/halanych/chaetognath.html) which consist in sequencing collections of expressed sequence tags (ESTs) could produce a large number of new molecular markers. This might clarify the phylogenetic position of the chaetognaths. However, while awaiting these sequences, we hypothesize that research focusing on integrating molecular and paleontological data in analysis of the evolutionary history of the chaetognaths could reveal some interesting features, e.g., if all the paralog chaetognath genes known to date (rRNAs and actin) are investigated, this will constitute an unusual and fruitful approach.

Probably since the early evolution of life, gene duplications have played a determinant role in the genome evolution, provided material for the invention of new enzymatic or structural properties and more complex regulatory and developmental patterns (Taylor and Raes, 2004). Homology refers to two structures or sequences that evolved from a single ancestral structure or sequence and homology of sequences can be of two types: orthology or paralogy, which respectively differ in that one proceeds from speciation and the other from gene duplication. Divergent paralogs could, if undetected, misguide phylogenetic studies. The unsuccessful positioning of the chaetognath in the phylogenetical analyses could be due partially to the presence of paralogs. However, according to us, the use of paralogous genes could play a determinant role in the phylogenetical studies at least at two levels. On the one hand, if paralogs were already present in the last common ancestor, the phylogenies for each of the paralogs should be identical except for instance of horizontal gene transfer and additional later duplication events. The branch that connects the two sets of ancient paralogs indicates the placement of the last common ancestor. On the other hand, gene duplication times for paralogous genes could be predicted on the basis of a molecular clock based on orthologous gene comparisons in the same gene families. We hypothesize that the use of paralogs in a bilaterian phylogenetical investigation could permit to date the chaetognath rRNA genes divergence event. Moreover, this molecular dating analysis cannot be applied to the actin phylogeny because when two duplicate genes encoding proteins adapt to specialized functions, natural selection may favor amino acid changes that better adapt the proteins encoded to their specific functions (Goodman *et al.*, 1975). However, actin sequence analysis could give information concerning the rate of gene conversion and the arising of the two great types of chaetognath actins.

The 18S gene and the large-subunit rRNA gene (LSU or 28S), although less frequently used, are successful for molecular phylogenetic analysis, especially for evaluating deep-level relationships among organisms (Mallatt and Winchell, 2002). These two genes combined the advantages to be ubiquitous and homologous in all organisms. Unfortunately, use of rRNA sequences did not help convincingly to define the chaetognath affinities, due to the long-branch attraction artefact according various authors (Halanych, 1996; Mallatt and Winchell, 2002; Telford and Holland, 1993; Wada and Satoh, 1994). Moreover, the rDNA sequences represented in fully processed rRNA are essentially identical in most organisms, i.e., rRNA genes are subject to concerted evolution (Gonzalez and Sylvester, 2001). As a consequence, one observes that paralogous sequences in the same species are more similar than orthologous sequences of different species. Multiple molecular mechanisms may

account for this phenomenon: gene conversion, repeated unequal crossover and gene amplification (Liao, 1999). There are, however, exceptions to this rule: two classes of ancient paralogs of 18S-28S rRNA have been reported in chaetognaths (Telford and Holland, 1997; Papillon *et al.*, 2006). Similarly, paralog 18S rRNAs are known, e.g., in cephalopods (Bonnaud *et al.*, 2003), in flatworms (Carranza *et al.*, 1999) and in apicomplexans (Rooney, 2004).

Moreover, molecular analysis based on actin sequences could gain another perspective on the chaetognath evolutionary history. Actins are ubiquitous and highly conserved proteins that play key roles in several basic functions of the organism, such as cytoskeletal structure, cell division, cell motility and muscle contraction (Hooper and Thuma, 2005). Actins in most organisms can be divided into two groups: cytoplasmic and muscle actins. In mammals, for instance, at least six actin isoforms have been identified: two cytoplasmic actins (β and γ) and four isoforms in muscular cells, i.e., two specific for striated muscles (α -skeletal and α -cardiac) and two specific for smooth muscles (α -aortic and γ -enteric) (Vandekerckhove and Weber, 1981). Concerning Chaetognatha, three different actin genes have been found in the benthic species *Paraspadella gotoi* (Yasuda *et al.*, 1997); PgAct1 gene encode a cytoplasmic actin and PgAct2 and PgAct3 are expressed in muscle. Numerous works have investigated phylogenetic relationships using actin genes or proteins (Hooper and Thuma, 2005). In addition, many of the conclusions drawn from various studies about the evolutionary rate of actin are pertinent when using the amino acid sequences, whereas due to saturation and pronounced codon bias, the nucleotide sequences are unuseful (Li, 1997).

The aim of this study is thus to try to understand the major steps of the chaetognath evolutionary history, combining paralog genes analyses and paleontological data.

Materials and Methods

This study was undertaken owing to the following considerations: 1 mention of two classes of 18S-28S rRNA (Papillon *et al.*, 2006), that allowed now to dispose of three groups of paralogous genes (18S rRNA, 28S rRNA and actin); 2 paleontological data by two of us (J.V. and J.P.C.). Indeed, combination of these two different approaches gave opportunities to undertake the evolutionary history of the chaetognaths under a new angle.

Complete or nearly complete sequences have been used for 18S analyses, whereas for 28S rRNA investigations, only D2 variable regions have been used. Firstly, both primary and secondary structures of metazoan 18S and 28S have been obtained from the European ribosomal RNA database (<http://www.psb.ugent.be/rRNA/>) housed in Gent (Belgium) (Wuyts *et al.*, 2004). In this database, all sequences have been extracted in the FASTA format which preserves indications of secondary structure elements, based on the adopted rRNA secondary structure, which in turn is corroborated by the observation of compensating substitutions in the alignment. Additional sequences have been extracted from GenBank database. Sequence alignments are based on the secondary structure of the molecules, as determined by comparative sequence analysis. The final alignments were optimized manually using a multiple sequence alignment editor BioEdit version 7 (Hall, 1999; <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Table 1 lists the species used, their accession codes and the taxa they represent. In a previous analysis, molecular investigations with only chaetognath sequences have been performed in order to find chaetognath sequences which have the shorter branch lengths and to select one sequence by order (Aphragmophora and Phragmophora). In addition, due to possible intragenomic exchange between two classes of paralogous rRNA (Telford and Holland, 1997), for both 18S and 28S genes, the sequences chosen were originated from different genus, or when not possible for different species. For the other taxa, when more than one 18S or 28S sequences were available for a taxon, the representative with the slower evolutionary rate was used. For example, as already suggested by Aguinaldo *et al.* (1997), *Drosophila melanogaster* sequences have been excluded

Table 1: List of the 18S and 28S rRNA gene sequences analyzed in this study

Taxa	Species	Accession number				
		18S	28S			
Deuterostomia						
Chordata						
Craniata	Hyperotreti	Myxiniiformes	<i>Myxine glutinosa</i>	M97574		
	Vertebrata	Gnathostomata	Actinopterygii (ray-finned fishes)	<i>Epiplatretus stouti</i>		AF061797
		Sarcopterygii	Coelacanthimorpha	<i>Salmo trutta</i>	X98839	
			Tetrapoda	<i>Danio rerio</i>		BX537263
			Aves	<i>Latimeria chalumnae</i>	L11288	U34336
			Mammalia	<i>Gallus gallus</i>	AF173612	DQ018756
				<i>Homo sapiens</i>	K03432	U13369
				<i>Mus musculus</i>	X00686	BK000964
				<i>Rattus norvegicus</i>	K01593	V01270
				<i>Xenopus laevis</i>	X02995	X59734
		Hyperoartia	Petromyzontiformes	<i>Lampetra aepyptera</i>	M97573	
				<i>Petromyzon marinus</i>		AF061798
	Cephalochordata			<i>Branchiostoma floridae</i>	M97571	AF061796
	Urochordata			<i>Herdmania curvata</i>	AF165827	
				<i>Herdmania momus</i>		X53538
Echinodermata						
	Eleutherozoa	Asterozoa		<i>Ophiopholis aculeata</i>	L28055	
		Echinozoa		<i>Asterias forbesii</i>		AF212169
				<i>Echirus esculentus</i>	Z37125	AY026367
				<i>Arbacia punctulata</i>		
	Pelmatozoa	Crinoidea		<i>Amphipholis squamata</i>	X97156	AF212168
				<i>Florometra serratifima</i>		
Protostomia						
Lophotrochozoa						
	Mollusca			<i>Aplysia punctata</i>	AJ224919	
				<i>Aplysia californica</i>		AY026366
	Annelida	Clitellata		<i>Lumbricus rubellus</i>	Z83753	AY364866
				<i>Hirudo medicinalis</i>		
		Polychaeta		<i>Nereis virens</i>	Z83754	AY732225
				<i>Schizobranhia insignis</i>		
				<i>Siro rubens</i>	U36998	
				<i>Pandirus imperator</i>		AY210830
				<i>Limulus polyphemus</i>	L81949	AF212167
				<i>Scolopendra cirrigulata</i>	U29493	
				<i>Proterouidus fuscus</i>	AF173236	
				<i>Cormocephalus hartmeyeri</i>		AY210812
				<i>Limnadia lenticularis</i>	L81934	
				<i>Triops</i> sp.		AY210844
				<i>Pugettia quadridentis</i>	Z22518	
				<i>Lysmata wurdemanni</i>		DQ079794
				<i>Hypogastrura dolsana</i>	Z26765	
				<i>Poecura aquatica</i>		AY210838
				<i>Glauia terreia</i>	Z97568	
				<i>Apis mellifera</i>		AY703551
				<i>Sagitta elegans (type I)</i>	Z19551	Z77108
				<i>Krohmita pacifica (type II)</i>	DQ351891	
				<i>Sagitta bipunctata (type II)</i>		Z77127
				<i>Paraspadella gotoi (type I)</i>	D14362	
				<i>Eukrohnia hamata (type I)</i>		Z77105
				<i>Eukrohnia fowleri (type II)</i>	DQ351889	
				<i>Spadella cephaloptera (type II)</i>		Z77129
				<i>Stachyromyces cerevisiae</i>	Z75578	U53879
Fungi						

Table 2: List of the actin gene sequences analyzed in this study

Taxa	Species	Accession numbers
Deuterostomia Vertebrata Chordata	Mammalia	<i>Bos taurus</i>
		<i>Canis familiaris</i>
		<i>Homo sapiens</i>
		<i>Macaca mulatta</i>
		<i>Monodelphis domestica</i>
		<i>Mus musculus</i>
		<i>Pan troglodytes</i>
		<i>Rattus norvegicus</i>
	Aves	<i>Gallus gallus</i>
	Actinopterygii (ray-finned fished)	<i>Danio rerio</i>
		<i>Takifugu rubripes</i>
		<i>Tetraodon nigroviridis</i>
	Amphibia	<i>Xenopus tropicalis</i>
	Urochordata	<i>Ciona intestinalis</i>
	Protostomia	Insecta
		<i>Drosophila melanogaster</i>
		<i>Apis mellifera</i>
Nematoda		<i>Caenorhabditis elegans</i>
Chaetognatha		<i>Paraspadella gotoi</i>
Fungi		<i>Saccharomyces cerevisiae</i>

as insect representative. This criterion was used to reduce homoplasy and erroneous results due to large differences in divergence values (long branches attraction). Furthermore, in practice, the use of more slowly evolving representatives allows a larger portion of 18S rDNA molecule to be unambiguously aligned. In addition, only sequences from taxa which are not difficult to align have been used, thus excluding alignments considered as unreliable. In preliminary studies, phylogeny reconstructions were performed using a phenetic approach based on a matrix of pairwise nucleotide differences: the neighbor-joining (NJ) algorithm (Saitou and Nei, 1987) as implemented in Treecon (Van de Peer and De Wachter, 1997) as well as with a cladistic approach following the maximum parsimony (MP) criterion and using the heuristic algorithm of PAUP* 4.0 (Swofford, 2001). In order to date duplication events, Maximum Likelihood (ML) from the rRNA alignments were calculated using the DNAMLK program from PHYLIP package (Felsenstein, 2005) (<http://evolution.genetics.washington.edu/phylip.html>). DNAMLK applies ML assuming trees must be consistent with a molecular clock, i.e., the leaves are all equidistant from the root. As DNAMLK, for which no bootstrap values could be calculated, generated same main groups than Neighbor-Joining/UPGMA method version 3.6a3 (PHYLIP package), the robustness of the nodes evaluated by 100 bootstrap replicates using this last program have been reported on the DNAMLK phylograms.

For the analyses using actin sequences, all the predicted actin paralogs of whole metazoan genomes have been extracted from the ENSEMBL genome database (<http://www.ensembl.org>). The complete listing of species and taxa represented is found in Table 2. The sequences were aligned to the three chaetognath actin sequences (Yasuda *et al.*, 1997) using the multiple alignment program ClustalW (Higgins *et al.*, 1992). In preliminary studies, various methods contained within the PHYLIP package

(Felsenstein, 2005) have been used; for the final phylogenetic tree, distances were calculated using PROTDIST which calculates a distance matrix based on amino acid substitution and the results were used for tree construction using the neighbor-joining method by NEIGHBOR.

Results

rRNA Analyses

The alignment of thirty 18S sequences is based on secondary structure of complete or nearly complete sequences and displays 653 variable sites with 474 informative for parsimony. The study of the composition shows no particular deviation in the base frequency equilibrium, that is near of 23% for each type of nucleotide in the whole data set which is homogeneous (chi square test with $\alpha = 0.05$). Plotting of distance matrices based on "Transitions" versus "Transversions" substitutions displays no sign of saturation in the data set. Tree reconstructed using DNAMLK, where the ML method is employed under the constraint of a molecular clock, does not reveal aberrant groupings (Fig. 1A); however, Neighbor Joining (NJ) analysis which confirms this topology (data not shown) evidenced that most of the groupings are not supported by high bootstrap values. Our analysis suggests that Chaetognatha may be a sister-group to the rest of Bilateria. The other different groups of this analysis are presented briefly here. As expected, if excluded chaetognaths, the Protostomia and Deuterostomia are both monophyletic even if it is not statistically supported. Concerning the Deuterostomia, both Echinodermata and Eleutherozoa constitute monophyletic groups which have relatively high bootstrap values. The Echinodermata are the sister group of the Chordata. Chordate monophyly has a non-significant statistical support. In addition, within the Chordata, our analysis suggests that the tunicates and not the cephalochordates are the closest living relatives of vertebrates, agreeing thus with recent molecular investigations (Delsuc *et al.*, 2006). In addition, our analysis strongly supports both the monophyly of Tetrapoda and Amniota. Within Protostomia, two clades have been evidenced, Lophotrochozoa and Ecdysozoa (Halanych *et al.*, 1995; Aguinaldo *et al.*, 1997), but they are supported by weak bootstrap values. Similarly, if the tree suggests the monophylies of respectively Annelida, Chelicerata, Myriapoda, Crustacea and Hexapoda, all these groupings are not supported statistically. Literature data (Bromham and Hendy, 2000; Wang and Gu, 2000; Nei and Glazko, 2001; Aris-Brosou and Yang, 2002; Engel and Grimaldi, 2004; Peterson *et al.*, 2004; Pisani *et al.*, 2004; Blair and Hedges, 2005a, b; Glazko *et al.*, 2005; Hedges *et al.*, 2004; Waloszed *et al.*, 2005) allowed to date most of the divergence events among Bilateria. According to these data, estimation of chaetognath 18S rRNA gene duplication times assuming a molecular clock suggests that this molecular event would have occurred after the Amphibia/Amniota split which is estimated at 360 millions years ago (MYA) and before the Mammalia/Aves divergence time (310-326 MYA) (Table 3).

Concerning the 28S rRNA analyses, only a portion of approximately 500 bp have been analyzed, corresponding to the D2 expansion segment which is thought to encode a surface loop on the 28S rRNA molecule immediately 3' to a conserved region that interacts with 5.8 rRNA. The D2 expansion segment has been found to exhibit relatively high substitution rates in numerous taxa (Gillespie *et al.*, 2005). However, only this region has been sequenced in the study which has revealed the presence of two distinct classes of 28S rRNA gene in chaetognaths (Telford and Holland, 1997) and in spite of the fact that this region could not be appropriate to investigate the relationships between Bilateria, we have nevertheless undertaken a phylogenetic analysis. The alignment of thirty 28S D2 sequences based on secondary structure displays 317 variable sites with 131 informative for parsimony. Plotting of distance matrices based on "Transitions" versus "Transversions" substitutions displays no sign of saturation in the data set. All the phylogenetical methods reveal several aberrant groupings at different levels of the tree. This is probably due to the shorter length of the region studied, associated to a higher evolution rate. In the DNAMLK analysis, only the groupings of Chaetognatha,

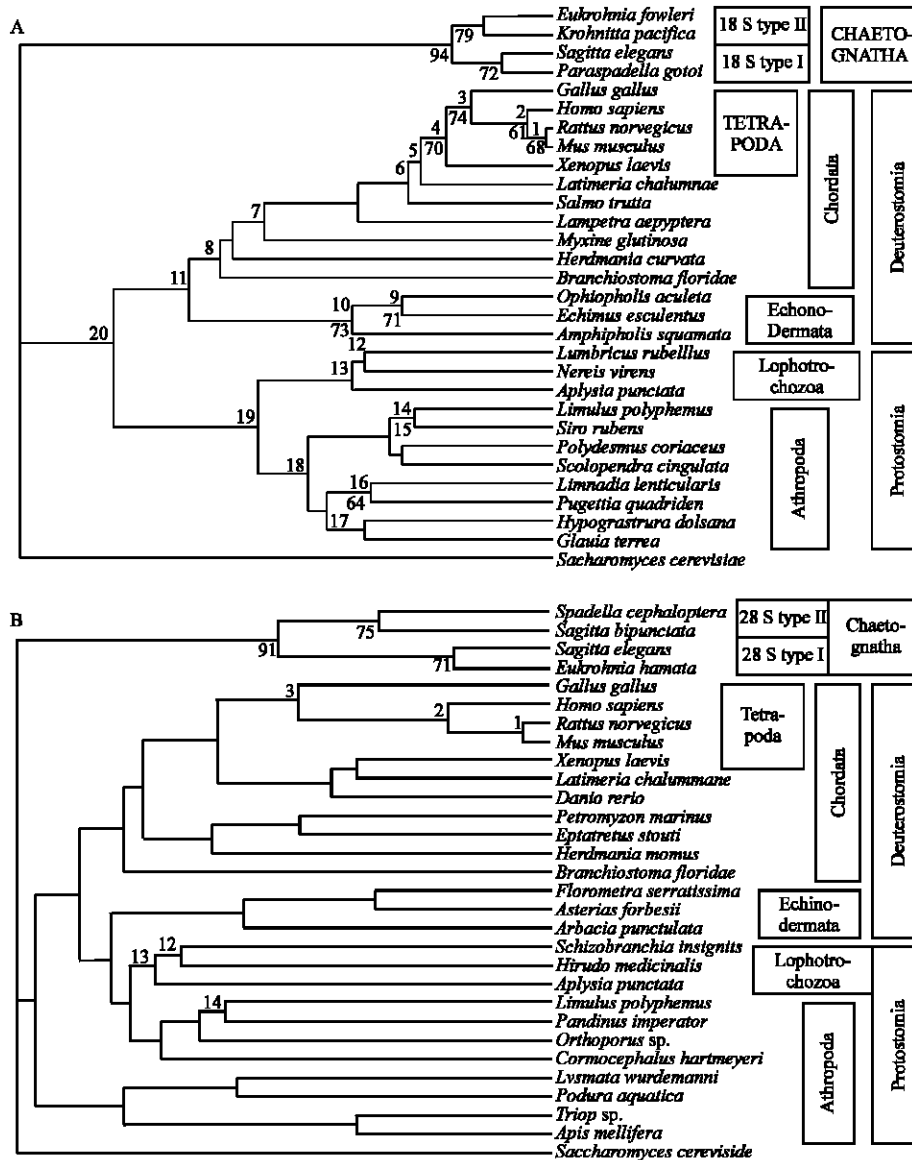


Fig. 1: Phylogenetic trees generated by using the DNAML program (DNA maximum likelihood with molecular clock) and based on 18S (A) and 28S (B) partial sequences respectively. Bootstrap values obtained using NJ analyses are given in italic below each node only if they exceed 60. Numbers above some nodes refer to estimated divergence times (Table 3). In the 28S rRNA genes tree, bold lines represent non aberrant phylogenetical positions

Amniota (and sub-taxa), Lophotrochozoa, Annelida and Chelicerata are in accordance with the traditional vision of the bilaterian taxonomy and only the Chaetognatha clade and the two chaetognath classes of 28S are statistically supported (Fig. 1B). If only the non aberrant groups are used to date the 28S duplication, this suggests that this event would have occurred after these three divergence times: Mollusca/Annelida (527-535 MYA), Clitellata/Polychaeta (505-508 MYA) and Arachnida/Merostomata (480-485 MYA) and before the Aves/Mammalia split (310-326 MYA).

Table 3: Fossil record ages and divergence time estimates (in Million years ago) for Deuterostomia and Protostomia taxa. The literature references from a to I (letters in superscript) are respectively: (Blair and Hedges, 2005a, b; Glazko *et al.*, 2005; Peterson *et al.*, 2004; Engel and Grimaldi, 2004; Waloszed *et al.*, 2005; Nei and Glazko, 2001; Pisani *et al.*, 2004; Wang and Gu, 2000; Hedges *et al.*, 2004; Aris-Brosou and Yang, 2002; Bromham and Hendy, 2000)

	Fossil	Means of the molecular estimates of divergence time		Corresponding number in the figures 1A and 1B
		Using various genes	Using 18S rRNA	
Mus/Rattus	12 ^a			1
Primates/Rodentia	65 ^a	107 ^g		2
Mammalia/Aves	310 ^b	326 ^a		3
Amniota/Amphibia	370 ^c	370 ^a		4
Tetrapoda/Coelacanthimorpha	390 ^a	430 ^a		5
Tetrapoda/Actinopterygii	425 ^c	476 ^a		6
Gnathostomata/Agnatha	520 ^a	652 ^a		7
Vertebrata/Cephalochordata	520 ^a	891 ^a		8
Asterozoa/Echinozoa	485 ^a	581 ^a		9
Eleutherozoa/Crinoidea	520 ^a	730 ^a		10
Chordata/Echinodermata	530 ^c	896 ^a	550 ^a -586 ^d	11
Clitellata/Polychaeta	505 ^d		508 ^a	12
Annelida-Mollusca	535 ^d		527 ^a	13
Arachnida-Merostomata	485 ^c	480 ^b		14
Chilopoda/Diplopoda		446 ^b		15
Branchiopoda/Malacostraca		618 ^b		16
Collembola/Insecta (true insects)	400 ^e		510 ^a	17
Chelicerata-Myriapoda/pancrustacea	543 ^e		522 ^a	18
Annelida-Mollusca/Arthropoda	543 ^e			19
Protostomia/Deuterostomia	542 ^f	833 ^e -993 ⁱ -976 ^j	560 ^a -588 ^d	20

These letters correspond in superscript are the literature references from a to I, respectively

Phylogenetical Analysis of Actin Paralogs

Phylogenetic analysis of amino acid sequences of three chaetognath actin isoforms, along with known actin isoform amino acid sequences from complete animal genomes, was conducted to examine the relationship between the three chaetognath actin paralogs and other metazoan actin paralogs. The alignment of sequences based on 379 amino-acids displays 151 variable positions with 73 informative for parsimony. The study of the composition in amino-acids shows no particular deviation in the data set which is quite homogeneous. Furthermore, plotting of distance matrices based on “Conservative differences” versus “No conservative differences” displays no sign of saturation in the data set. In addition, in the MP analysis, the tree-length distribution skewness (Huelsenbeck, 1991) $g1 = -0.872$ indicates a strong phylogenetic signal in the data set; however, the value of the consistency index (CI = 0.627) is indicative of some level of homoplasy. Based on 106 actin amino acid deduced sequences, molecular trees were constructed using NJ, ML and parcimony methods; all these methods produced similar topologies. For the tree which is shown in the Fig. 2, the program PROTDIST has been used to compute a distance matrix from the alignment and then the program NEIGHBOR has been used to calculate a Neighbor-joining tree based on this matrix. The tree evidenced that the number of actin isoforms varies in different lineages (Hooper and Thnma, 2005); generally, mammals possess six different isoforms, until nine different isoforms have been characterized in teleost fishes and diptera insects have been shown to have at least six actin genes.

Generally, in the actin molecular analyses, the gene types appear to be more conserved between species than within species, indicating that the divergence of these genes predates the divergence of the species; this is confirmed by our phylogenetic analyses for mammals and insects. The topology of the phylogenetic tree evidenced that animal actins are grouped in two distinct classes [(muscular actins of vertebrates)/(actins of invertebrates + cytoplasmic actins of vertebrates)], themselves subdivided in numerous groups. However, if this subdivision in groups is relatively easy for some taxa like mammals, it is more difficult, even impossible, for teleosteans due to the great number of actin isoforms, or for

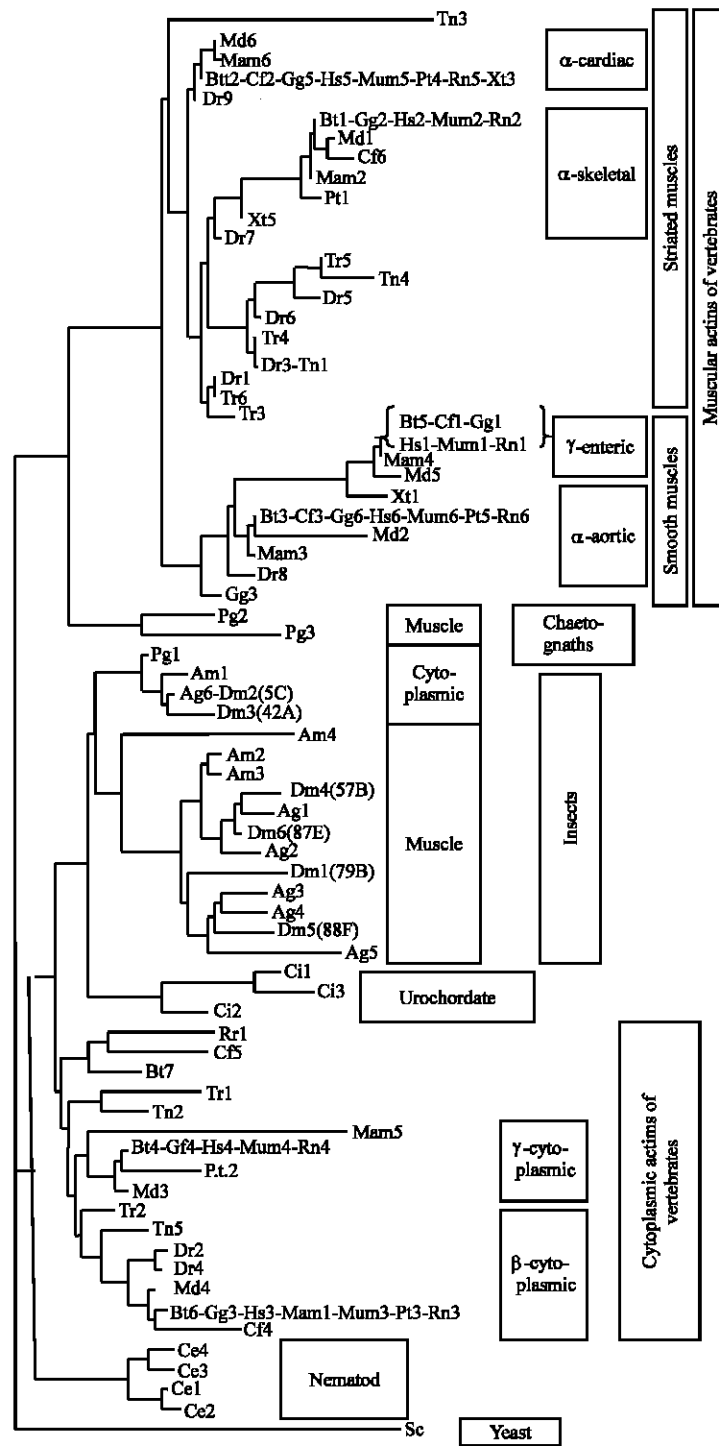


Fig. 2: Phylogenetic tree analysis of actin amino acid sequences

insects due to the lack of some orthologs principally in *Apis mellifera*. In addition, due to the relatively great number of sequences and of the relatively low number of informative sites, the bootstrap values are insignificant and thus not shown.

The first class contains all the muscular vertebrate actins which are divided in two subclasses themselves subdivided each in two groups of isoforms of muscular cells; two groups for striated muscles (α -skeletal and α -cardiac) and two groups for smooth muscles [vascular (α -aortic) and non vascular isoforms (γ -enteric)]. Except for *Pan troglodytes*, all the mammal species have an isoform in each of these four groups; on the other hand, the number of isoforms for *Xenopus tropicalis* and *Gallus gallus* is lower, while teleostean genomes contain a greater number of isoforms in this class. The two muscular chaetognath actins are sister group of the muscular vertebrate; however, due to the insignificant bootstrap value of this grouping, the chaetognath actins could also have a basal position in the animal actin tree. Interestingly, the position of muscle chaetognath sequences is not the result from the "long-branches attraction" artefact; indeed, when sequences with relatively long branches have been discarded from the dataset, the tree topology remains globally unchanged (data not shown).

The second class of actin could be divided itself into two subclasses. All the nematode actins group together and form a basal clade of this class. One of the subclasses contains all the vertebrate cytoplasmic actins and is subdivided in two groups containing respectively the β or the γ vertebrate cytoplasmic actins. However, some mammal and teleostean cytoplasmic actins have shifted to a basal position compared to these two groups of actins. The other subclass contains all the ascidian (urochordate) actins which form a sister clade of a group containing the chaetognath cytoplasmic actin and all the insect actins. The chaetognath cytoplasmic actin differs from some of the arthropod actins only by one amino acid (data not shown), suggesting a convergent evolution. Finally, the tree suggests that the six *D. melanogaster* isoforms have not their orthologous counterparts in *Anopheles gambiae* or *A. mellifera*; in addition, the number of isoforms is lower in Hymenoptera than in Diptera.

Discussion

As already evidenced by various authors, 18S phylogeny do not allow to resolve the enigma of the chaetognath origin (Telford and Holland, 1993; Wada and Satoh, 1994; Halanych, 1996; Peterson and Eernisse, 2001; Mallat and Winchell, 2002). However, we have hypothesized that rRNA duplication events could be timed reliably, assuming a molecular clock even if the molecular clock hypothesis has been criticized in the past, particularly because of the considerable error potentially associated with clock calibration, the large gap between extrapolation time and fossil records and potential heterogeneity in the rates of evolution of different species (Kumar, 2005).

Concerning rRNA genes, it may be reliable to use a molecular clock based on orthologues in timing the duplication of paralogues. Indeed, contrary to the genes encoding proteins, where generally natural selection favors amino acid changes that better adapt the proteins encoded to their specific functions, after duplication, rRNA genes preserve the same function. The 18S phylogenetical tree evidenced that this event would have occurred after numerous splits dated approximately at 400 MYA [within Deuterostomia: Tetrapoda/Coelacanthimorpha and Amphibia/Amniota and within Protostomia: Collembola/Insecta (true insects)]. Due to the difficulties to found 18S sequences which gave unambiguous secondary structure prediction alignment and for which fossil records are well known, it is less easy to found divergence event weakly posterior to the chaetognath rRNA duplication, only the divergence time between diapsid-synapsid (Aves/Mammalia), for which fossil evidence dates the split at approximately 310 MYA (Glazko *et al.*, 2005). Obviously, the length and the dates of the time window deduced from 18S analyses are criticisable; however, in spite of the fact that the length of the sequences are very short and exhibit significantly higher evolutionary rate, 28S investigations suggest a duplication date at 310-480 MYA. In addition, the similarity between the two time windows is in

favor of the hypothesis that the duplication had concerned the whole rRNA cluster. Interestingly, the wide distribution of both SSU and LSU rRNA classes across Chaetognatha and previous phylogenetic analyses (Telford and Holland, 1997; Papillon *et al.*, 2006), also corroborate an ancestral duplication of the whole ribosomal gene cluster, prior to the radiation of extant chaetognaths

Although gene redundancy could be explained by gene duplications, we proposed an alternative hypothesis to explain the presence of rRNA paralogs since several hundred million years without gene conversion. Indeed, allopolyploidy (genome combination after species hybridization) could provide a better explanation for all the gathered evidence. Present investigations suggest that the event leading to the presence of two classes of both 18 and 28S genes would have occurred at approximately the same period (of more than 300 million years). Interestingly, whereas the chaetognath fossils are relatively abundant since the Cambrian to the Carboniferous, the post-Carboniferous period (*c.a.* 300 million years) lacks any fossil record, suggesting that the paralog genes would have arisen during a crisis period. This absence of fossil is probably related to a drastic reduction of the number of individuals and species, an event increasing the risk of inter-specific matings, as it has been shown for the formation of plant allopolyploid species during the colder climates and changes in the Pleistocene period (Cronn *et al.*, 1999). In the same manner, the chaetognath paralogous rRNA genes would have occurred during the post-Carboniferous crisis as a result of allopolyploidy. Two classes of genes resulting from an allopolyploid event must be defined as homeologous; however, as this event remains putative, we still use the term paralogs. If allopolyploidy is a prominent mode of speciation in higher plants (Ma and Gustafson, 2005), it is not rare in animals; indeed, it may have been the predominant mode of polyploidization in fishes (Le comber and Smith, 2004) and has been described in other taxa such as Amphibia (Stock *et al.*, 2005) or rodents (Gallardo *et al.*, 2004). Chaetognaths are hermaphroditic, with paired ovaries in trunk and paired testes in tail (Ghirardelli and Gamulin, 2004). However, mating occurs allowing the allopolyploid hypothesis plausible. Moreover, other elements do not favour the duplication hypothesis. After duplication, the duplicates have identical sequences and if one or a few sequences accumulate specific mutations, as they are extremely minority compared to the pool of consensus sequences, they should, by genic conversion, find again the consensus sequence. In the allopolyploidy hypothesis, the number of "paralogs" is equivalent and they are on homologous (but not the same) chromosome, that reduces the risk of gene conversion. In addition, if the "paralogs", which origin is very ancient, are present in all the extent chaetognath investigated to date, as they are probably submitted to high selection pressure, this suggests an adaptative advantage for each rRNA classes.

The increasing availability of whole nuclear genome sequences added to powerful and fast bioinformatical methods are opportunities to reanalyse the chaetognath phylogenetic position due to the fact that phylogenetic analysis of gene families are less affected by unrecognized horizontal gene transfer, unrecognized paralogy, highly variable rates of gene evolution, or misalignment, than phylogenies based on single genes. Indeed, as expected, the phylogenetic analyses of the actin protein families have revealed some interesting features. In their study, Yasuda *et al.* (1997) have evidenced that the three *P. gotoi* actin genes have different expression patterns; our phylogenetical study clearly evidenced that these actin genes encode at least two different types of actin. The cytoplasmic actin (PgAct1) is seen to group within insect actins and the two muscle actins (PgAct2 and PgAct3) form a sister group with vertebrate muscle actins. Constraints on biological materials and adaptation to particular habits or habitats will produce widespread convergence. This result suggests that the chaetognath cytoplasmic and muscle actins are extremely divergent in comparison with actins within other invertebrate phyla. Indeed, probably due to gene conversion, in numerous invertebrates taxa, e.g., in ecdysozoan, the actin sequences are nearly identical in spite of different patterns of distribution and in phylogenetical analyses, for a given taxon, all the various actin isoforms could group together (Hooper and Thuma, 2005). This has been well evidenced for crustacean and nematodes; for example,

in a member of the latter (*Caenorhabditis elegans*), four actin genes have been identified and none which differs by more than three amino acids (Krause *et al.*, 1989). On the other hand, among the three *P. gotoi* actins, the amino acid identities were 91.5% between PgAct1 and PgAct2, 86.2% between PgAct1 and Pgact3 and 88.9% between PgAct2 and PgAct3 (Yasuda *et al.*, 1997), suggesting that in this species, gene conversion is an unlikely scenario in the molecular evolution of these genes; by comparison, all vertebrate actin isoforms are over 90% identical to each other. Moreover, our analyses suggest that muscle actins may have arisen at least three times in evolutionary history, in respectively, Chaetognatha, Vertebrata and Insecta. In addition, if the muscle actins of chaetognath are a sister group of those of vertebrates while the sequence cytoplasmic chaetognath actin is similar, except for a base, to some of the cytoplasmic actins of arthropods, this could suggest that the Chaetognath are a very ancient phylum that would have emerged at the time of the Protostomia/Deuterostomia split. Similarly, chaetognaths share embryological and morphological features with these two main branches of Bilateria, suggesting numerous bilaterian plesiomorphies. Mounier *et al.* (1992) have hypothesized that muscle actins arose from cytoplasmic isoforms at many times during animal evolution. The present results support this theory, chaetognath muscle actins would have arisen from cytoplasmic isoforms independently of the other taxa, just after the bilaterian radiation.

Interestingly, recent paleontological study evidenced that the chaetognath bodyplan is very ancient, the oldest attested chaetognaths occur in Lower Cambrian rocks from South China (Chengjiang biota ca. 525 million years old) (Chen and Huang, 2002; Vannier and Chen, 2005; Vannier *et al.*, 2005; Vannier *et al.*, 2006). In addition, protoconodonts which are millimeter-size hook-like, phosphatized microfossils are known from at least the basal Cambrian (e.g., China) and, in many aspects of their external morphology and internal structure, resemble chaetognath grasping spines (Szaniawski, 2002; Vannier *et al.*, 2005, 2006). Although the anatomy of the protoconodont animal remains unknown, these similarities suggest that protoconodont animals are possibly chaetognaths or animals closely related to this phylum. Two sets of fossil evidence: 1) the true chaetognaths from the Chengjiang biota and 2) the protoconodonts, give credit to the idea that chaetognaths have a Precambrian origin and are possibly among the oldest bilaterian animals. In addition, if the relationships of the Chaetognatha to other bilaterian animals remain rather uncertain, other molecular investigations also suggest that they would have originated in the Early Cambrian or before. In their molecular studies, Telford and Holland (1993) said the most likely position of the chaetognaths is as descendants from an early metazoan branch possibly originating prior to the radiation of the major coelomate groups. Wada and Satoh (1994) and later on Halanych (1996) came to similar conclusions. Further investigations by Telford and Holland (1997) confirmed the earlier results and led them to these statements: The chaetognaths are extraordinarily homogenous phylum of animals and the lineage leading to chaetognaths separated from other phyla early in metazoan radiation, probably in the Precambrian. All these conclusions are compatible with the hypothesis that the chaetognaths belong to one of the earliest bilaterians, which differentiated at the beginning, or even before, the great Cambrian evolutionary explosion.

The analysis of chaetognath actins and the finding of 18S-28S paralogs in this phylum suggest that gene conversion could occur less frequently than in other taxa. To date, only another taxon exhibits similar characteristics, the cephalopods, each of the three main actin clades of which are distinct from one another (Carlini *et al.*, 2000) and paralogous rRNA genes have been sequenced (Bonnaud *et al.*, 2003). The chaetognath actins have not very long branches compared to other taxa. This may probably reflect that, in spite of a long period of independent evolution of this phylum, there is no great differing rate of sequences divergence of actin coding regions. In addition, long branch lengths have been found both in the 18S and 28S rDNA analyses (Telford and Holland, 1993; Wada and Satoh, 1994; Halanych, 1996; Peterson and Eernisse 2001; Mallat and Winchell, 2002). We hypothesize that what is regarded as a faster rate of molecular evolution would probably be due to an earlier divergence associated with

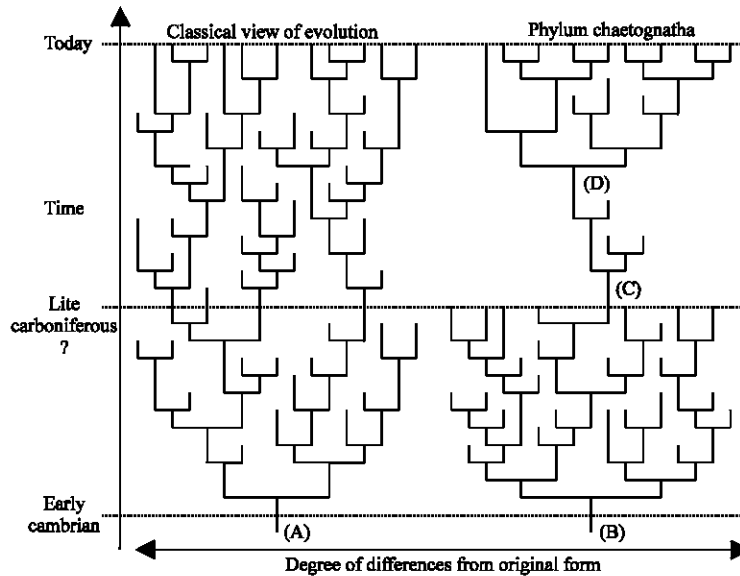


Fig. 3: Putative model of the evolution of the chaetognath lineage, (A) Classical view of the evolution of a phylum, (B) Evolution of the phylum chaetognatha, (C) Arising of the rRNA paralog genes, (D) Separation between Aphramaphora and Phragmaphora

a relatively low rate of gene conversion. And the low rate of gene conversion is the determining element in the preservation of the specificity of the different forms of rRNA paralogs since approximately 300 million years. Another remark confirms the great genomic stability in chaetognaths; whereas chromosome number can vary widely between closely related organisms, karyotypic analyses reveal that all the chaetognath species investigated to date (five *Sagitta* species and one *Spadella*) have the same chromosome number ($n = 9$) (Stevens, 1910; Ghirardelli, 1959; Arnaud, 1963). In addition, the haploid nuclear DNA contents, or “C-values” of *Sagitta enflata* and *Spadella cephaloptera* which are respectively 0.71 and 1.05 (Gregory T.R., unpublished data, <http://www.genomesize.com/>), are relatively stable within species belonging to two different orders. Relative stasis at a so high taxonomic level could suggest a relatively constrained genome sizes (Gregory, 2004). However, the hypothesis of low rate of gene conversion must be modulated. Both in our 18S and 28S analyses, the higher bootstrap values have been found for the groupings of the two classes of chaetognath rRNA sequences which evidenced that these genes are far closer to each other than to any other bilaterian rRNA sequences and that these groupings are unambiguous; this also suggests the isolement of this taxon among the bilaterians. But, in return, the bootstrap values of both the two classes of rRNA genes are weaker; this suggests possible intragenomic exchange between the classes of rRNA which agree with the results of Telford and Holland (1997) who have evidenced that 28S sequences of the genus *Eukrohnia* contain a similar-sized insertion in an identical position in both classes of genes.

Present results suggest that the evolutionary history of the chaetognaths involved at least three distinct stages. Firstly, the lineage leading to chaetognaths separated from other phyla early in bilaterian radiation, probably in Precambrian or in the Early Cambrian. This is supported by 18S and actin analyses and by paleontological data. Secondly, a molecular event occurred probably only once in the ancestor of all the actual chaetognaths and led to the presence of rRNA paralog genes. The relatively low rate of gene conversion in this taxon would have allowed the conservation of the paralogs

in all the chaetognath investigated to date. Lastly, more recently, the chaetognath taxon was subdivided into two orders, Aphragmophora and Phragmophora. These hypotheses are schematized in Fig. 3. On the contrary to the classical view of a tree of life, all the deepest branches of the tree of chaetognath, except one, have been pruned, since the chaetognath are relatively abundant in the fossil record until the latter Carboniferous (Vannier *et al.*, 2006); this is probably the result of differential extinction, a concept developed by Gould (1982). The duplication of whole rRNA clusters or an allopolyploidy event occurred during this crisis period; our molecular data date this period at approximately 300-400 million years. In the future, careful calibration using various several molecular markers will be investigated as well as the physiological role of the rRNA paralogs which remains to be addressed.

References

- Aguinaldo, A.M., J.M. Turbeville, L.S. Linford, M.C. Rivera, J.R. Garey, R.A. Raff and L.A. Lake, 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 387: 489-493.
- Aris-Brosou, S. and Z. Yang, 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18 S ribosomal RNA phylogeny. *Syst. Biol.*, 51: 703-714.
- Arnaud, J., 1963. Sur l'emploi de l'hématoxyline ferrique de Reagaud pour l'étude de la garniture chromosomique chez les chaetognathes. *Comm. Intl. Explor. Sci. Mer Méditer., Rapp. P.V.*, 17: 617-620.
- Blair, J.E. and S.B. Hedges, 2005a. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.*, 22: 2275-2284.
- Blair, J.E. and S.B. Hedges, 2005b. Molecular clocks do not support the Cambrian Explosion. *Mol. Biol. Evol.*, 22: 387-390.
- Bonnaud, L., A. Saihi and R. Boucher-Rodoni, 2003. Are 28S rDNA and 18S rDNA informative for cephalopod phylogeny? *Bull. Mar.*, 71: 197-208.
- Bromham, L.D. and M.D. Hendy, 2000. Can fast early rates reconcile molecular dates with the Cambrian explosion? *Proc. R. Soc. Lond. B Biol. Sci.*, 267: 1041-1047.
- Carlini, D.B., K.S. Reece and J.E. Grave, 2000. Actin gene family evolution and the phylogeny of coleoid cephalopods (Mollusca: Cephalopoda). *Mol. Biol. Evol.*, 17: 1353-1370.
- Carranza, S., J. Bagna and M. Riutort, 1999. Origin and evolution of paralogous rRNA gene clusters within the flatworm family dugesiidae (platyhelminthes, tricladida). *J. Mol. Evol.*, 49: 250-259.
- Casanova, J.P., 1999. Chaetognatha. In: *South Atlantic Zooplankton*. Boltovskoy, D. (Ed.), Backhuys Publishers, Leiden, pp: 1353-1374.
- Chen, J.Y. and D.Y. Huang, 2002. A possible lower cambrian chaetognath (arrow worm). *Science*, 298: 187.
- Cronn, R.C., R.L. Small and J.F. Wendel, 1999. Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl. Acad. Sci. USA.*, 96: 14406-14411.
- Delsuc, F., H. Brinkmann, D. Chourrout and H. Philippe, 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, 439: 965-968.
- Engel, M.S. and D.A. Grimaldi, 2004. New light shed on the oldest insect. *Nature*, 427: 627-630.
- Erber, A., D. Riemer, M. Bovenschulte and K. Weber, 1998. Molecular phylogeny of metazoan intermediate filament proteins. *J. Mol. Evol.*, 47: 751-762.
- Felsenstein, J., 2005. PHYLIP: Phylogeny Inference Package, Version 3.65. University of Washington, Seattle, WA (distributed by the author).

- Gallardo, M.H., G. Kausel, A. Jimenez, C. Bacquet, C. Gonzalez, J. Figueroa, N. Kohler and R. Ojeda, 2004. Whole-genome duplications in South American desert rodents (Octodontidae). *Biol. J. Linn. Soc. Lond.*, 82: 443-451.
- Ghirardelli, E., 1959. L'apparato riproduttore femminile nei Chaetognati. *Rend. Acad. Nez.* XL, 10: 1-46.
- Ghirardelli, E., 1968. Some aspects of the biology of the Chaetognaths. *Adv. Mar. Biol.*, 6: 271-375.
- Ghirardelli, E. and T. Gamulin, 2004. Chaetognatha. *Fauna d'Italia*, 39: 1-155.
- Gillespie, J.J., J.B. Munro, J.M. Heraty, M.J. Yoder, A.K. Owen and A.E. Carmichael, 2005. A secondary structural model of the 28S rRNA expansion segments D2 and D3 for Chalcidoid wasps (Hymenoptera: Chalcidoidea). *Mol. Biol. Evol.*, 22: 1593-608.
- Glazko, G.V., E.V. Kookin and I.B. Rogozin, 2005. Molecular dating: Ape bones agree with chicken entrails. *Trends Genet.*, 21: 89-92.
- Gonzalez, I.L. and J.E. Sylvester, 2001. Human rDNA: Evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics*, 73: 255-263.
- Goodman, M., G.M., Moore and G. Matsuda, 1975. Darwinian evolution in the genealogy of haemoglobin, 253: 603-608.
- Gould, S.J., 1982. Change in Development Timing as a Mechanism of Macroevolution. In: *Evolution and Development*. Bonner, J.T. (Ed.), Springer-Verlag, Berlin, pp: 333-346.
- Gregory, T.R., 2004. Macroevolution, hierarchy theory and the C-value enigma. *Paleobiology*, 30: 179-202.
- Halanych, K.M., J.D. Bacheller, A.M.A. Aguinaldo, S.M. Liva, D.M. Hillis and J.A. Lake, 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science*, 267: 1641-1643.
- Halanych, K.M., 1996. Testing hypotheses of Chaetognath origins: Long branches revealed by 18S ribosomal DNA. *Syst. Biol.*, 45: 223-246.
- Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, 41: 95-98.
- Hedges, S.B., J.E. Blair, M.L. Venturi and J.L. Shoe, 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.*, 4: 2.
- Helfenbein, K.G., H.M. Fourcade, R.G. Vanjani and J.L. Boore, 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc. Natl. Acad. Sci. USA.*, 101: 10639-10643.
- Higgins, D.G., A.J. Bleasby and R. Fuchs, 1992. Clustal V-improved software for multiple sequence alignment. *Comput. Applied Biosci.*, 8: 189-191.
- Hooper, S.L. and J.B. Thuma, 2005. Invertebrate muscles: Muscle specific genes and proteins. *Physiol. Rev.*, 85: 1001-1060.
- Huelsenbeck, J.P., 1991. Tree-length distribution skewness - an indicator of phylogenetic information. *Syst. Zool.*, 40: 257-270.
- Krause, M., M. Wild, B. Rosenzweig and H. Hirsh, 1989. Wild-type and mutant actin genes in *Caenorhabditis elegans*. *J. Mol. Biol.*, 208: 381-392.
- Kumar, S., 2005. Molecular clocks: Four decades of evolution. *Nat. Rev. Genet.*, 6: 654-662.
- Le Comber, S.C. and C. Smith, 2004. Polyploidy in fishes: Patterns and processes. *Biol. J. Linn. Soc. Lond.*, 82: 431-442.
- Li, W.H., 1997. *Molecular Evolution*. Sinauer (Ed.), Sunderland, Massachusetts, pp:177-193.
- Liao, D.Q., 1999. Concerted Evolution: Molecular mechanism and biological implications, *Am. J. Hum. Gen.*, 64: 24-30.

- Ma, X.F. and J.P. Gustafson, 2005. Genome evolution of allopolyploids: A process of cytological and genetic diploidization. *Cytogenet. Genome Res.*, 109: 236-249.
- Mallatt, J. and C.J. Winchell, 2002. Testing the new animal phylogeny: First use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. *Mol. Biol. Evol.*, 19: 289-301.
- Mounier, N., M. Gouy, D. Mouchiroud and J.C. Prudhomme, 1992. Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. *J. Mol. Evol.*, 34: 406-415.
- Nei, M.P.X. and G. Glazko, 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. USA.*, 98: 2497-2502.
- Papillon, D., Y. Perez, L. Fasano, Y. LeParco 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. LeParco, 2004. Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. *Mol. Biol. Evol.*, 21: 2122-2129.
- Papillon, D., Y. Perez, X. Caubit and Y. LeParco, 2006. Systematics of Chaetognatha under the light of molecular data, using duplicated ribosomal 18S DNA sequences. *Mol. Phyl. Evol.*, 38: 621-634.
- 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.
- Peterson, K.J., J.B. Lyons, K.S. Nowak, C.M. Takacs, M.J. Wargo and M.A. McPeck, 2004. Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. USA.*, 101: 6536-6541.
- Pisani, D., L.L. Poling, M. Lyons-Weiler and S. Blair-Hedges, 2004. The colonization of land by animals: Molecular phylogeny and divergence times among arthropods. *BMC Biol.*, 2: 1.
- Rooney, A.P., 2004. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18 S rRNA genes in apicomplexans. *Mol. Biol. Evol.*, 21: 1704-1711.
- Saitou, N. and M. Nei, 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406-425.
- Stevens, N.M., 1910. Further studies on reproduction in *Sagitta*. *J. Morphol.*, 21: 279-319.
- Stock, M., C. Steinlein, D.K. Lamatsch, M. Schartl and M. Schmid, 2005. Multiple origins of tetraploid taxa in the Eurasian *Bufo viridis* subgroup. *Genetica*, 124: 255-272.
- Swofford, D.L., 2001. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b8. Sunderland, Massachusetts: Sinauer Associates.
- Szaniawski, H., 2002. New evidence for the protoconodont origin of chaetognaths. *Acta Palaeontol. Pol.*, 47: 405-419.
- Taylor, J.S. and J. Raes, 2004. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.*, 38: 615-643.
- 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 P.W.H. Holland, 1997. Evolution of 28S ribosomal DNA in Chaetognaths: Duplicate genes and molecular phylogeny. *J. Mol. Evol.*, 44: 135-144.
- Tokioka, T., 1965. The taxonomical outline of Chaetognatha. *Publ. Seto Mar. Biol. Lab.*, 12: 335-357.
- Van de Peer, Y. and R. de Wachter, 1997. Construction of evolutionary distance trees with TREECON for windows: Accounting for variation in nucleotide substitution rate among sites. *CABIOS*, 13: 227-230.
- Vandekerkhove, J. and K. Weber, 1981. Actin typing on total cellular extracts: A highly sensitive protein-chemical procedure able to distinguish different actins. *Eur. J. Biochem.*, 113: 595-603.

- Vannier, J. and J.Y. Chen, 2005. Early Cambrian Food Chain: New Evidence from Fossil Aggregates in the Maotianshan Shale Biota, SW China. *Palaios*, 20: 3-26.
- Vannier, J., M. Steiner, E. Renvoisé, H. Sixue and J.P. Casanova, 2005. Arrow worms: Small marine predators from the deep time. *Acta Micropalaeontol. Sin.*, pp: 189-190.
- Vannier, J., M. Steiner, E. Renovoise, S.-X. Hu and J.-P. Casanova, 2006. Predator arrow worms in the Early Cambrian food webs. *Proc. R. Soc. B*, In Press.
- 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.
- Waloszek, D., J.E. Repetski and A. Maas, 2005. A new Late Cambrian pentastomid and a review of the relationships of this parasitic group. *Trans. R. Soc. Edinb. Earth Sci.*, 96: 163-176.
- Wang, Y. and X. Gu, 2000. Evolutionary patterns of gene families generated in the early stage of vertebrates. *J. Mol. Evol.*, 51: 88-96.
- Wuyts, J., G. Perriere and Y. Van de Peer, 2004. The European ribosomal RNA database. *Nucleic Acids Res.*, 32: D101-D103.
- Yasuda, E., T. Goto, K.W. Makabe and N. Satoh, 1997. Expression of actin genes in the arrow worm *Paraspadella gotoi* (Chaetognatha). *Zool. Sci.*, 14: 953-960.